

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

UNIVERSITÉ M'HAMED BOUGARA – BOUMERDÈS



FACULTÉ DES SCIENCES

THÈSE DE DOCTORAT

Présentée par :

ZERARGA Loufi

Filière : Informatique

Option : Informatique

Extension Floue de l'Analyse de Concepts Formels pour la Recherche d'Information

Soutenue publiquement le **21/02/2018**, devant le jury :

M. MEZGHICHE	Mohamed	Prof-UMBB	Président
M. DJOUADI	Yassine	Prof-USTHB	Directeur de thèse
M. AHMED OUAMER	Rachid	Prof-UMMTO	Examineur
M. KECHID	Samir	Prof-USTHB	Examineur
M. IMACHE	Rabah	MCA-UMBB	Examineur
M. MAOUCHE	Amin Riad	MCA-UMBB	Examineur

Année Universitaire : 2017/2018

Résumé

Les travaux effectués dans le cadre de cette thèse reposent principalement sur l'idée d'utiliser le cadre théorique offert par l'analyse de concepts formels (ACF) dans le domaine de la recherche d'information (RI), domaine fort souvent abordé uniquement sous un volet expérimental. Estimant que la RI peut aussi tirer profit d'un cadre formel, nous proposons une théorie axiomatique multi-sortes qui formalise le processus de recherche et les objets liés à la RI. Cette théorie offre un cadre générique capable de reproduire et de représenter les modèles de RI existants. Cette généralité est mise en évidence en prouvant que trois interprétations différentes satisfont notre théorie ; à savoir le modèle Booléen, le modèle flou standard et le modèle vectoriel. L'utilité de notre théorie est aussi mise en évidence par le fait qu'elle permet d'enrichir le modèle Booléen et le modèle flou standard. En d'autres termes, nous démontrons des propriétés algébriques caractérisant les différentes formes de requêtes de ces modèles par les treillis de concepts formels et les treillis de NII-paires relatifs respectivement à l'ACF et à sa lecture possibiliste et floue. Ainsi enrichi, le modèle Booléen a été mis à contribution dans le cadre de la RI structurée en se basant sur l'extension triadique de l'ACF, permettant ainsi le traitement des requêtes orientées contenu et des requêtes orientées contenu et structure.

La quasi-majorité des approches existantes de RI basées sur l'ACF ne prennent pas en considération la pondération des requêtes. A partir de ce constat, nous proposons une approche de recherche par navigation dans les treillis relatifs à l'extension floue de l'ACF et à son interprétation possibiliste. L'ignorance partielle/totale étant une réalité forte souvent constatée dans les matrices d'incidence *Documents* \times *Termes*, nous proposons, dans le cadre de cette extension floue, de modéliser de pareilles réalités par des intervalles de vérité. Afin de faciliter la recherche par navigation dans les treillis inhérents à l'ACF, nous proposons une transformation linéaire de ces derniers en définissant un ordre total sur les ensembles de concepts formels flous et duallement sur les ensembles de NII-paires floues.

Mots clefs : *Analyse de Concepts Formels, Recherche d'Information, Théorie Axiomatique Multi-Sortes, Théorie des possibilités, Théorie des Ensembles Flous.*

Abstract

The works accomplished within this thesis is mainly based on the idea of using the theoretical framework of formal concept analysis (FCA) in the information retrieval (IR) field's, a field often addressed only under an experimental way. Estimating that IR can also benefit from a formal framework, we propose a many-sorted axiomatic theory that formalizes the retrieval process and the objects related to IR. The theory relies on a sound set of axioms driving the retrieval process as proof of theorems. The genericity of the proposed approach consists of a main motivation. As such, it will be proved that three IR models, namely the Boolean model; the fuzzy-set-based extension of the Boolean model; and the vector space model, satisfy the proposed theory, establishing then its consistency. The utility of our theory is also highlighted by the fact that it enriches both the Boolean model and the standard fuzzy model. In other words, we prove useful algebraic properties. These properties characterize the different queries forms of these models by lattices of formal concepts and lattices of NII-pairs, related to FCA and its possibility-theoretic extension in both crisp and fuzzy settings. Thus enriched, the Boolean model has been used as part of the structured IR based on triadic FCA, allowing to process both content-only queries and content-and-structure queries.

It may be remarked that almost all existing FCA-based IR approaches do not consider weighted queries. From this, we propose a retrieval-browsing approach in lattices related to fuzzy FCA settings and its possibility-theoretic extension. Partial/total ignorance being a strong reality often found in *Documents* × *Terms* matrix, we propose in fuzzy FCA settings, to model such reality by truth interval-valued. In order to facilitate the navigation retrieval in the lattices related to FCA, we propose a linear transformation of these lattices by defining a total order on the sets of fuzzy formal concepts and on the sets of fuzzy NII-pairs.

Keywords : *Formal Concept Analysis, Information Retrieval, Many-Sorted Axiomatic Theory, Possibility Theory, Fuzzy Sets Theory.*

Remerciements

Je voudrais tout d'abord remercier grandement et chaleureusement Monsieur DJOUADI Yassine, professeur à l'USTHB, de m'avoir proposé ce thème et diriger cette thèse. Je tiens à lui témoigner ma profonde gratitude pour l'aide précieuse qu'il m'a apporté, ses conseils avisés, sa prévoyance, sa disponibilité, sa générosité, son engagement et sa rigueur de travail. J'ai pris un grand plaisir à travailler avec lui car outre son appui scientifique, il a toujours été là pour me soutenir et me conseiller au cours de l'élaboration de ce travail.

Je tiens à remercier vivement Monsieur le président de jury MEZGHICHE Mohamed, Professeur à l'UMBB, pour l'honneur qu'il nous fait de présider le jury. Ses conseils et discussions sur la logique nous ont beaucoup aidé dans cette thèse. Je le remercie également de m'avoir accueilli dans le laboratoire LIMOSE.

Je tiens à remercier vivement

Monsieur AHMED-OMER Rachid, Professeur à l'UMMTO,

Monsieur KECHID Samir, Professeur à l'USTHB,

Monsieur IMACHE Rabah, Maître de conférences « A » à l'UMBB,

Monsieur MAOUCHE Amin Riad, Maître de Conférences « A » à l'UMBB,

qui ont accepté de juger ce travail et y avoir consacré de leurs précieux temps.

Messieurs, Je vous suis reconnaissant d'avoir porté vos regards d'experts sur ma thèse.

Je tiens à exprimer mes sincères remerciements à Monsieur le Recteur de l'UMBB, le Professeur BENTELLIS Hakim, pour son soutien et ses encouragements.

Les mots ne suffiraient pas pour exprimer mes profondes gratitude à Faiza (mon épouse) et à Ania (ma fille) pour avoir compris et endurer les impératifs de cette thèse. Ma reconnaissance s'adresse à mes sœurs et à mes frères, qui ont su m'apporter, sans relâche leurs soutiens durant l'accomplissement de ce travail et durant tout mon parcours universitaire. Merci à vous.

Je tiens à remercier mes collègues du département d'informatique et en particulier M. AHMED-NACER Messaoud, M. AIT-BOUZIAD Ahmed, Mme ALOUANE Bisma, M. AMIR Karim, M. CHAABANI Mohamed, M. SADOU Mohamed (Actuellement à Alger 1) et M. SALMI Chiekh.

Je tiens également à remercier mes collègues cadres de la direction générale des enseignements et de la formation supérieurs (MESRS, 2014-2016) et en particulier Mme ABABOU Assia (Actuellement Directrice de l'ESSS), M. AIT-ZAI Abdelhakim, M. ALLOUCHE Mohand, Mme AMINE Fiza et Mme. KOUCHEIDA Safia.

Je dédie ce travail,
À la mémoire de mes parents.
À Faiza, Ania et Aris.

Table des matières

Introduction Générale	1
1 Analyse de Concepts Formels (ACF)	6
1.1 Introduction	7
1.2 Notions de Base	7
1.3 Analyse de Concepts Formels	8
1.3.1 Contexte Formel	9
1.3.2 Opérateur de Dérivation et Connexion de Galois	9
1.3.3 Concept Formel	10
1.3.4 Treillis de Concepts Formels	11
1.4 Lecture Possibiliste de l'ACF	13
1.5 Limites de l'ACF Classique	14
1.6 Extension Triadique de l'ACF	15
1.7 Extension Floue de l'ACF	17
1.7.1 Ensembles Flous et Logique Floue	17
1.7.2 ACF Floue	20
1.8 Extension floue à Valeur d'Intervalle de Vérité	27
1.9 Conclusion	28
2 Recherche d'Information (RI)	29
2.1 Introduction	31
2.2 Les Modèles Ensemblistes	32
2.2.1 Modèle Booléen	32
2.2.2 Modèle Booléen basé sur les Ensembles Flous	33
2.2.3 Modèle P-norm	34
2.3 Les Modèles Algébriques	34
2.3.1 Modèle Vectoriel	34
2.3.2 Modèle d'Indexation Sémantique Latente (LSI)	36

2.4	Les Modèles Probabilistes	37
2.4.1	Le Modèle de Base	37
2.4.2	Le modèle de Langue	37
2.5	Les Modèles Logiques	38
2.5.1	Modèles basés sur la Logique Propositionnelle	39
2.5.2	Modèles basés sur des Systèmes d'Argumentation Probabiliste	39
2.5.3	Modèles basés sur la Logique Modale	40
2.5.4	Modèles basés sur la Logique Descriptive	41
2.5.5	Modèles basés sur des Graphes Conceptuels	43
2.5.6	Modèles basés sur la Théorie des Situations	44
2.5.7	Modèles basés sur Datalog Probabiliste	45
2.5.8	Les Méta-Modèles	45
2.6	Les Approches de RI basées sur l'ACF	46
2.6.1	Amélioration des Résultats de Requêtes	47
2.6.2	Web et Courrier Électronique	48
2.6.3	Logiciel, Image et Base de Connaissances	49
2.6.4	Traitement de Requêtes sur des Données Hétérogènes	50
2.6.5	Réponses Contextuelles et Classement des Résultats	51
2.6.6	Réinjection de Pertinence	52
2.6.7	Les Systèmes de RI basés sur l'ACF	53
2.6.8	Les Approches de RI basées sur l'ACF Floue	54
2.7	Conclusion	56
3	Théorie Multi-Sorte pour la Recherche d'Information	58
3.1	Introduction	59
3.2	Théorie Multi-Sorte	60
3.2.1	Vocabulaire	60
3.2.2	Axiomes	61
3.2.3	Inférence	61
3.3	Modèle Booléen	63
3.4	Modèle basé sur les Ensembles Flous	68
3.5	Modèle Vectoriel	72
3.6	Conclusion	76
4	Lecture Possibiliste de l'Analyse de Concepts Formels pour la Recherche d'Information	78
4.1	Introduction	79
4.2	Enrichissement du Modèle \mathcal{M}_B basé sur l'ACF	79
4.3	Recherche Orientée Contenu/Structure	84
4.3.1	Traitement des Requêtes Orientées Contenu	88
4.3.2	Traitement de Requêtes Orientées Contenu et Structure	94

4.4	Conclusion	95
5	Extension Floue de l'ACF pour la RI	97
5.1	Introduction	98
5.2	Enrichissement du Modèle \mathcal{M}_f basé sur l'ACF Floue	98
5.3	Pondération de Requêtes	107
5.3.1	Requêtes Disjonctives	109
5.3.2	Formes Négatives de Requêtes Pondérées	110
5.4	Ignorance Partielle/Totale dans la Matrice d'Incidence	110
5.5	Réduction Linéaire des Treillis	114
5.6	Conclusion	117
	Conclusion Générale et Perspectives	119

Table des figures

1.1	Diagramme de Hasse du Treillis de Concepts Formels Associé au Contexte de l'Éxemple.1.3.1.	12
4.1	Isomorphisme entre \hat{E} et $\mathcal{L}(\mathcal{K}_2) \setminus \{\langle \mathcal{D}_2, \emptyset \rangle\}$	85
4.2	Isomorphisme entre \hat{E}_\neg et $\mathcal{P}(\overline{\mathcal{K}}_2) \setminus \{\langle \emptyset, \emptyset \rangle\}$	86
4.3	Exemple de Document Structuré	87
4.4	Diagramme de Hasse du Treillis $\mathcal{L}(\mathfrak{K}_T)$	90
4.5	Diagramme de Hasse du Treillis $\mathcal{P}(\mathfrak{K}_T)$	92
5.1	(1) Isomorphisme entre $\hat{\mathbb{F}}$ et $\mathcal{B}(\tilde{\mathcal{K}}_e) \setminus \{\langle \{d_1^1, d_2^1, \dots, d_5^1\}, \emptyset \rangle\}$ (2) Isomorphisme entre $\check{\mathbb{F}}$ et $\mathcal{B}_{\text{NII}}(\tilde{\mathcal{K}}_e) \setminus \{\langle \{d_1^0, d_2^0, \dots, d_5^0\}, \emptyset \rangle\}$	108
5.2	Le Treillis de Concepts Formels Flous $\mathcal{L}(\tilde{\mathcal{K}}_e)$	111
5.3	Le Treillis de NII-paires $\mathcal{P}(\tilde{\mathcal{K}}_e)$	112
5.4	Les Treillis Linéaires Correspondants aux Treillis $\mathcal{P}(\tilde{\mathcal{K}}_e)$ (à gauche) et $\mathcal{L}(\tilde{\mathcal{K}}_e)$ (à droite)	116

Liste des tableaux

1.1	Exemple de Contexte Formel	9
1.2	Exemple de Contexte Formel Triadique	16
1.3	Exemples de T-normes et de T-conormes	19
1.4	Les S-implications Usuelles	19
1.5	Les R-implications Usuelles	20
3.1	Alphabet de la Théorie \mathfrak{T}	60
3.2	La Fonction d'Interprétation I_B	64
3.3	La Relation Booléenne \mathcal{R}_1	64
3.4	La Relation Booléenne \mathcal{R}_2	67
3.5	La Fonction d'Interprétation I_f	68
3.6	Le Contexte Formel Flou $\tilde{\mathcal{K}}_1$	69
3.7	Le Contexte Formel Flou $\tilde{\mathcal{K}}_2 = (\{d_1, \dots, d_6\}, \{t_1, \dots, t_6\}, \tilde{\mathcal{R}})$	71
3.8	La Fonction d'Interprétation I_v	73
3.9	Normalisation du Tableau.3.7	75
4.1	Le Contexte Formel Triadique $\mathfrak{K}_e = (\mathcal{D}_e, \mathcal{T}_e, \mathcal{E}_e, \mathfrak{R}_e)$	89
4.2	Les Concepts Formels Dyadiques du Contexte $(\mathcal{D}_e \times \mathcal{E}_e, \mathcal{T}_e, (\mathfrak{R}_e)_{\mathcal{T}})$	91
4.3	Les NII-Paires du Contexte Formel $(\mathcal{D}_e \times \mathcal{E}_e, \mathcal{T}_e, (\mathcal{R}_e)_{\mathcal{T}})$	93
5.1	Exemple de Contexte Formel Flou	107
5.2	Exemple de Contexte Formel Flou à Intervalle de Vérité	114
5.3	Les Distances Associée Aux Concepts Formels à Valeur d'Intervalle	118

Introduction Générale

Contexte

L'Analyse de Concepts Formels (ACF) a été introduite par le Wille [1982] en tant que méthode de représentation de connaissance et d'analyse de données. Les fondements théoriques de l'ACF reposent sur la théorie des treillis définie par G. Birkhoff et la théorie des ensembles. Wille a utilisé l'interprétation philosophique du concept comme une unité de pensée comprenant un ensemble d'objets et un ensemble de leurs attributs communs. Par la suite, l'ACF a été consolidée mathématiquement par [Ganter et Wille \[1999\]](#).

L'ACF repose sur la notion de “Contexte Formel” qui représente une relation binaire reliant un ensemble d'objets à un ensemble d'attributs (propriétés). Généralement un contexte formel est représenté par une matrice avec en ligne les objets et en colonne les attributs. Une marque (par exemple, le symbole “×”) placée sur l'intersection d'une ligne et d'une colonne signifie que l'objet de la ligne possède l'attribut de la colonne. L'ACF consiste à induire des paires de sous-ensembles d'objets et d'attributs $\langle \text{Objets}, \text{Attributs} \rangle$ appelées concepts formels, le sous-ensemble *Objets* est appelé extension et le sous-ensemble *Attributs* est appelé intension. Pour un concept formel donné, les objets de l'extension possèdent tous les attributs de son intension, De même tous les attributs de l'intension sont partagés par tous les objets de son extension. L'ACF permet de définir une hiérarchie dite subsomption entre les concepts formels qui peut être interprétée comme une relation de généralisation/spécialisation entre les concepts formels. Tous les concepts formels d'un contexte formel ainsi que toutes leurs relations peuvent être transformés en une structure de treillis de concepts qui peut être présentée sous forme de diagramme de Hasse. Mathématiquement, les concepts formels sont induits en utilisant l'opérateur de dérivation de Galois (noté Δ), appelé également opérateur de suffisance [[Düntsch et Gediga, 2003](#)]. Cet opérateur forme une connexion de Galois entre les sous-ensembles d'objets et d'attributs et sa composition ($\Delta \circ \Delta$) forme un opérateur de fermeture.

Une interprétation de l'ACF basée sur la théorie des possibilités [Zadeh, 1978] a été proposée par Dubois et al. [2007]. Ces auteurs ont mis en évidence l'intérêt d'introduire de nouveaux opérateurs de dérivation en plus de l'opérateur de suffisance. Ainsi, ils ont introduit les opérateurs de possibilité (Π), de nécessité (N) et de suffisance duale (∇).

La Recherche d'Information (RI) est l'une des applications naturelle de l'ACF, qui résulte de l'évidente analogie entre la relation $Objects \times Attributs$ liée à l'ACF et la matrice d'incidence $Documents \times Termes$ liée à la RI. Godin et al. [1986] ont montré l'intérêt de l'utilisation de l'ACF dans la RI. L'intension d'un concept formel est considérée comme une requête, et son extension est considérée comme l'ensemble des documents pertinents à cette requête. La consolidation mathématique de l'ACF a donné lieu à son utilisation effective et efficace en RI. Depuis, plusieurs approches ont été proposées [Carpineto et al., 2004; Messai et al., 2005; Nauer et Toussaint, 2007, 2009]. À l'instar de ces travaux, les recherche effectuées dans le cadre de cette thèse reposent sur l'utilisation de la théorie de l'analyse de concepts formels et de son extension floue dans le domaine de la recherche d'information.

Les applications de RI sont souvent amenées à manipuler des réalités concrètes où les entrées de la matrice d'incidence $Documents \times Termes$ sont données non plus par la marque "X" mais par une valeur matérialisant le poids d'un terme dans un document. Ce type de réalités nous a amené à considérer l'extension floue de l'ACF pour la RI.

L'ACF floue repose sur la notion de contexte formel flou, où les relations objet-attribut sont des degrés de vérité à partir de l'échelle $L = [0, 1]$. Depuis le premier article de Burusco et Fuentes-González [1994], de nombreuses approches ont été proposées : Pollandt [1997], Bělohávek [2001], Yahia et Jaoua [2001], [Djouadi et Prade, 2011], etc. Le point commun de ces approches et qu'elles se basent sur la théorie des sous-ensembles flous de Zadeh [1965] et la logique floue sous-jacente.

Problématique

L'état de l'art effectué dans le cadre de cette thèse, relativement aux différentes approches de RI basées sur l'ACF, a permis de faire ressortir les constatations suivantes :

1. La plupart des approches existantes se basent sur une relation $Documents \times Termes$ de type Booléenne. Dans pareil cas, les poids des termes d'indexation dans les documents ne sont pas considérés.
2. Les requêtes sont aussi Booléennes et elles ne permettent pas de considérer des requêtes pondérées ;
3. Les requêtes sont limitées à la forme conjonctive.

4. Les résultats d'une recherche ne sont pas ordonnés.

En résumé, ces approches sont équivalentes au modèle Booléen de [Salton \[1968\]](#), qui est le premier modèle de recherche d'information, et elles se limitent aux requêtes conjonctives. Il se pose alors les questions de comment 1) Prendre en charge les poids des termes d'indexation dans les documents et dans les requêtes, 2) Associer des scores de pertinence aux documents, 3) Considérer la négation et la forme disjonctive de requêtes.

Contributions

Bien que la recherche d'information ait été le plus souvent abordée sous le volet expérimental, nous estimons qu'elle doit être guidée par des modèles théoriques. Dans cette perspective, l'axiomatisation est une méthodologie approuvée pour l'étude des modèles mathématiques. C'est dans ce cadre que nous proposons, dans une première contribution, une théorie axiomatique (notée \mathfrak{T}) pour la RI. La théorie ainsi proposée est basée sur une logique multi-sorte. L'aspect multi-sorte de la proposition est motivé par la diversité des objets considérés dans les systèmes de RI, à savoir les documents, les termes d'indexation et les requêtes. L'objectif principal de la théorie est de définir des règles syntaxiques correspondant aux axiomes conduisant le processus de recherche d'information comme preuve de théorèmes. Une telle preuve consiste en toute séquence de formules telle que chaque formule soit un axiome, soit déduit des formules précédentes en utilisant les règles de déduction. Nous avons, également, prouvé que trois modèles de RI, à savoir le modèle Booléen [[Salton, 1968](#)]; le modèle flou standard [[Bookstein, 1980](#)]; et le modèle vectoriel [[Salton et al., 1975](#)] satisfont la théorie proposée, prouvant ainsi sa cohérence et établissant sa généralité.

Dans une seconde contribution, nous proposons l'enrichissement du modèle Booléen par une caractérisation des différentes formes de requêtes. En partant du modèle de Salton et la théorie \mathfrak{T} , nous démontrons des propriétés (relations) algébriques significatives (isomorphismes) entre des sous-ensembles de paires (*Documents, requête*) d'un côté, les treillis de concepts formels et/ou de NII-paires¹ relatifs à la lecture possibiliste de l'ACF de l'autre côté. Ces relations algébriques permettent d'approuver, d'une part, l'application de l'ACF dans la RI, et d'autre part, l'application de la lecture possibiliste de l'ACF pour le traitement de la forme disjonctive et la forme négation des requêtes.

Dans une démarche semblable à la précédente contribution, nous proposons une caractérisation des différentes formes de requêtes Booléennes du modèle flou standard de [[Bookstein, 1980](#)]. Pour cette contribution, nous considérons un contexte

1. Les NII-paires sont obtenues par les opérateurs de dérivations Π et N .

formel flou, et nous prouvons des isomorphismes entre des sous-ensembles de paires (*Documents, requête*) et des treillis de concepts formels semi-flous et des treillis de NII-paires semi-flous.

Nous avons jugé intéressant d’aborder la recherche d’information structurée basée sur l’ACF. Rappelons que la RI structurée a comme but de retrouver des parties de documents au lieu des documents entiers. Dans ce cadre, nous avons proposé de traiter les requêtes Booléennes orientées contenu et orientées structure sur la base de l’extension triadique de l’ACF, où la relation du contexte formel est ternaire entre les objets, les attributs et un troisième ensemble dit de conditions.

Dans une autre contribution, nous proposons d’étendre notre approche de RI basée sur l’ACF floue au traitement de requêtes pondérées (forme disjonctive et négation). Cette proposition utilise les treillis de NII-paires floues.

Nous proposons également l’application de l’extension floue de l’ACF basée sur les intervalles de vérité pour la RI. cette contribution permet, principalement, de traiter l’ignorance partielle/totale dans un contexte formel flou.

Dans une dernière contribution, nous proposons une méthode de linéarisation des treillis de concepts formels et des treillis de NII-paires. Les treillis ainsi obtenus définissent un ordre total sur les ensembles de concepts formels et les ensembles de NII-paires ce qui facilite le mode de recherche par navigation.

Les approches proposées et les contributions effectuées dans le cadre de cette thèse ont été ainsi validées par des publications. La théorie axiomatique, les modèles sémantiques y afférant et l’enrichissement du modèle \mathcal{M}_B ont été publiés dans [Zerarga et Djouadi, 2017]. L’approche triadique pour la recherche d’information structurée a été publiée dans [Zerarga et Djouadi, 2013]. Tandis que dans [Zerarga et Djouadi, 2012], nous avons publié l’approche basée sur les valeurs à intervalle de vérité pour le traitement de l’ignorance partielle/totale dans la matrice d’incidence $Documents \times termes$.

Organisation du Manuscrit

En plus d’une introduction générale et d’une conclusion, ce manuscrit est organisé en cinq chapitres. Les deux premiers sont dédiés à l’état de l’art, tandis que les trois autres rapportent nos différentes contributions.

Le Chapitre. 1 présente l’Analyse de Concepts Formels et ses différentes extensions. Nous rappelons dans un premier temps des notions relatives à la théorie des ensembles nécessaires à la compréhension de l’ACF. Nous présentons la théorie de l’ACF, sa lecture possibiliste et une description de l’état de l’art des extensions floues de l’ACF. Le Chapitre. 2 est dédié à la Recherche d’Information. Nous y rappelons les principaux modèles et nous y présentons un état de l’art exhaustif sur les modèles logiques et les modèle basés sur l’ACF.

Dans le Chapitre. 3, nous présentons notre théorie axiomatique multi-sortes pour

la RI et nous démontrons que les modèles Booléen, flou standard et vectoriel de la RI classique satisfont cette théorie. Dans le Chapitre. 4, nous proposons une caractérisation des requêtes Booléennes élémentaires (conjonctives, disjonctive et négations) par les treillis relatifs à l'ACF et sa lecture possibiliste. Nous présentons aussi une nouvelle approche de RI structurée basée sur l'ACF triadique. Le Chapitre. 5 décrit nos contributions dans le cadre de l'extension floue de l'ACF pour la RI. Nous présentons d'abord une caractérisation des requêtes élémentaires permettant l'enrichissement du modèle flou standard. Cette caractérisation repose sur les treillis de concepts formels semi-flous et les treillis de NII-paires semi-floues. Nous présentons aussi une nouvelle approche permettant la pondération des termes dans les requêtes et dans les documents. Notre contribution relative à l'expression de l'ignorance totale/partielle des poids d'indexation est aussi présentée dans ce chapitre. Nous y proposons enfin une démarche de linéarisation des treillis de concepts formels flous et des treillis de NII-paires floues.

Chapitre 1

Analyse de Concepts Formels (ACF)

Sommaire

1.1	Introduction	7
1.2	Notions de Base	7
1.3	Analyse de Concepts Formels	8
1.3.1	Contexte Formel	9
1.3.2	Opérateur de Dérivation et Connexion de Galois	9
1.3.3	Concept Formel	10
1.3.4	Treillis de Concepts Formels	11
1.4	Lecture Possibiliste de l'ACF	13
1.5	Limites de l'ACF Classique	14
1.6	Extension Triadique de l'ACF	15
1.7	Extension Floue de l'ACF	17
1.7.1	Ensembles Flous et Logique Floue	17
1.7.2	ACF Floue	20
1.8	Extension floue à Valeur d'Intervalle de Vérité	27
1.9	Conclusion	28

1.1 Introduction

Ce chapitre présente l'Analyse de Concepts Formels (ACF) et ses différentes extensions. Nous rappelons dans un premier temps des notions de base et les définitions mathématiques relatives à la théorie des ensembles nécessaires à la compréhension de l'ACF. Dans la Section. 1.3 nous présentons les notions et les définitions relatives à la théorie de l'ACF, nous donnerons également un aperçu des algorithmes de construction des treillis de concepts formels. Dans la Section. 1.4 nous détaillons la lecture possibiliste de l'ACF en présentant les différents opérateurs de dérivation. Nous discuterons par la suite, en Section. 1.5, les limites de l'ACF. Le reste du chapitre traite les différentes extensions de l'ACF. Nous commencerons par l'approche triadique, et nous dresserons dans la Section. 1.7 un état de l'art relatif aux extensions floues de l'ACF. Finalement, dans la Section. 1.8, nous présentons l'extension floue à valeur d'intervalle de vérité.

1.2 Notions de Base

Nous allons, dans cette section, rappeler des définitions et des concepts mathématiques nécessaires à l'introduction de l'ACF.

Définition 1.2.1. (*Relation Binaire*) Une relation binaire \mathcal{R} entre deux ensembles arbitraires \mathcal{X} et \mathcal{Y} est définie dans le produit Cartésien $\mathcal{X} \times \mathcal{Y}$, elle consiste en un ensemble de paires (x, y) avec $x \in \mathcal{X}$ et $y \in \mathcal{Y}$. On note $x\mathcal{R}y$ quand $(x, y) \in \mathcal{R}$.

Quand $\mathcal{X} = \mathcal{Y}$, \mathcal{R} est dite relation binaire sur \mathcal{X} .

Définition 1.2.2. (*Relation d'ordre*) La relation binaire \mathcal{R} sur un ensemble \mathcal{E} est dite relation d'ordre si elle satisfait les conditions suivantes :

1. (*réflexive*) $\forall x \in \mathcal{E} : x\mathcal{R}x$;
2. (*antisymétrique*) $\forall x, y \in \mathcal{E} : x\mathcal{R}y \wedge y\mathcal{R}x \Rightarrow x = y$;
3. (*transitive*) $\forall x, y, z \in \mathcal{E} : x\mathcal{R}y \wedge y\mathcal{R}z \Rightarrow x\mathcal{R}z$.

La relation d'ordre \mathcal{R} est dite totale si $\forall x, y \in \mathcal{E}, x\mathcal{R}y \vee y\mathcal{R}x$, autrement elle dite partielle.

Nous utilisons souvent la notation \leq pour désigner une relation d'ordre.

Définition 1.2.3. (*Ensemble ordonné*) On appelle ensemble ordonné tout couple (\mathcal{E}, \leq) formé d'un ensemble \mathcal{E} et d'une relation d'ordre \leq sur \mathcal{E} .

Définition 1.2.4. (*Supremum, Infimum*) Soit (\mathcal{E}, \leq) un ensemble ordonné et E un sous ensemble de \mathcal{E} (i.e. $E \subseteq \mathcal{E}$). l'élément s est appelé borne supérieure ou **supremum** de E (noté $\bigvee E$) si $\forall e \in E, e \leq s$. Duallement, l'élément i est appelé borne inférieure ou **infimum** de E (noté $\bigwedge E$) si $\forall e \in E, i \leq e$.

Définition 1.2.5. (*Treillis*) Un treillis est un ensemble partiellement ordonné (\mathcal{E}, \leq) tel que $x \vee y$ et $x \wedge y$ existent pour tout couple d'éléments $x, y \in \mathcal{E}$. Un treillis est dit complet si $\bigvee E$ et $\bigwedge E$ existent pour tout sous ensemble E de \mathcal{E} .

Définition 1.2.6. (*Fermeture*) On appelle opérateur de fermeture sur un ensemble ordonné (\mathcal{X}, \leq) toute application $\Phi : \mathcal{X} \rightarrow \mathcal{X}$ vérifiant les propriétés suivantes :

- (CL1) Φ est extensive : $\forall x \in \mathcal{X}, x \leq \Phi(x)$;
- (CL2) Φ est isotone : $\forall x, y \in \mathcal{X}, x \leq y \Rightarrow \Phi(x) \leq \Phi(y)$;
- (CL3) Φ est idempotent : $\forall x \in \mathcal{X}, \Phi(x) = \Phi(\Phi(x))$.

Définition 1.2.7. (*Ouverture*) On appelle opérateur d'ouverture sur un ensemble ordonné (\mathcal{X}, \leq) toute application $\Psi : \mathcal{X} \rightarrow \mathcal{X}$ vérifiant les propriétés suivantes :

- (OP1) Ψ est contractante : $\forall x \in \mathcal{X}, \Psi(x) \leq x$;
- (OP2) Ψ est isotone : $\forall x, y \in \mathcal{X}, x \leq y \Rightarrow \Psi(x) \leq \Psi(y)$;
- (OP3) Ψ est idempotent : $\forall x \in \mathcal{X}, \Psi(x) = \Psi(\Psi(x))$.

Définition 1.2.8. (*Connexion de Galois*) Soient $\Phi : \mathcal{A} \rightarrow \mathcal{B}$ et $\Psi : \mathcal{B} \rightarrow \mathcal{A}$ deux applications définies sur les ensembles ordonnés $(\mathcal{A}, \leq_{\mathcal{A}})$ et $(\mathcal{B}, \leq_{\mathcal{B}})$. Φ et Ψ forment une connexion de Galois (ou une correspondance de Galois) entre $(\mathcal{A}, \leq_{\mathcal{A}})$ et $(\mathcal{B}, \leq_{\mathcal{B}})$ si elles vérifient les conditions suivantes :

1. $\forall a_1, a_2 \in \mathcal{A}, a_1 \leq_{\mathcal{A}} a_2 \Rightarrow \Phi(a_2) \leq_{\mathcal{B}} \Phi(a_1)$;
2. $\forall b_1, b_2 \in \mathcal{B}, b_1 \leq_{\mathcal{B}} b_2 \Rightarrow \Psi(b_2) \leq_{\mathcal{A}} \Psi(b_1)$;
3. $\forall a \in \mathcal{A}, \forall b \in \mathcal{B}, a \leq_{\mathcal{A}} \Psi(\Phi(a)) \wedge b \leq_{\mathcal{B}} \Phi(\Psi(b))$.

1.3 Analyse de Concepts Formels

L'Analyse de concepts formels (ACF) [Wille, 1982; Ganter et Wille, 1999] a émergé dans les années 80 pour tenter de restructurer la théorie des treillis [Birkhoff, 1967] et donner un cadre pratique (ou applicatif) facilitant son utilisation dans les applications du monde réel. S'inscrivant dans le domaine des mathématiques appliquées, ainsi, depuis son avènement, l'ACF a trouvé usage dans plusieurs domaines comme la psychologie, la sociologie, l'anthropologie, la médecine, la biologie, la linguistique, l'informatique et les mathématiques [Wolff, 2003]. L'ACF est centrée autour de la notion de concept, qui est considéré, du point de vue philosophique, comme l'unité de base de la pensée humaine. En partant d'une représentation d'une réalité¹, appelée **contexte formel**, donnée par un ensemble d'objets, un ensemble d'attributs (ou propriétés) de ces objets et une relation binaire reliant ces objets à leurs attributs respectifs, l'ACF vise à extraire, en utilisant des **opérateurs** dédiés, les **concepts formels** sous forme de paire d'objets et de leurs attributs communs. Dans ce qui suit nous définissons ces notions de manière formelle.

1. Souvent cette représentation est partielle.

1.3.1 Contexte Formel

Définition 1.3.1. (*Contexte Formel*) Un **Contexte Formel** est un triplet $\mathcal{K} = (\mathcal{O}, \mathcal{A}, \mathcal{R})$ où \mathcal{O} est un ensemble d'objets, \mathcal{A} est un ensemble d'attributs et \mathcal{R} est une relation binaire entre \mathcal{O} et \mathcal{A} et vérifiant $\mathcal{R} \subseteq \mathcal{O} \times \mathcal{A}$. Un couple $(o, a) \in \mathcal{R}$ (noté aussi $o\mathcal{R}a$) signifie que l'objet o possède l'attribut a .

Nous noterons par $\bar{\mathcal{K}} = (\mathcal{O}, \mathcal{A}, \bar{\mathcal{R}})$ le contexte formel complémentaire de \mathcal{K} , où $\bar{\mathcal{R}}$ est la relation complémentaire de \mathcal{R} . Formellement, $\bar{\mathcal{R}} = \{(o, a) \in \mathcal{O} \times \mathcal{A} \mid (o, a) \notin \mathcal{R}\}$.

Exemple 1.3.1. Le Tableau 1.1 donne un exemple de contexte formel représentant un ensemble d'animaux et certains de leurs attributs (propriétés).

Tableau 1.1 – Exemple de Contexte Formel

\mathcal{R}	Ovipare	Denté	Volant	Aquatique	Mammifère
Lièvre		×			×
Chauve-souris		×	×		×
Canard	×		×	×	
Castor		×		×	×
Tortue	×			×	
Crocodile	×	×		×	

1.3.2 Opérateur de Dérivation et Connexion de Galois

Soit $\mathcal{K} = (\mathcal{O}, \mathcal{A}, \mathcal{R})$ un contexte formel, nous définissons la fonction polymorphe $R : \mathcal{O} \rightarrow 2^{\mathcal{A}}$ et $R : \mathcal{A} \rightarrow 2^{\mathcal{O}}$ par : $R(o) = \{a \in \mathcal{A} \mid o\mathcal{R}a\}$ et $R(a) = \{o \in \mathcal{O} \mid o\mathcal{R}a\}$, où $2^{\mathcal{A}}$ et $2^{\mathcal{O}}$ représentent respectivement l'ensemble des parties de \mathcal{A} et \mathcal{O} . Pour tout $O \subseteq \mathcal{O}$ et $A \subseteq \mathcal{A}$, l'opérateur de dérivation de Galois $(\cdot)'$ est défini par :

$$\begin{aligned} O' &= \{a \in \mathcal{A} \mid \forall o \in \mathcal{O} (o \in O \Rightarrow (o, a) \in \mathcal{R})\} \\ &= \begin{cases} \bigcap_{o \in O} R(o) & \text{si } O \neq \emptyset \\ \mathcal{A} & \text{sinon} \end{cases} \end{aligned} \quad (1.1)$$

$$\begin{aligned} A' &= \{o \in \mathcal{O} \mid \forall a \in \mathcal{A} (a \in A \Rightarrow (o, a) \in \mathcal{R})\} \\ &= \begin{cases} \bigcap_{a \in A} R(a) & \text{si } A \neq \emptyset \\ \mathcal{O} & \text{sinon} \end{cases} \end{aligned} \quad (1.2)$$

Intuitivement, O' est l'ensemble des attributs possédés par tous les objets de O , et A' est l'ensemble des objets possédant tous les attributs de A . Les opérateurs $(.)' : 2^{\mathcal{O}} \rightarrow 2^{\mathcal{A}}$ et $(.)' : 2^{\mathcal{A}} \rightarrow 2^{\mathcal{O}}$ sont appelés opérateurs de dérivation de Galois. Ils sont également appelés opérateurs de suffisance dans [Düntsch et Gediga, 2003; Dubois et al., 2007]. La composition des ces opérateurs $(.)'' : 2^{\mathcal{O}} \rightarrow 2^{\mathcal{O}}$ (resp. $(.)'' : 2^{\mathcal{A}} \rightarrow 2^{\mathcal{A}}$) définit une fermeture sur l'ensemble $2^{\mathcal{O}}$ (resp. $2^{\mathcal{A}}$). Les ensembles O'' et A'' sont des fermés pour les compositions $(.)''$. L'ensemble des fermés O'' de $2^{\mathcal{O}}$ muni de l'inclusion est un treillis complet. Duallement, l'ensemble des fermés A'' de $2^{\mathcal{A}}$ muni de l'inclusion est un treillis complet. Ainsi, $(.)'' : 2^{\mathcal{O}} \rightarrow 2^{\mathcal{O}}$ et $(.)'' : 2^{\mathcal{A}} \rightarrow 2^{\mathcal{A}}$ forment une connexion de Galois entre $(2^{\mathcal{O}}, \subseteq)$ et $(2^{\mathcal{A}}, \subseteq)$.

Exemple 1.3.2. *En considérant le contexte formel du Tableau. 1.1, nous donnons ci-dessous, quelques exemples de l'application de l'opérateur de suffisance ainsi que sur sa composition :*

- $\{\text{Lièvre, Chauve-souris}\}' = \{\text{Denté, Mammifère}\}$
- $\{\text{Lièvre, Chauve-souris, Canard, Castor, Tortue, Crocodile}\}' = \emptyset$
- $\{\text{Ovipare, Denté}\}' = \{\text{Crocodile}\}$
- $\{\text{Ovipare, Denté, Volant, Aquatique, Mammifère}\}' = \emptyset$
- $\{\text{Lièvre, Chauve-souris}\}'' = \{\text{Lièvre, Chauve-souris, Castor}\}$
- $\{\text{Ovipare, Denté}\}'' = \{\text{Ovipare, Denté, Aquatique}\}$

1.3.3 Concept Formel

Après avoir défini un contexte formel et l'opérateur de suffisance, nous donnons ci-après la définition d'un concept formel.

Définition 1.3.2. (*Concept Formel*) *Soit $\mathcal{K} = (\mathcal{O}, \mathcal{A}, \mathcal{R})$ un contexte formel. Un concept formel est une paire $\langle O, A \rangle$ avec $O \subseteq \mathcal{O}$ et $A \subseteq \mathcal{A}$, telle que $O' = A$ et $A' = O$.*

Autrement dit, $\langle O, A \rangle$ est un concept formel si est seulement si O est l'ensemble des objets possédant tous les attributs de l'ensemble A , et A est l'ensemble des attributs possédés par tous les objets de l'ensemble O . Dans ce cas, O est appelé l'extension (*extent*) et A est appelé l'intension (*intent*) du concept formel $\langle O, A \rangle$.

Nous pouvons, également, définir un concept formel visuellement dans la représentation graphique du contexte formel (i.e. le tableau) comme étant le rectangle maximal contenant uniquement des "x" et aucun vide.

Exemple 1.3.3. *En considérant le contexte formel du Tableau. 1.1, nous donnons ci-dessous, quelques exemples de paires $\langle O, A \rangle$ qui sont (ou non) des concepts formels :*

- $\langle \{\text{Lièvre, Chauve-souris, Castor}\}, \{\text{Denté, Mammifère}\} \rangle$ est un concept formel,

- $\langle \{Lièvre, Chauve-souris, Canard, Castor, Tortue, Crocodile\}, \emptyset \rangle$ est un concept formel,
- $\langle \{Canard, Crocodile\}, \{Ovipare, Denté\} \rangle$ n'est pas un concept formel.

Il est aussi important de préciser que les concepts formels peuvent être organisés hiérarchiquement par la relation de subsomption, que nous présentons dans la sous-section suivante.

1.3.4 Treillis de Concepts Formels

Soit $\mathcal{K} = (\mathcal{O}, \mathcal{A}, \mathcal{R})$ un contexte formel, et soient $\langle O_1, A_1 \rangle, \langle O_2, A_2 \rangle$ deux concepts formels de \mathcal{K} . Nous définissons la relation d'ordre partiel \leq , de subsomption, entre les concepts formels par :

$\langle O_1, A_1 \rangle \leq \langle O_2, A_2 \rangle$ si et seulement si $O_1 \subseteq O_2$ (et d'une manière équivalente $A_2 \subseteq A_1$).

Dans ce cas, $\langle O_2, A_2 \rangle$ est dite super-concept de $\langle O_1, A_1 \rangle$, Duallement, $\langle O_1, A_1 \rangle$ est dite sub-concept de $\langle O_2, A_2 \rangle$.

Exemple 1.3.4. À partir du contexte formel du Tableau 1.1, nous avons :

- $\langle \{Chauve-souris\}, \{Denté, Volant, Mammifère\} \rangle$ est un sub-concept de $\langle \{Lièvre, Chauve-souris, Castor\}, \{Denté, Mammifère\} \rangle$.
- $\langle \{Lièvre, Chauve-souris, Castor\}, \{Denté, Mammifère\} \rangle$ est un super-concept de $\langle \{Chauve-souris\}, \{Denté, Volant, Mammifère\} \rangle$

La relation \leq se base sur deux inclusions duales, entre les extensions et les intensions des concepts formels. Elle peut ainsi être interprétée comme une relation de généralisation/spécialisation entre les concepts formels. Un concept est plus général qu'un autre concept s'il contient plus d'objets dans son extension. De façon duale, un concept est plus spécifique qu'un autre s'il contient moins d'objets dans son extension.

L'ensemble de tous les concepts formels d'un contexte formel \mathcal{K} dénoté par $\mathcal{B}(\mathcal{K})$ équipé par la relation d'ordre partiel \leq (c-à-d $(\mathcal{B}(\mathcal{K}), \leq)$) forme un treillis complet, appelé treillis de concepts formels. Le théorème fondamental suivant, montre que le treillis de concepts formels est un treillis complet et caractérise l'infimum et le supremum de sous-ensemble de concepts formels.

Théorème 1.3.1. [Ganter et Wille, 1999]

Soit $\mathcal{K} = (\mathcal{O}, \mathcal{A}, \mathcal{R})$ un contexte formel et $\mathcal{B}(\mathcal{K})$ l'ensemble de tous les concepts formels de \mathcal{K} . Le treillis $(\mathcal{B}(\mathcal{K}), \leq)$ est un treillis complet dans lequel l'infimum et le supremum sont donnés par :

$$\bigwedge_{j \in J} \langle O_j, A_j \rangle = \left\langle \bigcap_{j \in J} O_j, \left(\bigcup_{j \in J} A_j \right)'' \right\rangle$$

$$\bigvee_{j \in J} \langle O_j, A_j \rangle = \left\langle \left(\bigcup_{j \in J} O_j \right)'' , \bigcap_{j \in J} A_j \right\rangle$$

La Figure 1.1 illustre le treillis de concepts formels associé au Tableau.1.1. Les objets Lièvre, Chauve-souris, Canard, Castor, Tortue et Crocodile sont désignés resp. par les objets o_1, o_2, o_3, o_4, o_5 et o_6 . De même, les attributs Ovipare, Denté, Volant, Aquatique et Mammifère sont désignés resp. par a_1, a_2, a_3, a_4 et a_5 . Dans ce dia-

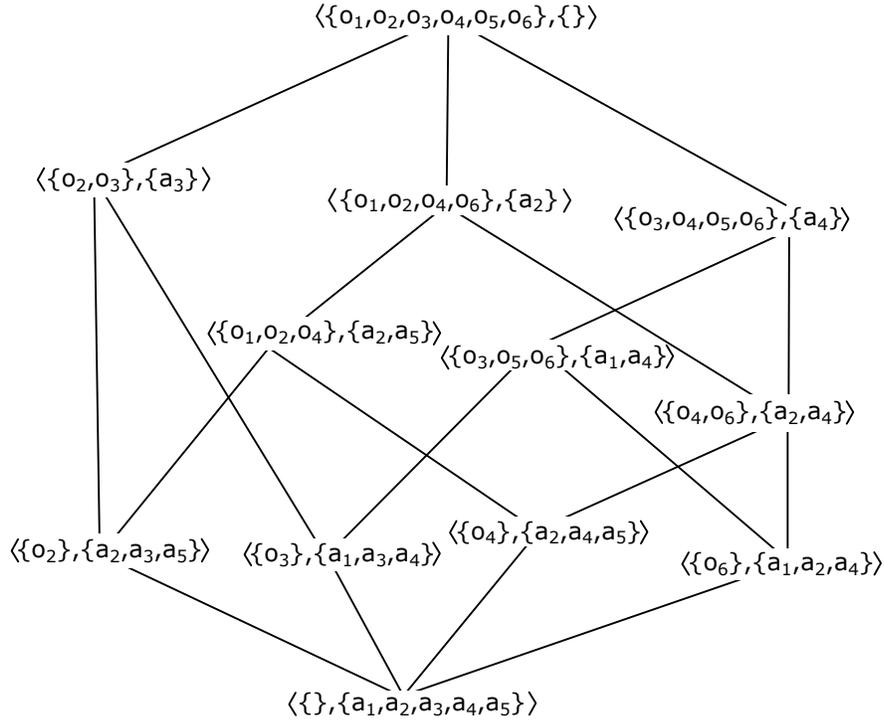


FIGURE 1.1 – Diagramme de Hasse du Treillis de Concepts Formels Associé au Contexte de l'Exemple.1.3.1.

gramme (appelé diagramme de Hasse) chaque noeud dénote un concept formel, quant aux ligne du diagramme, elles dénotent les relations directes entre les noeuds (les concepts formels). La lecture descendante des ces lignes correspond à la relation de super-concept. Duallement la lecture ascendante correspond à la relation de sub-concept. Notons que le treillis de concepts formels est isomorphe au contexte formel. De plus, la représentation graphique du treillis sous forme de diagramme de HASSE, facilite la compréhension et l'interprétation de la relation de subsumption entre les concepts formels ainsi que les relations entre les objets et les attributs au sein du même concept formel.

La construction du treillis de concepts d'un contexte formel peut être décomposée en deux parties. La première concerne le calcul des concepts formels, et la deuxième vise la représentation graphique du treillis ou la construction du diagramme de HASSE correspondant à ce treillis [Guénoche et Mechelen, 1993].

Le problème de calcul des concepts formels² à partir d'un contexte formel a fait l'objet de nombreux travaux de recherche, ainsi, plusieurs algorithmes ont été proposés dans la littérature. Parmi ces travaux, nous citons : Chein [Chein, 1969], Norris [Norris, 1978], NextClosure [Ganter, 1984; Ganter et Reuter, 1991], Bordat [Bordat, 1986], Close-by-One [Kuznetsov, 1993], Godin [Godin et al., 1995], Galois [Carpineto et Romano, 1996], Nourine [Nourine et Raynaud, 1999], Titanic [Stumme et al., 2000, 2002], Divide&Conquer [Valtchev et Missaoui, 2001; Valtchev et al., 2002] et AddIntent [van der Merwe et al., 2004].

1.4 Lecture Possibiliste de l'ACF

Dans une lecture possibiliste de l'ACF, inspirée de la théorie des possibilités [Zadeh, 1978], Dubois et al. [2007] ont mis en évidence l'introduction de nouveaux opérateurs (de dérivation) ensemblistes dans le paradigme de l'ACF. Ainsi, les auteurs ont défini quatre opérateurs ensemblistes Π , Δ , N et ∇ désignant respectivement la possibilité potentielle, la suffisance, la nécessité potentielle et la suffisance duale. Comme l'opérateur de suffisance Δ est le même qui est déjà utilisé en AFC (Dénaté par $(.)'$ dans la Section.1.3.2), nous donnons ci-dessous la définition des trois autres opérateurs.

Définition 1.4.1. Soit $\mathcal{K} = (\mathcal{O}, \mathcal{A}, \mathcal{R})$ un contexte formel. Étant donné un ensemble $O \subseteq \mathcal{O}$, les opérateurs Π , N et ∇ sont définis par :

- O^Π désigne l'ensemble des attributs possédés par au moins un objet $o \in O$:

$$\begin{aligned} O^\Pi &= \{a \in \mathcal{A} \mid \exists o \in O, o\mathcal{R}a\} \\ &= \{a \in \mathcal{A} \mid O \cap R(a) \neq \emptyset\} \end{aligned} \quad (1.3)$$

- O^N désigne l'ensemble des attributs possédés uniquement par les objets de O :

$$\begin{aligned} O^N &= \{a \in \mathcal{A} \mid \forall o \in \mathcal{O} (o\mathcal{R}a \Rightarrow o \in O)\} \\ &= \{a \in \mathcal{A} \mid R(a) \subseteq O\} \end{aligned} \quad (1.4)$$

- O^∇ désigne l'ensemble des attributs qui ne sont possédés par aucun objet de O :

$$\begin{aligned} O^\nabla &= \{a \in \mathcal{A} \mid O \cup R(a) \neq \mathcal{O}\} \\ &= \{a \in \mathcal{A} \mid \exists o \in \mathcal{O} \setminus O, o\overline{\mathcal{R}}a\} \end{aligned} \quad (1.5)$$

2. Cette partie peut être décomposée en deux étapes [Guénoche et Mechelen, 1993], 1- l'énumération des rectangles maximaux (les fermés), et 2- la recherche de la relation d'ordre partiel entre ces rectangles.

Où $\mathcal{O} \setminus O$ est le complément³ de l'ensemble O et $\overline{\mathcal{R}}$ est le complémentaire de \mathcal{R} (§1.3.1).

A^Π , A^N et A^∇ sont duallement définis.

Exemple 1.4.1. *En considérant le contexte formel du Tableau 1.1, nous donnons ci-après des exemples de dérivations des opérateurs Δ , Π , N et ∇ :*

- $\{\text{Canard, Tortue, Crocodile}\}^\Delta = \{\text{Ovipare, Aquatique}\}$
- $\{\text{Canard, Tortue, Crocodile}\}^\Pi = \{\text{Ovipare, Denté, Volant, Aquatique}\}$
- $\{\text{Canard, Tortue, Crocodile}\}^N = \{\text{Ovipare}\}$
- $\{\text{Canard, Tortue, Crocodile}\}^\nabla = \{\text{Ovipare, Denté, Volant, Aquatique}\}$

Nous terminons cette section en rappelant certaines propriétés des opérateurs Δ , Π , N et ∇ .

Proposition 1.4.1. *[Djouadi et Prade, 2011] Soit $\mathcal{K} = (\mathcal{O}, \mathcal{A}, \mathcal{R})$ un contexte formel, pour tous ensembles $X, Y \subseteq \mathcal{O}$ (et d'une manière équivalente $X, Y \subseteq \mathcal{A}$), les propriétés suivantes sont vérifiées :*

$$(P.1.1) : (X \cup Y)^\Pi = X^\Pi \cup Y^\Pi.$$

$$(P.1.2) : (X \cap Y)^\Pi \subseteq X^\Pi \cap Y^\Pi.$$

$$(P.1.3) : (X \cup Y)^N \supseteq X^N \cup Y^N.$$

$$(P.1.4) : (X \cap Y)^N = X^N \cap Y^N.$$

En plus nous avons :

$$(P.1.5) : ((.)^\Pi)^N \text{ est un opérateur de fermeture.}$$

$$(P.1.6) : ((.)^N)^\Pi \text{ est un opérateur d'ouverture.}$$

$$(P.1.7) : (.)^\Pi \text{ et } (.)^N \text{ sont des opérateurs isotones, c-à-d } X \subseteq Y \Leftrightarrow X^\Pi \subseteq Y^\Pi \Leftrightarrow X^N \subseteq Y^N.$$

La preuve de ces propriétés peut être trouvée dans [Djouadi et Prade, 2011].

1.5 Limites de l'ACF Classique

Bien que l'ACF offre un cadre efficient pour la representation de connaissances et de nombreux domaines d'application en font utilisation, il reste que son paradigme présente des limites qui ont ouvert des perspectives d'amélioration et d'extension de l'ACF.

3. “\” est la complémentarité absolue des ensembles. $\mathcal{O} \setminus O$ est constitué de tous les éléments de \mathcal{O} n'appartenant pas à O .

L'AFC est centrée autour de la notion de concept, qui est un regroupement d'individus et leurs attributs communs. Toutefois, dans la réalité ce regroupement peut être sujet à certains critères. D'où la nécessité de considérer des contextes formels comportant plusieurs axes. Dans ce contexte, [Lehmann et Wille \[1995\]](#) ont proposé de considérer un troisième axe dans le contexte formel pour représenter les conditions, ce qui a donné naissance à l'ACF triadique. L'un des exemple de ces relations ternaires est celui des index personnalisés du système Bibsonomy [[Benz et al., 2010](#)]. Ce système organise les index des utilisateurs selon une relation ternaire. Cette relation relie les éléments de trois ensembles distincts, à savoir, 1) les utilisateurs, 2) les marque-pages ou signets (*bookmarks*) et 3) les sujets (*topics ou tags*).

Un second constat est que l'ACF classique est limitée à la représentation de relations Booléennes. Cependant, certaines applications concrètes, où les données d'entrée des contextes formels sont obtenues par des mesures, des observations et des jugements humains, peuvent donner lieu aux notions d'imprécision, d'incertitude et de gradualité, et ont conduit à considérer des contextes formels floue, où les relations objet-attribut sont des valeurs de vérité à partir de l'échelle $L = [0, 1]$ qui estime dans quelle mesure un objet possède un attribut.

En ce qui concerne notre domaine d'application (i.e. la recherche d'information), une façon plus rigoureuse de représenter la relation entre les documents et les termes d'indexation est de considérer les fréquences d'apparition ou les poids des termes dans les documents au lieu de la simple information binaire indiquant la présence ou l'absence d'un terme dans un document.

Le reste de ce chapitre sera consacré aux extensions triadique et floue de l'ACF.

1.6 Extension Triadique de l'ACF

L'approche triadique de l'AFC a été introduite par [Lehmann et Wille \[1995\]](#). Elle formalise une relation ternaire entre des objets, des attributs et des conditions. Le contexte formel triadique est un quadruplet $(\mathcal{O}, \mathcal{A}, \mathcal{C}, \mathcal{R}_3)$, où \mathcal{O} est l'ensemble des objets, \mathcal{A} est l'ensemble des attributs, \mathcal{C} est l'ensemble des conditions, et \mathcal{R}_3 est une relation ternaire entre \mathcal{O} , \mathcal{A} et \mathcal{C} ($\mathcal{R}_3 \subseteq \mathcal{O} \times \mathcal{A} \times \mathcal{C}$). Un élément $(o, a, c) \in \mathcal{R}_3$ signifie que l'objet o possède l'attribut a sous la condition c (ou relativement à la condition c).

Exemple 1.6.1. *Nous illustrons dans le Tableau. 1.2 un exemple de contexte formel triadique*

Pour définir les opérateurs de dérivation dans l'ACF triadique, [Wille \[1995\]](#) a considéré la représentation sous-jacente alternative d'un contexte formel triadique donnée par $\mathcal{K}_t = (\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3, \mathcal{Y})$. Pour $\{i, j, k\} = \{1, 2, 3\}$ avec $j < k$ et pour $X \subseteq \mathcal{K}_i$

Tableau 1.2 – Exemple de Contexte Formel Triadique

\mathcal{R}_3	c_1			c_2			c_3			c_4		
	a_1	a_2	a_3									
o_1	×	×	×	×		×	×	×	×	×	×	×
o_2	×		×	×		×		×		×		×
o_3			×			×						
o_4			×			×				×		
o_5			×			×		×			×	
o_6			×		×	×	×		×			×

et $Z \subseteq \mathcal{K}_j \times \mathcal{K}_k$, les (i) –opérateurs de dérivation sont définis par :

$$X^{(i)} = \{(a_j, a_k) \in \mathcal{K}_j \times \mathcal{K}_k \mid \forall a_i \in X, (a_i, a_j, a_k) \in \mathcal{Y}\}$$

$$Z^{(i)} = \{a_i \in \mathcal{K}_i \mid \forall (a_j, a_k) \in Z, (a_i, a_j, a_k) \in \mathcal{Y}\}$$

Cette définition produit les opérateurs de dérivation des contextes formels dyadiques⁴ définis par :

$$\mathcal{K}_i^{(1)} = (\mathcal{K}_1, \mathcal{K}_2 \times \mathcal{K}_3, \mathcal{Y}^{(1)})$$

$$\mathcal{K}_i^{(2)} = (\mathcal{K}_2, \mathcal{K}_1 \times \mathcal{K}_3, \mathcal{Y}^{(2)})$$

$$\mathcal{K}_i^{(3)} = (\mathcal{K}_3, \mathcal{K}_1 \times \mathcal{K}_2, \mathcal{Y}^{(3)})$$

Où $a_1 \mathcal{Y}^{(1)}(a_2, a_3) \Leftrightarrow a_2 \mathcal{Y}^{(2)}(a_1, a_3) \Leftrightarrow a_3 \mathcal{Y}^{(3)}(a_1, a_2) \Leftrightarrow (a_1, a_2, a_3) \in \mathcal{Y}$

La construction de concepts formels triadiques nécessite d'autres opérateurs de dérivation [Wille, 1995]. Pour $\{i, j, k\} = \{1, 2, 3\}$ et pour $X_i \subseteq \mathcal{K}_i$, $X_j \subseteq \mathcal{K}_j$, et $A_k \subseteq \mathcal{K}_k$, le (i, j, A_k) –opérateur de dérivation est défini par :

$$X_i^{(i,j,A_k)} = \{a_j \in \mathcal{K}_j \mid \forall (a_i, a_k) \in X_i \times A_k, (a_i, a_j, a_k) \in \mathcal{Y}\}$$

$$X_j^{(i,j,A_k)} = \{a_i \in \mathcal{K}_i \mid \forall (a_j, a_k) \in X_j \times A_k, (a_i, a_j, a_k) \in \mathcal{Y}\}$$

Cette définition produit les opérateurs de dérivation des contextes dyadiques définis par :

$$\mathcal{K}_{A_k}^{(ij)} = (\mathcal{K}_i, \mathcal{K}_j, \mathcal{Y}_{A_k}^{ij})$$

Avec, $(a_i, a_j) \in \mathcal{Y}_{A_k}^{ij}$ si et seulement si $\forall a_k \in A_k, (a_i, a_j, a_k) \in \mathcal{Y}$.

Pour calculer un concept formel triadique à partir d'un ensemble d'objets X_1 (comme extension) par rapport à un ensemble de conditions A_3 , nous générons

4. Nous utilisons le terme dyadique pour désigner le contexte formel classique contenant uniquement les objets et les attributs

d'abord un concept formel dyadique dans le contexte $\mathcal{K}_{A_3}^{12}$ en considérant l'extension X_1 , puis, étendre ce concept dyadique au concept formel triadique en utilisant (3)–opérateur de dérivation correspondant dans $\mathcal{K}^{(3)}$. Ceci est formellement représenté par :

$$(X_1^{(1,2,A_k)(1,2,A_k)}, X_1^{(1,2,A_k)}, (X_1^{(1,2,A_k)(1,2,A_k)} \times X_1^{(1,2,A_k)})^{(3)})$$

Un triplet $\langle C_1, C_2, C_3 \rangle$, avec $C_i \subseteq \mathcal{K}_i$ pour $i \in \{1, 2, 3\}$, est un concept formel triadique si et seulement si $C_i = (A_j \times A_k)^{(i)}$ pour tous $\{i, j, k\} = \{1, 2, 3\}$ et $j < k$. Notons, également, que les trois triplets $\langle (\mathcal{K}_2 \times \mathcal{K}_3)^{(1)}, \mathcal{K}_2, \mathcal{K}_3 \rangle$, $\langle \mathcal{K}_1, (\mathcal{K}_1 \times \mathcal{K}_3)^{(2)}, \mathcal{K}_3 \rangle$ et $\langle \mathcal{K}_1, \mathcal{K}_2, (\mathcal{K}_1 \times \mathcal{K}_2)^{(3)} \rangle$ sont des concepts formels triadiques appelés “*les concepts triadiques extremes*”. L'ensemble de tous les concepts formels triadiques induit à partir du contexte formel triadique $\mathcal{K}_t = (\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3, \mathcal{Y})$ est structuré par les relation de préordre⁵ \lesssim_i et ses classes d'équivalence \sim_i données par :

$$\langle A_1, A_2, A_3 \rangle \lesssim_i \langle B_1, B_2, B_3 \rangle \Leftrightarrow A_i \subseteq B_i \text{ et}$$

$$\langle A_1, A_2, A_3 \rangle \sim_i \langle B_1, B_2, B_3 \rangle \Leftrightarrow A_i = B_i \quad (i = 1, 2, 3)$$

1.7 Extension Floue de l'ACF

Avant de présenter les extensions floues de l'ACF, nous rappelons d'abord des notions de base sur les ensembles flous et la logique sous-jacente.

1.7.1 Ensembles Flous et Logique Floue

La théorie des ensembles flous à été introduite par Zadeh [1965] dans son papier fondateur “*Fuzzy Sets*”. Il définit un ensemble flou comme une collection telle que l'appartenance d'un élément à cette collection peut prendre un degré varié de 0 à 1. Contrairement aux ensembles classiques ou crisp (théorie de Cantor) ou l'appartenance ne peut prendre que deux valeurs 0 ou 1. Entre autre, cette théorie permet de représenter mathématiquement l'imprécision relative à certaines classes d'objets et sert de fondement à la logique floue. La base de cette théorie est l'ensemble flou que nous définissons ci-dessous.

Définition 1.7.1. (*Ensembles Flous*) Soit \mathcal{U} un univers de discours, un ensemble flou \tilde{F} dans \mathcal{U} est défini par la fonction d'appartenance $\mu_{\tilde{F}}$, qui associe pour chaque element $x \in \mathcal{U}$ son degré d'appartenance $\mu_{\tilde{F}}(x)$ à l'ensemble \tilde{F} .

5. Une relation de préordre est une relation réflexive et transitive

Notation : Dans la suite de cette thèse, pour un univers de discours $\mathcal{U} = \{x_i, i = 1 \dots n\}$ et un sous-ensemble flou \tilde{F} dans \mathcal{U} , nous utiliserons la notation $\tilde{F} = \{x_1^{\alpha_1}, x_2^{\alpha_2}, \dots, x_i^{\alpha_i}, \dots\}$, où $\alpha_i = \mu_{\tilde{F}}(x_i), \forall x_i \in \mathcal{U}$.

Les opérations usuelles sur les ensembles classiques ont été généralisées aux ensemble flous. Ces généralisations sont basées sur la fonction d'appartenance et deux opérateurs flous; la norme triangulaire (t-norme, notée \top) et la conorme triangulaire (t-conorme, noté \perp) avec, respectivement, une sémantique de conjonction et de disjonction. Les deux définitions qui suivent rappellent les propriétés algébriques de ces opérateurs. Des exemples de ces opérateurs sont donnés dans le Tableau. 1.3.

Définition 1.7.2. (*T-norme*) Une t-norme (\top), ou norme triangulaire, est une fonction $\top : [0, 1] \times [0, 1] \rightarrow [0, 1]$ telle que pour tous $x, y, z, t \in [0, 1]$:

- $\top(x, y) = \top(y, x)$ (commutativité),
- $\top(x, \top(y, z)) = \top(\top(x, y), z)$ (associativité),
- $\top(x, y) \leq \top(z, t)$ si $x \leq z$ et $y \leq t$ (monotonie),
- $\top(x, 1) = x$ (1 est l'élément neutre).

Définition 1.7.3. (*T-conorme*) Une t-conorme (\perp), ou conorme triangulaire, est une fonction $\perp : [0, 1] \times [0, 1] \rightarrow [0, 1]$ telle que pour tous $x, y, z, t \in [0, 1]$:

- $\perp(x, y) = \perp(y, x)$ (commutativité),
- $\perp(x, \perp(y, z)) = \perp(\perp(x, y), z)$ (associativité),
- $\perp(x, y) \leq \perp(z, t)$ si $x \leq z$ et $y \leq t$ (monotonie),
- $\perp(x, 0) = x$ (0 est l'élément neutre).

Les t-normes (resp. t-conormes) permettent la généralisation de l'opération ensembliste d'intersection (resp. d'union) des ensembles classiques sur les ensembles flous. Nous présentons ci-dessous la généralisation floue de l'intersection, l'union et le complément :

Définition 1.7.4. (*Généralisation des Opérations Ensemblistes*) Soit \mathcal{U} un univers de discours, et soient \tilde{F} et \tilde{G} deux ensembles flous définis dans \mathcal{U} . Les opérations ensemblistes floues "intersection", "union" et complément sont définies par :

- Intersection : $\forall x \in \mathcal{U}, \mu_{\tilde{F} \cap \tilde{G}}(x) = \mu_{\tilde{F}}(x) \top \mu_{\tilde{G}}(x)$
- Union : $\forall x \in \mathcal{U}, \mu_{\tilde{F} \cup \tilde{G}}(x) = \mu_{\tilde{F}}(x) \perp \mu_{\tilde{G}}(x)$
Où \top (resp. \perp) désigne une t-norme (resp. t-conorme).
- Complément : $\forall x \in \mathcal{U}, \mu_{\tilde{F}^c}(x) = 1 - \mu_{\tilde{F}}(x)$

Au delà de la généralisation de l'union et de l'intersection des ensembles, l'égalité et l'inclusion ont été, également, généralisées sur les ensembles flous, elles sont données par :

Tableau 1.3 – Exemples de T-normes et de T-conormes

	t-norme : $\top(x, y)$	t-conorme : $\perp(x, y)$
Zadeh	$Min(x, y)$	$Max(x, y)$
Lukasiewicz	$Max(x + y - 1, 0)$	$Min(x + y, 1)$
Probabiliste	$x \cdot y$	$x + y - x \cdot y$
Weber	$\begin{cases} a & \text{si } b = 1 \\ b & \text{si } a = 1 \\ 0 & \text{sinon} \end{cases}$	$\begin{cases} a & \text{si } b = 0 \\ b & \text{si } a = 0 \\ 1 & \text{sinon} \end{cases}$

Tableau 1.4 – Les S-implications Usuelles

Nom	Notation	Valeur de vérité	t-conorme sous-jacente
Kleene-Dienes	\rightarrow_{K-D}	$Max(1 - p, q)$	$p \perp q = Max(p, q)$
Reichenback	\rightarrow_{Rb}	$1 - p + p \cdot q$	$p \perp q = p + q - p \cdot q$
Lukaseiwicz	\rightarrow_{Lu}	$Min(1 - p + q, 1)$	$p \perp q = Min(p + q, 1)$

Égalité floue : $\tilde{F} = \tilde{G} \Leftrightarrow \forall x \in \mathcal{U}, \mu_{\tilde{F}}(x) = \mu_{\tilde{G}}(x)$,

Inclusion floue : $\tilde{F} \subseteq \tilde{G} \Leftrightarrow \forall x \in \mathcal{U}, \mu_{\tilde{F}}(x) \leq \mu_{\tilde{G}}(x)$.

Dans la logique classique, une proposition atomiques (p) et l'implication entre deux propositions $p \Rightarrow q$ ne prend que deux valeurs Booléennes, *vraie* ou *fausse*. Dans la logique floue, les propositions atomiques et les proposition composées par des implications floues prennent des valeurs de vérité dans l'intervalle $[0, 1]$. Dans la littérature, on distingue principalement deux familles d'implications floues, les r-implications et les s-implications, que nous présentons ci-après. Ces implications sont des fonctions (\rightarrow) définies par :

$$\begin{aligned} \rightarrow : [0, 1] \times [0, 1] &\rightarrow [0, 1] \\ (p, q) &\mapsto (p \rightarrow q) \end{aligned}$$

L'appellation s-implication vient de l'anglais (*strong implication*). Les implications floues de cette famille sont définies à partir de la généralisation de l'expression $((\neg p) \vee q)$ qui est donnée par : $p \rightarrow q = (1 - p) \perp q$, où la disjonction " \vee " est généralisée par une conorme triangulaire. Dans la Table 1.4, nous donnons des exemples de s-implications.

La deuxième famille d'implications floues englobe les r-implications. Cette dénomination vient du fait que ces implications vérifiées le principe de *résiduation*. Pour rappel, un couple (\rightarrow, \top) est dit *residue* si est seulement si $\forall x, y, z \in [0, 1], x \top y \leq$

Tableau 1.5 – Les R-implications Usuelles

Nom	Notation	Valeur de vérité	t-norme sous-jacente
Gödel	\rightarrow_{Go}	$\begin{cases} 1 & \text{si } p \leq q \\ q & \text{sinon} \end{cases}$	$p \top q = \text{Min}(p, q)$
Goguen	\rightarrow_{Gg}	$\begin{cases} 1 & \text{si } p \leq q \\ q/p & \text{sinon} \end{cases}$	$p \top q = p \cdot q$
Lukaseiwicz	\rightarrow_{Lu-R}	$\begin{cases} 1 & \text{si } p \leq q \\ 1 - p + q & \text{sinon} \end{cases}$	$p \top q = \text{Max}(p + q - 1, 0)$

$z \Leftrightarrow y \leq x \rightarrow z$. Les r-implications sont définies comme suit :

$$p \rightarrow q = \text{Sup}\{x \in [0, 1] \mid p \top x \leq q\}$$

Nous donnons dans la Table 1.5 quelques exemples de r-implications.

Comme les implications floues généralisent l'implication matérielle (\Rightarrow) ou l'inclusion (\supset), la valeur de vérité de $p \rightarrow q$, quelque soit l'implication floue \rightarrow considérée, est équivalente à celle de $p \Rightarrow q$ quand les propositions p et q prennent les valeurs de vérité dans $\{1, 0\}$.

1.7.2 ACF Floue

L'ACF consiste à induire à partir d'un contexte formel binaire des concepts formels $\langle \text{Objets}, \text{Attributs} \rangle$. Ces concepts formels se basent sur la formule suivante : "Un objet appartient au sous-ensemble *Objets* s'il possède tous les attributs du sous-ensemble *Attributs*". Et comme la relation considérée dans le contexte formel est Booléenne, un objet possède totalement un attribut ou ne le possède pas, qui est un cadre restrictif. En effet, dans la réalité l'ACF est amenée à considérer des contextes formels de natures diverses, obtenus par des mesures, des observations, des jugements,...etc, où la relation entre un objet et un attribut n'est plus Booléenne (0 ou 1) mais elle peut être incertaine, graduelle ou imprécise.

Diverses approches ont été proposées dans la littérature pour étendre l'ACF au cas flou. Le point commun entre ces approches est la notion de contexte formel flou. Ces approches se basent sur des structures algébriques de la forme $\mathcal{L} = (L, \dots)$, où L est l'intervalle des degrés de vérité, généralement $L = [0, 1]$. Ainsi, le contexte formel flou (appelé aussi L -contexte) est donné par le triplet $\tilde{\mathcal{K}} = (\mathcal{O}, \mathcal{A}, \mathcal{I})$, où \mathcal{O} et \mathcal{A} sont respectivement les ensembles d'objets et d'attributs, et $\mathcal{I} : \mathcal{O} \times \mathcal{A} \rightarrow L$ est une relation floue (appelé aussi L -relation) tel que $\mathcal{I} \subseteq L^{\mathcal{O} \times \mathcal{A}}$. Ainsi, la valeur de $\mathcal{I}(o, a) \in L$ est interprétée comme le degré pour lequel l'objet o possède l'attribut a . Nous présentons ci-après l'ensemble de ces approches.

1.7.2.1 L'approche de Burusco et Fuentes-González

La première extension floue de l'ACF a été proposée par [Burusco et Fuentes-González \[1994\]](#). Les auteurs considèrent la structure algébrique $\mathcal{L} = (L, \leq, ', \perp, 0, 1)$ tel que $(L, \leq, 0, 1)$ est un treillis complet délimité par l'infimum 0 et le supremum 1, $'$ est l'opérateur unaire de complémentation (i.e. $\forall x \in L, x' = 1 - x$) et \perp est une t-conorme dans L . Pour un L -contexte $(\mathcal{O}, \mathcal{A}, \mathcal{I})$, les auteurs définissent les deux images : $\uparrow : L^{\mathcal{O}} \rightarrow L^{\mathcal{A}}$ et $\downarrow : L^{\mathcal{A}} \rightarrow L^{\mathcal{O}}$, pour $O \in L^{\mathcal{O}}$ et $A \in L^{\mathcal{A}}$ par :

$$O^{\uparrow}(a) = \bigwedge_{o \in \mathcal{O}} (O(o)' \perp \mathcal{I}(o, a))$$

$$A^{\downarrow}(o) = \bigwedge_{a \in \mathcal{A}} (A(a)' \perp \mathcal{I}(o, a))$$

Les auteurs ont montré que l'ensemble $\mathcal{B}(\mathcal{O}, \mathcal{A}, \mathcal{I}) = \{\langle O, A \rangle \in L^{\mathcal{O}} \times L^{\mathcal{A}} \mid O^{\uparrow} = A, A^{\downarrow} = O\}$ muni de la relation d'ordre \leq (\leq est définie par : $\langle O_1, A_1 \rangle \leq \langle O_2, A_2 \rangle$ ssi $O_1 \subseteq O_2$ (ssi $B_2 \subseteq B_1$)) est un treillis complet. Cependant, la composition des opérateurs de dérivation \uparrow et \downarrow ne préservent pas la fermeture, c-à-d $O \subseteq O^{\uparrow\downarrow}$ et $A \subseteq A^{\downarrow\uparrow}$, qui est une propriété fondamentale dans l'analyse de concepts formels classique. Toutefois, pour $L = \{0, 1\}$, $'$ est la négation classique ($1' = 0$ et $0' = 1$) et \perp correspond à la disjonction ($\forall x, y \in L, x \perp y = \text{Max}(x, y)$). Le treillis $(\mathcal{B}(\mathcal{O}, \mathcal{A}, \mathcal{I}), \leq)$ est identique au treillis classique de concepts formels .

1.7.2.2 L'approche de Pollandt et Bělohlávek

[Pollandt \[1997\]](#) et, indépendamment, [Bělohlávek \[1998\]](#) ont élaboré une approche basée sur une structure algébrique vérifiant la propriété de résiduation. L'algèbre⁶ considérée dans cette approche est donnée par la structure $\mathcal{L} = (L, \wedge, \vee, \otimes, \rightarrow, 0, 1)$, où $(L, \wedge, \vee, 0, 1)$ est un treillis complet avec 1 et 0 sont respectivement le plus grand majorant et le plus petit minorant de L ; $(L, \otimes, 1)$ est un monoïde commutatif (\otimes est commutative, associative, et $\forall a \in L, 1 \otimes a = a \otimes 1 = a$); La t-norme \otimes et l'implication floue \rightarrow satisfont le principe de résiduation ($\forall x, y, z \in L, x \otimes y \leq z$ ssi $x \leq y \rightarrow z$). En considérant le L -contexte $(\mathcal{O}, \mathcal{A}, \mathcal{I})$, pour les ensembles flous $O \in L^{\mathcal{O}}$ et $B \in L^{\mathcal{A}}$, les auteurs définissent l'extension de l'opérateur de dérivation $(.)'$ par :

$$O^{\uparrow}(a) = \bigwedge_{o \in \mathcal{O}} (O(o) \rightarrow \mathcal{I}(o, a))$$

$$A^{\downarrow}(o) = \bigwedge_{a \in \mathcal{A}} (A(a) \rightarrow \mathcal{I}(o, a))$$

6. Une algèbre est une structure formée d'un ensemble combiné à une ou plusieurs lois de composition et éventuellement complétées par une relation ordre.

Dans [Bělohlávek et Vychodil, 2005], $O^\uparrow(a)$ est interprété comme le degré de vérité de l'expression “l'attribut ' a ' est possédé par tous les objets de O ” et duallement, $A^\downarrow(o)$ est interprété comme le degré de vérité de l'expression “l'objet ' o ' possède tous les attributs de A ”. Contrairement à l'approche de Burusco et Fuentes-González, cette approche préserve les fermetures $O \subseteq O^{\uparrow\downarrow}$ et $A \subseteq A^{\downarrow\uparrow}$. Elle préserve, également, les relations super-concept et sub-concept définies dans l'ACF classique. Le théorème fondamental de l'ACF (§1.3.1) a été généralisé au L -contexte dans [Bělohlávek, 2001, 2004]. Remarquant aussi, que pour $L = \{0, 1\}$, le treillis obtenu (l'ensemble $\mathcal{B}(\mathcal{O}, \mathcal{A}, \mathcal{I}) = \{\langle O, A \rangle \in L^{\mathcal{O}} \times L^{\mathcal{A}} \mid O^\uparrow = A, A^\downarrow = O\}$ muni de la relation d'ordre \leq) est identique à celui de l'ACF classique.

1.7.2.3 Les approches de Yahia & Jaoua et Krajčí

Les approches proposées indépendamment par Yahia et Jaoua [2001] et Krajčí [2003] sont similaires. Les auteurs proposent une paire (f, h) d'opérateurs de dérivation, définis tels que :

$$\begin{aligned} f(O)a &= \bigwedge_{o \in O} \mathcal{I}(o, a) \\ h(\tilde{A}) &= \{o \in \mathcal{O} \mid \forall a \in \mathcal{A}, \tilde{A}(a) \leq \mathcal{I}(o, a)\} \end{aligned}$$

Il en résulte des concepts formels où soit l'extension est un ensemble classique (crip) et l'intension est un ensemble flou, soit l'extension est un ensemble flou et l'intension est un ensemble classique (crisp).

L'ensemble $\mathcal{B}(\mathcal{O}, \mathcal{A}, \mathcal{I}) = \{\langle O, \tilde{A} \rangle \in 2^{\mathcal{O}} \times L^{\mathcal{A}} \mid f(O) = \tilde{A} \text{ et } h(\tilde{A}) = O\}$ muni de la relation d'ordre partiel \leq est appelé treillis de concepts formels semi-flous⁷, qui est aussi un treillis complet. Les concepts ainsi calculés ont des extensions classiques (crisp) et des intensions floues.

1.7.2.4 L'approche de génération classique des concepts flous

Afin de réduire le nombre (qui est souvent très important) de concepts formels des treillis $\mathcal{B}(\mathcal{O}, \mathcal{A}, \mathcal{I})$ relatifs aux approches de Pollandt [1997] et Bělohlávek [1998], l'approche développée dans [Bělohlávek et al., 2005] propose de considérer une partie du $\mathcal{B}(\mathcal{O}, \mathcal{A}, \mathcal{I})$, notée par $\mathcal{B}_c(\mathcal{O}, \mathcal{A}, \mathcal{I})$, qui consiste en l'ensemble de concepts formels flous $\langle \tilde{O}, \tilde{A} \rangle$ généré à partir des ensembles d'attributs ordinaires (ou classiques). C'est à dire, $\langle \tilde{O}, \tilde{A} \rangle \in \mathcal{B}_c(\mathcal{O}, \mathcal{A}, \mathcal{I})$ ssi $\exists A_c \subseteq \mathcal{A}$, $\tilde{O} = A_c^\downarrow$ et $\tilde{A} = A_c^{\downarrow\uparrow}$. Remarquant que par rapport aux approches de Yahia et Jaoua [2001] et Krajčí [2003], où les extensions sont des ensembles flous et les intensions sont des ensembles discrets, cette approche génère des concepts formels avec des extensions et des intensions floues.

7. One-sided fuzzy concept lattice.

1.7.2.5 Les approches à seuil

Dans plusieurs approches, les auteurs proposent des discrétisations du contexte formel flou. L'idée générale est d'obtenir un (des) contexte(s) formel(s) binaire(s) à partir d'un contexte flou, et de réduire le nombre de concepts formels dans le treillis flou. Nous présentons dans ce paragraphe les deux principales approches.

L'approche proposé dans [D \acute{u} ráková et al., 2002] est basé sur les α -coupes, elle consiste à considérer, pour un L -contexte $(\mathcal{O}, \mathcal{A}, \mathcal{I})$, les contextes formels binaires $(\mathcal{O}, \mathcal{A}, {}^\alpha\mathcal{I})$ pour des valeurs $\alpha \in C$, tel que C est un sous-ensemble dans L . Pour une valeur donnée α , la relation ${}^\alpha\mathcal{I}$ est donnée par : $\{(o, a) \mid \mathcal{I}(o, a) \geq \alpha\}$. De cette manière, le contexte formel $(\mathcal{O}, \mathcal{A}, {}^\alpha\mathcal{I})$ obtenu est binaire. En appliquant l'ACF classique sur ce contexte on obtient le treillis classique $\mathcal{B}(\mathcal{O}, \mathcal{A}, {}^\alpha\mathcal{I})$. L'ensemble des treillis $\mathcal{B}(\mathcal{O}, \mathcal{A}, {}^\alpha\mathcal{I})$ relative à un sous-ensemble C est fusionné dans un nouveau treillis. L'approche proposé dans [Zhao et al., 2007], considère deux bornes b_{inf} et b_{sup} formant un intervalle fermé dans L , elle transforme les valeurs des relations floues comprises entre b_{inf} et b_{sup} en ' \times '. La relation binaire ainsi obtenue est donnée par : $\{(o, a) \mid \mathcal{I}(o, a) \geq b_{inf} \wedge \mathcal{I}(o, a) \leq b_{sup}\}$.

1.7.2.6 Approche de Georgescu et Popescu

Une approche, inspirée de l'approche de Pollandt et Bělohlávek, a été proposée par Georgescu et Popescu [2004]. Les auteurs proposent l'utilisation d'une algèbre $(L, \wedge, \vee, *, \overset{G}{\rightarrow}, \overset{D}{\rightarrow}, 0, 1)$ non commutative. Ils considèrent alors, une conjonction (t-norme) $*$ non commutative, et deux implications résiduées, une à gauche ($\overset{G}{\rightarrow}$) et une autre à droite ($\overset{D}{\rightarrow}$) vérifiant respectivement les propriétés suivantes $\forall x, y$ et $z \in L$:

$$x \leq y \overset{G}{\rightarrow} z \text{ ssi } y * x \leq z$$

$$x \leq y \overset{D}{\rightarrow} z \text{ ssi } x * y \leq z$$

Les auteurs définissent deux opérateurs de dérivation sur l'ensemble des objets (respectivement attributs) comme suit :

$$\uparrow, \uparrow: L^{\mathcal{O}} \rightarrow L^{\mathcal{A}}$$

$$\tilde{\mathcal{O}}^\uparrow(a) = \bigwedge_{o \in \mathcal{O}} (\tilde{\mathcal{O}}(o) \overset{G}{\rightarrow} I(o, a))$$

$$\tilde{\mathcal{O}}^\uparrow(a) = \bigwedge_{o \in \mathcal{O}} (\tilde{\mathcal{O}}(o) \overset{D}{\rightarrow} I(o, a))$$

$$\downarrow, \downarrow: L^{\mathcal{A}} \rightarrow L^{\mathcal{O}}$$

$$\tilde{\mathcal{A}}^\downarrow(o) = \bigwedge_{a \in \mathcal{A}} (\tilde{\mathcal{A}}(a) \overset{G}{\rightarrow} I(o, a))$$

$$\tilde{A}^\downarrow(o) = \bigwedge_{a \in \mathcal{A}} (\tilde{A}(a) \xrightarrow{D} I(o, a))$$

Un concept formel flou est défini par ces auteurs par un triplet $\langle \tilde{O}, \tilde{A}_G, \tilde{A}_D \rangle$; une extension et deux intentions, avec : $\tilde{O}^\uparrow = \tilde{A}_G$, $\tilde{A}_G^\downarrow = \tilde{O}$, $\tilde{O}^\downarrow = \tilde{A}_D$, $\tilde{A}_D^\uparrow = \tilde{O}$.

1.7.2.7 L'approche par couvertures (*Hedge*)

Une approche dite “par couverture” est proposée par [Bělohlávek et Vychodil \[2005\]](#), elle est considérée comme une généralisation des approches de Pollandt, Bělohlávek et de l'approche de génération classique des concepts flous, elle utilise des opérations (fonction) unaires, dites couvertures (hedge), sur un treillis résidué complet L . La couverture, notée $*$ est une fonction unaire qui satisfait pour tous x et y de L les propriétés suivantes :

- 1) $1^* = 1$;
- 2) $x^* \leq x$;
- 3) $(x \rightarrow y)^* \leq x^* \rightarrow y^*$;
- 4) $a^{**} = a$.

Étant donné le L -contexte $(\mathcal{O}, \mathcal{A}, \mathcal{I})$, soient les couvertures $*_{\mathcal{O}}$ et $*_{\mathcal{A}}$, les opérateurs de dérivations \uparrow et \downarrow sont définis par :

$$O^\uparrow(a) = \bigwedge_{o \in \mathcal{O}} (O(o)^{*_{\mathcal{O}}} \rightarrow \mathcal{I}(o, a))$$

$$A^\downarrow(o) = \bigwedge_{a \in \mathcal{A}} (A(a)^{*_{\mathcal{A}}} \rightarrow \mathcal{I}(o, a))$$

Les auteurs de [[Bělohlávek et Vychodil, 2005](#)] ont montré que les opérateurs \uparrow et \downarrow forment une connexion de Galois. Ils ont montré également que l'ensemble $\mathcal{B}(\mathcal{O}^{*_{\mathcal{O}}}, \mathcal{A}^{*_{\mathcal{A}}}, \mathcal{I}) = \{ \langle O, A \rangle \in L^{\mathcal{O}} \times L^{\mathcal{A}} \mid O^\uparrow = A, A^\downarrow = O \}$ muni par la relation d'ordre partiel \leq forme un treillis complet.

1.7.2.8 L'approche du treillis de concepts formels généralisés

L'approche proposée dans [[Krajčí, 2005](#)] suggère de considérer trois ensembles de degrés de vérité, $L_{\mathcal{O}}$ (pour les objets), $L_{\mathcal{A}}$ (pour les attributs) et L (pour les relations floues *objet – attribut*). L'auteur suppose que les ensembles $L_{\mathcal{O}}$ et $L_{\mathcal{A}}$ muni de la relation \leq sont des treillis complets et L un ensemble partiellement ordonné. Pour définir les opérateurs de dérivations, Krajčí suppose l'existence d'un opérateur $\otimes : L_{\mathcal{O}} \times L_{\mathcal{A}} \rightarrow L$ satisfaisant pour toute famille d'indices J , tous $x, x_j \in L_{\mathcal{O}}$, $y, y_j \in L_{\mathcal{A}}$ et $z \in L$:

$$- \quad x_1 \leq x_2 \Rightarrow x_1 \otimes y \leq x_2 \otimes y$$

- $x_1 \leq x_2 \Rightarrow y \otimes x_1 \leq y \otimes x_2$
- $\forall j \in J, x_j \otimes y \leq z \Rightarrow (\bigvee_{j \in J} x_j) \otimes y \leq z$
- $\forall j \in J, x \otimes y_j \leq z \Rightarrow x \otimes (\bigvee_{j \in J} y_j) \leq z$

En se basant sur la structure de degrés de vérités $(L_{\mathcal{O}}, L_{\mathcal{A}}, L, \otimes, \leq, \dots)$, l'auteur introduit les opérateurs de dérivation $\nearrow : L_{\mathcal{O}}^{\mathcal{O}} \rightarrow L_{\mathcal{A}}^{\mathcal{A}}$ et $\searrow : L_{\mathcal{A}}^{\mathcal{A}} \rightarrow L_{\mathcal{O}}^{\mathcal{O}}$, qui sont donnés par :

$$O^{\nearrow}(a) = \bigvee \{y \in L_{\mathcal{A}} \mid \forall o \in \mathcal{O} : O(o) \otimes y \leq \mathcal{I}(o, a)\}$$

$$A^{\searrow}(o) = \bigvee \{x \in L_{\mathcal{O}} \mid \forall a \in \mathcal{A} : x \otimes A(a) \leq \mathcal{I}(o, a)\}$$

Ainsi, Krajčiči définit un concept formel comme étant une paire $\langle O, A \rangle \in L_{\mathcal{O}}^{\mathcal{O}} \times L_{\mathcal{A}}^{\mathcal{A}}$ satisfaisant $O^{\nearrow} = A$ et $A^{\searrow} = O$.

1.7.2.9 L'approche de Djouadi et Prade

Nous avons pu voir que certaines approches qui proposent des extensions floues de l'ACF ne maintiennent pas la propriété topologique de fermeture de la connexion de Galois (Burusco et Fuentes-González, Yahia et Jaoua, Krajčiči). Pour cela, toutes les autres approches utilisent des algèbres résiduées comme des structures algébriques pour maintenir cette propriété de fermeture. Il s'avère que la propriété de résiduation est "très forte" et réduit, en conséquence, le nombre d'implications éligibles pour l'ACF floue.

Afin d'élargir l'ensemble des implications floues éligibles pour la définition d'opérateurs de dérivation tout en garantissant la fermeture de la connexion de Galois, Djouadi et Prade [2011] ont introduit le principe de minimalité. Ce principe définit une condition suffisante, plus faible que le principe de résiduation. Cette condition est décrite dans le théorème 1.7.1.

Djouadi et Prade [2011] proposent, également, une extension floues des opérateurs de dérivations possibilistes. Pour cela, ils considèrent une algèbre exprimée sous forme de treillis flou distributif. Ce treillis est donné par la structure $\mathbf{L} = (L, \wedge, \vee, 0_L, 1_L, \rightarrow, *, \sim)$, où, L est un ensemble totalement ordonné, 0_L (resp. 1_L)⁸ est la borne inférieure (resp. supérieure) de L , \wedge et \vee dénotent l'infimum et le supremum, \sim est un opérateur de négation floue vérifiant $\sim 0_L = 1_L$ et $\sim 1_L = 0_L$, \rightarrow est une implication floue vérifiant les conditions de limites ($0_L \rightarrow 0_L = 0_L \rightarrow 1_L = 1_L \rightarrow 1_L = 1_L$ et $1_L \rightarrow 0_L = 0_L$), $*$ est une conjonction floue qui n'est pas nécessairement une t-norme et la paire $(\rightarrow, *)$ n'est pas nécessairement une paire résiduée.

Soit $\mathbf{L} = (L, \wedge, \vee, 0_L, 1_L, \rightarrow, *, \sim)$ un treillis distributif et $(\mathcal{O}, \mathcal{A}, \mathcal{I})$ un L -contexte, l'extension floue des opérateurs de suffisance (Δ), de possibilité (Π), de nécessité (\mathcal{N}) et de suffisance duale (∇) sont donnés pour $\tilde{O} \in L^{\mathcal{O}}$ et $\tilde{A} \in L^{\mathcal{A}}$ par :

8. 0_L (resp. 1_L) n'est pas forcément égale à 0 (resp. 1).

$$\tilde{O}^\Delta(a) = \bigwedge_{o \in \mathcal{O}} (\tilde{O}(o) \rightarrow \mathcal{I}(o, a)) \quad (1.6)$$

$$\tilde{A}^\Delta(o) = \bigwedge_{a \in \mathcal{A}} (\tilde{A}(a) \rightarrow \mathcal{I}(o, a))$$

$$\tilde{O}^\Pi(a) = \bigvee_{o \in \mathcal{O}} (\tilde{O}(o) * \mathcal{I}(o, a)) \quad (1.7)$$

$$\tilde{A}^\Pi(o) = \bigvee_{A \in \mathcal{A}} (\tilde{A}(A) * \mathcal{I}(o, A))$$

$$\tilde{O}^N(a) = \bigwedge_{o \in \mathcal{O}} (\mathcal{I}(o, a) \rightarrow \tilde{O}(o)) \quad (1.8)$$

$$\tilde{A}^N(o) = \bigwedge_{a \in \mathcal{A}} (\mathcal{I}(o, a) \rightarrow \tilde{A}(a))$$

$$\tilde{O}^\nabla(a) = \bigvee_{o \in \mathcal{O}} (\sim \tilde{O}(o) * \sim \mathcal{I}(o, a)) \quad (1.9)$$

$$\tilde{A}^\nabla(o) = \bigvee_{A \in \mathcal{A}} (\sim \tilde{A}(A) * \sim \mathcal{I}(o, A))$$

Théorème 1.7.1. *Principe de minimalité [Djouadi et Prade, 2011]*

Soit $\mathbf{L} = (L, \wedge, \vee, 0_L, 1_L, \rightarrow)$ un treillis distributif, la composition Δ^Δ est un opérateur de fermeture flou si la propriété suivante est vérifiée :

$$\forall p, q \in L, p \leq (p \rightarrow q) \rightarrow q$$

Exemple 1.7.1. *Considérons l'implication de Dienes donnée par :*

$$p \rightarrow q = \begin{cases} 1 & \text{si } p \leq q \\ (1 - p) \vee q & \text{sinon} \end{cases}$$

Cette implication ne vérifie pas le principe de résiduation. De ce fait, les approches existantes excluent son utilisation pour l'ACF floue. Pourtant, du fait qu'elle vérifie le principe de minimalité du théorème 1.7.1, elle est correctement éligible pour l'ACF floue car elle vérifie la propriété de fermeture.

Djouadi et Prade [2011] élargissent leur étude aux autres compositions d'opérateurs. Dans ce cadre, ils ont démontré que, sous certaines conditions, les compositions d'opérateurs forment des connexions de Galois. Nous présentons ci-après deux de leurs propositions que nous utiliserons plus tard dans cette thèse.

Théorème 1.7.2. [Djouadi et Prade, 2011] Soit $\mathbf{L} = (L, \wedge, \vee, 0_L, 1_L, \rightarrow, *)$ un treillis distributif, la composition ${}^{\text{NII}}$ (i.e. $\Pi \circ \text{N}$) est un opérateur de fermeture flou si la propriété suivante est satisfaite pour tous p et q de L :

$$p \rightarrow (p * q) \geq q$$

Théorème 1.7.3. [Djouadi et Prade, 2011] Soit $\mathbf{L} = (L, \wedge, \vee, 0_L, 1_L, \rightarrow, *)$ un treillis distributif, la composition ${}^{\text{IIN}}$ (i.e. $\text{N} \circ \Pi$) est un opérateur d'ouverture flou si la propriété suivante est satisfaite pour tous p et q de L :

$$(p \rightarrow q) * p \leq q$$

1.8 Extension floue à Valeur d'Intervalle de Vérité

Une des sémantiques qui peut être associée à l'extension floue de l'ACF est que les valeurs des relations floues *objet-attribut* peuvent être considérées comme un raffinement du symbole “ \times ” dans les contextes formels classiques. Ce raffinement représente le degré avec lequel l'objet possède l'attribut (ou l'objet satisfait la propriété). Cependant il n'est pas toujours facile de mesurer précisément ce degré. Et plus généralement, ces degrés sont attribués d'une manière approximative. Il convient, alors, d'attribuer un sous-intervalle dans $[0, 1]$ à la relation objet-attribut plutôt qu'une valeur dans $[0, 1]$. Surtout lorsque l'évaluation de ce degré est soumise à une certaine variation, ou est partiellement inconnue. Ceci a donné lieu à l'analyse de concepts formels à intervalles flous. Un ensemble d'intervalles flous U dans un univers \mathcal{U} est défini par $U = \{\langle u, \underline{\mu}_U(u), \bar{\mu}_U(u) \rangle, u \in \mathcal{U}\}$, où $\underline{\mu}_U, \bar{\mu}_U : \mathcal{U} \rightarrow L$ sont les fonctions d'appartenance minimale et maximale des éléments de l'univers \mathcal{U} à l'ensemble U . Ces fonctions vérifient pour tout élément $u \in \mathcal{U}$, $0 \leq \underline{\mu}_U(u) \leq \bar{\mu}_U(u) \leq 1$.

L'extension floue à intervalles de vérités de l'ACF⁹ proposée dans [Djouadi et Prade, 2009] considère le treillis $(\mathcal{L}[L], \leq_{\mathcal{I}})$, avec :

- $L = [0, 1]$,
- $\mathcal{L}[L] = \{x \mid x = [\alpha, \beta], \alpha, \beta \in [0, 1] \text{ et } \alpha \leq \beta\}$,
- $[0, 0]$ est le plus petit élément de $\mathcal{L}[L]$,
- $[1, 1]$ est le plus grand élément de $\mathcal{L}[L]$,
- $\leq_{\mathcal{I}}$ est une relation d'ordre partiel définie par : $[\alpha_1, \alpha_2] \leq_{\mathcal{I}} [\beta_1, \beta_2] \Leftrightarrow \alpha_1 \leq \beta_1 \text{ et } \alpha_2 \leq \beta_2$,
- L'infimum \wedge est donné par : $[\alpha_1, \alpha_2] \wedge [\beta_1, \beta_2] = [\min(\alpha_1, \beta_1), \min(\alpha_2, \beta_2)]$
- Le supremum \vee est donné par : $[\alpha_1, \alpha_2] \vee [\beta_1, \beta_2] = [\max(\alpha_1, \beta_1), \max(\alpha_2, \beta_2)]$

9. Interval-valued fuzzy formal concept analysis.

La définition des opérateurs logiques dans la logique d'intervalles flous résulte naturellement de leur homologue en logique floue. Ainsi, la généralisation des opérateurs d'implication a été définie dans [Alcalde et al., 2005; Gasse et al., 2008] par :

Définition 1.8.1. (*Implication d'intervalles flous*) Une implication d'intervalles flous est une image $Imp : \mathcal{L}[L] \times \mathcal{L}[L] \rightarrow \mathcal{L}[L]$ satisfaisant :

1. $Imp(x, y) \leq_{\mathcal{I}} Imp(x', y)$ pour $x' \leq_{\mathcal{I}} x$
2. $Imp(x, y) \leq_{\mathcal{I}} Imp(x, y')$ pour $y \leq_{\mathcal{I}} y'$
3. $Imp([0, 0], [0, 0]) = Imp([0, 0], [1, 1]) Imp([1, 1], [1, 1]) = [1, 1]$
4. $Imp([1, 1], [0, 0]) = [0, 0]$

Djouadi et Prade [2009] définissent un $\mathcal{L}[L]$ -contexte par le quadruplet $(\mathcal{L}[L], \mathcal{O}, \mathcal{A}, \mathcal{I})$, où $\mathcal{L}[L]$ est un treillis complet, \mathcal{O} est l'ensemble d'objets, \mathcal{A} est l'ensemble d'attributs et \mathcal{I} est une relation d'intervalles flous ($\mathcal{I} : \mathcal{O} \times \mathcal{A} \rightarrow \mathcal{L}[L]$). Ils étendent l'opérateur de suffisance (noté $*$) aux intervalles flous par les expressions suivantes :

$$\begin{aligned} \tilde{O}^*(a) &= \bigwedge_{o \in \mathcal{O}} Imp(O(o), \mathcal{I}(o, a)) \\ \tilde{A}^*(o) &= \bigwedge_{a \in \mathcal{A}} Imp(A(a), \mathcal{I}(o, a)) \end{aligned} \quad (1.10)$$

Les auteurs de cette approche établissent la condition minimale pour le maintien de la fermeture de la connexion de Galois $**$ que nous rappelons dans le théorème suivant :

Théorème 1.8.1. *Principe de minimalité [Djouadi et Prade, 2009]* L'opérateur de dérivation de Galois $**$ défini sur l'ensemble $\mathcal{L}[L]^{\mathcal{O}}$ (et duallement sur l'ensemble $\mathcal{L}[L]^{\mathcal{A}}$) est un opérateur de fermeture si la propriété suivante est satisfaite :

$$\forall x, y \in \mathcal{L}[L], x \leq_{\mathcal{I}} Imp(Imp(x, y), y)$$

1.9 Conclusion

Ce chapitre a été consacré à la théorie de l'ACF et ses extensions. Dans un premier temps, nous avons rappelé des notions mathématiques qui sont à la base de la théorie de l'ACF et sa lecture possibiliste. Par la suite, nous avons élaboré un état d'art des principales extensions de l'ACF. Nous avons ainsi passé en revue l'extension triadique, l'extension floue et enfin l'extension floue à valeur d'intervalle de vérité.

Rappelons que ce travail de thèse s'inscrit dans l'extension floue de l'ACF pour la recherche d'information. A ce titre, le chapitre suivant présentera la recherche d'information et mettra particulièrement l'accent sur la recherche d'information basée sur l'ACF.

Chapitre 2

Recherche d'Information (RI)

Sommaire

2.1	Introduction	31
2.2	Les Modèles Ensemblistes	32
2.2.1	Modèle Booléen	32
2.2.2	Modèle Booléen basé sur les Ensembles Flous	33
2.2.3	Modèle P-norm	34
2.3	Les Modèles Algébriques	34
2.3.1	Modèle Vectoriel	34
2.3.2	Modèle d'Indexation Sémantique Latente (LSI)	36
2.4	Les Modèles Probabilistes	37
2.4.1	Le Modèle de Base	37
2.4.2	Le modèle de Langue	37
2.5	Les Modèles Logiques	38
2.5.1	Modèles basés sur la Logique Propositionnelle	39
2.5.2	Modèles basés sur des Systèmes d'Argumentation Probabiliste	39
2.5.3	Modèles basés sur la Logique Modale	40
2.5.4	Modèles basés sur la Logique Descriptive	41
2.5.5	Modèles basés sur des Graphes Conceptuels	43
2.5.6	Modèles basés sur la Théorie des Situations	44
2.5.7	Modèles basés sur Datalog Probabiliste	45
2.5.8	Les Méta-Modèles	45
2.6	Les Approches de RI basées sur l'ACF	46
2.6.1	Amélioration des Résultats de Requêtes	47
2.6.2	Web et Courrier Électronique	48

2.6.3	Logiciel, Image et Base de Connaissances	49
2.6.4	Traitement de Requêtes sur des Données Hétérogènes	50
2.6.5	Réponses Contextuelles et Classement des Résultats	51
2.6.6	Réinjection de Pertinence	52
2.6.7	Les Systèmes de RI basés sur l'ACF	53
2.6.8	Les Approches de RI basées sur l'ACF Floue	54
2.7	Conclusion	56

2.1 Introduction

La recherche d'information (RI) est le domaine qui traite la représentation, le stockage, l'organisation et la manière de retrouver des informations dans une collection de documents numériques, appelée corpus. Retrouver des informations c'est le processus de sélection (élection) de documents (ou des parties de documents) répondant à des besoins informationnels exprimés dans une requête utilisateur. Ceci revient alors, à mesurer la pertinence des documents de la collection par rapport à la requête. Dans un système de RI (SRI) cette mesure est réalisée par la fonction d'appariement ou de correspondance *document-requête*. Pour réaliser cela, le SRI représente les documents et les requêtes dans un formalisme unifié puis compare les deux représentations. Cette comparaison se traduit par un score ou un degré de similitude entre un document et une requête. La fonction d'appariement entre un document et une requête est désignée par $RSV(d, q)$ (pour *Relevance Status Value*), où d est la représentation du document et q est la représentation de la requête dans le même formalisme.

La représentation de documents est basée sur les termes d'indexation issus du processus d'indexation. Ce dernier consiste, principalement, à déterminer les termes représentatifs dans les documents et de leur associer, si le SRI le permet, un poids d'importance (ou de pondération) par le biais de la fonction de pondération. Le modèle permet de donner une interprétation des termes choisis pour représenter le contenu d'un document.

De manière générale, les fonctions de pondération, à l'exemple des mesures TF-IDF (*Term Frequency-Inverse Document Frequency*) [Salton et Yang, 1973] et BM25 [Robertson et Walker, 1997], sont basées sur l'agrégation de deux paramètres, l'un local et l'autre global. Le paramètre local quantifie la représentativité du terme dans le document, tandis que le paramètre global quantifie la représentativité du terme dans le corpus.

Le modèle joue un rôle central dans la RI. Il a la charge de créer les représentations internes des documents et des requêtes basées sur les termes d'indexation, et de définir la méthode de comparaison entre ces représentations (de définir la fonction RSV).

Selon Baeza-Yates et Ribeiro-Neto [1999], les modèles de RI peuvent être regroupés en trois catégories, qui sont les modèles ensemblistes, algébriques et probabilistes. Dans les sections 2.2, 2.3 et 2.4, nous décrivons les principaux modèles de ces trois catégories. Pour chacun des modèles présentés, les deux points importants seront définis : la représentation et la comparaison des documents et des requêtes.

Ensuite, nous présentons dans la Section 2.5 une description de l'état de l'art des modèles logiques. Ces modèles logiques peuvent être considérés comme des formalisations logiques (ou des modélisations logiques) des modèles des précédentes catégories.

Enfin, dans la Section.2.6 nous dressons un état de l'art des approches de recherche d'information basées sur l'ACF classées selon leurs fonctionnalités.

2.2 Les Modèles Ensemblistes

2.2.1 Modèle Booléen

Le modèle Booléen proposé par Salton [1968], est le premier modèle de recherche d'information. Ce modèle est basé sur la théorie des ensembles et assume que les documents sont représentés par un ensemble de termes d'indexation. Dans ce modèle, une requête q est une combinaison Booléenne de ces termes d'indexation en utilisant les opérateurs logiques classiques *et* (\wedge), *ou* (\vee) et *non* (\neg). Formellement, les requêtes q sont construites sur l'ensemble des termes d'indexation $\{t_i, i = 1, 2, 3, \dots\}$ selon la grammaire suivante :

$$\begin{aligned} q &::= r \mid \emptyset \\ r &::= t_i \mid (r \wedge r) \mid (r \vee r) \mid \neg r \end{aligned}$$

La requête " $t_1 \wedge t_2$ " est satisfaite par un document d si et seulement si d contient les deux termes t_1 et t_2 . De même, la requête " $t_1 \vee t_2$ " est satisfaite par le document d si et seulement si il contient t_1 ou t_2 ou les deux termes en même temps. La requête " $t_1 \wedge (\neg t_2)$ " est satisfaite par le document d si et seulement si il contient t_1 et ne contient pas t_2 . Des requêtes Booléennes plus complexes peuvent être construites à partir de ces opérateurs et évaluées selon les règles classiques de l'algèbre Booléenne. De ce fait, représenter un document par un ensemble de termes est suffisant pour le traitement des requêtes.

Pour une requête donnée, le processus de recherche consiste à assigner les valeurs Booléenne $\{1, 0\}$ à chacun des documents de la collection. Ainsi, un document est retrouvé (pertinent) si ses termes d'indexation satisfont l'expression logique de la requête. Formellement, l'évaluation de la pertinence d'un document d à une requête donnée est définie précisément par :

$$RSV(d, t) = 1 \text{ si } t \in d, 0 \text{ sinon} \quad (2.1)$$

$$RSV(d, q_1 \wedge q_2) = RSV(d, q_1) \wedge RSV(d, q_2) \quad (2.2)$$

$$RSV(d, q_1 \vee q_2) = RSV(d, q_1) \vee RSV(d, q_2) \quad (2.3)$$

$$RSV(d, \neg q) = 1 - RSV(d, q) \quad (2.4)$$

Malgré la large utilisation de ce modèle (MISTRAL de CII/Bull, GOLEM-PASSAT de Siemens, STAIRS d'IBM), il présente un certain nombre d'inconvénients :

- La sélection d'un document est basée sur une décision binaire,

- Les documents retournés à l'utilisateur ne sont pas ordonnés selon leur pertinences,
- Problème de collections volumineuses : le nombre de documents retournés peut être considérable.

Afin de remédier à ces problèmes, certaines extensions ont été proposées, comme le modèle Booléen basé sur la théorie des ensembles flous [Maron et Kuhns, 1960; Noreault et al., 1977; Bookstein, 1980; Kraft et Buell, 1983] et le modèle Booléen étendu [Salton et al., 1983].

2.2.2 Modèle Booléen basé sur les Ensembles Flous

Cette extension du modèle Booléen standard vise à tenir compte de la pondération des termes dans les document et/ou les requêtes en utilisant la théories des ensembles flous. Dans la littérature, on peut distinguer trois familles d'approches [Kraft et Buell, 1983] qui généralisent le model Booléen standard :

1. Documents pondérés, requêtes Booléens et le RSV calculé en utilisant les règles des ensembles flous [Maron et Kuhns, 1960], [Salton, 1975];
2. Documents non pondérés, requêtes pondérés (floues) et le RSV peut être classique ou flou [Angione, 1975], [Koll, 1978], [Noreault et al., 1977];
3. Documents pondérés et requêtes pondérés (poids ou seuils des termes de la requête et/ou sur des sous-expressions Booléennes), le RSV est calculé par une fonction générale [Bookstein, 1980], [Buell et Kraft, 1981], [Buell et Kraft, 1982]

À titre d'exemple, l'extension floue du modèle Booléen proposé dans [Bookstein, 1980] (également appelée modèle flou standard) permet d'associer une valeur pondérée floue à des termes d'indexation et, par conséquent, permet d'obtenir un classement du résultat d'une recherche selon la pertinence des documents. L'extension floue du RSV est formellement donnée par :

$$RSV(d, t) = \alpha \quad (\alpha \in [0, 1]) \quad (2.5)$$

$$RSV(d, q_1 \wedge q_2) = \top(RSV(d, q_1), RSV(d, q_2)) \quad (2.6)$$

$$RSV(d, q_1 \vee q_2) = \perp(RSV(d, q_1), RSV(d, q_2)) \quad (2.7)$$

$$RSV(d, \neg q) = 1 - RSV(d, q) \quad (2.8)$$

Où, α est le poids du terme t dans le document d , et l'opérateur \top (resp. \perp) est une t-norme (resp. t-conorme). Remarquant que l'opérateur Min (resp. Max) a été habituellement choisi par les approches existantes en tant que t-norme (resp. t-conorme).

2.2.3 Modèle P-norm

L'un des inconvénient des approches basés sur les ensembles flous (les extensions floues du modèle Booléen) et que le processus d'évaluation de la pertinence des documents est discriminatoire par rapports aux poids des termes d'indexation. En effet, pour une requête donnée, le rang associé à un document dépend uniquement du poids le plus bas (resp. le plus haut) des termes d'indexation dans les requêtes conjonctive (resp. disjonctive). Pour remédier à cet inconvénient, [Salton et al. \[1983\]](#) ont proposé le modèle P -norm (modèle Booléen étendu). Ce modèle permet d'évaluer la pertinence d'un document d'une manière plus conforme au jugement d'un utilisateur humain [[Lee, 1994](#)]. La requête demeure une expression booléenne classique. Tandis que les termes d'un document sont pondérés. En général le poids d'un terme d'indexation dans un document est en fonction du nombre d'occurrences de ce terme d'indexation dans le document.

Le modèle P -norm définit des fonctions de similarité pour les “*et*” et “*ou*” Booléens généralisés. Ces fonctions ont un paramètre noté P qui varie de 1 à ∞ . La pertinence est donnée comme une valeur de similarité entre une requête et un document. Étant donné l'ensemble des termes d'indexation $\{t_i, i = 1 \dots n\}$, chaque document d est donné sous forme d'un ensemble de termes pondérés $d = \{t_i^{w_i}, i = 1 \dots N\}$ (w_i correspond au poids du terme t_i dans le document d). La similarité entre le document d et la requête disjonctive $Q_{ou(P)} = t_1^{u_1} \vee^P t_2^{u_2} \vee^P \dots \vee^P t_n^{u_n}$ (respectivement, la requête conjonctive $Q_{et(P)} = t_1^{u_1} \wedge^P t_2^{u_2} \wedge^P \dots \wedge^P t_n^{u_n}$) est donnée par :

$$Sim(d, Q_{ou(P)}) = \sqrt[P]{\frac{w_1^P u_1^P + w_2^P u_2^P + \dots + w_n^P u_n^P}{u_1^P + u_2^P + \dots + u_n^P}}$$

$$Sim(d, Q_{et(P)}) = 1 - \sqrt[P]{\frac{(1 - w_1)^P u_1^P + (1 - w_2)^P u_2^P + \dots + (1 - w_n)^P u_n^P}{u_1^P + u_2^P + \dots + u_n^P}}$$

Remarquant que ce modèle se ramène au modèle Booléen à $P = \infty$, et il est équivalent, pour $P = 1$, au modèle vectoriel [[Salton et al., 1975](#)], que nous présentons dans la section suivante.

2.3 Les Modèles Algébriques

2.3.1 Modèle Vectoriel

Le modèle vectoriel de la recherche d'information a été introduit par [Salton et al. \[1975\]](#) et implémenté dans le système SMART. En se basant sur une approche algébrique, le modèle de base représente les documents et les requêtes sous forme de vecteur dans l'espace vectoriel des termes d'indexations ($\vec{t}_1, \vec{t}_2, \dots$). Pour n termes

d'indexations, et sous l'hypothèse qu'ils sont indépendants, chaque document d de la collection est représenté par son vecteur de poids $\vec{d} = [w_1, w_2, \dots, w_i, \dots, w_n]$, où w_i correspond au poids du terme t_i dans le document d . D'une manière similaire, une requête q est représentée par le vecteur $\vec{q} = [w_1^q, w_2^q, \dots, w_i^q, \dots, w_n^q]$, où w_i^q correspond au poids associé au terme t_i dans la requête q . Le processus de recherche consiste alors à associer à chaque document de la collection la similarité (ou le degré de corrélation) entre son vecteur et le vecteur de la requête. Un document est d'autant plus pertinent à une requête que le vecteur associé est similaire à celui de la requête. Cette similarité peut être exprimée par l'une des mesures suivantes :

- Le produit scalaire :

$$Sim(d, q) = \sum_{i=1}^n (w_i \cdot w_i^q) \quad (2.9)$$

- Cosinus :

$$Sim(d, q) = \frac{\sum_{i=1}^n w_i \times w_i^q}{\sqrt{\sum_{i=1}^n (w_i)^2} \times \sqrt{\sum_{i=1}^n (w_i^q)^2}} \quad (2.10)$$

- Coefficient de Dice :

$$Sim(d, q) = \frac{2 \cdot \sum_{i=1}^n w_i \times w_i^q}{\sum_{i=1}^n (w_i)^2 + \sum_{i=1}^n (w_i^q)^2} \quad (2.11)$$

- La mesure de Jaccard :

$$Sim(d, q) = \frac{\sum_{i=1}^n w_i \times w_i^q}{\sum_{i=1}^n (w_i)^2 + \sum_{i=1}^n (w_i^q)^2 - \sum_{i=1}^n w_i \times w_i^q} \quad (2.12)$$

- Mesure de superposition :

$$Sim(d, q) = \frac{\sum_{i=1}^n w_i \times w_i^q}{Min(\sum_{i=1}^n (w_i)^2, \sum_{i=1}^n (w_i^q)^2)} \quad (2.13)$$

2.3.2 Modèle d'Indexation Sémantique Latente (LSI)

Dans le modèle vectoriel décrit ci-dessus, un espace vectoriel est défini pour une collection de documents de telle sorte que chaque dimension de l'espace est un terme occurant dans la collection, et chaque document est spécifié comme un vecteur avec une coordonnée pour chaque terme d'indexation de la collection. Ce modèle définit une approximation efficace des propriétés statistiques de la collection, cependant, il suppose que les termes sont indépendants, et qu'ils sont orthogonaux dans l'espace vectoriel. L'ajout d'un nouveau terme à l'espace, par exemple, un terme qui a été précédemment omis parce qu'il n'était pas considéré comme un bon discriminateur, n'a aucun effet sur les termes existants définissant l'espace. Et l'ajout d'un nouveau document à la collection n'ajoute pas seulement de nouveaux termes d'indexation dans l'espace, mais affecte également les poids des termes existants car cela affecte leurs IDF. Par conséquent, il s'agit, uniquement, d'une relation terme-document (elle n'est pas une relation terme-terme) et elle ne permet pas de considérer des relations entre termes. Or, dans la pratique des termes peuvent être des synonymes et font référence aux mêmes concepts. Ainsi, un nouveau modèle vectoriel, appelé l'Indexation Sémantique Latente (LSI), a été proposé dans [Deerwester et al., 1990] pour prendre en considération les relations statistiques entre les termes d'indexation. LSI est une technique qui projette les requêtes et les documents dans un espace réduit (k -dimensions ou k -facteurs) pour construire des index conceptuels portant sur la sémantique des mots dans les documents. Ce modèle se base sur la décomposition en valeurs singulières (SVD) de la matrice termes-documents initiale. À partir de la matrice *Documents – Termes* (noté V) décrivant la relation initiale entre les documents de la collection et les termes d'indexation dans l'espace vectoriel initial de n dimensions, LSI produit trois matrices T , S et D tel que le produit $T * S * D$ soit le plus proche possible à la matrice V avec :

D : la matrice *Documents–Concepts* dans le k -espace, où les colonnes représentent les concepts, et les lignes correspondent aux vecteurs documents spécifiés en fonctions des k concepts.

T : la matrice *Terme – Concepts*, où les colonnes représentent les concepts, et les lignes représentent les vecteurs des termes initiales en fonction des nouveaux concepts.

S : Les éléments diagonaux de la matrice S attribuent des poids (appelés «valeurs singulières») aux facteurs k en fonction de leur importance. Cela permet à l'utilisateur d'avoir un certain contrôle sur la dimension du k -espace.

En conséquence, dans l'espace sémantique latent, une requête et un document peuvent avoir une forte similitude de cosinus, même s'ils ne partagent aucun terme, tant que leurs termes sont sémantiquement semblables.

2.4 Les Modèles Probabilistes

2.4.1 Le Modèle de Base

Plusieurs approches [Maron et Kuhns, 1960; Robertson et al., 1982; Bookstein, 1983; Fuhr, 1989] ont tenté de définir la pondération des termes d'indexation et pertinence de façon plus formelle en s'appuyant sur la théorie des probabilités. Étant donné une requête utilisateur q et un document d , le modèle probabiliste tente d'estimer la probabilité que le document d appartienne à la classe des documents pertinents (ou non pertinents). Un document est alors sélectionné si la probabilité qu'il soit pertinent à q , notée $P(Pert/d, q)$, est supérieure à la probabilité qu'il soit non pertinent à q , notée $P(Npert/d, q)$. Le score d'appariement entre le document d et la requête q , noté $RSV(d, q)$, est donné [Robertson et Walker, 1994] par :

$$RSV(d, q) = \frac{P(Pert/d, q)}{P(Npert/d, q)}$$

Plus cette proportion est élevée pour un document, plus ce document est pertinent pour la requête. Cependant, les deux probabilités nécessaires ne sont pas directement calculables. Pour cela, on utilise les règles de Bayes suivantes :

$$P(Pert/d, q) = \frac{P(d, q/Pert) * P(Pert)}{P(d, q)}$$

$$P(Npert/d, q) = \frac{P(d, q/Npert) * P(Npert)}{P(d, q)}$$

$$RSV(d, q) = \frac{P(Pert/d, q)}{P(Npert/d, q)} = \frac{P(d, q/Pert) * P(Pert)}{P(d, q/Npert) * P(Npert)} = C * \frac{P(d/Pert, q)}{P(d/Npert, q)}$$

Où

- $P(Pert)$ est la probabilité qu'un document est pertinent,
- $P(Npert)$ est la probabilité qu'un document n'est pas pertinent,
- $C = \frac{P(Pert)}{P(Npert)}$;
- $P(d/Pert, q)$ (resp. $P(d/Npert, q)$) est la probabilité d'observer d sachant que l'on observe la pertinence (resp. non pertinence) en présence de q ;
- $P(d, q)$ est la probabilité conjointe du couple (d, q) .

2.4.2 Le modèle de Langue

Le modèle de langue [Ponte et Croft, 1998; Boughanem et al., 2006] considère que la pertinence d'un document pour une requête est en rapport avec la probabilité que

la requête puisse être générée par le modèle de langue du document. Ainsi, il considère qu'un document d incarne un sous-langage, pour lequel il tente de construire un modèle de langue M_d . Le score du document d pour une requête q est déterminé par la probabilité que son modèle génère la requête. Ce score est donné par :

$$Score(q, d) = P(q/M_d)$$

Et comme une requête q peut être vue comme une suite de mots $t_1 t_2 \dots t_n$, le score $Score(q, d)$ est donné par :

$$Score(q, d) = P(t_1 t_2 \dots t_n / M_d) = \prod_{i=1}^n P(t_i / M_d)$$

$P(t_i / M_d)$ peut être estimé en se basant sur l'estimation maximale de vraisemblance (maximum likelihood estimation), donnée par :

$$P(t_i / M_d) = \frac{tf(t_i, d)}{\sum_t tf(t, d)}$$

Où $tf(t_i/d)$ est la fréquence du terme t_i dans le document d . Il est remarquable que lorsqu'un terme t_j de la requête q n'appartient pas au document d , le score $Score(q, d)$ est nul, vu que la fréquence $tf(t_j, d)$ est nulle. Afin de pallier à cet inconvénient, des techniques, dites de lissages [Boughanem et al., 2006], ont été proposées pour assigner des valeurs non nulles aux probabilités $P(t_j / M_d)$ même si t_j n'appartient pas à M_d .

2.5 Les Modèles Logiques

Dans cette section nous mettons particulièrement l'accent sur les approches logiques de RI, étant donné que notre proposition retient le formalisme logique pour utiliser l'ACF dans la RI.

Dans la littérature, de nombreuses approches ont proposé l'utilisation des logiques formelles dans la RI. Ces approches sont basées sur la proposition fondatrice de van Rijsbergen [1986]. L'idée générale de l'auteur est que si un document d et une requête q sont des formules dans une même logique, la pertinence de d par rapport à q est alors formalisée par l'implication $d \rightarrow q$. Autrement dit, pour une requête q , le processus de recherche consiste à retrouver des documents d pour lesquels la formule logique $d \rightarrow q$ est vraie ou valide, où " \rightarrow " n'est pas nécessairement l'implication matérielle " \supset ". van Rijsbergen [1986] propose également de mesurer le score de similitude document-requête par l'incertitude de l'implication $d \rightarrow q$ (noté $U(d \rightarrow q)$). Dans ce qui suit, nous présentons une description de l'état de l'art de ces approches.

2.5.1 Modèles basés sur la Logique Propositionnelle

Losada et Barreiro [2001] utilisent la logique propositionnelle pour construire un modèle de RI. Chaque terme d'indexation est considéré comme une proposition atomique, qui peut être vrai ou faux dans un document ou une requête. Les documents et les requêtes sont des expressions logiques formées en utilisant les termes d'indexation.

La pertinence est considérée comme une conséquence logique, où d est pertinent pour q si et seulement si $d \models q$. Comme le modèle est proposé dans le cadre de la logique propositionnelle alors $d \models q$ est équivalent à $\models d \supset q$. Dans la sémantique formelle de la logique propositionnelle, $d \models q$ signifie que tout modèle de d est aussi un modèle de q , ou en d'autres termes, pour toute interprétation dans laquelle d est vraie, q est également vrai.

Pour estimer l'incertitude de $d \models q$ (qui correspond à la distance entre d et q), Losada et Barreiro [2001] utilisent la révision des croyances (*Belief Revision*), qui est une technique permettant d'exprimer formellement la notion de proximité entre expressions logiques. En d'autres termes, le processus de révision de croyances traite la mise à jour d'une connaissance existante K avec une nouvelle information s , notée $K \circ s$, où, s'il n'y a pas de contradiction entre K et s , la connaissance mise à jour devient $K \circ s = K \wedge s$, sinon le processus traite du changement minimal qui devrait être fait sur K afin de construire des connaissances mises à jour K' qui ne se contredisent pas avec s . Ainsi, la connaissance mise à jour devient $K \circ s = K' \wedge s$.

Du fait que les propositions atomiques correspondent aux termes d'indexation, la distance entre un document d et une requête q correspond au nombre de propositions atomiques différentes entre les expressions logiques de d et de q . Losada et Barreiro [2001] utilisent cette distance dans une fonction ad-hoc pour mesurer le score $U(d \rightarrow q)$ et prétendent que leur modèle est équivalent au modèle vectoriel.

2.5.2 Modèles basés sur des Systèmes d'Argumentation Probabiliste

Un système d'argumentation est obtenu à partir de la logique propositionnelle en considérant deux ensembles disjoints ; un ensemble de variables d'intérêt (noté I) et un ensemble de variables hypothèses (ou suppositions) (noté H). Un système d'argumentation est un triplet (I, H, ξ) , où ξ représente la base de connaissances du système qui est l'ensemble de formules (formules bien formées) propositionnelles construites sur l'alphabet $I \cup H$. Lorsqu'on modélise la connaissance dans un système d'argumentation propositionnelle, les arguments peuvent être déduits de la base de connaissances en supprimant ou en actualisant certaines hypothèses. Une hypothèse h est une formule logique bien formée avec des symboles dans $I \cup H$. Un littéral α est un support d'argument pour l'hypothèse h si $\alpha \wedge \xi \models h$. Ceci dit que h est

la conséquence logique de $\alpha \wedge \xi$. Le quasi-support de l'hypothèse h dans la base de connaissances ξ est donné par la formule :

$$qs(h, \xi) = \forall \alpha_i, \alpha_i \wedge \xi \models h$$

Les systèmes d'argumentation probabiliste augmentent la logique propositionnelle par un mécanisme pour capturer l'incertitude en utilisant les probabilités. Soit (I, H, ξ) un système d'argumentation propositionnelle, et X l'ensemble des probabilités assignées aux hypothèses de H , le quadruplet (I, H, ξ, X) est appelé "système d'argumentation probabiliste".

Deux autres fonctions sont utilisées dans les systèmes d'argumentation probabiliste, le support d'une hypothèse (notée sp) et le degré du support d'une hypothèse (notée dsp) qui sont définies par :

$$sp(h, \xi) = qs(h, \xi) \wedge \neg qs(\perp, \xi)$$

$$dsp(h, \xi) = \frac{P(qs(h, \xi) / \neg qs(\perp, \xi))}{1 - P(qs(\perp, \xi))}$$

Picard et Savoy [2000] utilisent les systèmes d'argumentation probabiliste pour construire un modèle de RI. Ils considèrent que les termes d'indexation, les documents et les requêtes comme étant des propositions atomiques. Le contenu des documents et des requêtes est représenté par l'ensembles des règles de la forme :

$$a_{ij} \wedge d_i \rightarrow t_j$$

Ceci signifie que le document d_i est indexé par le terme t_j avec le poids (incertitude) $P(a_{ij}) = P(t_j/d_i)$. Les relations sémantiques entre les termes sont représentées par des règles de type :

$$b_{ij} \wedge t_i \rightarrow t_j$$

Cette règle signifié qu'il y a une relation sémantique entre le terme t_i et t_j avec une grandeur (*strength*) donnée par la probabilité $P(b_{ij})$. L'ensemble de ces règles forme la base de connaissance su système de recherche d'information (notée K). Ainsi, Picard et Savoy [2000] considèrent que \rightarrow est une implication matérielle dans l'expression $d \rightarrow q$. Ils estiment l'incertitude $U(d \supset q)$ de deux façons :

$$\begin{array}{ll} \text{Symbolique :} & U(d \supset q) = sp(q, K \wedge d) \\ \text{Numérique :} & U(d \supset q) = dsp(q, K \wedge d) \end{array}$$

2.5.3 Modèles basés sur la Logique Modale

La logique modale considère deux opérateurs (modalités), nécessité (\square) et possibilité (\diamond) qui peuvent être ajoutés à d'autres formalismes logiques comme les logiques

propositionnelle et de premier ordre. Cette logique adopte la sémantique des mondes possibles [Kripke, 1963] (appelé également la sémantique de Kripke). La sémantique de Kripke est une structure $\langle W, R \rangle$, où W est un ensemble non-vidé appelé l'ensemble des mondes possibles (*possible worlds*), et R une relation binaire ($R \subseteq W \times W$) appelée relation d'accessibilité.

La logique modale a été utilisée par Nie [1989] pour développer un modèle logique pour la recherche d'information. Les documents sont des mondes, et les requêtes sont des formules. Un document représenté par un monde d est pertinent pour une requête représentée par une formule q si q est vraie dans le monde d , ou si q est vraie dans un monde d_n accessible à partir de d . La relation d'accessibilité capture la transformation des documents. Le fait que le monde d est connecté au monde d_n est interprété comme étant la transformation de d en d_n . Par exemple, d_n contient des synonymes des termes d'indexation de d . L'implication $d \rightarrow q$ est donc équivalente à trouver un chemin du monde initial possible d à un autre monde possible d_n qui satisfait la requête q . Dans ce modèle, l'incertitude $U(d \rightarrow q)$ est donnée par la distance totale (ou le coût) de ce chemin.

La relation d'accessibilité présente diverses propriétés dans le domaine de la recherche d'information. À titre d'exemple, les liens hypertexte peuvent être modélisés par ces relations. Ainsi, si un document d_1 possède un lien hypertexte d'un document d_2 il peut y avoir une relation d'accessibilité entre ces deux documents. D'une manière formelle, Nie [1989] interprète l'implication $d \rightarrow q$ de la RI par la conséquence logique \models ,

1. si $d \models q$ alors d est une réponse possible à q ,
2. si $d \models \diamond q$, alors d est une bonne réponse pour q après une possible transformation de d (réponse moins favorable),
3. si $d \models \square q$ alors d est une bonne réponse de q pour toutes les transformations possibles de d ,
4. dans le cas particulier où $\forall d, d \models q$, alors toute d est une réponse possible pour q . Et la requête q est considérée comme une requête générale pour la collection. Duallement, si $\forall d, d \models \neg q$ alors la requête q est trop spécifique pour la collection.

Un autre modèle de la RI basé sur la logique modale à été proposé dans [Crestani et Rijsbergen, 1995]. Ce modèle représente chaque terme d'indexation t par un monde possible, un document d est vrai dans le monde t si t apparaît dans d . De même pour les requête, une requête q est vraie dans un monde t si t apparaît dans q .

2.5.4 Modèles basés sur la Logique Descriptive

Les logiques de description ou logiques descriptives (*LD*) sont une famille de langages de représentation de connaissance qui peuvent être utilisés pour représenter

la connaissance terminologique d'un domaine d'application d'une manière formelle et structurée. La syntaxe du langage d'une *LD* se caractérise par un ensemble de symboles et des règles de constructions permettant d'obtenir des formules à partir de ces symboles et de constructeurs spécifiques au langage. Les *LD* utilisent les notions de concept, de rôle et d'individu (appelé aussi blocs de construction)

- Un individu correspond à un élément d'un univers donné ;
- Un concept correspond à une classe d'éléments et est interprété comme un ensemble dans un univers donné. Les concepts peuvent être atomiques (correspondant à un symbole de type concept), ou complexes obtenus en utilisant les constructeurs de concepts ;
- Un rôle d'une *LD* correspond à un lien entre deux éléments et est interprété comme une relation binaire sur un univers donné. Les rôles peuvent être atomiques, ou complexes obtenus en utilisant les constructeurs de rôles.

Les concepts sont reliés entre eux par l'intersection (\cap) ou l'union (\cup). Les quantificateurs universel (\forall) et existentiel (\exists) permettent de relier les rôles avec les concepts. Deux types de raisonnement sont aussi définis dans les *LD* ; la subsomption (\sqsubseteq) entre les concepts ou les rôles, l'assertion ($:$) qui permet de relier les concepts et les rôles aux individus. *LD* divise la connaissance en deux parties :

1. Les informations terminologiques (T-boxe) : un ensemble de formules relatives à la représentation et la manipulation des concepts et des rôles. Ces formules sont génériques (ou globales), vraies dans tous les modèles et pour tous les individus.
2. Les informations assertionnelles (A-boxe) : un ensemble de formules relatives à la description et la manipulation des individus, vraies pour certains individus.

Meghini et al. [1993] utilisent une *LD* spéciale, appelé *MIRTL* (*Terminological Logic for Modelling Information Retrieval*), pour formaliser l'implication $d \rightarrow q$ et construire un modèle de recherche d'information. Un document d est représenté en tant qu'individu qui est une instance d'un concept D . De même, une requête q est représentée en tant que concept. La pertinence d'un document d par rapport à une requête q est donnée par :

- Soit, l'individu d est une instance du concept q , désigné par $d : q$.
- Ou, le concept D , qui ne contient que le document d , est subsumé par le concept q , désigné par $D \sqsubseteq q$.

Les deux décisions $d : q$ et $D \sqsubseteq q$ sont des décisions binaires, c'est-à-dire que la base de connaissances les satisfait ou non (Ceci dit, ce modèle ne traite pas l'incertitude). Une extension de *MIRTL*, appelée *P - MIRTL*, a été proposé dans [Sebastiani, 1994], en ajoutant deux types de probabilités, à savoir le degré de croyance/confiance des assertions et les statistiques des concepts et des rôles. Ainsi, *P - MIRTL* permet d'évaluer par le biais de ces probabilités l'incertitude de l'implication $d \rightarrow q$.

2.5.5 Modèles basés sur des Graphes Conceptuels

Un *graphe conceptuel* (GC) est un graphe bipartie avec de sortes deux nœuds : concept et relation. Un concept est représenté par un couple [*type*, *réfèrent*]. Les nœuds sont liés par des arcs orientés, et un arc lie toujours un concept à une relation (un nœud concept peut ne pas être rattaché à une relation). Initialement, les graphes conceptuels ont été proposés par Sowa [1984] comme un formalisme de représentation de connaissances. Il existe une spécification (sous-type) et une hiérarchie de généralisation (super-type) entre les types de concepts, où un type de concept C_1 est un sous-type de concept C_2 si et seulement si chaque instance de C_1 est également une instance de C_2 , notée $C_1 \leq C_2$. Dans la théorie initiale des GCs , Sowa [1984] présente quatre opérateurs, appelés règles de formation canonique :

1. La copie d'un graphe : elle permet de dupliquer un graphe afin de l'utiliser sans modifier l'original,
2. La simplification : efface une composante d'un graphe, si une relation est dupliquée, alors on peut supprimer l'un des duplicata et tous ses arcs,
3. La restriction : permet de remplacer le type d'un concept du graphe par un de ses sous-types, tant que le réfèrent du concept reste conforme à ce sous-type,
4. La jointure réunit deux graphes à partir d'un concept dans chaque graphe, si ces deux concepts ont les mêmes référents et le même type, alors on copie l'un d'eux dans un nouveau graphe auquel on lie les relations liées aux deux concepts des graphes originaux.

Une autre opération (non-canonique), dite projection, qui peut être obtenue en combinant les opérations précédentes a été définie dans [Sowa, 1984]. La projection recherche dans un graphe une spécialisation d'un autre graphe. Cet opérateur permet de retrouver des définitions au sein d'un graphe et sera donc particulièrement utile dans l'application des définitions, ou pour la recherche d'un schéma dans une base de graphes. La projection du graphe G (notée $\pi(G)$) sur le graphe G' est un sous-graphe de G' , avec $\pi(G) \leq G'$.

Chevallet et Chiaramella [1995] ont montré que le formalisme GC est équivalent à la logique de premier ordre, où chaque GC G est transformé à une formule logique $\Phi(G)$, et la relation d'ordre partielle entre les GCs correspond à la validité de l'implication matérielle [Chevallet et Chiaramella, 1995] suivant :

$$G_1 \leq G_2 \Rightarrow \models [\Phi(G_1) \supset \Phi(G_2)]$$

En utilisant les graphes conceptuels Chevallet et Chiaramella [1995] ont proposé un modèle de recherche d'information. Dans ce modèle, chaque document est un GC , et d'une manière équivalente, une formule logique du premier ordre, une requête q est aussi un GC . Un document d est pertinent pour une requête q s'il existe une

projection $\pi(q)$ de q sur d . Autrement dit, si d contient un sous-graphe $\pi(q)$ qui est une spécialisation de q .

À partir de la définition de l'opération de projection, la décision $d \rightarrow q$ équivaut à $\models (\Phi(d) \supset \Phi(q))$. L'incertitude de cette implication est estimé en utilisant la sémantique de Kripke (§2.5.3), où un document est un monde possible et la relation accessible est l'une des opérations sur les *GCs*.

2.5.6 Modèles basés sur la Théorie des Situations

La *théorie de la situation* (*TS*) a été formalisée dans les années 80 par [Barwise et Perry \[1983\]](#) comme étant une théorie mathématique fournissant un cadre formel de modélisation et de représentation d'information. L'idée de base de TS est d'étudier ce qui rend l'information valide au lieu d'étudier si une information est vraie ou fausse. Le concept de base de cette théorie est l'*Infon*. Un Infon est une structure $\langle\langle R, o_1, o_2, \dots, o_n; v \rangle\rangle$, où R est la relation qui relie les objets o_1, o_2, \dots, o_n si $v = 1$. Une situation est une description partielle du monde définie comme un ensemble d'Infons. Une situation s supporte un infon i , noté $s \models i$, si i est rendu vrai par s , ou équivalent, si i peut être déduit de l'ensemble des infons de s .

TS généralise un ensemble de situations ayant des caractéristiques communes dans un type de situation. La notation $s : A$ signifie que la situation s est du type A . Une contrainte est une relation entre deux types de situation A et A' , notée $A \Rightarrow A'$. Une contrainte comme $A \Rightarrow A'$ signifie que l'apparition d'une situation $s : A$ implique l'existence d'une situation $s' : A'$. L'incertitude est représentée comme une contrainte conditionnelle, notée $A \Rightarrow A'|B$, ce qui signifie que la contrainte $A \Rightarrow A'$ est satisfaite sous certaines conditions B , où B est un type de situation.

[Lalmas et van Rijsbergen \[1993\]](#) ont construit un modèle de recherche d'information basé sur la théorie des situations. Ils considèrent que chaque document est une situation et une requête comme étant un Infon ou un ensemble d'Infons. Le document d est pertinent pour la requête q si est seulement si d supporte q ($d \models q$). L'incertitude de cette décision est donnée par :

$$U(d \models q) = \begin{cases} 1 & \text{si } d \models q \\ \text{Max}_{\{D'|D \Rightarrow D'|B, \exists d':D', d' \models q\}} \delta(D, D') & \text{sinon} \end{cases}$$

Où D est le type de d , D' est un autre type relatif à D sous certain condition B et $\delta(D, D')$ est défini par :

$$\delta(D, D') = \begin{cases} 1 & \text{si } D \Rightarrow D' \\ 0 < \alpha < 1 & \text{si } D \Rightarrow D'|B \\ 0 & \text{sinon} \end{cases}$$

2.5.7 Modèles basés sur Datalog Probabiliste

Datalog probabiliste (pDatalog) [Fuhr, 2000] est une extension de Datalog¹, basé sur les clauses de Horn sans symbole de fonction en argument des prédicats. Dans pDatalog, les faits et les règles sont préfixés par un poids de probabilité α , avec $0 < \alpha \leq 1$.

pDatalog a été utilisé dans la recherche d'information [Fuhr, 1995] [Lalmas et Bruza, 1998], où un document d est un ensemble de faits probabilistes de la forme $\alpha \textit{term}(d, t)$, qui signifie que le document d est indexé par le terme t et α correspond à la probabilité que le prédicat $\textit{term}(d, t)$ soit vrai. Fuhr interprète d'une manière équivalente α comme étant la probabilité que le document d se rapporte au terme t (d is about t). Il étend, en plus, la définition précédente de d , où les termes d'indexation de d peuvent être déduits en utilisant le prédicat *about* et les règles d'inférence suivantes :

$$\textit{about}(d, t) : \neg \textit{term}(d, t)$$

$$\textit{about}(d, t) : \neg \textit{link}(d, d') \wedge \textit{about}(d', t)$$

où le prédicat $\textit{link}(d, d')$ fait référence à ce que d et d' sont explicitement ou implicitement liés. Le processus de recherche (ou la décision $d \rightarrow q$) est défini comme étant une règle d'inférence. Cependant, le modèle définit deux formes de règles d'inférence en relation avec les connecteurs logiques entre les requêtes :

- Conjonction : $q = t \wedge t'$:

$$q(d) : \neg \textit{about}(d, t) \wedge \textit{about}(d, t')$$

- Disjonction : $q = t \vee t'$:

$$q(d) : \neg \textit{term}(d, t)$$

$$q(d) : \neg \textit{term}(d, t')$$

où d est une variable, t et t' sont des constantes.

2.5.8 Les Méta-Modèles

L'objectif de l'élaboration d'un méta-modèle pour la recherche d'information est d'étudier formellement, dans un cadre unifié, les propriétés et les caractéristiques des systèmes de recherche d'information. L'avantage est qu'il sera alors possible de comparer les systèmes de recherche d'information.

L'origine de l'utilisation des preuves logiques pour conduire formellement le processus de recherche en RI remonte aux travaux de Nie [1989], où il a été montré que le modèle logique basé la logique modale est une forme générale des autres modèles. L'idée a ensuite, été approfondie dans [Bruza, 1994][Huibers, 1996], où les

1. Datalog est un langage de requête et de règles pour les bases de données déductives basé sur la logique des prédicats.

auteurs ont proposé un cadre formel dans lequel les différents modèles peuvent être théoriquement exprimés, étudiés et comparés. Ce cadre formel définit la relation “à propos”, notée $| \sim$ (*aboutness relationship*). Étant donnés deux objets a et b , $a| \sim b$ signifie que l'objet a est à propos de l'objet b . Plusieurs axiomes sont alors définis par les auteurs pour caractériser les modèles de RI. À titre d'exemple, nous citons ci-dessous quelques un de ces axiomes :

- i) Réflexivité : $a| \sim a$;
- ii) Symétrie : si $a| \sim b$ alors $b| \sim a$;
- iii) Transitivité : si ($a| \sim b$ et $b| \sim c$) alors $a| \sim c$;

La plupart des modèles de recherche d'information semblent satisfaire la réflexivité, car si les objets sont des documents, alors si un document est soumis en tant que requête, ce document doit être retrouvé. Le modèle d'espace vectoriel satisfait la symétrie. Pour être formellement étudié, un modèle de recherche d'information doit être exprimé dans le cadre ci-dessus. Les objets et la relation de base doivent être instanciés. Pour le modèle Booléen, les objets sont des formules et $| \sim$ est l'implication matérielle. Le modèle résultant est basé sur la logique classique avec l'hypothèse du monde clos [Reiter, 1987]. Il est facile de prouver que ce modèle satisfait la réflexivité mais pas la symétrie.

Notons que dans Bruza [1994] les auteurs considèrent un système d'axiomes qui ne doit pas être interprété dans un sens logique strict, mais plutôt d'une manière plus intuitive .

2.6 Les Approches de RI basées sur l'ACF

Il y'a lieu de remarquer que la recherche d'information est l'une des applications naturelle de l'ACF. Ceci résulte de l'évidente analogie entre la relation *Objets* \times *Attributs* (notion étroitement à l'ACF) et la matrice d'incidence *Documents* \times *Termes* (notion étroitement liée à la RI). Godin et al. [1986] ont précocement mis en évidence l'intérêt de l'utilisation de l'ACF pour la RI. L'intension d'un concept formel est alors considéré comme une requête, et son extension est considérée comme un ensemble de documents pertinents potentiels. Par la suite, les premiers travaux [Carpineto et Romano, 1993; Godin et al., 1993] ont étudié la possibilité d'utiliser les treillis de concepts comme base au développement de système de recherche d'information. L'idée est de représenter le corpus de (une collection) documents sous la forme de contexte formel. Où les documents sont les objets, les termes d'indexation sont les attributs du contexte formel, la relation *Objets* \times *Attributs* correspond à la matrice d'incidence *Documents* \times *Terms*. Chaque élément du treillis de concepts formels associé à ce contexte formel est vu comme un couple formé par une requête conjonctive Booléenne, dont les mots clés sont l'ensemble des termes contenus dans l'intension

du concept formel, et l'ensemble de documents pertinents pour cette requête sont les documents contenus dans l'extension du concept formel. Pour un concept formel donné tout document appartenant à son extension est considéré pertinent pour la requête conjonctive Booléenne formée par l'ensemble des termes de l'intension de ce concept formel. Et d'une manière duale, un document n'appartenant pas à l'extension de ce concept formel est considéré non pertinent pour la requête Booléenne associée.

L'évaluation d'une requête conjonctive donnée, dans la plupart des approches de RI qui s'appuient sur l'ACF, revient à rechercher dans le treillis, le concept formel dont l'intension contient tous les termes de cette requête. Les relations de spécialisation et de généralisation (sub-concept/super-concept) entre les concepts formels sont interprétées comme des changements mineurs sur les requêtes et permettent d'effectuer des recherches progressives dans le treillis. Ceci suppose que la requête existe déjà dans le treillis. L'insertion d'une requête dans ce treillis est réalisée par l'un des algorithmes de construction incrémentale de treillis de concepts. Ce mode de recherche est appelé "recherche par interrogation". Un deuxième mode, appelé "recherche par navigation", est utilisé dans certaines approches en exploitant la structure hiérarchique des treillis de concepts.

Le mode de recherche par interrogation utilise les algorithmes de construction incrémentale des treillis de concepts. La définition de requête consiste à spécifier directement les termes d'indexation qui décrivent les documents recherchés. Le concept formel (souvent virtuel) décrivant la requête est ensuite inséré dans le treillis. La recherche des documents pertinents revient à localiser le concept le plus général incluant les termes spécifiés dans la requête. Le mode de recherche par navigation exploite la structure hiérarchique des treillis. Il consiste à explorer les concepts formels en visualisant le diagramme de Hasse du treillis de concepts. Plusieurs approches se basent sur ces deux modes de recherche simultanément, la recherche se ramène ainsi *i)* à effectuer une combinaison libre de requête, résultant en un positionnement sur le concept le plus général incluant les termes de la requête, *ii)* et à explorer (naviguer) en suivant les liens entre les concepts du treillis.

Nous présentons dans le reste de cette section l'état de l'art des approches de RI basées sur l'ACF, cataloguées selon leurs domaines d'application. Notons que cet état de l'art est inespéré principalement des travaux de [Codochedo et Napoli \[2015\]](#) et de [Poelmans et al. \[2012\]](#).

2.6.1 Amélioration des Résultats de Requêtes

L'une des applications les plus naturelles de l'ACF et des treillis de concepts sous-jacents est le raffinement de requêtes et l'amélioration du résultat de recherche. Dans ce cadre, plusieurs travaux ont été menés comme les travaux de [\[Messai et al., 2006a\]](#) et [\[Martinovic et Gajdos, 2005\]](#). Les auteurs de ces travaux proposent des

techniques de raffinement et d'expansion de requêtes en s'appuyant sur les ontologies de domaines, ce qui permet l'amélioration du rappel par généralisation ou par spécialisation, tout en se basant sur la structure du treillis de concepts.

Le scénario typique de l'amélioration de résultats d'une requête [Godin et al., 1989, 1993] commence par la visualisation de la partie du treillis de concepts centrée sur la requête utilisateur, le treillis de concept associé à une collection de documents est considéré alors comme un espace de recherche qui peut être exploré par l'utilisateur. Ce dernier peut alors consulter les documents de chacun des noeuds (concepts formels) de la partie du treillis visualisée. Il peut, également, se déplacer à d'autres parties du treillis en formulant d'autres requêtes.

Cette technique a ensuite été dotée, dans [Carpineto et Romano, 1996], de la possibilité pour l'utilisateur de limiter et/ou de modifier dynamiquement le treillis de recherche en incluant de nouveaux termes d'indexation ou en excluant d'autres.

Dans le cadre de la recherche sur les résumés médicaux, Stojanovic [2005] a présenté une approche de navigation spatiale basée sur l'ACF pour le raffinement des requêtes. L'approche fournit à un utilisateur les requêtes qui sont "proches" de la requête initiale.

2.6.2 Web et Courrier Électronique

Le potentiel de visualisation et de compréhension des treillis de concepts formels ont été exploités par plusieurs auteurs pour le développement d'applications pour le Web et la messagerie électronique. L'idée principale est de créer un contexte formel à partir des résultats fournis par les moteurs de recherche sur le web, ou les messages électroniques d'un compte de messagerie, de construire le treillis correspondant à ce contexte, puis de proposer à l'utilisateur un classement (une classification) des résultats suivant le treillis de concepts formels relatif à ce contexte.

Généralement, pour une requête utilisateur, un moteur de recherche retourne une liste d'URLs, de titres et de descriptions pour chacun des documents jugés pertinents. Des applications², telles que CREDO [Carpineto et al., 2004], FooCA [Koester, 2006], SearchSleuth [Ducrou et Eklund, 2007] et CRECHAINDO [Nauer et Toussaint, 2009], ont été développées pour l'analyse et l'exploration de ces résultats. Ces applications construisent un contexte formel pour chaque requête utilisateur, les documents sont des objets, les mots clés des titres et des résumés correspondent aux attributs. Elles offrent à l'utilisateur une vue globale des résultats de sa requête et lui permettent de naviguer à travers les concepts formels. Kim et Compton [2004] ont présenté le système "KAnavigator" pour les petites communautés web des domaines spécialisés. Ce système offre la possibilité aux utilisateurs d'annoter les documents qu'ils jugent pertinents. Messai et al. [2005] ont utilisé les treillis de concepts pour la

2. Ces applications seront présentées dans la Section 2.6.7.

découverte et l'interrogation de ressources génomiques sur le web. Nauer et Toussaint [2009] proposent, également, une navigation interactive où l'utilisateur peut raffiner sa requête en sélectionnant les documents qu'il juge pertinents ou non. De cette manière, le treillis est dynamiquement modifié et reconstruit. Cigarrán et al. [2005] utilisent la sélection automatique des phrases nominales en tant que descripteurs³ de documents pour construire un treillis de concepts formels optimisé. L'optimisation de descripteurs de documents permet de produire des treillis de concepts plus petits, plus clairs qui offrent une navigation facile. Recio-García et al. [2006] utilisent l'ACF et les ontologies de domaine pour associer des annotations sémantiques aux pages Web. Les techniques d'appariement et de raisonnement à partir de cas (relatif à l'intelligence artificielle) peuvent être appliquées pour retrouver ces pages annotées en tant que cas. Liu et al. [2007] utilisent l'ACF pour optimiser un moteur de recherche d'informations (News) personnalisé afin d'aider les utilisateurs à obtenir rapidement le contenu d'informations dont ils ont besoin. Les applications *Conceptual Email Manager* [Cole et al., 2003] et *Mail-Sleuth* [Eklund et Wormuth, 2005] permettent aux utilisateurs de naviguer à travers leurs messages électroniques en construisant un treillis de concepts à partir de contextes formels, où les messages représentent les objets et les descripteurs des messages (en plus des mots clés du contenu des messages) représentent les attributs.

2.6.3 Logiciel, Image et Base de Connaissances

Peng et al. [2005] présentent une méthode pour générer un treillis de concepts représentant des relations conceptuelles entre les services Web et retrouver avec précision les services Web à partir de ce treillis. Bruno et al. [2005] proposent une approche combinant l'ACF et Support Vector Machines pour identifier automatiquement les concepts clés dans la documentation textuelle des services Web, construire un treillis de concepts basé sur ces annotations de service et classer les services dans des domaines spécifiques. Poshyvanyk et Marcus [2007] ont combiné l'ACF et l'Indexation Sémantique Latente (LSI) pour la localisation des concepts (descripteurs ou mots clés) de RI dans le code source d'Eclipse. LSI est utilisé pour associer les concepts exprimés dans les requêtes aux parties pertinentes du code source. Le résultat est fourni sous forme d'une liste classée des éléments du code source, organisée dans un treillis de concepts formels. Muangon et Intakosum [2009] ont combiné l'ACF avec le raisonnement à partir de cas (CBR) pour choisir des *modèles de patron* de conception appropriés pour les problèmes de conception spécifique. Cette approche résout certains problèmes des méthodes de recherche de patrons de conception en utilisant la recherche avec des mots clés. Peng et Zhao [2007] proposent une méthode pour la construction incrémentale d'une ontologie de recherche de composants logi-

3. Un descripteur permet de caractériser les informations contenues dans un document.

ciel basée sur l'ACF. L'ontologie contient les caractérisations des composants stockés dans les dépôts de fournisseurs.

Ahamd et Jang [2003] construisent des treillis de concepts à partir de descriptions associées à des images. Les images pertinentes, une fois localisées sur le treillis, sont ensuite recherchées dans une bases de données. Le projet ImageSleuth [Ducrou et al., 2006] utilise l'ACF pour le classification (clustering) et la navigation à travers des collections d'images annotées. Le diagramme de Hasse associé au treillis de concepts formels n'est pas directement montré à l'utilisateur.

La vue rendue à l'utilisateur est composée du concept formel relatif à la requête et une liste de voisins supérieurs et inférieurs. L'extension de ce concept formel contient des miniatures d'images, et son intention contient des descriptions d'image.

Le système DVDSleuth [Ducrou, 2007] construit un espace d'information dans le site de vente en ligne *Amazon.com* et utilise l'ACF pour découvrir des DVDs similaires et explorer leur voisinage. Amato et Meghini [2008] partent d'une image initiale fournie par l'utilisateur et utilisent un treillis de concepts pour retrouver des images similaires. Les attributs dans ce treillis sont des facettes qui correspondent aux critères de similarités d'image (texture, couleur, forme, ...). Les valeurs dans le contexte formel indiquent pour chaque facette la similarité des images de la base de données par rapport à l'image fournie par l'utilisateur. Cette approche offre, également, la possibilité à l'utilisateur de naviguer entre les concepts formels.

Dans le domaine des base de données, Ducrou et al. [2005] ont développé une application, appelé D-SIFT, qui permet l'exploration des schémas d'une BDD relationnelle. Tane et al. [2006] ont introduit la théorie de requête multi-contexte. En utilisant l'ACF, Cette théorie permet de définir un espace virtuel de vues (pertinentes pour l'utilisateur) de base de connaissance. Hachani et al. [2009] ont proposé une approche coopérative, basée sur l'ACF et la logique floue, pour traiter les réponses vides des requêtes conjonctives floues. Pour une requête avec une réponse vide, l'approche génère les sous-requêtes les plus proches avec des réponses non-vides.

2.6.4 Traitement de Requêtes sur des Données Hétérogènes

Plusieurs travaux ont été entrepris pour définir et traiter des requêtes complexes et pour intégrer des données provenant de sources hétérogènes. Sampaio de Souza et Davis [2004] utilisent l'ACF pour traiter les requêtes des utilisateurs sur un ensemble d'ontologies superposées, qui ont été créées par des groupes indépendants. Messai et al. [2005] présentent une approche basée sur l'ACF guidée par les connaissances de domaine pour la découverte de sources de données biologiques sur le Web. Les ontologies de domaine sont utilisées pour représenter un ensemble de connaissances qui reflètent le contenu et la qualité des sources de données. Les ressources génomiques sont organisées dans un treillis de concepts en fonction de leurs caractéristiques communes. Le treillis de concept constitue le support de la découverte

par navigation ou par interrogation. Nafkha et Jaoua [2005] proposent un système fondé sur l'ACF pour la recherche d'objets similaires dans des sources d'information hétérogènes. Polaillon et al. [2007] ont présenté une approche de recherche d'information qui utilise l'ACF en conjonction avec la sémantique pour fournir des réponses contextuelles aux requêtes des utilisateurs. L'utilisateur formule une requête sur un ensemble de sources de données hétérogènes, cet ensemble est sémantiquement unifié par la notion de "contexte conceptuel" et fournit ainsi un espace commun de navigation à l'utilisateur. Ceravolo et al. [2007] utilisent l'ACF pour correspondre et mapper des éléments provenant de sources de données hétérogènes à une ontologie commune.

2.6.5 Réponses Contextuelles et Classement des Résultats

Le principal avantage de l'ACF pour la RI est la possibilité de fournir des réponses contextuelles aux requêtes de l'utilisateur [Carpineto et Romano, 2005]. Un nombre important de travaux en ACF-RI traite de la contextualisation de résultats d'une requête. D'une manière générale, tous les travaux utilisant les ontologies de domaine ou la classification de documents [Sampaio de Souza et Davis, 2004; Nafkha et Jaoua, 2005; Tane et al., 2006; Polaillon et al., 2007; Ceravolo et al., 2007] traitent (implicitement) la contextualisation de résultats. Cependant, le classement des résultats de requête reste un objet non (ou presque non) traité dans la littérature. Cela est dû au fait que ces travaux se basent sur l'analyse de concepts formels classique. Or, l'ACF classique ne permet pas d'associer des poids et des degrés de pertinence aux objets et/ou attributs. Néanmoins, quelques travaux utilisent le classement par rapport au treillis de concepts (*concept lattice-based ranking*) [Carpineto et Romano, 1998], c'est à dire, le classement des résultats d'une requête se fait selon la distance dans le treillis par rapport au concept formel le plus pertinent. L'idée de base de ce classement est que les documents d'un corpus peuvent être organisés dans un treillis de concepts selon les termes d'indexation partagés.

La première utilisation de ce principe remonte aux travaux de Carpineto et Romano [1998] dans le système REFINER, où la requête utilisateur et ses résultats sont représentés par un concept formel virtuel dans le treillis de concepts formels. Une série d'anneaux concentriques autour de ce concept virtuel, sont ajoutés sur le diagramme de Hasse du treillis. Les anneaux représentent les distances des documents de la collection par rapport à la requête utilisateur. Messai et al. [2008] classent en partie l'ensemble des attributs en fonction de leur importance. Cette hiérarchie représente la connaissance du domaine utilisée pour l'amélioration (raffinement) des requêtes et la navigation sur le treillis de concepts. Les hiérarchies des attributs sont utilisées pour définir des requêtes complexes contenant des attributs de différents niveaux d'importance. Afin de faciliter la navigation dans la collection de documents, Zhang et al. [2008] proposent une méthode, basée sur l'ACF, pour construire une

hiérarchie à deux niveaux en réponse à une requête. Une fois que tous les concepts formels sont extraits à l'aide de l'ACF, les concepts les plus pertinents pour la requête sont alors mises en avant. [Ignatov et Kuznetsov \[2009\]](#) utilisent l'ACF pour détecter les documents dupliqués (ou presque dupliqués) dans les résultats de la recherche Web en calculant les cluster de documents similaire. Récemment, [Codocedo et al. \[2014\]](#) ont présenté un système de classement de documents en utilisant des notions de raisonnement à partir de cas. Cette approche inspirée du classement basé sur les treillis de concepts utilise le treillis de concepts pour trouver des *modifications de requêtes* appropriées à travers des éléments pivots appelés *concepts cousins*. Les modifications de requêtes sont évaluées en fonction de la distance sémantique à la requête originale donnant, ainsi, un classement automatique des documents.

2.6.6 Réinjection de Pertinence

Un facteur important dans l'efficacité d'un système de recherche d'information est d'aider l'utilisateur à réduire ce que nous appelons "*l'écart cognitif*". Cet écart décrit la distance entre l'espace d'information recherché par l'utilisateur et l'espace occupé par les capacités à décrire ses besoins en information [[Ingwersen, 1996](#)]. Afin d'optimiser cet écart, [Rocchio \[1971\]](#) a introduit le mécanisme de Réinjection de Pertinence "Relevance Feedback", qui peut être vue comme un système de «calibrage de requête» en utilisant des entrées d'utilisateurs supplémentaires.

Une approche de Réinjection de Pertinence dans la RI basée sur l'ACF a été proposée par [Carpineto et Romano \[1998\]](#) dans leur système REFINER. Pour une requête utilisateur, le système génère un "concept virtuel dans le treillis de concepts". En utilisant les sub-concepts et les super-concepts du concept virtuel, REFINER est en mesure de proposer des raffinements et des élargissements minimaux pour la requête utilisateur. [Grootjen et van der Weide \[2002\]](#) ont proposé une approche presque similaire appelé "Réinjection de Pertinence Conceptuelle" (*conceptual relevance feedback*). Plus tard les même auteurs [[Grootjen et van der Weide, 2006](#)] ont réadapté cette approche en une méthode d'expansion de requête, ils ont montré que le treillis de concept résultant après une requête à deux termes, contient des raffinements qui fonctionnent aussi bien qu'une formulation plus précise de la requête.

[Nauer et Toussaint \[2007\]](#) ont présenté une technique de réinjection de pertinence dynamique, appelé Réinjection de Pertinence Explicite, appliquée sur un moteur de recherche Web standard (tel que Google). La technique consiste à itérer sur le contexte formel initial par des procédures d'extension et de réduction. Des extensions du contexte formel sont réalisées à chaque fois que l'utilisateur soumet une nouvelle requête ou donne une évaluation positive sur les résultats retournés. Les réductions sont effectuées à chaque fois que l'utilisateur donne une évaluation négative. Cette technique de réinjection de pertinence explicite est aussi prise en charge par l'approche développée, pour la recherche d'images, par [Martinez et Loi-](#)

sant [2002]. Ducrou et al. [2006] ont proposé la réinjection de pertinence implicite. Les auteurs utilisent une procédure appelée “Requête par l'exemple”. Au lieu que l'utilisateur évalue explicitement la pertinence des documents, la requête est modifiée par un ensemble de terme d'indexation créé manuellement par l'utilisateur.

2.6.7 Les Systèmes de RI basés sur l'ACF

Nous présentons dans cette section un aperçu de l'ensemble des principales applications et systèmes de recherche d'information basés sur l'analyse de concepts formels.

FaIR [Priss, 2000] : Il s'agit d'un système de facettes⁴ dans lequel les concepts formels de documents et de descripteurs sont mappés aux entrées de dictionnaire (thésaurus). FaIR dispose d'un langage de requête superposant l'ensemble des concepts formels avec les opérateurs logiques *et*, *ou*, *non*.

REFINER [Carpineto et Romano, 1998] : Il combine la recherche d'informations Booléenne et la navigation basée sur le contenu en utilisant les treillis de concepts. Lors de la soumission d'une requête, REFINER crée et affiche une partie du treillis de concepts associée aux documents recherchés, et centrés autour de cette requête. Le treillis de concepts affiché par REFINER montre le résultat de la requête avec l'ensemble de changements minimaux possibles (raffinement/élargissement de la requête initiale).

CREDO [Carpineto et al., 2004] : Il se présente comme une interface superposant les moteurs de recherche Web (Google ou Yahoo). Il implémente principalement les fonctions d'expansion de requêtes présentées dans REFINER.

JBrainDead [Cigarrán et al., 2004] : Il combine dynamiquement un moteur de recherche de texte libre avec l'ACF pour organiser les résultats d'une requête. Ce système n'est pas restreint à des domaines spécifiques et n'utilise pas de descripteurs attribués manuellement pour des documents ni des dictionnaires spécifiques au domaine. En exploitant la liste classée des documents retournés par un moteur de recherche, le système détermine dynamiquement les attributs les plus appropriés pour l'ensemble des documents et génère le treillis de concepts correspondant.

Mail-Sleuth [Eklund et Wormuth, 2005] : Est un outil commercial de recherche et de navigation dans les collections de courriers électroniques personnels sous l'hypothèse que les utilisateurs sont en mesure de gérer un diagramme de Hasse du treillis de concepts. Les auteurs ont étendu cette idée à différents domaines d'application : ImageSleuth [Ducrou et al., 2006] pour la navigation

4. La recherche à facettes donne aux utilisateurs les moyens de filtrer une collection de données en choisissant un ou plusieurs critères (les facettes).

et la recherche d'images, DVDSleuth [Ducrou, 2007] pour l'exploration de catalogues Web de DVD, SearchSleuth [Ducrou et Eklund, 2007] implémente une interface qui étend un moteur de recherche Web standard et AnnotationSleuth [Eklund et Ducrou, 2009] pour l'exploration de collection de musée virtuel.

FooCA [Koester, 2006] : Il utilise les titres, les descriptions et l'URL des documents retournés par le moteur de recherche pour créer un contexte formel et le treillis de concepts associé. Ce contexte formel généré pour une requête peut être exploré et modifié de façon itérative.

BR-Explorer [Messai et al., 2006b] : Initialement BR-Explorer construit un contexte formel représentant la relation entre les sources de données biologiques et leurs méta-données, ainsi que le treillis de concepts correspondant. Pour une requête donnée, BR-Explorer génère le concept formel associé à cette requête et l'insère dans le treillis de concepts construit précédemment. BR-Explorer fournit, alors, l'ensemble des sources de données pertinentes classées en fonction de leur pertinence par rapport à la requête considérée. Une procédure d'affinement de requête basée sur une ontologie (de données biologiques) est intégrée dans ce système.

Camelis [Ferré, 2007] : Il est basé sur une généralisation logique de l'ACF, appelée analyse de concepts logiques (*Logical Concept Analysis-LCA*), où les attributs sont remplacés par des formules logiques. Il offre aux utilisateurs une organisation dynamique de photos à partir des métadonnées, et couvre les aspects suivants : (1) un langage expressif pour décrire des photos et interroger la collection (Logical Information System Query Language), (2) les métadonnées (manuelles et automatiques) sont intégrées et (3) une recherche hybride qui permet l'interrogation (requête) de la collection et la navigation dans la collection.

CreChainDo [Nauer et Toussaint, 2009] : Est un système implémentant la réinjection de pertinence explicite, où l'utilisateur peut évaluer positivement ou négativement les résultats d'une recherche sur le Web. C'est ce qui fait de CreChainDo un système itératif et interactif pour la recherche sur Internet en fournissant à l'utilisateur un résultat structuré et synthétique organisé dans un treillis de concept, ainsi que la possibilité d'évaluer les résultats de recherche d'une requête. Ces choix d'utilisateurs sont convertis en opérations sur le treillis de concepts. Le treillis de concepts est mis à jour de manière dynamique pour une meilleure adaptation aux demandes d'utilisateurs.

2.6.8 Les Approches de RI basées sur l'ACF Floue

Djouadi a proposé dans [Djouadi, 2011, 2012] une approche de traitement des requêtes pondérées basée sur l'extension floue de l'ACF. L'approche repose sur un

contexte formel flou $(\mathcal{D}, \mathcal{T}, \mathcal{R})$, où D est l'ensemble de documents, \mathcal{T} est l'ensemble de termes d'indexation et \mathcal{R} une relation binaire floue dans $L^{\mathcal{D} \times \mathcal{T}}$ qui associe à toute paire $(d, t) \in \mathcal{D} \times \mathcal{T}$ le poids $(\mathcal{R}(d, t))$ du terme t dans le document d .

L'approche utilise la lecture possibiliste de l'ACF (§1.4) pour le traitement des différentes formes de requêtes. Ainsi, pour un ensemble flou de termes $\tilde{T} = \{t_1^{w_1}, t_2^{w_2}, \dots, t_k^{w_k}\}$:

- L'ensemble de documents flous \tilde{D} satisfaisant la requête conjonctive pondérée $\hat{Q}(\tilde{T})$, de la forme $t_1^{w_1} \wedge t_2^{w_2} \wedge \dots \wedge t_k^{w_k}$, est donné par \tilde{T}^Δ .
- L'ensemble de documents flous \tilde{D} satisfaisant la requête disjonctive pondérée $\check{Q}(\tilde{T})$, de la forme $t_1^{w_1} \vee t_2^{w_2} \vee \dots \vee t_k^{w_k}$, est donné par \tilde{T}^Π .

Pour l'expression de la négation dans les requêtes, Djouadi a démontré que que l'ensemble \tilde{D} de documents satisfaisant la requête $\hat{Q}(\tilde{T})$ (resp. $\check{Q}(\tilde{T})$) est donnée par $\tilde{T}^{\frac{\Pi}{\mathcal{R}}}$ (resp. $\tilde{T}^{\frac{\Delta}{\mathcal{R}}}$), où $\frac{\Pi}{\mathcal{R}}$ et $\frac{\Delta}{\mathcal{R}}$ font référence respectivement aux opérateurs Π et Δ appliqués à la relation $\overline{\mathcal{R}}$ complémentaire de \mathcal{R} (i.e $\overline{\mathcal{R}}(d, t) = 1 - \mathcal{R}(d, t)$).

Pour une requête Q et sa réponse \tilde{D} , l'approche considère le degré d'appartenance d'un document d à l'ensemble \tilde{D} (i.e $\tilde{D}(d)$) comme étant le score de satisfaction de ce document à la requête Q . Par conséquent, les documents satisfaisants la requête Q peuvent être classés par ordre décroissant de leurs scores.

Maio et al. [2012] ont proposé une approche de recherche hiérarchique sur les ressources web. Les auteurs se basent sur l'extension floue à seuil de l'ACF pour construire un treillis de concepts formels classique, puis ils remplacent les intensions de ces concepts par des sous-ensembles flous. Un concept formel classique $\langle D, T \rangle$ est remplacé dans le treillis par le concept $\langle \tilde{D}, T \rangle$ tel que :

$$\tilde{D}(d) = \begin{cases} \min_{t \in T} \mathcal{I}(d, t) & \text{si } d \in D \\ 0 & \text{sinon} \end{cases}$$

La même approche a été utilisée par la suite dans [Maio et al., 2015] pour le traitement de requêtes conjonctives en ligne sur des microblogs. Il est remarquable que cette approche se limite aux requêtes conjonctives non pondérées, et la discrétisation par seuil du contexte formel flou engendre une perte d'informations.

Dans une démarche similaire, une approche de RI basée sur l'ACF floue a été proposée dans [Aswani Kumar et al., 2015]. Elle génère des concepts formels flous $\langle D, \tilde{T} \rangle$, où D est un ensemble ordinaire de documents, et \tilde{T} est un ensemble flou de termes. Cela permet de traiter des requêtes conjonctives pondérées, cependant elle ne permet pas le classement des résultats de recherche.

2.7 Conclusion

Dans ce chapitre nous avons donné une description exhaustive de l'état de l'art de la recherche d'information. Nous avons d'abord présenté les approches classiques (ensemblistes, algébriques et probabilistes) suivant la taxonomie de [Baeza-Yates et Ribeiro-Neto \[1999\]](#). Ensuite, nous avons détaillé les approches logiques de la RI ainsi que les approches de RI basées sur l'ACF.

De nombreux travaux ont porté sur l'utilisation de la logique pour modéliser le processus de recherche d'information. L'objectif principal, de ces travaux, est de créer des modèles expressifs et uniformes capables d'améliorer l'efficacité de la recherche. La diversité de ces approches reflète l'importance des formalismes logiques dans la modélisation de la RI. Néanmoins, aucun consensus n'a été atteint quant à savoir quelle est la meilleure technique [[Lalmas et Bruza, 1998](#)]. Toutefois, nous avons remarqué que :

- À l'exception du méta-modèle de [Bruza \[1994\]](#), ces approches sont des modèles (sémantiques ou interprétation) des logiques considérées.
- Le méta-modèle de [Bruza \[1994\]](#) définit des axiomes intuitifs (i.e. ne sont pas formels).
- L'incertitude $U(d \rightarrow q)$ n'est pas intrinsèquement évaluée dans le processus d'inférence, mais elle est plutôt calculée en utilisant des fonctions ad-hoc et externes à ces logiques.

Il existe donc une demande pour le développement d'un nouveau formalisme capable de modéliser les systèmes de RI de manière plus générique. Pour cela, nous proposons dans le [Chapitre 3](#) une théorie axiomatique, et nous démontrons qu'elle est satisfaite par trois modèles classiques de la RI, à savoir le modèle Booléen, le modèle flou standard et le modèle vectoriel.

L'ACF est centrée sur la notion de concept formel, qui est un regroupement d'objets et de leurs attributs communs. Cette notion est à la base de l'utilisation de l'ACF dans le domaine de la recherche d'information, où on essaye de regrouper des documents et leurs descripteurs ou leurs termes d'indexation. Le potentiel de l'ACF dans l'organisation en treillis de concepts formels a également trouvé usage dans la recherche d'information, où les treillis sont exploités pour le raffinement ou l'expansion de requêtes. Il est remarquable que la plupart des approches de RI basées sur l'ACF supposent que :

- La relation $Documents \times Termes$ est Booléenne ;
- Les requêtes sont aussi Booléennes et elles ne permettent pas de considérer des requêtes pondérées ;

- Les requêtes sont limitées à la forme conjonctive, à l'exception du système *FaIR* [Priss, 2000], qui superpose le treillis de concepts formels par un langage de requête.
- Les résultats d'une recherche ne sont pas ordonnés vu qu'aucun score n'est associé aux documents pertinents.

Sur la base de l'ensemble de ces remarques,

1. Nous proposons dans le Chapitre 4, un enrichissement du modèle Booléen en s'appuyant, d'un coté, sur la théorie axiomatique (présenté dans le Chapitre suivant) et, de l'autre coté, sur une lecture possibiliste de l'ACF,
2. Nous proposons dans le Chapitre 5, certaines extensions floues de l'ACF pour la RI, qui permettront *i*) de considérer des relations *Documents* \times *Termes* floues, *ii*) de classer les résultats de recherche et *iii*) de traiter des requêtes pondérées. En s'appuyant sur une lecture possibiliste de l'ACF floue nous prendrons en considération les formes disjonctives et négatives des requêtes.

Chapitre 3

Théorie Multi-Sorte pour la Recherche d'Information

Sommaire

3.1	Introduction	59
3.2	Théorie Multi-Sorte	60
3.2.1	Vocabulaire	60
3.2.2	Axiomes	61
3.2.3	Inférence	61
3.3	Modèle Booléen	63
3.4	Modèle basé sur les Ensembles Flous	68
3.5	Modèle Vectoriel	72
3.6	Conclusion	76

3.1 Introduction

Dans de nombreuses branches des mathématiques et de l'informatique, on formalise des déclarations sur plusieurs types d'objets. Ainsi, les langages logiques et les structures utilisées pour les interpréter sont conçus comme étant multi-sortes ; c'est-à-dire que l'ensemble des variables du langage s'étend sur plus d'un univers de domaine d'objets. La notion de multi-sorte¹ (*many-sorted*) a été introduite par Wang [1952] pour étudier des systèmes axiomatiques implicants des objets de différents types ou catégories.

Il existe de nombreux exemples où la sémantique des formules sont des structures multi-sortes ; En géométrie, pour prendre un exemple simple et ancien, on utilise des univers de points, de lignes, d'angles, de triangles, etc. Dans la théorie du type simple on utilise un univers pour les individus, plus des univers d'ensembles d'individus, des ensembles d'ensembles d'individus, etc. Dans la logique de second ordre, nos besoins sont : un univers pour les individus, un univers pour une relation unaire sur les individus, un univers pour les relations binaires sur les individus, etc. En raisonnement sur les calculs (Reasoning about Computation), on utilise un univers d'états, où un état peut être un ensemble de registres de mémoire utilisés par l'ordinateur, et un autre univers pour le temps. On peut ajouter des univers pour des constructions syntaxiques telles que des programmes et leurs spécifications.

En se basant sur cette notion de “multi-sortes”, nous proposons dans ce chapitre, une théorie axiomatique pour la RI. Ce choix est motivé par les différents types d'objets considérés (manipulés et utilisés) dans les système de RI, à savoir les documents, les termes d'indexation et les requêtes. L'objectif principal de la théorie est de définir des règles syntaxiques correspondant aux axiomes conduisant le processus de RI comme preuve de théorèmes. Contrairement aux approches logiques existantes de la RI, notre théorie (contribution) définit un système d'axiomes formels et permet, dans certaines interprétations, de déduire explicitement les scores de pertinence document-requête.

Le reste du chapitre est organisé comme suit. Dans la section 3.2, nous présentons le vocabulaire, les axiomes et le système d'inférence de la théorie. Nous démontrons, respectivement dans les sections 3.3, 3.4 et 3.5, que le modèle Booléen de Salton [1968], le modèle flou standard de Bookstein [1980] et le modèle vectoriel [Salton et al., 1975] satisfont notre théorie, prouvant ainsi sa cohérence.

1. Sorte (Larousse) : Caractéristique commune qui définit, à l'intérieur d'un ensemble d'êtres ou de choses de même nature, un groupe, un genre, un type identique.

3.2 Théorie Multi-Sorte

La théorie axiomatique multi-sorte (dénnoté par \mathfrak{T}) que nous proposons dans ce chapitre est définie sur un univers de discours multi-sorte $\mathfrak{D} = \{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$ où \mathcal{X}, \mathcal{Y} et \mathcal{Z} sont des ensembles quelconques avec $\mathcal{X} \subseteq \mathcal{Z}$. Le choix d'une logique multi-sorte est motivé par le fait que la recherche d'information traite des objets de différents types, à savoir, des documents, des termes d'indexation et des requêtes. Il y'a lieu aussi de constater que même ces objets peuvent avoir différentes représentations. Par exemple un document est représenté différemment d'un modèle à un autre (un ensemble de termes dans le modèle Booléen, un vecteur dans le modèle vectoriel). Dans la suite de cette section, nous présentons le vocabulaire et les axiomes de \mathfrak{T} et nous montrons son mécanisme d'inférence.

3.2.1 Vocabulaire

Le langage de la théorie \mathfrak{T} est l'ensemble de formules bien formées construites sur la base de l'alphabet illustré dans le Tableau 3.1

Tableau 3.1 – Alphabet de la Théorie \mathfrak{T}

Constantes	$y_{\perp} \in \mathcal{Y}, z_{\perp} \in \mathcal{Z}$
Variables	$x, y, y_1, y_2, z, z_1, z_2$
Connecteurs	\wedge, \Rightarrow
Fonctions	$\oplus, \otimes : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathcal{Y}$ $\sqcap, \sqcup : \mathcal{Z} \times \mathcal{Z} \rightarrow \mathcal{Z}$ $\ominus : \mathcal{Y} \rightarrow \mathcal{Y}$ $\sim : \mathcal{Z} \rightarrow \mathcal{Z}$ $f : \mathcal{X} \rightarrow \mathcal{Y}$
Prédicats	$\mathbf{S}(\cdot, \cdot)$

Les formules de \mathfrak{T} , sont inductivement construites suivant les règles ci-dessous. Pour la commodité de lecture, nous utilisons la notation $v : \mathcal{S}$, qui signifie que la variable v est de sorte \mathcal{S} (i.e. $v \in \mathcal{S}$).

- Chaque variable est un terme,
- Si t est un terme alors $f(t)$ est un terme,
- Si t_1, t_2 sont des termes alors $\ominus t_1, t_1 \otimes t_2, t_1 \oplus t_2$ sont des termes,
- Si t_1, t_2 sont des termes alors $\sim t_1, t_1 \sqcap t_2, t_1 \sqcup t_2$ sont des termes,
- Si t_1, t_2 sont des termes alors $\mathbf{S}(t_1, t_2)$ est une formule atomique,

- Chaque formule atomique est une formule bien formée,
- Si A et B sont des formules bien formées alors $A \wedge B$ et $A \Rightarrow B$ sont des formules bien formées,
- Si v est une variable, \mathcal{S} est une sorte et A une formule bien formée, alors $\forall v : \mathcal{S}, A$ est une formule bien formée.

3.2.2 Axiomes

La théorie \mathfrak{T} est composée des cinq axiomes suivants :

$$(A1) \mathbf{S}(y_{\perp}, z_{\perp})$$

$$(A2) \forall x : \mathcal{X} \mathbf{S}(f(x), x)$$

$$(A3) \forall y_1, y_2 : \mathcal{Y}, \forall z_1, z_2 : \mathcal{Z} (\mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2) \Rightarrow \mathbf{S}((y_1 \otimes y_2), (z_1 \sqcap z_2)))$$

$$(A4) \forall y_1, y_2 : \mathcal{Y}, \forall z_1, z_2 : \mathcal{Z} (\mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2) \Rightarrow \mathbf{S}((y_1 \oplus y_2), (z_1 \sqcup z_2)))$$

$$(A5) \forall y : \mathcal{Y}, \forall z : \mathcal{Z} (\mathbf{S}(y, z) \Rightarrow \mathbf{S}(\ominus y, \sim z))$$

La théorie est complétée par la règle d'inférence du Modus Ponens (MP) et la règle d'introduction de la conjonction (IC), données respectivement par :

$$(MP) \frac{A \quad A \Rightarrow B}{B}$$

$$(IC) \frac{A \quad B}{A \wedge B}$$

3.2.3 Inférence

La théorie proposée \mathfrak{T} permet d'inférer de nouveaux théorèmes. En règle générale, la démonstration d'un théorème ϕ peut être considérée comme le résultat d'une séquence finie de formules $\psi_1, \psi_2, \dots, \psi_n$ où $\psi_n = \phi$ et chaque formule ψ_i est soit un axiome, soit elle est déduite des formules précédentes en utilisant la règle d'introduction de la conjonction ou la règle d'inférence du Modus Ponens. En conséquence, et comme les règles IC et MP conservent la validité des formules inférées, si tous les axiomes sont vrais dans une interprétation alors un simple raisonnement par déduction sur la séquence montre que ψ_n est vraie dans cette interprétation.

La proposition suivante donne des théorèmes particuliers de \mathfrak{T} qui seront utilisés dans le reste du document.

Proposition 3.2.1. *Les formules ϕ_1, ϕ_2, ϕ_3 et ϕ_4 sont des théorèmes de \mathfrak{T} :*

$$(\phi_1) \forall x_1, x_2, \dots, x_r : \mathcal{X}, \mathbf{S}(f(x_1) \otimes f(x_2) \otimes \dots \otimes f(x_r), x_1 \sqcap x_2 \sqcap \dots \sqcap x_r)$$

$$(\phi_2) \forall x_1, x_2, \dots, x_r : \mathcal{X}, \mathbf{S}(f(x_1) \oplus f(x_2) \oplus \dots \oplus f(x_r), x_1 \sqcup x_2 \sqcup \dots \sqcup x_r)$$

$$(\phi_3) \forall x_1, x_2, \dots, x_r : \mathcal{X}, \mathbf{S}(\ominus(f(x_1) \otimes f(x_2) \otimes \dots \otimes f(x_r)), \sim (x_1 \sqcap x_2 \sqcap \dots \sqcap x_r))$$

$$(\phi_4) \forall x_1, x_2, \dots, x_r : \mathcal{X}, \mathbf{S}(\ominus(f(x_1) \oplus f(x_2) \oplus \dots \oplus f(x_r)), \sim (x_1 \sqcup x_2 \sqcup \dots \sqcup x_r))$$

Preuve.

(ϕ_1)

$$\begin{array}{c}
\mathbf{S}(f(x_1), x_1) \quad (A2) \\
\mathbf{S}(f(x_2), x_2) \quad (A2) \\
\hline
\mathbf{S}(f(x_1), x_1) \wedge \mathbf{S}(f(x_2), x_2) \quad (IC) \\
\mathbf{S}(f(x_1), x_1) \wedge \mathbf{S}(f(x_2), x_2) \Rightarrow \mathbf{S}(f(x_1) \otimes f(x_2), x_1 \sqcap x_2) \quad (A3) \\
\hline
\mathbf{S}(f(x_1) \otimes f(x_2), x_1 \sqcap x_2) \quad (MP) \\
\mathbf{S}(f(x_3), x_3) \quad (A2) \\
\hline
\mathbf{S}(f(x_1) \otimes f(x_2), x_1 \sqcap x_2) \wedge \mathbf{S}(f(x_3), x_3) \quad (IC) \\
\mathbf{S}(f(x_1) \otimes f(x_2), x_1 \sqcap x_2) \wedge \mathbf{S}(f(x_3), x_3) \Rightarrow \\
\mathbf{S}(f(x_1) \otimes f(x_2) \otimes f(x_3), x_1 \sqcap x_2 \sqcap x_3) \quad (A3) \\
\hline
\mathbf{S}(f(x_1) \otimes f(x_2) \otimes f(x_3), x_1 \sqcap x_2 \sqcap x_3) \quad (MP) \\
\vdots \\
\hline
\mathbf{S}(f(x_1) \otimes f(x_2) \otimes f(x_3) \otimes \cdots \otimes f(x_r), x_1 \sqcap x_2 \sqcap x_3 \sqcap \cdots \sqcap x_r) \quad (MP)
\end{array}$$

(ϕ_2)

$$\begin{array}{c}
\mathbf{S}(f(x_1), x_1) \quad (A2) \\
\mathbf{S}(f(x_2), x_2) \quad (A2) \\
\hline
\mathbf{S}(f(x_1), x_1) \wedge \mathbf{S}(f(x_2), x_2) \quad (IC) \\
\mathbf{S}(f(x_1), x_1) \wedge \mathbf{S}(f(x_2), x_2) \Rightarrow \mathbf{S}(f(x_1) \oplus f(x_2), x_1 \sqcup x_2) \quad (A4) \\
\hline
\mathbf{S}(f(x_1) \oplus f(x_2), x_1 \sqcup x_2) \quad (MP) \\
\mathbf{S}(f(x_3), x_3) \quad (A2) \\
\hline
\mathbf{S}(f(x_1) \oplus f(x_2), x_1 \sqcup x_2) \wedge \mathbf{S}(f(x_3), x_3) \quad (IC) \\
\mathbf{S}(f(x_1) \oplus f(x_2), x_1 \sqcup x_2) \wedge \mathbf{S}(f(x_3), x_3) \Rightarrow \\
\mathbf{S}(f(x_1) \oplus f(x_2) \oplus f(x_3), x_1 \sqcup x_2 \sqcup x_3) \quad (A4) \\
\hline
\mathbf{S}(f(x_1) \oplus f(x_2) \oplus f(x_3), x_1 \sqcup x_2 \sqcup x_3) \quad (MP) \\
\vdots \\
\hline
\mathbf{S}(f(x_1) \oplus f(x_2) \oplus f(x_3) \oplus \cdots \oplus f(x_r), x_1 \sqcup x_2 \sqcup x_3 \sqcup \cdots \sqcup x_r) \quad (MP)
\end{array}$$

(ϕ_3)

$$\begin{array}{c}
\mathbf{S}(f(x_1) \otimes f(x_2) \otimes f(x_3) \otimes \cdots \otimes f(x_r), x_1 \sqcap x_2 \sqcap x_3 \sqcap \cdots \sqcap x_r) \quad (\phi_1) \\
\mathbf{S}(f(x_1) \otimes f(x_2) \otimes f(x_3) \otimes \cdots \otimes f(x_r), x_1 \sqcap x_2 \sqcap x_3 \sqcap \cdots \sqcap x_r) \Rightarrow \\
\mathbf{S}(\ominus(f(x_1) \otimes f(x_2) \otimes f(x_3) \otimes \cdots \otimes f(x_r)), \sim (x_1 \sqcap x_2 \sqcap x_3 \sqcap \cdots \sqcap x_r)) \quad (A5) \\
\hline
\mathbf{S}(\ominus(f(x_1) \otimes f(x_2) \otimes f(x_3) \otimes \cdots \otimes f(x_r)), \sim (x_1 \sqcap x_2 \sqcap x_3 \sqcap \cdots \sqcap x_r)) \quad (MP)
\end{array}$$

(ϕ_4)

$$\begin{array}{c}
\mathbf{S}(f(x_1) \oplus f(x_2) \oplus f(x_3) \oplus \cdots \oplus f(x_r), x_1 \sqcup x_2 \sqcup x_3 \sqcup \cdots \sqcup x_r) \quad (\phi_2) \\
\mathbf{S}(f(x_1) \oplus f(x_2) \oplus f(x_3) \oplus \cdots \oplus f(x_r), x_1 \sqcup x_2 \sqcup x_3 \sqcup \cdots \sqcup x_r) \Rightarrow \\
\mathbf{S}(\ominus(f(x_1) \oplus f(x_2) \oplus f(x_3) \oplus \cdots \oplus f(x_r)), \sim (x_1 \sqcup x_2 \sqcup x_3 \sqcup \cdots \sqcup x_r)) \quad (A5) \\
\hline
\mathbf{S}(\ominus(f(x_1) \oplus f(x_2) \oplus f(x_3) \oplus \cdots \oplus f(x_r)), \sim (x_1 \sqcup x_2 \sqcup x_3 \sqcup \cdots \sqcup x_r)) \quad (MP)
\end{array}$$

□

Dans les sections qui suivent, nous proposons trois sémantiques associées à la théorie \mathfrak{T} . Ces sémantiques sont basées sur des approches de RI. Il sera également prouvé que chaque modèle de RI sous-jacent satisfait la théorie \mathfrak{T} prouvant ainsi sa cohérence.

3.3 Modèle Booléen

Étant donné que la matrice d'incidence liée à la RI et le contexte formel lié à l'ACF sont des notions équivalentes, nous préférons nous référer à cette dernière dans le cadre du modèle Booléen de RI.

Soit $\mathcal{K} = (\mathcal{D}, \mathcal{T}, \mathcal{R})$ un contexte formel avec $\mathcal{D} = \{d_1, d_2, \dots, d_m\}$ un ensemble fini de documents, $\mathcal{T} = \{t_1, t_2, \dots, t_n\}$ un ensemble fini de termes d'indexation et \mathcal{R} une relation binaire Booléen dans $\mathcal{D} \times \mathcal{T}$ ($(d_i, t_j) \in \mathcal{R}$ signifie que le document d_i est indexé par le terme t_j). Sous l'Hypothèse du *Monde Clos* (*Closed World Assumption*) [Reiter, 1978], $(d_i, t_j) \notin \mathcal{R}$ signifie que d_i n'est pas indexé par le terme t_j . Ainsi, si d_i est indexé par le terme t_j alors le *RSV* associé au document d_i pour la requête atomique t_j est égale à 1 (i.e. $RSV(d_i, t_j) = 1$ si $(d_i, t_j) \in \mathcal{R}$).

La sémantique sous-jacente du modèle Booléen associé à la théorie \mathfrak{T} est donnée par la structure $\langle \llbracket \mathcal{D} \rrbracket^{I_B}, I_B \rangle$, tel que $\llbracket \mathcal{D} \rrbracket^{I_B} = \llbracket \{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\} \rrbracket^{I_B} = \{\mathcal{T}, 2^{\mathcal{D}}, \mathcal{Q}\}$, où \mathcal{Q} correspond à l'ensemble de requêtes Booléennes construites sur l'ensemble de termes d'indexation \mathcal{T} . Ces requêtes sont construites avec la grammaire donnée en §2.2.1. La fonction d'interprétation I_B est illustrée dans le Tableau 3.2.

Nous pouvons remarquer que la sémantique associée au prédicat $\mathbf{S}(y, z)$ permet de retrouver le plus large ensemble de documents $\llbracket y \rrbracket^{I_B}$ (dans le sens de l'inclusion) satisfaisants la requête $\llbracket z \rrbracket^{I_B}$. L'exemple 3.3.1 illustre cette sémantique en considérant le contexte formel du Tableau 3.3.

Exemple 3.3.1. *Étant donné le contexte formel $\mathcal{K}_1 = (\{d_1, d_2, d_3\}, \{t_1, t_2\}, \mathcal{R}_1)$, où \mathcal{R}_1 est représentée par le Tableau 3.3. Soit la requête $\llbracket z \rrbracket^{I_B} = t_1 \vee t_2$. D'après la sémantique associée au prédicat \mathbf{S} (Expression.(B10)), $\{d_1, d_2, d_3\}$ est le plus large ensemble de documents satisfaisants la requête $t_1 \vee t_2$. Si on considère l'ensemble $\llbracket y \rrbracket^{I_B} = \{d_1\}$ alors $\llbracket \mathbf{S}(y, z) \rrbracket^{I_B}$ est faux, bien que $RSV(d_1, \llbracket z \rrbracket^{I_B}) = 1$ du moment que $RSV(d_2, \llbracket z \rrbracket^{I_B}) = 1 \not\Rightarrow d_2 \in \llbracket y \rrbracket^{I_B}$. De la même manière, nous avons :*

- Pour $\llbracket y \rrbracket^{I_B} = \{d_2\}$, $\llbracket \mathbf{S}(y, z) \rrbracket^{I_B}$ est faux,*
- Pour $\llbracket y \rrbracket^{I_B} = \{d_3\}$, $\llbracket \mathbf{S}(y, z) \rrbracket^{I_B}$ est faux,*
- Pour $\llbracket y \rrbracket^{I_B} = \{d_1, d_2\}$, $\llbracket \mathbf{S}(y, z) \rrbracket^{I_B}$ est faux,*
- Pour $\llbracket y \rrbracket^{I_B} = \{d_1, d_3\}$, $\llbracket \mathbf{S}(y, z) \rrbracket^{I_B}$ est faux,*
- Pour $\llbracket y \rrbracket^{I_B} = \{d_2, d_3\}$, $\llbracket \mathbf{S}(y, z) \rrbracket^{I_B}$ est faux.*

Tableau 3.2 – La Fonction d'Interprétation I_B

Syntaxe	Interprétation	#
y_{\perp}	\emptyset (i.e. l'ensemble vide de documents)	(B1)
z_{\perp}	$\emptyset_{\mathcal{Q}}$ (i.e. la requête vide)	(B2)
$f(x)$	$R(\llbracket x \rrbracket^{I_B})$ où $R(\cdot)$ est la fonction définie dans §1.3.2 i.e $R(\llbracket x \rrbracket^{I_B}) = \{d \in \mathcal{D} \mid (d, \llbracket x \rrbracket^{I_B}) \in \mathcal{R}\}$	(B3)
$y_1 \otimes y_2$	$\llbracket y_1 \rrbracket^{I_B} \cap \llbracket y_2 \rrbracket^{I_B}$	(B4)
$y_1 \oplus y_2$	$\llbracket y_1 \rrbracket^{I_B} \cup \llbracket y_2 \rrbracket^{I_B}$	(B5)
$\ominus y$	$\mathcal{D} \setminus \llbracket y \rrbracket^{I_B}$ (i.e $\llbracket \ominus y \rrbracket^{I_B}$ est le complement absolu de $\llbracket y \rrbracket^{I_B}$)	(B6)
$z_1 \sqcap z_2$	$\llbracket z_1 \rrbracket^{I_B} \wedge \llbracket z_2 \rrbracket^{I_B}$	(B7)
$z_1 \sqcup z_2$	$\llbracket z_1 \rrbracket^{I_B} \vee \llbracket z_2 \rrbracket^{I_B}$	(B8)
$\sim z$	$\neg \llbracket z \rrbracket^{I_B}$	(B9)
$\mathbf{S}(y, z)$	$(\forall d \in \llbracket y \rrbracket^{I_B}, RSV(d, \llbracket z \rrbracket^{I_B}) = 1) \wedge$ $(\forall d \in \mathcal{D}, RSV(d, \llbracket z \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y \rrbracket^{I_B})$	(B10)

Tableau 3.3 – La Relation Booléenne \mathcal{R}_1

\mathcal{R}_1	t_1	t_2
d_1	×	
d_2		×
d_3	×	×

Notons par \mathcal{M}_B la structure $\{\llbracket \mathcal{D} \rrbracket^{I_B}, I_B\}$ (i.e. $\mathcal{M}_B = \{\llbracket \mathcal{D} \rrbracket^{I_B}, I_B\}$). Le théorème suivant établit que \mathcal{M}_B est une modèle de \mathfrak{T} :

Théorème 3.3.1. $\mathcal{M}_B \models \mathfrak{T}$

Preuve. Pour prouver $\mathcal{M}_B \models \mathfrak{T}$, Il suffit de prouver que tous les axiomes de la théorie \mathfrak{T} sont satisfaits dans le modèle \mathcal{M}_B .

(A1) Il est bien connu que, dans un système de recherche d'informations, le résultat d'une requête vide (i.e. $\emptyset_{\mathcal{Q}}$ dans notre cas) est un ensemble vide de documents (i.e. $\emptyset_{\mathcal{D}}$), Alors nous avons : $\mathcal{M}_B \models \mathbf{S}(y_{\perp}, z_{\perp})$.

(A2) D'une part, à partir de la définition de $R(\cdot)$, on a :

- i) $\forall d \in R(t) : (d, t) \in \mathcal{R}$
- ii) $\forall d \in \mathcal{D} : (d, t) \in \mathcal{R} \Rightarrow d \in R(t)$

d'autre part, en utilisant (2.1), on a :

iii) $RSV(d, t) = 1$ si $(d, t) \in \mathcal{R}$

de i), ii) et iii) on a :

$$\begin{aligned}
& (\forall d \in R(t) : RSV(d, t) = 1) \wedge (\forall d \in \mathcal{D} : RSV(d, t) = 1 \Rightarrow d \in R(t)) \\
\Rightarrow & (\forall d \in R(\llbracket x \rrbracket^{I_B}) : RSV(d, \llbracket x \rrbracket^{I_B}) = 1) \wedge \\
& (\forall d \in \mathcal{D} : RSV(d, \llbracket x \rrbracket^{I_B}) = 1 \Rightarrow d \in R(\llbracket x \rrbracket^{I_B})) \quad \text{en substituant } t \text{ par } \llbracket x \rrbracket^{I_B} \\
\Rightarrow & \mathcal{M}_B \models \forall x : \mathcal{X} \mathbf{S}(f(x), x) \quad \text{en utilisant (B10)}
\end{aligned}$$

(A3) Supposons que $\mathcal{M}_B \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2)$, alors nous avons :

$$\begin{aligned}
i) & \mathcal{M}_B \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2) \\
\Rightarrow & (\forall d \in \llbracket y_1 \rrbracket^{I_B} : RSV(d, q_1) = 1) \wedge (\forall d \in \llbracket y_2 \rrbracket^{I_B} : RSV(d, q_2) = 1) \\
\Rightarrow & \forall d \in \llbracket y_1 \rrbracket^{I_B} \cap \llbracket y_2 \rrbracket^{I_B} : RSV(d, \llbracket z_1 \rrbracket^{I_B}) = 1 \wedge RSV(d, \llbracket z_2 \rrbracket^{I_B}) = 1 \\
\Rightarrow & \forall d \in \llbracket y_1 \rrbracket^{I_B} \cap \llbracket y_2 \rrbracket^{I_B} : RSV(d, \llbracket z_1 \rrbracket^{I_B} \wedge \llbracket z_2 \rrbracket^{I_B}) = 1 \\
& \hspace{15em} \text{en utilisant l'Expression.2.2} \\
\Rightarrow & \forall d \in \llbracket y_1 \otimes y_2 \rrbracket^{I_B} : RSV(d, \llbracket z_1 \sqcap z_2 \rrbracket^{I_B}) = 1 \\
& \hspace{15em} \text{en utilisant (B4) et (B7)}
\end{aligned}$$

$$\begin{aligned}
ii) & \mathcal{M}_B \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2) \\
\Rightarrow & (\forall d \in \mathcal{D} : RSV(d, \llbracket z_1 \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y_1 \rrbracket^{I_B}) \\
& \wedge (\forall d \in \mathcal{D} : RSV(d, \llbracket z_2 \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y_2 \rrbracket^{I_B}) \\
\Rightarrow & \forall d \in \mathcal{D} : RSV(d, \llbracket z_1 \rrbracket^{I_B}) = 1 \\
& \wedge RSV(d, \llbracket z_2 \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y_1 \rrbracket^{I_B} \wedge d \in \llbracket y_2 \rrbracket^{I_B} \\
\Rightarrow & \forall d \in \mathcal{D} : RSV(d, \llbracket z_1 \rrbracket^{I_B} \wedge \llbracket z_2 \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y_1 \rrbracket^{I_B} \cap \llbracket y_2 \rrbracket^{I_B} \\
& \hspace{15em} \text{en utilisant l'Expression. 2.2} \\
\Rightarrow & \forall d \in \mathcal{D} : RSV(d, \llbracket z_1 \sqcap z_2 \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y_1 \otimes y_2 \rrbracket^{I_B} \\
& \hspace{15em} \text{en utilisant (B4) et (B7)}
\end{aligned}$$

De i) et ii) nous obtenons :

$$\begin{aligned}
& \mathcal{M}_B \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2) \\
\Rightarrow & \forall d \in \llbracket y_1 \otimes y_2 \rrbracket^{I_B} : RSV(d, \llbracket z_1 \sqcap z_2 \rrbracket^{I_B}) = 1 \\
& \wedge \forall d \in \mathcal{D} : RSV(d, \llbracket z_1 \sqcap z_2 \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y_1 \otimes y_2 \rrbracket^{I_B} \\
\Rightarrow & \mathcal{M}_B \models \mathbf{S}(y_1 \otimes y_2, z_1 \sqcap z_2) \quad \text{en utilisant (B10)}
\end{aligned}$$

(A4) Supposons que $\mathcal{M}_B \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2)$, alors nous avons :

$$\begin{aligned}
& i) \quad \mathcal{M}_B \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2) \\
\Rightarrow & (\forall d \in \llbracket y_1 \rrbracket^{I_B} : RSV(d, \llbracket z_1 \rrbracket^{I_B}) = 1) \\
& \quad \wedge (\forall d \in \llbracket y_2 \rrbracket^{I_B} : RSV(d, \llbracket z_2 \rrbracket^{I_B}) = 1) \\
\Rightarrow & \forall d \in \llbracket y_1 \rrbracket^{I_B} \cup \llbracket y_2 \rrbracket^{I_B} : RSV(d, \llbracket z_1 \rrbracket^{I_B}) = 1 \vee RSV(d, \llbracket z_2 \rrbracket^{I_B}) = 1 \\
\Rightarrow & \forall d \in \llbracket y_1 \rrbracket^{I_B} \cup \llbracket y_2 \rrbracket^{I_B} : RSV(d, \llbracket z_1 \rrbracket^{I_B} \vee \llbracket z_2 \rrbracket^{I_B}) = 1 \\
& \hspace{15em} \text{en utilisant l'Expression. 2.3} \\
\Rightarrow & \forall d \in \llbracket y_1 \oplus y_2 \rrbracket^{I_B} : RSV(d, \llbracket z_1 \sqcup z_2 \rrbracket^{I_B}) = 1 \quad \text{en utilisant (B5) et (B8)}
\end{aligned}$$

$$\begin{aligned}
& ii) \quad \mathcal{M}_B \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2) \\
\Rightarrow & (\forall d \in \mathcal{D} : RSV(d, \llbracket z_1 \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y_1 \rrbracket^{I_B}) \\
& \quad \wedge (\forall d \in \mathcal{D} : RSV(d, \llbracket z_2 \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y_2 \rrbracket^{I_B}) \\
\Rightarrow & \forall d \in \mathcal{D} : RSV(d, \llbracket z_1 \rrbracket^{I_B}) = 1 \vee RSV(d, \llbracket z_2 \rrbracket^{I_B}) = 1 \\
& \quad \Rightarrow d \in \llbracket y_1 \rrbracket^{I_B} \vee d \in \llbracket y_2 \rrbracket^{I_B} \\
\Rightarrow & \forall d \in \mathcal{D} : RSV(d, \llbracket z_1 \rrbracket^{I_B} \vee \llbracket z_2 \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y_1 \rrbracket^{I_B} \cup \llbracket y_2 \rrbracket^{I_B} \\
& \hspace{15em} \text{en utilisant l'Expression. 2.3} \\
\Rightarrow & \forall d \in \mathcal{D} : RSV(d, \llbracket z_1 \sqcup z_2 \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y_1 \oplus y_2 \rrbracket^{I_B} \\
& \hspace{15em} \text{en utilisant (B5) et (B8)}
\end{aligned}$$

De *i)* et *ii)* nous obtenons :

$$\begin{aligned}
& \mathcal{M}_B \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2) \\
\Rightarrow & (\forall d \in \llbracket y_1 \oplus y_2 \rrbracket^{I_B} : RSV(d, \llbracket z_1 \sqcup z_2 \rrbracket^{I_B}) = 1) \\
& \quad \wedge (\forall d \in \mathcal{D} : RSV(d, \llbracket z_1 \sqcup z_2 \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y_1 \oplus y_2 \rrbracket^{I_B}) \\
\Rightarrow & \mathcal{M}_B \models \mathbf{S}(y_1 \oplus y_2, z_1 \sqcup z_2) \quad \text{en utilisant (B10)}
\end{aligned}$$

Tableau 3.4 – La Relation Booléenne \mathcal{R}_2

\mathcal{R}_2	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
d_1		×			×	×		
d_2			×			×	×	
d_3			×			×	×	
d_4	×			×			×	
d_5	×			×				
d_6	×				×			×

(A5) Supposons que $\mathcal{M}_B \models \mathbf{S}(y, z)$, alors nous avons :

$$\begin{aligned}
& \mathcal{M}_B \models \mathbf{S}(y, z) \\
\Rightarrow & \forall d \in \llbracket y \rrbracket^{I_B} : RSV(d, \llbracket z \rrbracket^{I_B}) = 1 \wedge \forall d \in \mathcal{D} : RSV(d, \llbracket z \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket y \rrbracket^{I_B} \\
\iff & \forall d \notin \llbracket y \rrbracket^{I_B} : RSV(d, \llbracket z \rrbracket^{I_B}) = 0 \wedge \forall d \in \mathcal{D} : RSV(d, \llbracket z \rrbracket^{I_B}) = 0 \Rightarrow d \notin \llbracket y \rrbracket^{I_B} \\
\Rightarrow & (\forall d \in \mathcal{D} \setminus \llbracket y \rrbracket^{I_B} : RSV(d, \llbracket z \rrbracket^{I_B}) = 0) \\
& \wedge (\forall d \in \mathcal{D} : RSV(d, \llbracket z \rrbracket^{I_B}) = 0 \Rightarrow d \in \mathcal{D} \setminus \llbracket y \rrbracket^{I_B}) \\
\Rightarrow & (\forall d \in \mathcal{D} \setminus \llbracket y \rrbracket^{I_B} : RSV(d, \neg \llbracket z \rrbracket^{I_B}) = 1) \\
& \wedge (\forall d \in \mathcal{D} : RSV(d, \neg \llbracket z \rrbracket^{I_B}) = 1 \Rightarrow d \in \mathcal{D} \setminus \llbracket y \rrbracket^{I_B}) \\
& \hspace{15em} \text{en utilisant l'Expression. 2.4} \\
\Rightarrow & (\forall d \in \llbracket \ominus y \rrbracket^{I_B} : RSV(d, \llbracket \sim z \rrbracket^{I_B}) = 1) \\
& \wedge (\forall d \in \mathcal{D} : RSV(d, \llbracket \sim z \rrbracket^{I_B}) = 1 \Rightarrow d \in \llbracket \ominus y \rrbracket^{I_B}) \\
& \hspace{15em} \text{en utilisant (B6) et (B9)} \\
\Rightarrow & \mathcal{M}_B \models \mathbf{S}(\ominus y, \sim z) \\
& \hspace{15em} \text{en utilisant (B10)}
\end{aligned}$$

□

Exemple 3.3.2. Soit le contexte formel $\mathcal{K}_2 = (\mathcal{D}_2, \mathcal{T}_2, \mathcal{R}_2)$, où $\mathcal{D}_2 = \{d_1, \dots, d_6\}$, $\mathcal{T}_2 = \{t_1, \dots, t_8\}$ et \mathcal{R}_2 est donnée dans le Tableau. 3.4. Nous présentons ci-après des exemples dans lesquelles le prédicat \mathbf{S} est interprété à "vrai" dans le modèle \mathcal{M}_B , en considérant les axiomes A1-A5 de \mathfrak{T} et les théorèmes $\Phi_1 - \Phi_4$ de la Proposition. 3.2.1.

- Par A1 : $\llbracket \mathbf{S} \rrbracket^{I_B} (\emptyset, \emptyset_{\mathcal{Q}})$;
Par A2 : $\llbracket \mathbf{S} \rrbracket^{I_B} (\{d_4, d_5, d_6\}, t_1)$;
Par A3 : $\llbracket \mathbf{S} \rrbracket^{I_B} (\{d_6\}, t_1 \wedge t_5)$;
Par A4 : $\llbracket \mathbf{S} \rrbracket^{I_B} (\{d_1, d_4, d_5, d_6\}, t_1 \vee t_2)$;
Par A5 : $\llbracket \mathbf{S} \rrbracket^{I_B} (\{d_1, d_2, d_3, d_4, d_5\}, \neg(t_1 \wedge t_5))$;
Par Φ_1 : $\llbracket \mathbf{S} \rrbracket^{I_B} (\{d_1\}, t_2 \wedge t_5 \wedge t_6)$;
Par Φ_2 : $\llbracket \mathbf{S} \rrbracket^{I_B} (\{d_1, d_2, d_3\}, t_2 \vee t_5 \vee t_6)$;

Par $\Phi_3 : \llbracket \mathbf{S} \rrbracket^{IB} (\{d_2, d_3, d_4, d_5, d_6\}, \neg(t_2 \wedge t_5 \wedge t_6))$;

Par $\Phi_4 : \llbracket \mathbf{S} \rrbracket^{IB} (\{d_4, d_5, d_6\}, \neg(t_2 \vee t_5 \vee t_6))$;

3.4 Modèle basé sur les Ensembles Flous

Dans cette section, nous démontrons que le modèle de la RI basé sur les ensemble flous, proposé par Bookstein [1980], satisfait notre théorie \mathfrak{T} . Dans ce qui suit, nous présentons la matrice d'incidence relative à ce modèle par le contexte formel flou $\mathcal{K} = (\mathcal{D}, \mathcal{T}, \mathcal{I})$, où \mathcal{D} est un ensemble fini de documents (soit $\mathcal{D} = \{d_1, d_2, \dots, d_m\}$), \mathcal{T} est un ensemble fini de termes d'indexation (soit $\mathcal{T} = \{t_1, t_2, \dots, t_m\}$) et \mathcal{I} est une relation binaire floue dans $[0, 1]^{\mathcal{D} \times \mathcal{T}}$ qui associe à chaque paire (d_i, t_j) , pour $i = 1 \dots m$ et $j = 1 \dots n$, le poids du terme t_j dans le document d_i (à titre d'exemple, ce poids peut correspondre à la mesure $TF - IDF$). Pour une requête q donnée (requête Booléenne générée par la grammaire donnée dans §2.2.1 dans l'ensemble \mathcal{T}), le processus de recherche flou associe à tout document d_i , $i = 1 \dots m$ son score de pertinence (RSV flou) α_i . Ainsi, le résultat de recherche de la requête q est un ensemble flou de document \tilde{D}_q . Nous utiliserons la notation $\tilde{D} = \{d_1^{\alpha_1}, \dots, d_i^{\alpha_i}, \dots, d_m^{\alpha_m}\}$, où α_i est le score de pertinence du document d_i pour la requête q , il correspond aussi au degré d'appartenance du document d_i à l'ensemble flou \tilde{D} .

Soit \mathcal{M}_f la structure $\langle \llbracket \mathcal{D} \rrbracket^{I_f}, I_f \rangle$, où $\llbracket \mathcal{D} \rrbracket^{I_f} = \llbracket \{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\} \rrbracket^{I_f} = \{\mathcal{T}, [0, 1]^{\mathcal{D}}, \mathcal{Q}\}$, et \mathcal{Q} est l'ensemble de requêtes Booléennes construites dans \mathcal{T} par la grammaire donnée dans §2.2.1. La fonction d'interprétation I_f est détaillée dans le Tableau 3.5.

Tableau 3.5 – La Fonction d'Interprétation I_f

Syntaxe	Interprétation	#
y_{\perp}	$\{d_i^0, i = 1 \dots m\}$	(f1)
z_{\perp}	$\emptyset_{\mathcal{Q}}$ (i.e. la requête vide)	(f2)
$f(x)$	$\{d_i^{\alpha_i}, i = 1 \dots m \mid \alpha_i = RSV(d_i, \llbracket x \rrbracket^{I_f})\}$ (RSV est donné par les expressions 2.5-2.8)	(f3)
$y_1 \otimes y_2$	$\llbracket y_1 \rrbracket^{I_f} \cap \llbracket y_2 \rrbracket^{I_f}$	(f4)
$y_1 \oplus y_2$	$\llbracket y_1 \rrbracket^{I_f} \cup \llbracket y_2 \rrbracket^{I_f}$	(f5)
$\ominus y$	$\mathcal{D} \setminus \llbracket y \rrbracket^{I_f}$	(f6)
$z_1 \sqcap z_2$	$\llbracket z_1 \rrbracket^{I_f} \wedge \llbracket z_2 \rrbracket^{I_f}$	(f7)
$z_1 \sqcup z_2$	$\llbracket z_1 \rrbracket^{I_f} \vee \llbracket z_2 \rrbracket^{I_f}$	(f8)
$\sim z$	$\neg \llbracket z \rrbracket^{I_f}$	(f9)
$\mathbf{S}(y, z)$	$\forall d^{\alpha} \in \llbracket y \rrbracket^{I_f}, RSV(d, \llbracket z \rrbracket^{I_f}) = \alpha$	(f10)

Tableau 3.6 – Le Contexte Formel Flou $\tilde{\mathcal{K}}_1$

\mathcal{D}	\mathcal{T}	
	t_1	t_2
d_1	0.8	0.2
d_2	0.1	0.9
d_3	0.7	0.6

Dans ce modèle, on considère pour l'intersection floue (resp. l'union floue), le Minimum (resp. Maximum) en tant que t-norme (resp. t-conorme). Ces deux opérateurs flous et la complémentation floue (\setminus) sont donnés comme suit :

$$\begin{aligned}\tilde{D}_1 \cap \tilde{D}_2 &= \{d_i^{\alpha_i}, i = 1, \dots, m \mid \alpha_i = \min(\tilde{D}_1(d_i), \tilde{D}_2(d_i))\}. \\ \tilde{D}_1 \cup \tilde{D}_2 &= \{d_i^{\alpha_i}, i = 1, \dots, m \mid \alpha_i = \max(\tilde{D}_1(d_i), \tilde{D}_2(d_i))\}. \\ \mathcal{D} \setminus \tilde{D} &= \{d_i^{\alpha_i}, i = 1, \dots, m \mid \alpha_i = 1 - \tilde{D}(d_i)\}.\end{aligned}$$

La sémantique proposée du prédicat \mathbf{S} aborde directement le plus large ensemble flou de documents (dans le sens des scores de pertinence les plus élevés) $\llbracket y \rrbracket^{If}$ satisfaisant La requête $\llbracket z \rrbracket^{If}$, comme illustré dans l'exemple suivant :

Exemple 3.4.1. *En considérant le contexte formel flou $\tilde{\mathcal{K}}_1$ du Tableau. 3.6 et la requête $\llbracket z \rrbracket^{If} = t_1 \vee t_2$, nous avons alors :*

$$\begin{aligned}\text{Pour } \llbracket y \rrbracket^{If} &= \{d_1^{0.8}, d_2^{0.9}, d_3^{0.7}\}, \llbracket \mathbf{S}(y, z) \rrbracket^{If} \text{ est vraie;} \\ \text{Pour } \llbracket y \rrbracket^{If} &= \{d_1^{0.8}, d_2^{0.9}, d_3^{0.6}\}, \llbracket \mathbf{S}(y, z) \rrbracket^{If} \text{ est faus;} \\ \text{Pour } \llbracket y \rrbracket^{If} &= \{d_1^{0.2}, d_2^{0.9}, d_3^{0.5}\}, \llbracket \mathbf{S}(y, z) \rrbracket^{If} \text{ est faus.}\end{aligned}$$

Le théorème suivant établit que \mathcal{M}_f est également un modèle de la théorie \mathfrak{T} .

Théorème 3.4.1. $\mathcal{M}_f \models \mathfrak{T}$

Preuve. Il suffit de prouver que tous les axiomes de \mathfrak{T} sont satisfaits par le modèle \mathcal{M}_f .

(A1) Le modèle de RI basé sur les ensembles flou suppose que l'ensemble $\{d_i^0, i = 1 \dots m\}$ est le résultat de la requête vide \emptyset_Q . Donc nous avons :

$$\begin{aligned}\forall d \in \mathcal{D}, RSV(d, \emptyset_Q) &= 0 \\ \implies \llbracket \mathbf{S}(y_\perp, z_\perp) \rrbracket^{If} &\text{ est vraie} \\ \implies \mathcal{M}_f \models \mathbf{S}(y_\perp, z_\perp)\end{aligned}$$

(A2) Nous avons $\forall \llbracket x \rrbracket^{I_f}$:

$$\begin{aligned} & \llbracket f(x) \rrbracket^{I_f} = \{d_i^{\alpha_i}, i = 1 \dots m \mid \alpha_i = RSV(d_i, \llbracket x \rrbracket^{I_f})\} && \text{en utilisant (f3)} \\ \implies & \forall d^\alpha \in \llbracket f(x) \rrbracket^{I_f} : RSV(d, \llbracket x \rrbracket^{I_f}) = \alpha \\ \implies & \llbracket \mathbf{S}(f(x), x) \rrbracket^{I_f} \text{ est vraie} && \text{en utilisant (f10)} \\ \implies & \mathcal{M}_f \models \mathcal{X} \mathbf{S}(f(x), x) \end{aligned}$$

(A3) Supposant que $\mathcal{M}_f \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2)$ alors nous avons :

$$\begin{aligned} & \mathcal{M}_f \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2) \\ \implies & (\forall d^\alpha \in \llbracket y_1 \rrbracket^{I_f}, RSV(d, \llbracket z_1 \rrbracket^{I_f}) = \alpha) \wedge (\forall d^\beta \in \llbracket y_2 \rrbracket^{I_f} : RSV(d, \llbracket z_2 \rrbracket^{I_f}) = \beta) \\ \implies & \forall d^\alpha \in \llbracket y_1 \rrbracket^{I_f}, \forall d^\beta \in \llbracket y_2 \rrbracket^{I_f} : RSV(d, \llbracket z_1 \rrbracket^{I_f}) = \alpha \wedge RSV(d, \llbracket z_2 \rrbracket^{I_f}) = \beta \\ \implies & \forall d^\alpha \in \llbracket y_1 \rrbracket^{I_f}, \forall d^\beta \in \llbracket y_2 \rrbracket^{I_f} : RSV(d, \llbracket z_1 \rrbracket^{I_f} \wedge \llbracket z_2 \rrbracket^{I_f}) = \min(\alpha, \beta) \\ & \text{en utilisant l'Expression.2.6 et l'opérateur t-norme pris comme min} \end{aligned}$$

En prenant $\gamma = \min(\alpha, \beta)$, nous obtenons :

$$\begin{aligned} & \forall d^\gamma \in \llbracket y_1 \rrbracket^{I_f} \cap \llbracket y_2 \rrbracket^{I_f} : RSV(d, \llbracket z_1 \rrbracket^{I_f} \wedge \llbracket z_2 \rrbracket^{I_f}) = \gamma \\ \implies & \forall d^\gamma \in \llbracket y_1 \otimes y_2 \rrbracket^{I_f} : RSV(d, \llbracket z_1 \sqcap z_2 \rrbracket^{I_f}) = \gamma && \text{en utilisant (f4) et (f7)} \\ \implies & \llbracket \mathbf{S}(y_1 \otimes y_2, z_1 \sqcap z_2) \rrbracket^{I_f} \text{ est vraie} && \text{en utilisant (f10)} \\ \implies & \mathcal{M}_f \models \mathbf{S}(y_1 \otimes y_2, z_1 \sqcap z_2) \end{aligned}$$

(A4) Supposant que $\mathcal{M}_f \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2)$ alors nous avons :

$$\begin{aligned} & \mathcal{M}_f \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2) \\ \implies & (\forall d^\alpha \in \llbracket y_1 \rrbracket^{I_f}, RSV(d, \llbracket z_1 \rrbracket^{I_f}) = \alpha) \wedge (\forall d^\beta \in \llbracket y_2 \rrbracket^{I_f} : RSV(d, \llbracket z_2 \rrbracket^{I_f}) = \beta) \\ \implies & \forall d^\alpha \in \llbracket y_1 \rrbracket^{I_f}, \forall d^\beta \in \llbracket y_2 \rrbracket^{I_f} : RSV(d, \llbracket z_1 \rrbracket^{I_f}) = \alpha \wedge RSV(d, \llbracket z_2 \rrbracket^{I_f}) = \beta \\ \implies & \forall d^\alpha \in \llbracket y_1 \rrbracket^{I_f}, \forall d^\beta \in \llbracket y_2 \rrbracket^{I_f} : RSV(d, \llbracket z_1 \rrbracket^{I_f} \vee \llbracket z_2 \rrbracket^{I_f}) = \max(\alpha, \beta) \\ & \text{en utilisant (2.7) et en considérant max comme t-conorme} \end{aligned}$$

En prenant $\gamma = \max(\alpha, \beta)$, nous obtenons :

$$\begin{aligned} & \forall d^\gamma \in \llbracket y_1 \rrbracket^{I_f} \cup \llbracket y_2 \rrbracket^{I_f} : RSV(d, \llbracket z_1 \rrbracket^{I_f} \vee \llbracket z_2 \rrbracket^{I_f}) = \gamma \\ \implies & \forall d^\gamma \in \llbracket y_1 \oplus y_2 \rrbracket^{I_f} : RSV(d, \llbracket z_1 \sqcup z_2 \rrbracket^{I_f}) = \gamma && \text{en utilisant (f5) et (f8)} \\ \implies & \llbracket \mathbf{S}(y_1 \oplus y_2, z_1 \sqcup z_2) \rrbracket^{I_f} \text{ est vraie} && \text{en utilisant (f10)} \\ \implies & \mathcal{M}_f \models \mathbf{S}(y_1 \oplus y_2, z_1 \sqcup z_2) \end{aligned}$$

(A5) Supposant que $\mathcal{M}_f \models \mathbf{S}(y, z)$ nous avons alors :

$$\begin{aligned}
& \mathcal{M}_f \models \mathbf{S}(y, z) \\
\Rightarrow & \forall d^\alpha \in \llbracket y \rrbracket^{I_f} : RSV(d, \llbracket z \rrbracket^{I_f}) = \alpha \\
\Rightarrow & \forall d^\alpha \in \llbracket y \rrbracket^{I_f} : RSV(d, \neg \llbracket z \rrbracket^{I_f}) = 1 - \alpha && \text{en utilisant (2.8)} \\
\Rightarrow & \forall d^{1-\alpha} \in \mathcal{D} \setminus \llbracket y \rrbracket^{I_f} : RSV(d, \neg \llbracket z \rrbracket^{I_f}) = 1 - \alpha \\
& \quad (\text{since } \mathcal{D} \setminus \tilde{D} = \{d_i^{\alpha_i}, i = 1, \dots, m \mid \alpha_i = 1 - \tilde{D}(d_i)\}) \\
\Rightarrow & \forall d^\beta \in \llbracket \ominus y \rrbracket^{I_f} : RSV(d, \llbracket \sim z \rrbracket^{I_f}) = \beta \\
& \quad \text{en prenant } \beta = 1 - \alpha \text{ \& } \text{et en utilisant (f6),(f9)} \\
\Rightarrow & \llbracket \mathbf{S}(\ominus y, \sim z) \rrbracket^{I_f} \text{ est vraie} && \text{en utilisant (f10)} \\
\Rightarrow & \mathcal{M}_f \models \mathbf{S}(\ominus y, \sim z)
\end{aligned}$$

□

Exemple 3.4.2. *Étant donnée le contexte formel flou $\tilde{\mathcal{K}}_2 = (\{d_1, \dots, d_6\}, \{t_1, \dots, t_6\}, \tilde{\mathcal{R}})$, où $\tilde{\mathcal{R}}$ est donnée dans la Table 3.7, Nous présentons ci-après des exemples d'interprétation "vrai" du prédicat \mathbf{S} dans le modèle \mathcal{M}_f , en considérant les axiomes A1-A5 de \mathfrak{T} et les théorèmes $\Phi_1 - \Phi_4$ de la Proposition. 3.2.1.*

- Par A1 : $\llbracket \mathbf{S} \rrbracket^{I_f} (\{d_1^{0.0}, d_2^{0.0}, d_3^{0.0}, d_4^{0.0}, d_5^{0.0}, d_6^{0.0}\}, \emptyset_{\mathcal{Q}})$;
Par A2 : $\llbracket \mathbf{S} \rrbracket^{I_f} (\{d_1^{0.6}, d_2^{0.9}, d_3^{0.1}, d_4^{0.0}, d_5^{0.5}, d_6^{0.5}\}, t_1)$;
Par A3 : $\llbracket \mathbf{S} \rrbracket^{I_f} (\{d_1^{0.2}, d_2^{0.9}, d_3^{0.1}, d_4^{0.0}, d_5^{0.4}, d_6^{0.4}\}, t_1 \wedge t_5)$;
Par A4 : $\llbracket \mathbf{S} \rrbracket^{I_f} (\{d_1^{0.6}, d_2^{0.9}, d_3^{0.1}, d_4^{0.0}, d_5^{0.5}, d_6^{0.5}\}, t_1 \vee t_2)$;
Par A5 : $\llbracket \mathbf{S} \rrbracket^{I_f} (\{d_1^{0.8}, d_2^{0.1}, d_3^{0.9}, d_4^{1.0}, d_5^{0.6}, d_6^{0.6}\}, \neg(t_1 \wedge t_5))$;
Par Φ_1 : $\llbracket \mathbf{S} \rrbracket^{I_f} (\{d_1^{0.0}, d_2^{0.3}, d_3^{0.0}, d_4^{0.0}, d_5^{0.3}, d_6^{0.2}\}, t_2 \wedge t_5 \wedge t_6)$;
Par Φ_2 : $\llbracket \mathbf{S} \rrbracket^{I_f} (\{d_1^{0.5}, d_2^{0.9}, d_3^{0.7}, d_4^{0.7}, d_5^{0.9}, d_6^{0.5}\}, t_2 \vee t_5 \vee t_6)$;
Par Φ_3 : $\llbracket \mathbf{S} \rrbracket^{I_f} (\{d_1^{1.0}, d_2^{0.7}, d_3^{1.0}, d_4^{1.0}, d_5^{0.7}, d_6^{0.8}\}, \neg(t_2 \wedge t_5 \wedge t_6))$;
Par Φ_4 : $\llbracket \mathbf{S} \rrbracket^{I_f} (\{d_1^{0.5}, d_2^{0.1}, d_3^{0.3}, d_4^{0.3}, d_5^{0.1}, d_6^{0.5}\}, \neg(t_2 \vee t_5 \vee t_6))$.

Tableau 3.7 – Le Contexte Formel Flou $\tilde{\mathcal{K}}_2 = (\{d_1, \dots, d_6\}, \{t_1, \dots, t_6\}, \tilde{\mathcal{R}})$

$\tilde{\mathcal{R}}$	t_1	t_2	t_3	t_4	t_5	t_6
d_1	0.6	0.5	0.6	0.4	0.2	0.0
d_2	0.9	0.3	0.8	0.5	0.9	0.4
d_3	0.1	0.0	0.5	0.8	0.5	0.7
d_4	0.0	0.0	0.9	0.1	0.5	0.7
d_5	0.5	0.3	0.7	0.7	0.4	0.9
d_6	0.5	0.2	0.5	0.6	0.4	0.5

3.5 Modèle Vectoriel

Soit \mathcal{V} un espace vectoriel de n -dimension ayant comme base la famille de vecteurs termes $(\vec{t}_1, \vec{t}_2, \dots, \vec{t}_n)$. Chaque document d_i de la collection \mathcal{D} est représenté dans \mathcal{V} par le vecteur $\vec{d}_i = [w_1^i, w_2^i, \dots, w_n^i]$, de même toute requête q sur cette collection est représentée dans \mathcal{V} par son vecteur $\vec{q} = [w_1^q, w_2^q, \dots, w_n^q]$, où chaque composante w_j^i (resp. w_j^q) correspond au poids du terme t_j dans le document d_i (resp. la requête q). En normalisant les vecteurs de documents et de requêtes (la normalisation d'un vecteur \vec{v} consiste à diviser chaque composante de ce vecteur par sa norme i.e. $\vec{v} / \|\vec{v}\|$), la fonction de similarité de *cosinus*, donnée dans l'Expression. 2.10, Page.35, est identique à la fonction de similarité du produit scalaire de ces vecteurs normalisés, donnée dans l'Expression. 2.9.

Comme hypothèses de travail, nous admettons que (1) le nombre de documents de \mathcal{D} est égal au nombre de termes d'indexation (soit n) (2) les vecteurs de documents sont linéairement indépendants. Nous désignons par \mathbf{D} la matrice d'incidence dans laquelle chaque ligne i correspond au vecteur normalisé du document d_i . Pour tout vecteur de requête normalisé q , nous associons le vecteur de similarités \vec{C}_q de n -dimension où chaque composante i correspond à la similarité $Cos(\vec{d}_i, \vec{q})$, notée $Sim(d_i, q)$. En utilisant le calcul matriciel, \vec{C}_q est exprimé par :

$$\vec{C}_q = \begin{bmatrix} \vec{d}_1 \cdot \vec{q} \\ \vec{d}_2 \cdot \vec{q} \\ \vdots \\ \vec{d}_n \cdot \vec{q} \end{bmatrix} = \mathbf{D} \cdot \vec{q} = \begin{bmatrix} w_1^1 & w_2^1 & \dots & w_n^1 \\ w_1^2 & w_2^2 & \dots & w_n^2 \\ \vdots & \vdots & & \vdots \\ w_1^n & w_2^n & \dots & w_n^n \end{bmatrix} \cdot \begin{bmatrix} w_1^q \\ w_2^q \\ \vdots \\ w_n^q \end{bmatrix} \quad (3.1)$$

Dans la suite de cette section, $\vec{O}_{\mathcal{V}}$ désignera le vecteur null dans l'espace vectoriel \mathcal{V} , \mathcal{C} désignera l'ensemble des vecteurs de similarités \vec{C}_q et $\vec{O}_{\mathcal{C}}$ est le vecteur nul de cet ensemble.

Soit la structure $\mathcal{M}_v = \langle \llbracket \mathcal{D} \rrbracket^{I_v}, I_v \rangle$ où $\llbracket \mathcal{D} \rrbracket^{I_v} = \llbracket \{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\} \rrbracket^{I_v} = \{\mathcal{V}, \mathcal{C}, \mathcal{V}\}$ et la fonction d'interprétation I_v est donnée dans le Tableau. 3.8.

Comme le modèle d'espace vectoriel n'autorise pas la négation dans les requêtes, il suffit d'interpréter \sim et \ominus comme fonction identité.

Le théorème donné ci-dessous établit qu'un troisième modèle, à savoir le modèle d'espace vectoriel, satisfait également notre théorie \mathfrak{T} , mettant ainsi en évidence l'aspect générique de cette théorie.

Tableau 3.8 – La Fonction d’Interprétation I_v

Syntaxe	Interprétation	#
y_{\perp}	$\vec{0}_c$	(v1)
z_{\perp}	$\vec{0}_v$	(v2)
$f(x)$	$\mathbf{D} \cdot \eta(\llbracket x \rrbracket^{I_v})$	(v3)
$y_1 \otimes y_2$	$\mathbf{D} \cdot \eta(\mathbf{D}^{-1} \cdot \llbracket y_1 \rrbracket^{I_v} \Downarrow \mathbf{D}^{-1} \cdot \llbracket y_2 \rrbracket^{I_v})$	(v4)
$y_1 \oplus y_2$	$\mathbf{D} \cdot \eta(\mathbf{D}^{-1} \cdot \llbracket y_1 \rrbracket^{I_v} \Uparrow \mathbf{D}^{-1} \cdot \llbracket y_2 \rrbracket^{I_v})$	(v5)
$\ominus y$	$\llbracket y \rrbracket^{I_v}$	(v6)
$z_1 \sqcap z_2$	$\eta(\llbracket z_1 \rrbracket^{I_v} \Downarrow \llbracket z_2 \rrbracket^{I_v})$	(v7)
$z_1 \sqcup z_2$	$\eta(\llbracket z_1 \rrbracket^{I_v} \Uparrow \llbracket z_2 \rrbracket^{I_v})$	(v8)
$\sim z$	$\llbracket z \rrbracket^{I_v}$	(v9)
$\mathbf{S}(y, z)$	$\forall i \in \{1 \dots n\} \llbracket y \rrbracket^{I_v}[i] = \text{Sim}(d_i, \llbracket z \rrbracket^{I_v})$	(v10)

Où,
$$\eta(\vec{v}) = \begin{cases} \vec{v} / \|\vec{v}\| & \text{si } \vec{v} \neq \vec{0} \\ \vec{0} & \text{sinon} \end{cases};$$

$$\llbracket w_1, \dots, w_n \rrbracket \Uparrow \llbracket u_1, \dots, u_n \rrbracket = [\max(w_1, u_1), \dots, \max(w_n, u_n)];$$

$$\llbracket w_1, \dots, w_n \rrbracket \Downarrow \llbracket u_1, \dots, u_n \rrbracket = [\min(w_1, u_1), \dots, \min(w_n, u_n)].$$

Théorème 3.5.1. $\mathcal{M}_v \models \mathfrak{T}$ *Preuve.*

(A1) D’après (v2), nous avons :

$$\begin{aligned} \llbracket z_{\perp} \rrbracket^{I_v} &= \vec{0}_v \\ \implies \forall i \in \{1 \dots n\}, \vec{d}_i \cdot \llbracket z_{\perp} \rrbracket^{I_v} &= 0 \\ \implies \forall i \in \{1 \dots n\}, \text{Sim}(d_i, \llbracket z_{\perp} \rrbracket^{I_v}) &= \llbracket y_{\perp} \rrbracket^{I_v}[i] && \text{en utilisant 3.1 et (v1)} \\ \implies \llbracket \mathbf{S}(y_{\perp}, z_{\perp}) \rrbracket^{I_v} &\text{est vraie} && \text{en utilisant (v10)} \\ \implies \mathcal{M}_v \models \mathbf{S}(y_{\perp}, z_{\perp}) &&& \end{aligned}$$

(A2) À partir de (v3), nous avons :

$$\begin{aligned} \forall \llbracket x \rrbracket^{I_v}, \llbracket f(x) \rrbracket^{I_v} &= \mathbf{D} \cdot \eta(\llbracket x \rrbracket^{I_v}) && \text{en utilisant (v3)} \\ \implies \forall \llbracket x \rrbracket^{I_v}, \forall i \in \{0 \dots 1\}, \llbracket f(x) \rrbracket^{I_v}[i] &= \vec{d}_i \cdot \eta(\llbracket x \rrbracket^{I_v}) && \text{en utilisant (3.1)} \\ \implies \forall \llbracket x \rrbracket^{I_v}, \forall i \in \{0 \dots 1\}, \llbracket f(x) \rrbracket^{I_v}[i] &= \text{Sim}(d_i, \llbracket x \rrbracket^{I_v}) \\ \implies \llbracket \mathbf{S}(f(x), x) \rrbracket^{I_v} &\text{est vraie} && \text{en utilisant (v10)} \\ \implies \mathcal{M}_v \models \mathbf{S}(f(x), x) &&& \end{aligned}$$

(A3) Admettant que $\mathcal{M}_v \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2)$, nous avons donc :

$$\begin{aligned}
& \mathcal{M}_v \models \mathbf{S}(y_1, z_1) \wedge \mathbf{S}(y_2, z_2) \\
\implies & \forall i \in \{1 \dots n\} \llbracket y_1 \rrbracket^{I_v}[i] = \text{Sim}(d_i, \llbracket z_1 \rrbracket^{I_v}) \\
& \quad \wedge \llbracket y_2 \rrbracket^{I_v}[i] = \text{Sim}(d_i, \llbracket z_2 \rrbracket^{I_v}) \quad \text{en utilisant (v10)} \\
\implies & \llbracket y_1 \rrbracket^{I_v} = \mathbf{D} \cdot \llbracket z_1 \rrbracket^{I_v} \wedge \llbracket y_2 \rrbracket^{I_v} = \mathbf{D} \cdot \llbracket z_2 \rrbracket^{I_v} \quad \text{en utilisant (3.1)} \\
\implies & \mathbf{D}^{-1} \cdot \llbracket y_1 \rrbracket^{I_v} = \llbracket z_1 \rrbracket^{I_v} \wedge \mathbf{D}^{-1} \cdot \llbracket y_2 \rrbracket^{I_v} = \llbracket z_2 \rrbracket^{I_v}
\end{aligned}$$

En substituant $\mathbf{D}^{-1} \cdot \llbracket y_1 \rrbracket^{I_v}$ (resp. $\mathbf{D}^{-1} \cdot \llbracket y_2 \rrbracket^{I_v}$) par $\llbracket z_1 \rrbracket^{I_v}$ (resp. $\llbracket z_2 \rrbracket^{I_v}$) dans (v4) nous obtenons $\llbracket y_1 \otimes y_2 \rrbracket^{I_v} = \mathbf{D} \cdot \eta(\llbracket z_1 \rrbracket^{I_v}) \Downarrow \llbracket z_2 \rrbracket^{I_v}$. Ainsi :

$$\begin{aligned}
& \forall i \in \{0 \dots n\} \llbracket y_1 \otimes y_2 \rrbracket^{I_v}[i] = \vec{d}_i \cdot \eta(\llbracket z_1 \rrbracket^{I_v} \Downarrow \llbracket z_2 \rrbracket^{I_v}) \quad \text{en utilisant (3.1)} \\
\implies & \forall i \in \{0 \dots n\} \llbracket y_1 \otimes y_2 \rrbracket^{I_v}[i] = \text{Sim}(\vec{d}_i, \llbracket z_1 \sqcap z_2 \rrbracket^{I_v}) \quad \text{en utilisant (v7)} \\
\implies & \llbracket \mathbf{S}(y_1 \otimes y_2, z_1 \sqcap z_2) \rrbracket^{I_v} \text{ est vrai} \quad \text{en utilisant (v10)} \\
\implies & \mathcal{M}_v \models \mathbf{S}(y_1 \otimes y_2, z_1 \sqcap z_2)
\end{aligned}$$

(A4) Similaire à la preuve de (A3).

(A5) La preuve découle directement des interprétations données dans (v6) et (v9) à \ominus et à \sim qui correspondent à la fonction identité, ainsi,
 $\mathcal{M}_v \models \mathbf{S}(y, z) \implies \mathcal{M}_v \models \mathbf{S}(\ominus y, \sim z)$. □

Exemple 3.5.1. En considérons la matrice d'incidence \mathbf{D} donnée dans le Tableau 3.9 qui correspond à la normalisation du Tableau 3.7. Nous illustrons ci-après quelques interprétations vraies du prédicat \mathbf{S} dans le modèle \mathcal{M}_v :

Par A1 : $\llbracket \mathbf{S} \rrbracket^{I_v} ((0., 0., 0., 0., 0., 0.), (0., 0., 0., 0., 0., 0.))$;

Par A2 : $\llbracket \mathbf{S} \rrbracket^{I_v} ([.55, .54, .08, .00, .33, .44], [1., 0., 0., 0., 0., 0.])$ (E-1) ;

$\llbracket \mathbf{S} \rrbracket^{I_v} ([.18, .54, .39, .40, .26, .35], [0., 0., 0., 0., 1., 0.])$ (E-2) ;

$\llbracket \mathbf{S} \rrbracket^{I_v} ([.94, .74, .34, .31, .62, .69], [.76, .57, .43, .21, 0., 0.])$ (E-3) ;

$\llbracket \mathbf{S} \rrbracket^{I_v} ([.73, .95, .81, .68, .89, .49], [.60, 0., .30, .36, .48, .42])$ (E-4).

À titre d'exemple, (E-4) signifie que la similarité entre le document d_1 (resp. d_2, d_3, d_4, d_5, d_6) et la requête $\vec{q} = [0.60, 0.00, 0.30, 0.36, 0.48, 0.42]$ est 0.73 (resp. 0.95, 0.81, 0.68, 0.89, 0.49).

Par A3 : en utilisant (E-1) et (E-2), nous avons :

$$[.55, .54, .08, 0., .33, .44] \llbracket \otimes \rrbracket^{I_v} [.18, .54, .39, .40, .26, .35] = [0., 0., 0., 0., 0., 0.] ;$$

$$\text{et } \eta([1., 0., 0., 0., 0., 0.] \Downarrow [0., 0., 0., 0., 1., 0.]) = [0., 0., 0., 0., 0., 0.] ;$$

alors nous obtenons $\llbracket \mathbf{S} \rrbracket^{I_v} ([0., 0., 0., 0., 0., 0.], [0., 0., 0., 0., 0., 0.])$.

Par A4 : en utilisant (E-1) et (E-2), nous avons :

$$[.55, .54, .08, 0., .33, .44] \llbracket \oplus \rrbracket^{I_v} [.18, .54, .39, .40, .26, .35] = [.52, .77, .33, .28, .42, .5] ;$$

$$\text{et } \eta([1., 0., 0., 0., 0., 0.] \Uparrow [0., 0., 0., 0., 1., 0.]) = [.71, 0., 0., 0., .71, 0.] ;$$

Tableau 3.9 – Normalisation du Tableau.3.7

D	t_1	t_2	t_3	t_4	t_5	t_6
d_1	0.55	0.46	0.55	0.37	0.18	0.00
d_2	0.54	0.18	0.48	0.30	0.54	0.24
d_3	0.08	0.00	0.39	0.62	0.39	0.55
d_4	0.00	0.00	0.72	0.08	0.40	0.56
d_5	0.33	0.20	0.46	0.46	0.26	0.59
d_6	0.44	0.17	0.44	0.52	0.35	0.44

alors nous obtenons $\llbracket \mathbf{S} \rrbracket^{I_v} ([.52, .77, .33, .28, .42, .56], [.71, 0., 0., 0., .71, 0.])$.

Par ϕ_1 : en utilisant (E-1), (E-3) et (E-4), nous avons :

$([.55, .54, .08, .00, .33, .44] \llbracket \otimes \rrbracket^{I_v} [.94, .74, .34, .31, .62, .69]) \llbracket \otimes \rrbracket^{I_v} [.73, .95, .81, .68, .89, .49] =$
 $[.55, .54, .08, .00, .33, .44]$; et

$\eta(\eta([1., 0., 0., 0., 0., 0.] \Downarrow [.76, .57, .43, .21, 0., 0.] \Downarrow [.60, 0., .30, .36, .48, .42]) =$
 $[1., 0., 0., 0., 0., 0.]$; Alors nous obtenons $\llbracket \mathbf{S} \rrbracket^{I_v} ([.55, .54, .08, .00, .33, .44], [1., 0., 0., 0., 0., 0.])$;

Par ϕ_2 : en utilisant (E-1), (E-3) et (E-4), nous avons :

$([.55, .54, .08, 0., .33, .44] \llbracket \oplus \rrbracket^{I_v} [.94, .74, .34, .31, .62, .69]) \llbracket \oplus \rrbracket^{I_v} [.73, .95, .81, .68, .89, .49] =$
 $[.91, .73, .30, .26, .59, .67] \llbracket \oplus \rrbracket^{I_v} [.73, .95, .81, .68, .89, .49] = [.87, .94, .68, .57, .86, .91]$;

et $\eta(\eta([1., 0., 0., 0., 0., 0.] \Uparrow [.76, .57, .43, .21, 0., 0.] \Uparrow [.60, 0., .30, .36, .48, .42]) =$
 $\eta([.80, .46, .34, .17, 0., 0.] \Uparrow [.60, 0., .30, .36, .48, .42]) = [.65, .38, .28, .29, .39, .34]$;

Alors nous obtenons $\llbracket \mathbf{S} \rrbracket^{I_v} ([.87, .94, .68, .57, .86, .91], [.65, .38, .28, .29, .39, .34])$.

Rappelons que le modèle \mathcal{M}_v suppose que le nombre de documents (m) correspond au nombre de termes d'indexation (n) (c.-à-d. $n = m$). Dans la pratique, on peut y avoir deux autres cas ; 1) $n > m$ et 2) $m > n$.

- 1) Lorsque $n > m$, le modèle \mathcal{M}_v n'est plus valable, du fait que les interprétations (v4) et (v5) relatives aux opérateurs \otimes et \oplus ne sont plus valables en raison de l'inexistence de la matrice inverse \mathbf{D}^{-1} .
- 2) À l'inverse, Lorsque $m > n$, ce qui est le cas le plus usuel, la matrice d'incidence (qui est une matrice non carrée) correspondante peut être décomposée en plusieurs matrices carrées de vecteurs de documents linéairement indépendants. De cette manière, pour une requête donnée, plusieurs vecteurs de similarités peuvent être calculé indépendamment (un vecteur de similarité pour chacune des matrices carrées) et la réponse globale peut être reconstituée par concaténation ou fusion de ces vecteurs.

3.6 Conclusion

Dans ce chapitre, nous avons proposé une théorie axiomatique pour la RI. La théorie est définie sur un univers de discours multi-sortes qui permet de distinguer les différentes catégories d'objets considérés dans la RI. Nous avons conçu la théorie sur un système composé de cinq axiomes et nous avons considéré deux règles d'inférence qui permettent de déduire de nouveaux théorèmes, toute en respectant le langage de la théorie.

Ensuite, nous avons démontré que trois modèles classiques de la RI satisfont la théorie \mathfrak{T} ; Le modèle Booléen, le modèle flou standard et le modèle vectoriel. Pour chacun de ces modèles, nous avons défini une structure multi-sortes et une fonction d'interprétation, et nous avons prouvé que les axiomes de la théorie sont satisfaits dans ces trois modèles.

Dans les trois modèles \mathcal{M}_B , \mathcal{M}_f et \mathcal{M}_v nous avons interprété la sorte \mathcal{Y} par trois ensembles différents :

- Dans \mathcal{M}_B , la sorte \mathcal{Y} est interprétée par l'ensemble des parties de l'ensemble de documents $2^{\mathcal{D}}$. Dans ce modèle, aucun score de pertinence n'est associé aux documents.
- Dans \mathcal{M}_f , la sorte \mathcal{Y} est interprétée par l'ensemble des sous-ensembles flous de documents $L^{\mathcal{D}}$. Le score de pertinence d'un document appartenant à l'ensemble flou $\llbracket y \rrbracket^{I_f}$ par rapport à une requête $\llbracket z \rrbracket^{I_f}$, tel que $\llbracket \mathbf{S}(y, z) \rrbracket^{I_f}$ est vraie, est donnée par le degré d'appartenance de ce document à $\llbracket Y \rrbracket^{I_f}$. Ceci dit, contrairement à la plus part² des approches logiques de RI existantes, le score de pertinence est calculé explicitement.
- Dans \mathcal{M}_v , la sorte \mathcal{Y} est interprétée par l'ensemble des vecteurs de similarité \mathcal{C} . Pour un vecteur requête $\llbracket z \rrbracket^{I_v}$ et un vecteur de similarité $\llbracket y \rrbracket^{I_v}$ tel que $\llbracket \mathbf{S}(y, z) \rrbracket^{I_v}$ est vraie, toute $i^{\text{ème}}$ composante du vecteur $\llbracket y \rrbracket^{I_v}$ correspond à la similarité (cosinus) entre le $i^{\text{ème}}$ document et la requête $\llbracket z \rrbracket^{I_v}$.

D'une part, notre théorie présente une nouvelle vision par rapport aux approches logiques de RI existantes. En effet, la plupart des modèles existant tentent de formaliser l'implication de [van Rijsbergen \[1986\]](#) ($d \rightarrow q$) et d'évaluer son incertitude. Notre approche dérive une expression logique valide de la forme $\mathbf{S}(y, z)$ et son interprétation dans un modèle permet de retrouver, d'une manière explicite, les documents et les scores associés satisfaisant une requête. D'autre part, cette théorie offre un cadre générique de RI capable de reproduire et de représenter les modèles classiques, comme les modèles Booléen, flou standard et vectoriel.

Nous avons, en particulier, considéré quatre théorèmes de la théorie (prouvés dans la Section.3.2.3). Les interprétations de ces théorèmes dans les modèles \mathcal{M}_B et \mathcal{M}_f font ressortir des formes particulières de requêtes ; à savoir les requêtes conjonc-

2. À notre connaissance, la seule approche qui calcul le degré de pertinence d'une manière explicite (sans utilisé une fonction ad hoc) est l'approche basée sur pDatalog.

tives, disjonctives et leurs deux formes négatives. Nous utiliserons principalement ces quatre théorèmes et leurs interprétations dans les chapitres suivants, où ils ne serviront comme passerelles entre les modèles Booléen et flou standard d'un côté et la l'analyse de concepts formels classique et floue de l'autre côté.

Chapitre 4

Lecture Possibiliste de l'Analyse de Concepts Formels pour la Recherche d'Information

Sommaire

4.1	Introduction	79
4.2	Enrichissement du Modèle \mathcal{M}_B basé sur l'ACF	79
4.3	Recherche Orientée Contenu/Structure	84
4.3.1	Traitement des Requêtes Orientées Contenu	88
4.3.2	Traitement de Requêtes Orientées Contenu et Structure	94
4.4	Conclusion	95

4.1 Introduction

En tirant profit d'une lecture possibiliste de l'Analyse de Concepts Formels, nous proposons dans ce chapitre, d'une part, d'enrichir le modèle \mathcal{M}_B (présenté dans le précédent chapitre), et d'autre part, d'appliquer cette lecture à la recherche d'information structurée.

La première contribution (Section.4.2) concerne la formalisation des relations algébriques entre la théorie de l'ACF classique et le modèle \mathcal{M}_B . Plus précisément, cette contribution caractérise les requêtes Booléennes élémentaires (conjonctives, disjonctives et négations) par les treillis de concepts formels et les treillis de NII-paires.

La seconde contribution (Section.4.3) concerne l'application de l'extension triadique de l'ACF pour la recherche d'information structurée, qui a comme but de retrouver des parties de documents au lieu des documents entiers. Notons que dans l'extension triadique de l'ACF, la matrice d'incidence est représentée sous forme d'un contexte formel triadique (§1.6) où nous considérons :

- 1) Un troisième ensemble pour les éléments (granules) des documents structurés,
- 2) Une relation ternaire entre les documents, leurs éléments et les termes.

4.2 Enrichissement du Modèle \mathcal{M}_B basé sur l'ACF

Nous proposons dans cette section une caractérisation des requêtes élémentaires du modèle \mathcal{M}_B présenté dans le précédent chapitre. Cette caractérisation permet de considérer des requêtes conjonctives, disjonctives ainsi que leurs négations. Rappelons que toutes les approches existantes de RI basées sur l'ACF sont limitées à l'expression de requêtes conjonctives uniquement.

Soit $\mathcal{K} = (\mathcal{D}, \mathcal{T}, \mathcal{R})$ un contexte formel avec $\mathcal{D} = \{d_1, d_2, \dots, d_m\}$ un ensemble fini de documents, $\mathcal{T} = \{t_1, t_2, \dots, t_n\}$ un ensemble fini de termes d'indexation et \mathcal{R} une relation binaire Booléenne dans $\mathcal{D} \times \mathcal{T}$ ($(d_i, t_j) \in \mathcal{R}$ signifie que le document d_i est indexé par le terme t_j). Nous désignerons par \mathcal{Q} l'ensemble des toutes les requêtes construites sur l'ensemble de termes \mathcal{T} .

Pour un ensemble $T = \{t_1, t_2, \dots, t_r\}$ de termes, une requête conjonctive élémentaire $\hat{Q}(T)$ est de la forme $t_1 \wedge t_2 \wedge \dots \wedge t_r$, de la même manière, une requête disjonctive élémentaire $\check{Q}(T)$ est de la forme $t_1 \vee t_2 \vee \dots \vee t_r$. Les négations de ces deux formes de requêtes sont donnée par $\neg\hat{Q}(T)$ ($\neg(t_1 \wedge t_2 \wedge \dots \wedge t_r)$) et $\neg\check{Q}(T)$ ($\neg(t_1 \vee t_2 \vee \dots \vee t_r)$). Dans la proposition suivante nous caractérisons, dans le modèle \mathcal{M}_B (§3.3), les sous-ensemble de l'extension¹ du prédicat **S** en considérant ces quatres formes de requêtes.

Proposition 4.2.1. *Les interpretations suivantes sont vraies dans le modèle \mathcal{M}_B :*

1. L'extension du prédicat **S** est l'ensemble des paires (D, Q) tel que $\llbracket \mathbf{S} \rrbracket^{I_B}(D, Q)$ est vraie dans le modèle \mathcal{M}_B .

- (I1) $\forall T \neq \emptyset, \llbracket \mathbf{S} \rrbracket^{I_B}(\bigcap_{t_i \in T} R(t_i), \hat{Q}(T))$
- (I2) $\forall T \neq \emptyset, \llbracket \mathbf{S} \rrbracket^{I_B}(\bigcup_{t_i \in T} R(t_i), \check{Q}(T))$
- (I3) $\forall T \neq \emptyset, \llbracket \mathbf{S} \rrbracket^{I_B}(\mathcal{D} \setminus \bigcap_{t_i \in T} R(t_i), \neg \hat{Q}(T))$
- (I4) $\forall T \neq \emptyset, \llbracket \mathbf{S} \rrbracket^{I_B}(\mathcal{D} \setminus \bigcup_{t_i \in T} R(t_i), \neg \check{Q}(T))$

Preuve. La preuve résulte naturellement du théorème 3.3.1 - Page. 64 et la fonction d'interprétation I_B définie dans §3.3. Ainsi, nous avons à interpréter chaque variable x_i par t_i pour obtenir respectivement (I1), (I2), (I3), (I4). □

Soit \hat{P} (resp. \check{P} , \hat{P}_\neg , \check{P}_\neg) l'ensemble des paires (D, Q) tel que l'interprétation $\llbracket \mathbf{S} \rrbracket^{I_B}(D, Q)$ est vraie et Q est n'importe quelle requête conjunctive (resp. disjunctive, negative conjunctive, negative disjunctive) construite sur l'ensemble \mathcal{T} . Formellement, ces ensembles sont données par :

$$\begin{aligned} \hat{P} &= \{(D, \hat{Q}(T)) \in 2^{\mathcal{D}} \times \mathcal{Q} \mid T \neq \emptyset \text{ et } \llbracket \mathbf{S} \rrbracket^{I_B}(D, \hat{Q}(T)) = \text{vraie}\} \\ \check{P} &= \{(D, \check{Q}(T)) \in 2^{\mathcal{D}} \times \mathcal{Q} \mid T \neq \emptyset \text{ et } \llbracket \mathbf{S} \rrbracket^{I_B}(D, \check{Q}(T)) = \text{vraie}\} \\ \hat{P}_\neg &= \{(D, \neg \hat{Q}(T)) \in 2^{\mathcal{D}} \times \mathcal{Q} \mid T \neq \emptyset \text{ et } \llbracket \mathbf{S} \rrbracket^{I_B}(D, \neg \hat{Q}(T)) = \text{vraie}\} \\ \check{P}_\neg &= \{(D, \neg \check{Q}(T)) \in 2^{\mathcal{D}} \times \mathcal{Q} \mid T \neq \emptyset \text{ et } \llbracket \mathbf{S} \rrbracket^{I_B}(D, \neg \check{Q}(T)) = \text{vraie}\}. \end{aligned}$$

Pour la suite de cette section, nous considérons les notations suivantes :

- $\mathcal{L}(\mathcal{K})$ (resp. $\mathcal{L}(\overline{\mathcal{K}})$) l'ensemble de concepts formels $\langle D, T \rangle$ induit à partir de \mathcal{K} (resp. $\overline{\mathcal{K}}$),
- $\mathcal{P}(\mathcal{K})$ (resp. $\mathcal{P}(\overline{\mathcal{K}})$) l'ensemble des NII-paires induit à partir de \mathcal{K} (resp. $\overline{\mathcal{K}}$),
- $R_{\overline{\mathcal{K}}}(\cdot)$ désigne le complément de de la relation \mathcal{R} donnée par $\overline{\mathcal{R}} = \mathcal{D} \times \mathcal{T} \setminus \mathcal{R}$,
- $(\cdot)_{\overline{\mathcal{K}}}^\Delta$ désigne l'application de l'opérateur $(\cdot)^\Delta$ dans le contexte formel $\overline{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \overline{\mathcal{R}})$,
- $(\cdot)_{\overline{\mathcal{K}}}^\Pi$ désigne l'application de l'opérateur $(\cdot)^\Pi$ dans le contexte formel $\overline{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \overline{\mathcal{R}})$.

La proposition suivante établit des correspondances significatives entre les structures algébriques (treillis) $\mathcal{L}(\mathcal{K})$, $\mathcal{P}(\mathcal{K})$, $\mathcal{P}(\overline{\mathcal{K}})$, $\mathcal{L}(\overline{\mathcal{K}})$ est les ensembles \hat{P} , \check{P} , \hat{P}_\neg , \check{P}_\neg

Proposition 4.2.2. *Soit les contextes formels $\mathcal{K} = (\mathcal{D}, \mathcal{T}, \mathcal{R})$ et $\overline{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \overline{\mathcal{R}})$, alors nous avons les propriétés suivantes :*

- (P1) $\forall D \forall T \neq \emptyset, (\langle D, T \rangle \in \mathcal{L}(\mathcal{K}) \Rightarrow (D, \hat{Q}(T)) \in \hat{P})$
- (P2) $\forall D \forall T \neq \emptyset, (\langle D, T \rangle \in \mathcal{P}(\mathcal{K}) \Rightarrow (D, \check{Q}(T)) \in \check{P})$
- (P3) $\forall D \forall T \neq \emptyset, (\langle D, T \rangle \in \mathcal{P}(\overline{\mathcal{K}}) \Rightarrow (D, \neg \hat{Q}(T)) \in \hat{P}_\neg)$
- (P4) $\forall D \forall T \neq \emptyset, (\langle D, T \rangle \in \mathcal{L}(\overline{\mathcal{K}}) \Rightarrow (D, \neg \check{Q}(T)) \in \check{P}_\neg)$

Preuve.

(P1) Puisque $T \neq \emptyset$ et $\langle D, T \rangle \in \mathcal{L}(\mathcal{K})$, nous aurons alors $D = \bigcap_{t_i \in T} R(t_i)$. Ainsi, la preuve est faite en remplaçant $\bigcap_{t_i \in T} R(t_i)$ par D dans (I1) de la Proposition. 4.2.1.

(P2) Puisque $T \neq \emptyset$ et $\langle D, T \rangle \in \mathcal{P}(\mathcal{K})$, nous aurons alors $D = \bigcup_{t_i \in T} R(t_i)$. Ainsi, la preuve est faite en remplaçant $\bigcup_{t_i \in T} R(t_i)$ par D dans (I2) de la Proposition. 4.2.1.

(P3) Rappelons d'abord la lois de De Morgan :

$$\text{a) Pour toute famille de sous-ensembles } \{D_j\}_{j \in J} \subseteq \mathcal{D}, \quad \mathcal{D} \setminus \bigcup_j D_j = \bigcap_j \mathcal{D} \setminus D_j$$

$$\text{b) Pour toute famille de sous-ensembles } \{D_j\}_{j \in J} \subseteq \mathcal{D}, \quad \mathcal{D} \setminus \bigcap_j D_j = \bigcup_j \mathcal{D} \setminus D_j$$

$$\begin{aligned} \forall D \forall T \neq \emptyset, \langle D, T \rangle \in \mathcal{P}(\overline{\mathcal{K}}) &\implies D = T_{\overline{\mathcal{K}}}^{\Pi} \\ &\implies D = \bigcup_{t \in T} R_{\overline{\mathcal{K}}}(t) \\ &\implies D = \bigcup_{t \in T} (\mathcal{D} \setminus R(t)) \\ &\quad (\text{Puisque } \mathcal{D} \setminus R(t) = R_{\overline{\mathcal{K}}}(t)) \\ &\implies D = \mathcal{D} \setminus \bigcap_{t \in T} R(t) \quad (\text{\AA partir de b}) \end{aligned}$$

En remplaçant $\mathcal{D} \setminus \bigcap_{t \in T} R(t)$ par D dans (I3) de la Proposition. 4.2.1, il s'ensuit que la propriété (P3) est vérifiée.

(P4) Nous avons :

$$\begin{aligned} \forall D \forall T \neq \emptyset, \langle D, T \rangle \in \mathcal{L}(\overline{\mathcal{K}}) &\implies D = (T)_{\overline{\mathcal{K}}}^{\Delta} \\ &\implies D = \bigcap_{t \in T} R_{\overline{\mathcal{K}}}(t) \\ &\implies D = \bigcap_{t \in T} (\mathcal{D} \setminus R(t)) \\ &\implies D = \mathcal{D} \setminus \bigcup_{t \in T} R(t) \quad (\text{\AA partir de a}) \end{aligned}$$

En remplaçant $\mathcal{D} \setminus \bigcup_{t \in T} R(t)$ par D dans (I4) de la Proposition. 4.2.1, il s'ensuit que la propriété (P4) est vérifiée.

□

Nous pouvons remarquer que des différentes paires $(D_1, Q_1), (D_2, Q_2), \dots$ de l'extension du prédicat \mathbf{S} dans le modèle \mathcal{M}_B (i.e. $\llbracket \mathbf{S} \rrbracket^{I_B}(D_i, Q_i) = \text{vraie}$, pour $i = 1, 2, \dots$) peuvent avoir le même ensemble de documents D (i.e. $D_1 = D_2 = \dots = D$) et des requêtes syntaxiquement différentes. Dans un système de classification de documents, ceci engendre un inconvénient vu qu'il existe une redondance dans la formulation de requêtes et, ainsi, différentes requêtes peuvent avoir la même réponse (le

même ensemble de documents). Pour cela, nous proposons de raffiner les ensembles $\hat{P}, \check{P}, \hat{P}_-$ and \check{P}_- en considérant exclusivement les paires (D, Q) distinctes. Nous dirons que deux paires (D_1, Q_1) et (D_2, Q_2) sont distinctes si $D_1 \neq D_2$, autrement elles sont équivalentes (cf. Exemple. 4.2.1).

Etant donné l'ensemble \hat{P} (resp. \check{P}, \hat{P}_- et \check{P}_-), nous considérons l'ensemble des paires distinctes $\hat{E} \subseteq \hat{P}$ (resp. $\check{E} \subseteq \check{P}, \hat{E}_- \subseteq \hat{P}_-, \check{E}_- \subseteq \check{P}_-$) défini comme suit :

$$\forall (D_1, Q_1), (D_2, Q_2) \in \hat{E} : D_1 \neq D_2 \text{ et } \forall (D, Q) \in \hat{P}, \exists! (D', Q') \in \hat{E} : D = D'$$

$$\forall (D_1, Q_1), (D_2, Q_2) \in \check{E} : D_1 \neq D_2 \text{ et } \forall (D, Q) \in \check{P}, \exists! (D', Q') \in \check{E} : D = D'$$

$$\forall (D_1, Q_1), (D_2, Q_2) \in \hat{E}_- : D_1 \neq D_2 \text{ et } \forall (D, Q) \in \hat{P}_-, \exists! (D', Q') \in \hat{E}_- : D = D'$$

$$\forall (D_1, Q_1), (D_2, Q_2) \in \check{E}_- : D_1 \neq D_2 \text{ et } \forall (D, Q) \in \check{P}_-, \exists! (D', Q') \in \check{E}_- : D = D'$$

Nous illustrons ci-après des exemples de paires (*Documents, Requête*) distinctes et équivalentes.

Exemple 4.2.1. *Étant donné le contexte formel $\mathcal{K}_2 = (\mathcal{D}_2, \mathcal{T}_2, \mathcal{R}_2)$ illustré dans le Tableau. 3.4 - Page. 67, où $\mathcal{D}_2 = \{d_1, d_2 \dots d_6\}$ et $\mathcal{T}_2 = \{t_1, t_2 \dots t_8\}$, nous avons :*

- $(\{d_2, d_3\}, t_3), (\{d_2, d_3\}, t_6 \wedge t_7)$ et $(\{d_2, d_3\}, t_3 \wedge t_6 \wedge t_7)$ sont des paires équivalentes,
- $(\{d_1, d_2, d_3\}, t_6), (\{d_1, d_2, d_3\}, t_2 \vee t_6), (\{d_1, d_2, d_3\}, t_3 \vee t_6), (\{d_1, d_2, d_3\}, t_2 \vee t_3)$ et $(\{d_1, d_2, d_3\}, t_2 \vee t_3 \vee t_6)$ sont des paires équivalentes,
- $(\{d_2, d_3, d_4, d_5, d_6\}, \neg t_2)$ et $(\{d_2, d_3, d_4, d_5, d_6\}, \neg(t_2 \wedge t_5 \wedge t_6))$ sont des paires équivalentes,
- $(\{d_1, d_2, d_3\}, \neg t_1), (\{d_1, d_2, d_3\}, \neg(t_4 \vee t_8)), (\{d_1, d_2, d_3\}, \neg(t_1 \vee t_4 \vee t_8))$ sont des paires équivalentes,
- $(\{d_1\}, t_2 \wedge t_5 \wedge t_6)$ et $(\{d_6\}, t_1 \wedge t_5 \wedge t_8)$ sont des paires distinctes,
- $(\{d_1\}, t_2), (\{d_1, d_4, d_5, d_6\}, t_1 \vee t_2)$ et $(\{d_1, d_2, d_3\}, t_2 \vee t_3)$ sont des paires distinctes.

La proposition suivante établit des propriétés utiles entre les ensembles de paires distinctes définies ci-dessus et les treillis de concepts formels et de NII-paires dans les contextes formels \mathcal{K} , et $\bar{\mathcal{K}}$:

Proposition 4.2.3. *Soit le contexte formel $\mathcal{K} = (\mathcal{D}, \mathcal{T}, \mathcal{R})$. Les propriétés suivantes sont vérifiées $\forall T \subseteq \mathcal{T}, T \neq \emptyset$,*

$$(P1') \forall (D, \hat{Q}(T)) \in \hat{E}, \exists! \langle D', T' \rangle \in \mathcal{L}(\mathcal{K}) \text{ tel que } D' = D.$$

$$(P2') \forall (D, \check{Q}(T)) \in \check{E}, \exists! \langle D', T' \rangle \in \mathcal{P}(\mathcal{K}) \text{ tel que } D' = D.$$

$$(P3') \forall (D, \neg \hat{Q}(T)) \in \hat{E}_-, \exists! \langle D', T' \rangle \in \mathcal{P}(\bar{\mathcal{K}}) \text{ tel que } D' = D.$$

$$(P4') \forall (D, \neg \check{Q}(T)) \in \check{E}_-, \exists! \langle D', T' \rangle \in \mathcal{L}(\bar{\mathcal{K}}) \text{ tel que } D' = D.$$

Preuve. Soit $T \subseteq \mathcal{T}$ et $T \neq \emptyset$,

(P1') d'une part, nous avons :

$$\begin{aligned} (D, \hat{Q}(T)) \in \hat{E} &\implies D = \bigcap_{t \in T} R(t) \\ &\implies D = T^\Delta \end{aligned}$$

d'autre part, comme $\langle T^\Delta, (T^\Delta)^\Delta \rangle$ est un concept formel de \mathcal{K} , En prenant $T' = (T^\Delta)^\Delta$, nous obtenons : $\langle D, T' \rangle \in \mathcal{L}(\mathcal{K})$.

(P2') Nous avons :

$$\begin{aligned} (D, \check{Q}(T)) \in \check{E} &\implies D = \bigcup_{t \in T} R(t) \\ &\implies D = T^\Pi \end{aligned}$$

Et comme $\langle T^\Pi, (T^\Pi)^N \rangle$ est une NII-paire de \mathcal{K} , en prenant $T' = (T^\Pi)^N$, nous obtenons : $\langle D, T' \rangle \in \mathcal{P}(\mathcal{K})$.

(P3') Nous avons :

$$\begin{aligned} (D, \neg\hat{Q}(T)) \in \hat{E}_- &\implies D = \mathcal{D} \setminus \bigcap_{t \in T} R(t) \\ &\implies D = \bigcup_{t \in T} \mathcal{D} \setminus R(t) \\ &\implies D = \bigcup_{t \in T} R_{\overline{\mathcal{K}}}(t) \\ &\implies D = (T)_{\overline{\mathcal{K}}}^\Pi \end{aligned}$$

Et puisque $\langle (T)_{\overline{\mathcal{K}}}^\Pi, ((T)_{\overline{\mathcal{K}}}^\Pi)^N_{\overline{\mathcal{K}}} \rangle$ est une NII-paire de $\overline{\mathcal{K}}$, en prenant $T' = ((T)_{\overline{\mathcal{K}}}^\Pi)^N_{\overline{\mathcal{K}}}$, nous obtenons : $\langle D, T' \rangle \in \mathcal{P}(\overline{\mathcal{K}})$.

(P4') d'une part nous avons :

$$\begin{aligned} (D, \neg\check{Q}(T)) \in \check{E}_- &\implies D = \mathcal{D} \setminus \bigcup_{t \in T} R(t) \\ &\implies D = \bigcap_{t \in T} \mathcal{D} \setminus R(t) \\ &\implies D = \bigcap_{t \in T} R_{\overline{\mathcal{K}}}(t) \\ &\implies D = (T)_{\overline{\mathcal{K}}}^\Delta \end{aligned}$$

d'autre part, comme $\langle (T)_{\overline{\mathcal{K}}}^{\Delta}, ((T)_{\overline{\mathcal{K}}})_{\overline{\mathcal{K}}}^{\Delta} \rangle$ est un concept formel de $\overline{\mathcal{K}}$, en prenant $T' = ((T)_{\overline{\mathcal{K}}})_{\overline{\mathcal{K}}}^{\Delta}$ nous obtenons $\langle D, T' \rangle \in \mathcal{L}(\overline{\mathcal{K}})$. \square

Comme conséquence aux deux Propositions. 4.2.2 et 4.2.3, le théorème suivant donne des propriétés algébriques significatives :

Théorème 4.2.1. *Soient les contextes formels $\mathcal{K} = (\mathcal{D}, \mathcal{T}, \mathcal{R})$ et $\overline{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \overline{\mathcal{R}})$ alors,*

(T1) \hat{E} est isomorphe à $\mathcal{L}(\mathcal{K}) \setminus \{\langle \mathcal{D}, \emptyset \rangle\}$

(T2) \check{E} est isomorphe à $\mathcal{P}(\mathcal{K}) \setminus \{\langle \emptyset, \emptyset \rangle\}$

(T3) \hat{E}_{\neg} est isomorphe à $\mathcal{P}(\overline{\mathcal{K}}) \setminus \{\langle \emptyset, \emptyset \rangle\}$

(T4) \check{E}_{\neg} est isomorphe à $\mathcal{L}(\overline{\mathcal{K}}) \setminus \{\langle \mathcal{D}, \emptyset \rangle\}$

Preuve.

(T1) Considérons les applications

$f : \hat{E} \longrightarrow \mathcal{L}(\mathcal{K}) \setminus \{\langle \mathcal{D}, \emptyset \rangle\}$ et $g : \mathcal{L}(\mathcal{K}) \setminus \{\langle \mathcal{D}, \emptyset \rangle\} \longrightarrow \hat{E}$ définies respectivement par :

$$f\left(\langle D, \hat{Q}(T) \rangle\right) = \langle D, (T^{\Delta})^{\Delta} \rangle$$

$$g(\langle D, T \rangle) = \langle D, \hat{Q}(T_1) \rangle$$

i) À partir de (P1'), chaque élément de \hat{E} est associé (en utilisant f) exactement à un et un seul concept formel (une seule image) de $\mathcal{L}(\mathcal{K}) \setminus \{\langle \mathcal{D}, \emptyset \rangle\}$.

ii) Inversement, à partir de la propriété (P1), il s'ensuit que pour chaque concept formel de $\mathcal{L}(\mathcal{K}) \setminus \{\langle \mathcal{D}, \emptyset \rangle\}$ est associé (en utilisant g) à un et un seul élément de \hat{E} .

De *i)* et *ii)* nous déduisons que \hat{E} est isomorphe à $\mathcal{L}(\mathcal{K}) \setminus \{\langle \mathcal{D}, \emptyset \rangle\}$

(T2), (T3) et (T4) sont prouvées de la même manière. \square

Afin d'illustrer les isomorphismes du Théorème. 4.2.1, on considère le contexte formel \mathcal{K}_2 du Tableau.3.4. La Figure.4.1 (resp. Figure.4.2) illustre l'isomorphisme entre l'ensemble des concepts formels $\mathcal{L}(\mathcal{K}_2) \setminus \{\langle \mathcal{D}_2, \emptyset \rangle\}$ (resp. l'ensemble des NII - paires $\mathcal{P}(\overline{\mathcal{K}}_2) \setminus \{\langle \emptyset, \emptyset \rangle\}$) et l'ensemble des paires (*Documents*, *Requête*) distinctes \hat{E} (resp. \hat{E}_{\neg}).

4.3 Recherche Orientée Contenu/Structure

L'émergence et l'adoption de documents structurés, tels que le SGML (Standard Generalized Markup Language) et le XML (eXtensible Markup Language), dans les échanges de documents électroniques et (ses larges diffusions) sur le Web ont donnée

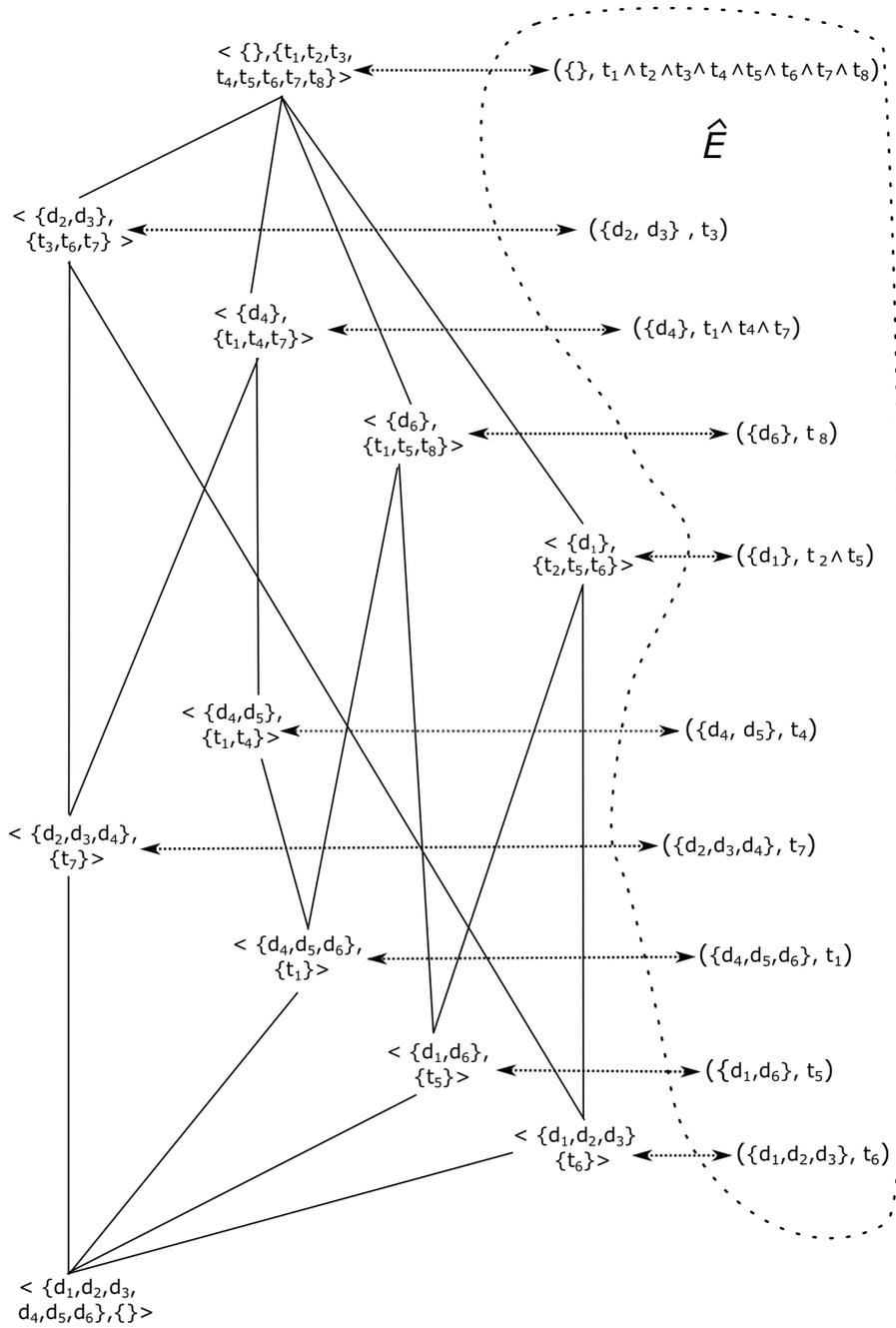


FIGURE 4.1 – Isomorphisme entre \hat{E} et $\mathcal{L}(\mathcal{K}_2) \setminus \{\langle \mathcal{D}_2, \emptyset \rangle\}$

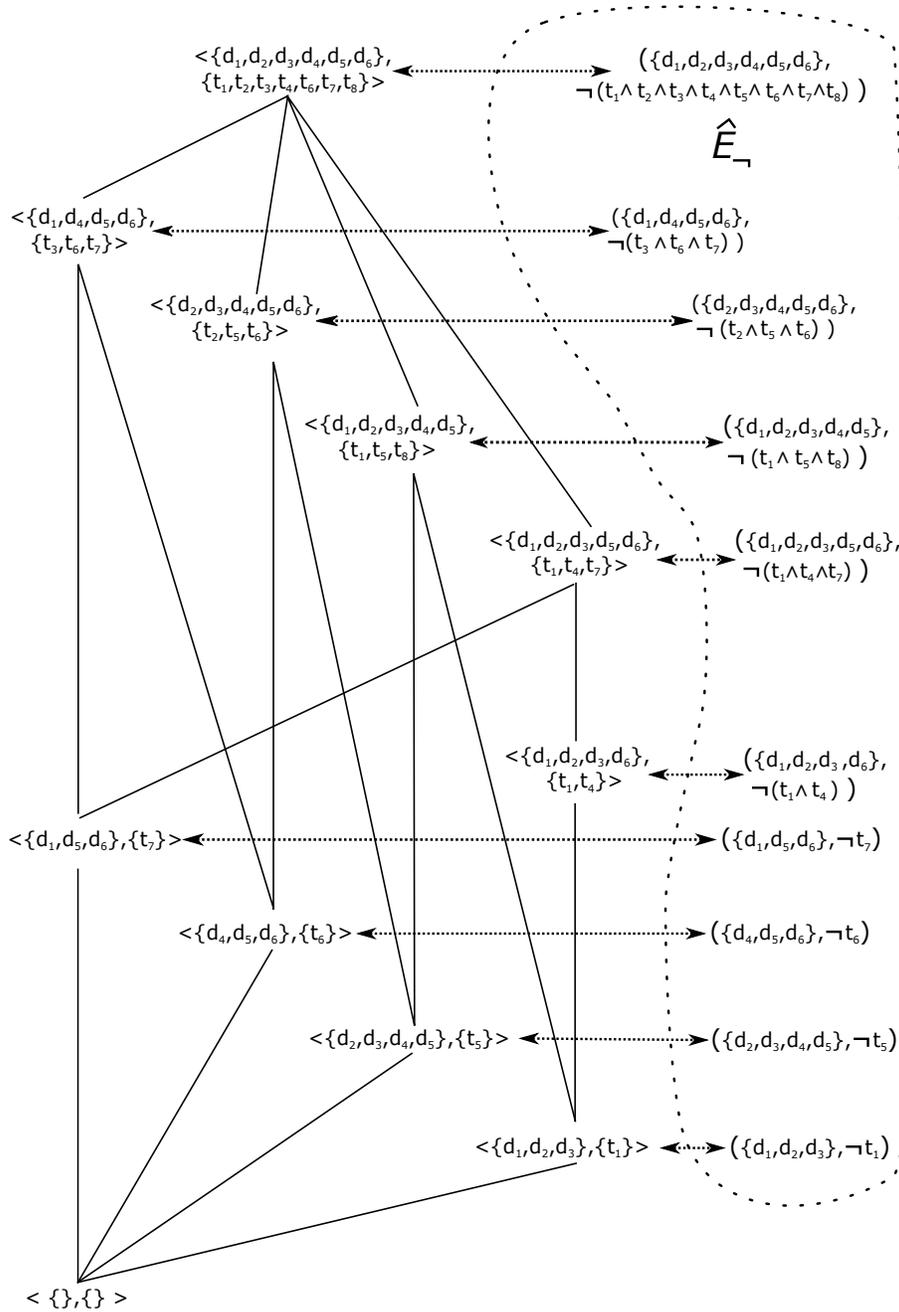


FIGURE 4.2 – Isomorphisme entre \hat{E}_- et $\mathcal{P}(\bar{\mathcal{K}}_2) \setminus \{\langle \emptyset, \emptyset \rangle\}$

lieu au développement de systèmes pour la recherche d'information structurée, dont le but est de retrouver les éléments (on parle également de granule) de documents pertinents (paragraphe, section, sous section, . . . etc.) et non pas des documents entiers. Ces nouveaux formats offrent la possibilité de représenter simultanément l'information textuelle et l'information relative à la structure du document. Ils permettent ainsi de représenter l'information sous une forme plus riche que le simple contenu et adaptée à des besoins spécifiques. À cette effet, nous présentons dans cette section une approche de recherche d'information structurée basée sur l'extension triadique de l'analyse de concepts formels.

Considérons le contexte formel triadique $\mathfrak{R} = (\mathcal{D}, \mathcal{T}, \mathcal{E}, \mathfrak{R})$, où \mathcal{D} est l'ensemble de documents, \mathcal{T} est l'ensemble de termes d'indexation, \mathcal{E} est l'ensemble d'éléments, et \mathfrak{R} est une relation triadique incluse dans $\mathcal{D} \times \mathcal{T} \times \mathcal{E}$. Contrairement à l'approche dyadique, nous admettons que les documents sont des ensembles d'éléments, et que ces éléments sont significativement exhaustifs et peuvent être retournés en tant que réponse aux requêtes d'utilisateurs. Dans la figure 4.3, nous donnons un exemple de document structuré. Ce document est structuré en quatre éléments ; *title*, *author*, *year* et *Body*.

```
<?xml version="1.0" encoding="UTF-8"?>
<article>
<title> A Triadic Approach to formal concept Analysis</title>
<author> F. Lehmann and R. Wille </author>
<year> 1995 </year>
<abstract> Formal concept analysis, developed during ....</abstract>
<body>
<sec >
<st> Triadic Contexts </st>
<p> Formal concept analysis views cocepts as </p>
</sec>
<sec >
<st> Triadic Concepts </st>
<p> concept are understoud as </p>
<p> as in the </p>
.
.
</sec>
.
.
</body>
</article>
```

FIGURE 4.3 – Exemple de Document Structuré

Un triplet $(d, t, e) \in \mathfrak{R}$ signifie que l'élément e du document d est indexé par le terme t . Nous soulignons que cette technique d'indexation est connue dans la littérature comme l'approche d'indexation basée sur des éléments (*element-based indexing approach*) [Lalmas, 2009] qui permet la recherche d'éléments à n'importe quel niveau de granularité. Notons que nous nous limitons dans cette section à l'indexation binaire (i.e un élément d'un document satisfait on non un terme).

4.3.1 Traitement des Requêtes Orientées Contenu

Les requêtes orientées contenu (Content-Only Queries) consistent en l'expression de termes sans spécifier de contraintes de structures sur les éléments recherchés. Nous nous intéressons dans cette section au traitement de ce type de requête en se basant sur l'extension triadique de l'ACF. Dans la suite de cette section nous utilisons la notion de "concept formel dyadique" pour le distinguer du concept formel triadique.

4.3.1.1 Requêtes Conjonctives

Soit $T = \{t_1, t_2, \dots, t_n\}$ un ensemble de termes ($T \subseteq \mathcal{T}$) et $\hat{Q}(T)$ la requête conjonctive orientée contenu construite sur T (i.e. $t_1 \wedge t_2 \wedge \dots \wedge t_n$). La réponse à cette requête consiste en l'ensemble d'éléments des documents indexés par tous les termes de T . Notons par \hat{P} l'ensemble de ces réponses, qui sont données sous la forme de couples (d, e) . \hat{P} peut être obtenu mathématiquement par l'opérateur de dérivation de suffisance appliqué à l'ensemble T . En se basant sur l'approche triadique de l'analyse de concepts formels (c.f. §1.6), \hat{P} est donné par :

$$\hat{P} = T^{\Delta\tau} = \{(d, e) \in \mathcal{D} \times \mathcal{E} \mid \exists t \in T, (d, t, e) \in \mathfrak{R}\} \quad (4.1)$$

Dans une notation polymorphe, $\hat{P}^{\Delta\tau}$ définit l'ensemble des termes satisfaits par tous les éléments de \hat{P} . $\hat{P}^{\Delta\tau}$ est donné par :

$$\hat{P}^{\Delta\tau} = \{t \in \mathcal{T} \mid \exists (d, e) \in \hat{P}, (d, t, e) \in \mathfrak{R}\} \quad (4.2)$$

Pour un ensemble de termes T , la proposition suivante établit l'unicité du contexte formel dyadique relatif à la requête $\hat{Q}(T)$:

Proposition 4.3.1. *Étant donné le contexte formel triadique $\mathfrak{R} = (\mathcal{D}, \mathcal{T}, \mathcal{E}, \mathfrak{R})$ et l'ensemble de termes T , alors il existe un et un seul concept formel dyadique relatif à la requête conjonctive orientée contenu $\hat{Q}(T)$ donné par :*

$$\langle T^{\Delta\tau}, (T^{\Delta\tau})^{\Delta\tau} \rangle$$

Preuve. La preuve résulte naturellement de la propriété de fermeture de la composition de l'opérateur de suffisance $((\cdot)^{\Delta\tau})^{\Delta\tau}$.

□

La Proposition. 4.3.1 nous permet de définir le mode de recherche par interrogation². Ainsi, pour une requête conjonctive $\hat{Q}(T)$, nous générons le concept formel dyadique $\langle T^{\Delta\tau}, (T^{\Delta\tau})^{\Delta\tau} \rangle$. Pour le mode de recherche par navigation, nous proposons

2. Ce mode de recherche est défini dans certaines approches de RI basées sur l'ACF classique (§2.6).

de construire le treillis de concepts formel dyadique de la forme $\langle \{(d_i, e_j) \mid d_i \in \mathcal{D}, e_j \in \mathcal{E}\}, T \rangle$ isomorphe au contexte triadique \mathfrak{K} . Pour ce faire, on derive à partir de \mathfrak{K} le contexte formel dyadique $\mathfrak{K}_{\mathcal{T}} = (\mathcal{D} \times \mathcal{E}, \mathcal{T}, \mathfrak{K}_{\mathcal{T}})$, où $\mathfrak{K}_{\mathcal{T}} = \{((d, e), t) \mid (d, t, e) \in \mathfrak{K}\}$. Le treillis de concepts formels induit à partir de $\mathfrak{K}_{\mathcal{T}}$ ($\mathcal{L}(\mathfrak{K}_{\mathcal{T}})$) permet de caractériser, conformément au Théorème.4.2.1, l'ensemble des requêtes orientées contenu construites sur l'ensemble \mathcal{T} .

Exemple 4.3.1. *Pour illustrer le traitement de requêtes orientées contenu, nous considérons le contexte triadique $\mathfrak{K}_e = (\mathcal{D}_e, \mathcal{T}_e, \mathcal{E}_e, \mathfrak{K}_e)$ du Tableau. 4.1.*

Tableau 4.1 – Le Contexte Formel Triadique $\mathfrak{K}_e = (\mathcal{D}_e, \mathcal{T}_e, \mathcal{E}_e, \mathfrak{K}_e)$

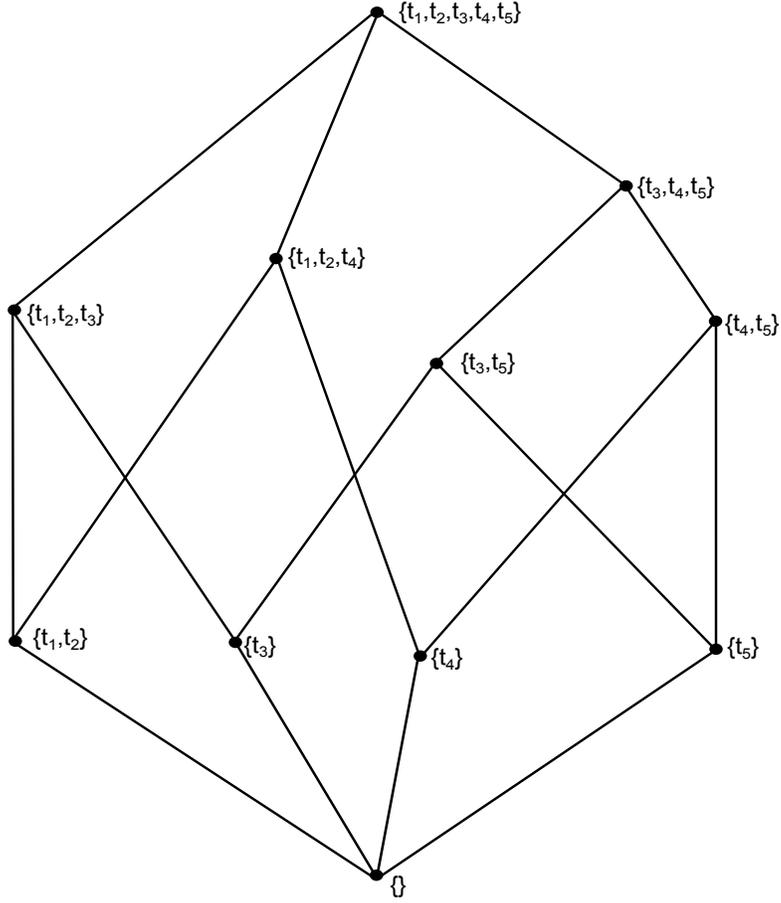
	$e_1 = \text{Title}$					$e_2 = \text{Authors}$				
	t_1	t_2	t_3	t_4	t_5	t_1	t_2	t_3	t_4	t_5
d_1			×	×	×	×	×		×	
d_2	×	×						×		
d_3	×	×	×							×
d_4				×	×			×		
d_5	×	×							×	×
d_6	×	×	×			×	×			
	$e_3 = \text{Abstract}$					$e_4 = \text{Body}$				
	t_1	t_2	t_3	t_4	t_5	t_1	t_2	t_3	t_4	t_5
d_1	×	×	×						×	
d_2				×	×			×		
d_3	×	×							×	
d_4	×	×	×						×	
d_5	×	×	×	×	×	×	×			
d_6			×					×		×

Si on considère la requête conjonctive $t_1 \wedge t_2$, le concept formel dyadique calculé est : $\langle \{t_1, t_2\}^{\Delta\tau}, (\{t_1, t_2\}^{\Delta\tau})^{\Delta\tau} \rangle = \langle \{t_1, t_2\}, \{(d_1, e_2), (d_1, e_3), (d_2, e_1), (d_3, e_1), (d_3, e_3), (d_4, e_3), (d_5, e_1), (d_5, e_3), (d_5, e_4), (d_6, e_1), (d_6, e_2)\} \rangle$

Le diagramme de Hasse du treillis de concepts formels associé au contexte formel dyadique $(\mathcal{D}_e \times \mathcal{E}_e, \mathcal{T}_e, (\mathfrak{K}_e)_{\mathcal{T}})$ est donné dans la Figure. 4.4 (Page. 90). Pour une meilleure lisibilité l'ensemble de concepts formels est énuméré dans le Tableau. 4.2 (Page. 91).

4.3.1.2 Requêtes Disjonctives

Considérons l'ensemble de termes $T = \{t_1, t_2, \dots, t_n\}$. La réponse à la requête disjonctive (orientée contenu) $\check{Q}(T)$, donnée par $t_1 \vee t_2 \vee \dots \vee t_n$, sont tous les éléments qui sont indexés par au moins un terme de T . L'ensemble de ces éléments (noté \check{P}

FIGURE 4.4 – Diagramme de Hasse du Treillis $\mathcal{L}(\mathfrak{K}_{\mathcal{T}})$

peut être exprimé par l'opérateur de dérivation de possibilité appliqué à l'ensemble T . Formellement, \check{P} est donné par :

$$\check{P} = T^{\Pi_{\mathcal{T}}} = \{(d, e) \in \mathcal{D} \times \mathcal{E} \mid \exists t \in T, (d, t, e) \in \mathfrak{A}\} \quad (4.3)$$

Cette expression nous permet de définir le mode de recherche par interrogation pour les requêtes disjonctives (orientées contenu). Pour la recherche par navigation, nous proposons de calculer toutes les NII-paires de la forme $\langle \check{P}, T \rangle$ tel que $\check{P} = T^{\Pi_{\mathcal{T}}}$ et $T = \check{P}^{\text{N}_{\mathcal{T}}}$, où $(.)^{\text{N}_{\mathcal{T}}}$ est défini par :

$$\check{P}^{\text{N}_{\mathcal{T}}} = \{t \in \mathcal{T} \mid \forall (d, e) \in \mathcal{D} \times \mathcal{E} ((d, t, e) \in \mathfrak{A} \Rightarrow (d, e) \in \check{P})\} \quad (4.4)$$

Soit $\mathcal{P}(\mathfrak{K}_{\mathcal{T}})$ l'ensemble des NII-paires induit à partir du contexte formel dyadique $\mathfrak{K}_{\mathcal{T}}$ par les deux opérateurs de dérivation $\Pi_{\mathcal{T}}$ et $\text{N}_{\mathcal{T}}$. $\mathcal{P}_{\mathcal{T}}(\mathfrak{K})$ est un treillis complet isomorphe³ au treillis de concepts formels dyadique $\mathcal{L}(\mathfrak{K}_{\mathcal{T}})$.

3. Cet isomorphisme a été prouvé dans [Djouadi et Prade, 2011].

Tableau 4.2 – Les Concepts Formels Dyadiques du Contexte $(\mathcal{D}_e \times \mathcal{E}_e, \mathcal{T}_e, (\mathfrak{R}_e)_{\mathcal{T}})$

#	Intension	Extension
1	\emptyset	$\{(d_1, e_1), (d_1, e_2), (d_1, e_3), (d_1, e_4), (d_2, e_1), (d_2, e_2), (d_2, e_3), (d_2, e_4), (d_3, e_1), (d_3, e_2), (d_3, e_3), (d_3, e_4), (d_4, e_1), (d_4, e_2), (d_4, e_3), (d_4, e_4), (d_5, e_1), (d_5, e_2), (d_5, e_3), (d_5, e_4), (d_6, e_1), (d_6, e_2), (d_6, e_3), (d_6, e_4)\}$
2	$\{t_5\}$	$\{(d_1, e_1), (d_2, e_3), (d_3, e_2), (d_4, e_1), (d_5, e_2), (d_5, e_3), (d_6, e_4)\}$
3	$\{t_4\}$	$\{(d_1, e_1), (d_1, e_2), (d_1, e_4), (d_2, e_3), (d_3, e_4), (d_4, e_1), (d_4, e_4), (d_5, e_2), (d_5, e_3)\}$
4	$\{t_3\}$	$\{(d_1, e_1), (d_1, e_3), (d_2, e_2), (d_2, e_4), (d_3, e_1), (d_4, e_2), (d_4, e_3), (d_5, e_3), (d_6, e_1), (d_6, e_3), (d_6, e_4)\}$
5	$\{t_4 t_5\}$	$\{(d_1, e_1), (d_2, e_3), (d_4, e_1), (d_5, e_2), (d_5, e_3)\}$
6	$\{t_3, t_5\}$	$\{(d_1, e_1), (d_5, e_3), (d_6, e_4)\}$
7	$\{t_1, t_2\}$	$\{(d_1, e_2), (d_1, e_3), (d_2, e_1), (d_3, e_1), (d_3, e_3), (d_4, e_3), (d_5, e_1), (d_5, e_3), (d_5, e_4), (d_6, e_1), (d_6, e_2)\}$
8	$\{t_3, t_4, t_5\}$	$\{(d_1, e_1), (d_5, e_3)\}$
9	$\{t_1, t_2, t_4\}$	$\{(d_1, e_2), (d_5, e_3)\}$
10	$\{t_1, t_2, t_3\}$	$\{(d_1, e_3), (d_3, e_1), (d_4, e_3), (d_5, e_3), (d_6, e_1)\}$
11	$\{t_1, t_2, t_3, t_4, t_5\}$	$\{(d_5, e_3)\}$

Exemple 4.3.2. *Pour illustrer le traitement de requêtes disjonctives orientées contenu, nous reconsidérons le contexte triadique du Tableau. 4.1. La Figure. 4.5 (Page. 92) illustre le diagramme de Hasse associé au treillis de NII-paires calculées à partir de ce contexte formel (i.e $\mathcal{P}(\mathfrak{R}_{\mathcal{T}})$). L'ensemble de NII-paires sont données dans le Tableau. 4.3 (Page. 93).*

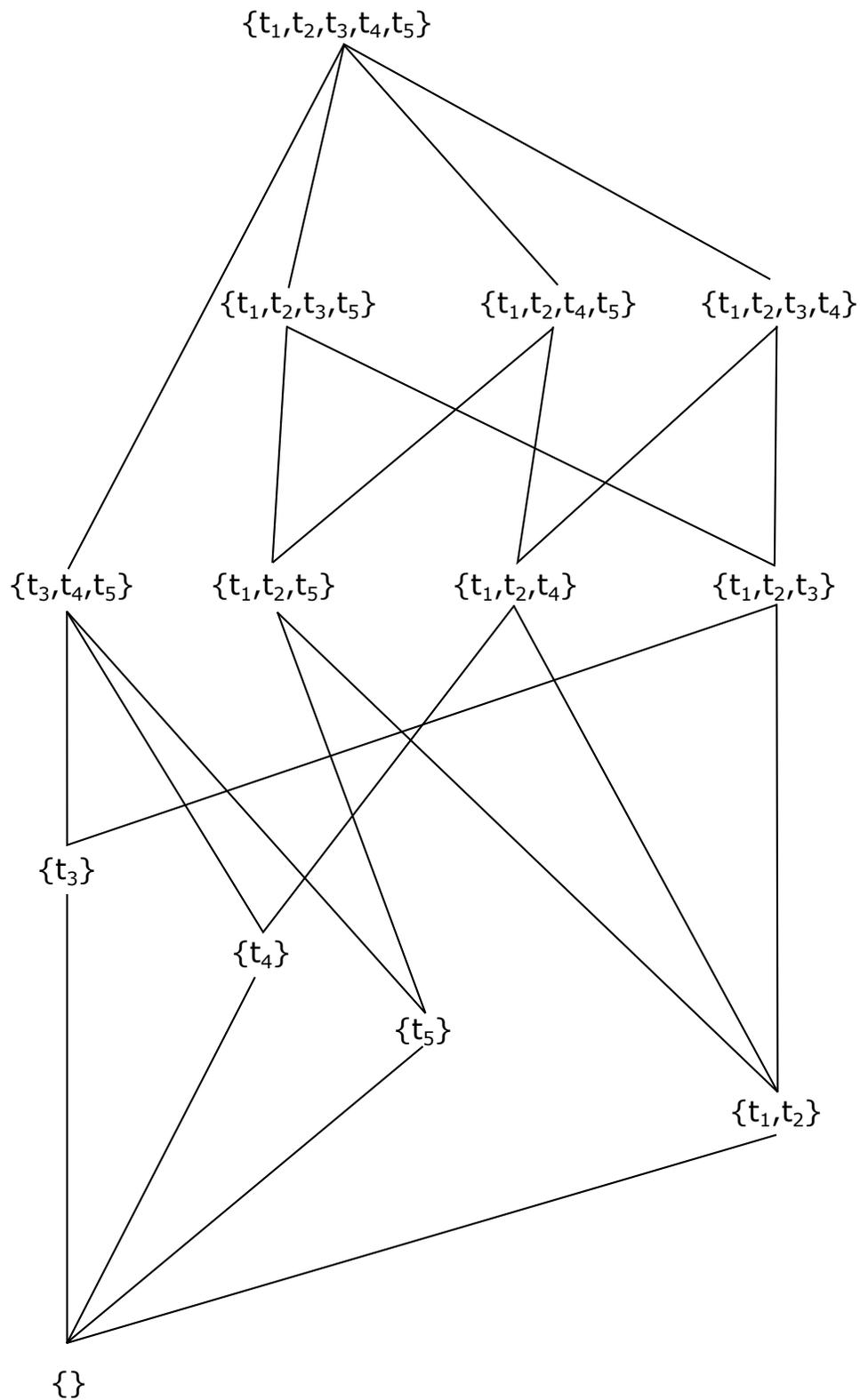


FIGURE 4.5 – Diagramme de Hasse du Treillis $\mathcal{P}(\mathfrak{R}_T)$

Tableau 4.3 – Les NII-Paires du Contexte Formel $(\mathcal{D}_e \times \mathcal{E}_e, \mathcal{T}_e, (\mathcal{R}_e)_{\mathcal{T}})$

Extension	Intension
$\{\}$	$\{\}$
$\{t_3\}$	$\{(d_1, e_1), (d_1, e_3), (d_2, e_2), (d_2, e_4), (d_3, e_1), (d_4, e_2), (d_4, e_3), (d_5, e_3), (d_6, e_1), (d_6, e_3), (d_6, e_4)\}$
$\{t_4\}$	$\{(d_1, e_1), (d_1, e_2), (d_1, e_4), (d_2, e_3), (d_3, e_4), (d_4, e_1), (d_4, e_4), (d_5, e_2), (d_5, e_3)\}$
$\{t_5\}$	$\{(d_1, e_1), (d_2, e_3), (d_3, e_2), (d_4, e_1), (d_5, e_2), (d_5, e_3), (d_6, e_4)\}$
$\{t_1, t_2\}$	$\{(d_1, e_2), (d_1, e_3), (d_2, e_1), (d_3, e_1), (d_3, e_3), (d_4, e_3), (d_5, e_1), (d_5, e_3), (d_5, e_4), (d_6, e_1), (d_6, e_2)\}$
$\{t_3, t_4\}$	$\{(d_1, e_1), (d_1, e_2), (d_1, e_3), (d_1, e_4), (d_2, e_2), (d_2, e_3), (d_2, e_4), (d_3, e_1), (d_3, e_4), (d_4, e_1), (d_4, e_2), (d_4, e_3), (d_4, e_4), (d_5, e_2), (d_5, e_3), (d_6, e_1), (d_6, e_3), (d_6, e_4)\}$
$\{t_3, t_5\}$	$\{(d_1, e_1), (d_1, e_3), (d_2, e_2), (d_2, e_3), (d_2, e_4), (d_3, e_1), (d_3, e_2), (d_4, e_1), (d_4, e_2), (d_4, e_3), (d_5, e_2), (d_5, e_3), (d_6, e_1), (d_6, e_3), (d_6, e_4)\}$
$\{t_4, t_5\}$	$\{(d_1, e_1), (d_1, e_2), (d_1, e_4), (d_2, e_3), (d_3, e_2), (d_3, e_4), (d_4, e_1), (d_4, e_4), (d_5, e_2), (d_5, e_3), (d_6, e_4)\}$
$\{t_3, t_4, t_5\}$	$\{(d_1, e_1), (d_1, e_2), (d_1, e_3), (d_1, e_4), (d_2, e_2), (d_2, e_3), (d_2, e_4), (d_3, e_1), (d_3, e_2), (d_3, e_4), (d_4, e_1), (d_4, e_2), (d_4, e_3), (d_4, e_4), (d_5, e_2), (d_5, e_3), (d_6, e_1), (d_6, e_2), (d_6, e_3), (d_6, e_4)\}$
$\{t_1, t_2, t_5\}$	$\{(d_1, e_1), (d_1, e_2), (d_1, e_3), (d_2, e_1), (d_2, e_3), (d_3, e_1), (d_3, e_2), (d_3, e_3), (d_4, e_1), (d_4, e_3), (d_5, e_1), (d_5, e_2), (d_5, e_3), (d_5, e_4), (d_6, e_1), (d_6, e_2), (d_6, e_4)\}$
$\{t_1, t_2, t_4\}$	$\{(d_1, e_1), (d_1, e_2), (d_1, e_3), (d_1, e_4), (d_2, e_1), (d_2, e_3), (d_3, e_1), (d_3, e_3), (d_3, e_4), (d_4, e_1), (d_4, e_3), (d_4, e_4), (d_5, e_1), (d_5, e_2), (d_5, e_3), (d_5, e_4), (d_6, e_1), (d_6, e_2)\}$
$\{t_1, t_2, t_3\}$	$\{(d_1, e_1), (d_1, e_2), (d_1, e_3), (d_2, e_1), (d_2, e_2), (d_2, e_4), (d_3, e_1), (d_3, e_3), (d_4, e_2), (d_4, e_3), (d_5, e_1), (d_5, e_3), (d_5, e_4), (d_6, e_1), (d_6, e_2), (d_6, e_3), (d_6, e_4)\}$
$\{t_1, t_2, t_3, t_5\}$	$\{(d_1, e_1), (d_1, e_2), (d_1, e_3), (d_2, e_1), (d_2, e_2), (d_2, e_3), (d_2, e_4), (d_3, e_1), (d_3, e_2), (d_3, e_3), (d_4, e_1), (d_4, e_2), (d_4, e_3), (d_5, e_1), (d_5, e_2), (d_5, e_3), (d_5, e_4), (d_6, e_1), (d_6, e_2), (d_6, e_3), (d_6, e_4)\}$
$\{t_1, t_2, t_3, t_4\}$	$\{(d_1, e_1), (d_1, e_2), (d_1, e_3), (d_1, e_4), (d_2, e_1), (d_2, e_3), (d_2, e_4), (d_3, e_1), (d_3, e_3), (d_3, e_4), (d_4, e_1), (d_4, e_2), (d_4, e_3), (d_4, e_4), (d_5, e_1), (d_5, e_2), (d_5, e_3), (d_5, e_4), (d_6, e_1), (d_6, e_2), (d_6, e_3), (d_6, e_4)\}$
$\{t_1, t_2, t_4, t_5\}$	$\{(d_1, e_1), (d_1, e_2), (d_1, e_3), (d_1, e_4), (d_2, e_1), (d_2, e_3), (d_3, e_1), (d_3, e_2), (d_3, e_3), (d_3, e_4), (d_4, e_1), (d_4, e_3), (d_4, e_4), (d_5, e_1), (d_5, e_2), (d_5, e_3), (d_5, e_4), (d_6, e_1), (d_6, e_2), (d_6, e_3), (d_6, e_4)\}$
\mathcal{T}_e	$\{(d_1, e_1), (d_1, e_2), (d_1, e_3), (d_1, e_4), (d_2, e_1), (d_2, e_2), (d_2, e_3), (d_2, e_4), (d_3, e_1), (d_3, e_2), (d_3, e_3), (d_3, e_4), (d_4, e_1), (d_4, e_2), (d_4, e_3), (d_4, e_4), (d_5, e_1), (d_5, e_2), (d_5, e_3), (d_5, e_4), (d_6, e_1), (d_6, e_2), (d_6, e_3), (d_6, e_4)\}$

4.3.2 Traitement de Requêtes Orientées Contenu et Structure

Les langages de requêtes orientés contenu et structure⁴ fournissent aux utilisateurs un moyen de spécifier, en plus des contenus, les contraintes sur les structures des informations recherchées. On peut distinguer trois catégories de ces langages [Lalmas, 2009] :

1. Langages basés sur les balises (Tag-based langages),
2. Langages basés sur les chemins (Path-based langages),
3. Langages basés sur les clauses (Clause-based langages).

Dans cette thèse, nous nous intéressons à la première catégorie, où ces langages (à l'exemple de XSEarch [Cohen et al., 2003]) offrent aux utilisateurs la possibilité d'accompagner (annoter) les termes de la requête avec une seule balise. Les balises permettent de spécifier les contraintes structurelles des résultats ciblés. À titre d'exemple, si l'information recherchée est “*les résumés sur l'analyse de concepts formels*”, la syntaxe de requête peut être donnée par : “*abstract : analyse de concepts formels*”.

Conformément à la notation que nous avons utilisé, les contraintes qui peuvent servir pour annoter les requêtes correspondent aux éléments de \mathcal{E} . Nous proposons ci-après une approche pour traiter les formes conjonctive et disjonctive de ce type de requête.

Soit $T = \{t_1, t_2, \dots, t_n\}$ un ensemble de termes ($T \subseteq \mathcal{T}$) et $\hat{Q}(e, T)$ la requête conjonctive orientée contenu et structure construite sur T (i.e. $e : t_1 \wedge t_2 \wedge \dots \wedge t_n$). Comme la contrainte sur la structure est connue, le processus de recherche se résume alors à rechercher la liste de documents contenant l'élément (ou la structure) e et satisfaisant tous les termes de cette requête. Pour ce faire, nous considérons le contexte formel dyadique $\mathfrak{R}_e = (\mathcal{D}, \mathcal{T}, \mathfrak{R}_e)$, où $\mathfrak{R}_e = \{(d, t) \mid (d, t, e) \in \mathfrak{R}\}$. La solution se ramène alors au cas classique, et l'ensemble de documents retourné est donné par :

$$T^{\Delta/\mathfrak{R}_e} = \{d \in \mathcal{D} \mid \forall t \in \mathcal{T} (t \in T \Rightarrow (d, t) \in \mathfrak{R}_e)\}$$

$(.)^{\Delta/\mathfrak{R}_e}$ permet aussi d'exprimer duallement l'ensemble de termes indexés par l'élément e de tous les documents d'un ensemble D :

$$D^{\Delta/\mathfrak{R}_e} = \{t \in \mathcal{T} \mid \forall d \in \mathcal{D} (d \in D \Rightarrow (d, t) \in \mathfrak{R}_e)\}$$

Le concept formel de la requête $\hat{Q}(e, T)$ est alors donné par la paire :

$$\langle T^{\Delta/\mathfrak{R}_e}, (T^{\Delta/\mathfrak{R}_e})^{\Delta/\mathfrak{R}_e} \rangle$$

4. C'est sur ce type de requête que se focalisent la plupart des langages de requêtes XML [Lalmas, 2009].

Dans le même contexte, pour la requête disjonctive orientée contenu et structure $\tilde{Q}(e, T)$ (i.e. $e : t_1 \vee t_2 \vee \dots t_n$), le processus de recherche se résume alors à retrouver les documents contenant l'élément e (la contrainte) indexés par au moins un terme de T . En utilisant la lecture possibiliste de l'analyse de concepts formels, l'ensemble de ces documents est exprimé par :

$$T^{\Pi/\mathfrak{R}_e} = \{d \in \mathcal{D} \mid \exists t \in \mathcal{T}, (d, e, t) \in \mathfrak{R}\}$$

Pour le mode de recherche par navigation, nous proposons la construction du treillis de concepts formels $\mathcal{L}(\mathfrak{R}_e)$ pour les requêtes conjonctives et le treillis de NII-paires $\mathcal{P}(\mathfrak{R}_e)$ pour les requêtes disjonctives orientées contenu et structure. Pour la construction du treillis $\mathcal{P}(\mathfrak{R}_e)$, l'opérateur de dérivation de nécessité N est défini par :

$$D^{N/\mathfrak{R}_e} = \{t \in \mathcal{T} \mid \forall d \in \mathcal{D} ((d, e, t) \in \mathfrak{R} \Rightarrow d \in D)\}$$

4.4 Conclusion

Dans ce chapitre nous avons présenté nos contributions dans le cadre de l'ACF Booléenne pour la RI. Dans la première contribution, en partant de la théorie axiomatique \mathfrak{T} et du modèle \mathcal{M}_B , nous avons démontré des propriétés algébriques significatives entre des ensembles de paires (*Documents – Requête*) relatives à la RI Booléenne et des treillis de paires $\langle \text{Documents}, \text{termes} \rangle$ liées à l'ACF et sa lecture possibiliste. Ces propriétés permettent d'approuver l'utilisation de la théorie de l'ACF et sa lecture possibiliste dans le domaine de la RI d'un côté. De l'autre, elles offrent un cadre théorique pour le développement d'approches de recherche par navigation sur les treillis de l'ACF. Par rapport aux approches existantes de RI basées sur l'ACF, notre contribution traite, en plus des requêtes conjonctives, les requêtes disjonctives et leurs deux formes négatives.

Dans la seconde contribution, nous avons proposé une approche de RI structurée Booléenne basée sur l'extension triadique de l'ACF. L'approche permet de traiter les différentes formes de requêtes orientées contenu et de requêtes orientées contenu et structure.

Bien que ces deux approches traitent, en plus des requêtes conjonctives, les requêtes disjonctive et la leurs formes négatives, il reste que :

1. Elles se basent sur les relations *Documents* \times *Termes* Booléennes, est les poids des termes d'indexation dans les documents ne sont pas considérés,
2. Elles traitent exclusivement des requêtes Booléennes (absence de pondération de termes de la requête),
3. Elle n'associent aucun score de pertinence aux documents et/ou éléments pour les requêtes fournies,

En tenant compte de ces remarque, nous proposons dans le chapitre suivant, nos contributions basées sur les extensions floues de l'ACF pour prendre en charge les poids des termes d'indexation dans les documents et dans les requêtes et associer des scores de pertinence aux documents.

Chapitre 5

Extension Floue de l'Analyse de Concepts Formels pour la Recherche d'Information

Sommaire

5.1	Introduction	98
5.2	Enrichissement du Modèle \mathcal{M}_f basé sur l'ACF Floue	98
5.3	Pondération de Requêtes	107
5.3.1	Requêtes Disjonctives	109
5.3.2	Formes Négatives de Requêtes Pondérées	110
5.4	Ignorance Partielle/Totale dans la Matrice d'Incidence	110
5.5	Réduction Linéaire des Treillis	114
5.6	Conclusion	117

5.1 Introduction

Nous avons vu dans le Chapitre. 2 que la plupart des approches de RI basées sur l'ACF ne tiennent pas compte de la pondération des termes dans les requête et/ou dans les documents (§2.6). Cela est dû au fait que la théorie de l'ACF classique considère des contextes formels Booléens, où les entrées ne peuvent prendre que les deux valeurs 0 ou 1. Or, la pondération des termes de requêtes et de documents est avérée primordiale dans l'évaluation de la pertinence des documents par rapport aux requêtes utilisateurs, dans le but de renvoyer aux utilisateurs des listes triées des résultats.

Afin de prendre en compte une telle pondération, nous considérons l'extension floue de l'ACF pour : 1) considérer la relation floue $Documents \times Termes$, où les entrées de la matrice d'incidence sont des poids ($\in [0, 1]$) des termes dans les documents, 2) le classement des résultats de recherche et 3) le traitement des requêtes pondérées.

Dans ce chapitre, nous proposons d'abord dans la Section.5.2, une caractérisation algébrique des requêtes du modèle M_f (§3.4) en se basant sur les treillis de concepts formels semi-flous et les treillis de NII-paires semi-floues. Notons que ce dernier type de treillis est originalement introduit dans le cadre de cette thèse. La Section. 5.3, élargit la contribution présentée dans la Section. 5.2 en vue de prendre en compte les requêtes pondérées.

L'absence d'information est une réalité fort souvent constatée dans des matrices d'incidence $Documents \times Termes$, notamment sur des gros corpus. À cette fin, nous proposons de modéliser cette absence (totale ou partielle) d'information par un intervalle de vérité. En conséquence, la Section. 5.4 propose une algèbre d'intervalles dans le cadre de la RI basée sur l'ACF floue.

Soucieux que l'ensemble (ou une partie) de nos propositions théoriques puissent être implémentées de manière à passer à l'échelle (scalable), notre dernière contribution vise à faciliter l'interrogation par navigation en transformant les treillis en structures linéaires. Cette aspect est présenté dans la Section. 5.5.

5.2 Enrichissement du Modèle \mathcal{M}_f basé sur l'ACF Floue

Dans cette section, nous nous intéressons au modèle de recherche d'information flou standard (§2.2.2). Pour cela, nous considérons l'intervalle $L = [0, 1]$ et le contexte formel flou $\tilde{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \mathcal{I})$, où \mathcal{D} est un ensemble fini de documents ($\mathcal{D} = \{d_1, d_2, \dots, d_m\}$), \mathcal{T} est un ensemble fini de termes d'indexation ($\mathcal{T} = \{t_1, t_2, \dots, t_n\}$) et \mathcal{I} est une relation binaire floue dans $L^{\mathcal{D} \times \mathcal{T}}$ qui associe à chaque couple $(d, t) \in \mathcal{D} \times \mathcal{T}$ ($\mathcal{I}(d, t)$) le poids du terme t dans le document d .

Soit \mathcal{Q} l'ensemble de toutes les requêtes Booléennes construites sur l'ensemble de termes d'indexation \mathcal{T} . Une requête élémentaire conjonctive $\hat{Q}(T)$, construite sur l'ensemble $T = \{t_1, t_2, \dots, t_r\}$ de termes d'indexation, est de la forme $t_1 \wedge t_2 \wedge \dots \wedge t_r$. Une requête disjonctive élémentaire $\check{Q}(T)$ est de la forme $t_1 \vee t_2 \vee \dots \vee t_r$. Les négations de ces deux formes de requêtes sont donnée par $\neg\hat{Q}(T)$ ($\neg(t_1 \wedge t_2 \wedge \dots \wedge t_r)$) et $\neg\check{Q}(T)$ ($\neg(t_1 \vee t_2 \vee \dots \vee t_r)$). Dans ce qui suit nous considérons l'image \tilde{f} défini par :

$$\begin{aligned} \tilde{f} : \mathcal{T} &\rightarrow L^{\mathcal{D}} \\ t &\mapsto \{d_i^{\alpha_i}, i = 1 \dots m \mid \alpha_i = \mathcal{I}(d_i, t)\} \end{aligned} \quad (5.1)$$

La proposition suivante caractérise, dans le modèle \mathcal{M}_f (§3.4), les sous-ensembles de l'extension du prédicat \mathbf{S} en considérant ces quatre formes élémentaires de requêtes.

Proposition 5.2.1. *Les interprétations suivantes sont vraies dans le modèle \mathcal{M}_f :*

- (IF1) $\forall T \neq \emptyset, \llbracket \mathbf{S} \rrbracket^{I_f}(\bigcap_{t_i \in T} \tilde{f}(t_i), \hat{Q}(T))$
- (IF2) $\forall T \neq \emptyset, \llbracket \mathbf{S} \rrbracket^{I_f}(\bigcup_{t_i \in T} \tilde{f}(t_i), \check{Q}(T))$
- (IF3) $\forall T \neq \emptyset, \llbracket \mathbf{S} \rrbracket^{I_f}(\mathcal{D} \setminus \bigcap_{t_i \in T} \tilde{f}(t_i), \neg\hat{Q}(T))$
- (IF4) $\forall T \neq \emptyset, \llbracket \mathbf{S} \rrbracket^{I_f}(\mathcal{D} \setminus \bigcup_{t_i \in T} \tilde{f}(t_i), \neg\check{Q}(T))$

Preuve. La preuve résulte naturellement du théorème 3.4.1 - Page. 69 et la fonction d'interprétation I_f définie dans §3.4. Ainsi, nous avons à interpréter chaque variable x_i par t_i pour obtenir respectivement (IF1), (IF2), (IF3), (IF4). □

Soit \hat{F} (resp. \check{F} , \hat{F}_\neg , \check{F}_\neg) l'ensemble des paires (D, Q) tel que l'interprétation $\llbracket \mathbf{S} \rrbracket^{I_f}(D, Q)$ est vraie et Q est une requête conjonctive (resp. disjonctive, négative conjonctive, négative disjonctive) construite sur l'ensemble \mathcal{T} . Formellement, ces ensembles sont donnés par :

$$\begin{aligned} \hat{F} &= \{(\tilde{D}, \hat{Q}(T)) \in L^{\mathcal{D}} \times \mathcal{Q} \mid T \neq \emptyset \text{ et } \llbracket \mathbf{S} \rrbracket^{I_f}(\tilde{D}, \hat{Q}(T)) = \text{vrai}\} \\ \check{F} &= \{(\tilde{D}, \check{Q}(T)) \in L^{\mathcal{D}} \times \mathcal{Q} \mid T \neq \emptyset \text{ et } \llbracket \mathbf{S} \rrbracket^{I_f}(\tilde{D}, \check{Q}(T)) = \text{vrai}\} \\ \hat{F}_\neg &= \{(\tilde{D}, \neg\hat{Q}(T)) \in L^{\mathcal{D}} \times \mathcal{Q} \mid T \neq \emptyset \text{ et } \llbracket \mathbf{S} \rrbracket^{I_f}(\tilde{D}, \neg\hat{Q}(T)) = \text{vrai}\} \\ \check{F}_\neg &= \{(\tilde{D}, \neg\check{Q}(T)) \in L^{\mathcal{D}} \times \mathcal{Q} \mid T \neq \emptyset \text{ et } \llbracket \mathbf{S} \rrbracket^{I_f}(\tilde{D}, \neg\check{Q}(T)) = \text{vrai}\}. \end{aligned}$$

Afin de caractériser ces ensembles, nous définissons ci-après des extensions des

opérateurs de dérivation de l'ACF et de sa lecture possibiliste :

$$T^\Delta(d) = \bigwedge_{t \in \mathcal{T}} \mathcal{I}(d, t) \quad (5.2)$$

$$\tilde{D}^{\Delta_c} = \{t \in \mathcal{T} \mid \forall d \in \mathcal{D}, \tilde{D}(d) \leq \mathcal{I}(d, t)\} \quad (5.3)$$

$$= \{t \in \mathcal{T} \mid \forall d \in \mathcal{D}, (\tilde{D}(d) \rightarrow_{RG} \mathcal{I}(d, t) = 1)\} \quad (5.4)$$

$$T^\Pi(d) = \bigvee_{t \in \mathcal{T}} \mathcal{I}(d, t) \quad (5.5)$$

$$\tilde{D}^{N_c} = \{t \in \mathcal{T} \mid \forall d \in \mathcal{D}, \mathcal{I}(d, t) \leq \tilde{D}(d)\} \quad (5.6)$$

$$= \{t \in \mathcal{T} \mid \forall d \in \mathcal{D}, (\mathcal{I}(d, t) \rightarrow_{RG} \tilde{D}(d) = 1)\} \quad (5.7)$$

Où “ \rightarrow_{RG} ” correspond à l'implication Rescher-Gaïnes donnée par :

$$x \rightarrow_{RG} y = \begin{cases} 1 & x \leq y \\ 0 & \text{sinon} \end{cases}$$

Les opérateurs $(.)^\Delta$ et $(.)^\Pi$ s'appliquent sur des ensembles classiques et génèrent des ensembles flous. Tandis que $(.)^{\Delta_c}$ et $(.)^{N_c}$ s'appliquent sur des ensembles flous et génèrent des ensembles classiques. Nous nous intéressons ci-après aux propriétés algébriques de certaines compositions de ces opérateurs.

Proposition 5.2.2. *la paire (Δ_c, Δ) forme une L-connection de Galois antitone.*

Preuve. La preuve vient du fait que :

- Δ est un opérateur flou antitone,
- Δ_c est un opérateur antitone,
- La composition $\Delta \circ \Delta_c$ est un opérateur de fermeture floue,
- La composition $\Delta_c \circ \Delta$ est un opérateur de fermeture.

□

Théorème 5.2.1. *Étant donné le L-contexte $\tilde{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \mathcal{I})$. Nous avons alors :*

I1) Π est isotone : $T_1 \subseteq T_2 \Rightarrow T_1^\Pi \subseteq T_2^\Pi$,

I2) N_c est isotone : $\tilde{D}_1 \subseteq \tilde{D}_2 \Rightarrow \tilde{D}_1^{N_c} \subseteq \tilde{D}_2^{N_c}$,

C3) La composition $N_c \circ \Pi$ (i.e. $((.)^\Pi)^{N_c}$) est un opérateur de fermeture flou,

O1) La composition $\Pi \circ N_c$ (i.e. $((.)^{N_c})^\Pi$) est un opérateur d'ouverture flou.

Preuve.

I1) $\forall d \in \mathcal{D}, T_1^\Pi(d) = \bigvee_{t \in T_1} \mathcal{I}(d, t), T_2^\Pi(d) = \bigvee_{t \in T_2} \mathcal{I}(d, t)$.

Pour d donnée, notons $\alpha_1 = \bigvee_{t \in T_1} \mathcal{I}(d, t)$ et $\alpha_2 = \bigvee_{t \in T_2} \mathcal{I}(d, t)$. Si $T_1 \subseteq T_2$ alors nous avons deux cas de figures :

cas N°1 : $\exists t \in T_2 \setminus T_1, \mathcal{I}(t, d) > \alpha_1$ alors $\alpha_2 > \alpha_1$ et par conséquence $T_1^{\text{II}}(d) \leq T_2^{\text{II}}(d)$.
cas N°2 : $\forall t \in T_2 \setminus T_1 \mathcal{I}(t, d) \leq \alpha_1$ alors $\alpha_2 = \alpha_1$ et par conséquence $T_1^{\text{II}}(d) \leq T_2^{\text{II}}(d)$.
Dans les deux cas $T_1^{\text{II}}(d) \leq T_2^{\text{II}}(d)$ alors $T_1 \subseteq T_2 \Rightarrow T_1^{\text{II}} \subseteq T_2^{\text{II}}$

I2) À partir de la définition de l'opérateur N_c , $\tilde{D}_1^{N_c} = \{t \mid \forall d \in \mathcal{D}, \mathcal{I}(d, t) \leq \tilde{D}_1(d)\}$
et $\tilde{D}_2^{N_c} = \{t \mid \forall d \in \mathcal{D}, \mathcal{I}(d, t) \leq \tilde{D}_2(d)\}$,
et puisque $\tilde{D}_1 \subseteq \tilde{D}_2 \Rightarrow \forall d \in \mathcal{D}, \tilde{D}_1(d) \leq \tilde{D}_2(d)$ alors $\forall t \in \tilde{D}_1^{N_c} \forall d \in \mathcal{D}, \mathcal{I}(d, t) \leq \tilde{D}_1(d) \leq \tilde{D}_2(d) \Rightarrow t \in \tilde{D}_2^{N_c} \Rightarrow \tilde{D}_1^{N_c} \subseteq \tilde{D}_2^{N_c}$

C3) Nous allons prouver les propriétés (CL1), (CL2) et (CL3) de la Définition. 1.2.6 (Page. 8)

(CL1) Extensivité : $T \subseteq (T^{\text{II}})^{N_c}$

$$\begin{aligned} T^{\text{II}} &= \{d^\alpha \mid \alpha = \bigvee_{t \in T} \mathcal{I}(d, t)\} \\ \Rightarrow \forall d \in \mathcal{D} \forall t \in T, \mathcal{I}(d, t) &\leq T^{\text{II}}(d) && \text{(Puisque } \alpha \text{ est le maximum)} \\ \Rightarrow \forall t \in T, t &\in (T^{\text{II}})^{N_c} && \text{(À partir de la défintion de } N_c) \\ \Rightarrow T &\subseteq (T^{\text{II}})^{N_c} \end{aligned}$$

(CL2) Isotonie : $T_1 \subseteq T_2 \Rightarrow (T_1^{\text{II}})^{N_c} \subseteq (T_2^{\text{II}})^{N_c}$

Par application successive de I1) et I2) nous obtenons : $T_1 \subseteq T_2 \Rightarrow T_1^{\text{II}} \leq T_2^{\text{II}} \Rightarrow (T_1^{\text{II}})^{N_c} \subseteq (T_2^{\text{II}})^{N_c}$

(CL3) Idempotence : $(T^{\text{II}})^{N_c} = (((T^{\text{II}})^{N_c})^{\text{II}})^{N_c}$

i) d'un coté, à partir de la définition de $(.)^{\text{II}}$, nous avons :

$$\begin{aligned} T^{\text{II}} &= \{d^\alpha \mid \alpha = \bigvee_{t \in T} \mathcal{I}(d, t)\} \\ \Rightarrow (T^{\text{II}})^{N_c} &= \{t \mid \forall d, \mathcal{I}(d, t) \leq T^{\text{II}}(d)\} \\ \Rightarrow ((T^{\text{II}})^{N_c})^{\text{II}} &= \{d^\alpha \mid \alpha = \bigvee_{t \in (T^{\text{II}})^{N_c}} \mathcal{I}(d, t)\} \\ \Leftrightarrow \forall d, ((T^{\text{II}})^{N_c})^{\text{II}}(d) &= \bigvee_{t \in (T^{\text{II}})^{N_c}} \mathcal{I}(d, t) \end{aligned}$$

de l'autre côté, $(T^{\text{II}})^{N_c} = \{t \mid \forall d, \mathcal{I}(d, t) \leq T^{\text{II}}(d)\} \Rightarrow \forall d \in \mathcal{D} \forall t \in (T^{\text{II}})^{N_c}, \mathcal{I}(d, t) \leq T^{\text{II}}(d)$, d'où $\forall d \in \mathcal{D}, ((T^{\text{II}})^{N_c})^{\text{II}}(d) = \bigvee_{t \in (T^{\text{II}})^{N_c}} \mathcal{I}(d, t) \leq T^{\text{II}}(d) \Rightarrow ((T^{\text{II}})^{N_c})^{\text{II}} \leq T^{\text{II}}$.
En appliquant I2) nous obtenons $((T^{\text{II}})^{N_c})^{\text{II}} \leq T^{\text{II}} \Rightarrow (((T^{\text{II}})^{N_c})^{\text{II}})^{N_c} \leq (T^{\text{II}})^{N_c}$

ii) En substituant, dans (CL1)- $T \subseteq (T^\Pi)^{N_c}$, T par $(T^\Pi)^{N_c}$ nous obtenons $:(T^\Pi)^{N_c} \subseteq (((T^\Pi)^{N_c})^\Pi)^{N_c}$

De i) et ii) nous concluons que $(T^\Pi)^{N_c} = (((T^\Pi)^{N_c})^\Pi)^{N_c}$

O1) Nous allons prouver les propriétés (OP1), (OP2) et (OP3) de la Définition. 1.2.7 (Page. 8)

(OP1) $\Pi \circ N_c$ est contractante : $\forall \tilde{D} \in L^{\mathcal{D}}, (\tilde{D}^{N_c})^\Pi \leq \tilde{D}$

$$\begin{aligned} & \forall \tilde{D} \in L^{\mathcal{D}}, \tilde{D}^{N_c} = \{t \in \mathcal{T} \mid \forall d \in \mathcal{D}, \mathcal{I}(d, t) \leq \tilde{D}(d)\} \\ \implies & \forall d \in \mathcal{D} \forall t \in \tilde{D}^{N_c}, \mathcal{I}(d, t) \leq \tilde{D}(d) \\ \implies & \forall d \in \mathcal{D} (\tilde{D}^{N_c})^\Pi(d) = \bigvee_{t \in \tilde{D}^{N_c}} \mathcal{I}(d, t) \leq \tilde{D}(d) \\ \implies & \forall d \in \mathcal{D} (\tilde{D}^{N_c})^\Pi(d) \leq \tilde{D}(d) \\ \implies & (\tilde{D}^{N_c})^\Pi \leq \tilde{D} \end{aligned}$$

(OP2) $\Pi \circ N_c$ est isotone : $\forall \tilde{D}_1, \tilde{D}_2 \in L^{\mathcal{D}}, \tilde{D}_1 \leq \tilde{D}_2 \Rightarrow (\tilde{D}_1^{N_c})^\Pi \leq (\tilde{D}_2^{N_c})^\Pi$
Par application successive de I2) et I1) nous obtenons : $\tilde{D}_1 \leq \tilde{D}_2 \Rightarrow \tilde{D}_1^{N_c} \subseteq \tilde{D}_2^{N_c}$
 $\Rightarrow (\tilde{D}_1^{N_c})^\Pi \leq (\tilde{D}_2^{N_c})^\Pi$

(OP3) $\Pi \circ N_c$ est idempotente : $(\tilde{D}_1^{N_c})^\Pi = (((\tilde{D}_1^{N_c})^\Pi)^{N_c})^\Pi$

i) En substituant, dans (CL1)- $T \subseteq (T^\Pi)^{N_c}$, T par $\tilde{D}_1^{N_c}$ nous obtenons $\tilde{D}_1^{N_c} \subseteq ((\tilde{D}_1^{N_c})^\Pi)^{N_c}$ et en appliquant I1) nous obtenons $(\tilde{D}_1^{N_c})^\Pi \leq (((\tilde{D}_1^{N_c})^\Pi)^{N_c})^\Pi$

ii) En substituant, dans (OP1)- $(\tilde{D}^{N_c})^\Pi \leq \tilde{D}$, \tilde{D} par $(\tilde{D}^{N_c})^\Pi$ nous obtenons $((\tilde{D}_1^{N_c})^\Pi)^{N_c} \subseteq ((\tilde{D}_1^{N_c})^\Pi)^{N_c}$

De i) et ii), nous concluons que $(\tilde{D}_1^{N_c})^\Pi = (((\tilde{D}_1^{N_c})^\Pi)^{N_c})^\Pi$ □

Corollaire 5.2.1. *la paire (N_c, Π) forme une L -connection de Galois isotone.*

Preuve. La prouve est déduite du théorème précédent. □

En utilisant ces opérateurs $(\Delta, \Delta_c, N_c$ et $\Pi)$, nous définissons les treillis de concepts formels semi-flous¹ (noté $\mathcal{B}(\tilde{\mathcal{K}})$) et de NII-paires semi-floues (noté $\mathcal{B}_{NII}(\tilde{\mathcal{K}})$) induit à partir du L -contexte $\tilde{\mathcal{K}}$ (et duallement à partir du L -contexte $\tilde{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \bar{\mathcal{I}})$, où $\bar{\mathcal{I}}(d, t) = 1 - \mathcal{I}(d, t)$), qui sont donnés par :

1. Ce treillis est le même généré par les approches de Yahia & Jaoua et Krajčí (§Page.22).

$$\mathcal{B}(\tilde{\mathcal{K}}) = \{\langle \tilde{D}, T \rangle \in L^{\mathcal{D}} \times 2^{\mathcal{T}} \mid \tilde{D} = T^{\Delta} \text{ et } T = \tilde{D}^{\Delta^c}\}$$

$$\mathcal{B}_{\text{NII}}(\tilde{\mathcal{K}}) = \{\langle \tilde{D}, T \rangle \in L^{\mathcal{D}} \times 2^{\mathcal{T}} \mid \tilde{D} = T^{\Pi} \text{ et } T = \tilde{D}^{\text{Nc}}\}$$

Proposition 5.2.3. *Soit les L -contextes $\tilde{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \mathcal{I})$ et $\bar{\tilde{\mathcal{K}}} = (\mathcal{D}, \mathcal{T}, \bar{\mathcal{I}})$, alors nous avons les propriétés suivantes :*

$$\begin{aligned} \text{(PF1)} \quad & \forall \tilde{D} \forall T \neq \emptyset, \left(\langle \tilde{D}, T \rangle \in \mathcal{B}(\tilde{\mathcal{K}}) \Rightarrow (\tilde{D}, \hat{Q}(T)) \in \hat{F} \right) \\ \text{(PF2)} \quad & \forall \tilde{D} \forall T \neq \emptyset, \left(\langle \tilde{D}, T \rangle \in \mathcal{B}_{\text{NII}}(\tilde{\mathcal{K}}) \Rightarrow (\tilde{D}, \check{Q}(T)) \in \check{F} \right) \\ \text{(PF3)} \quad & \forall \tilde{D} \forall T \neq \emptyset, \left(\langle \tilde{D}, T \rangle \in \mathcal{B}_{\text{NII}}(\bar{\tilde{\mathcal{K}}}) \Rightarrow (\tilde{D}, \neg \hat{Q}(T)) \in \hat{F}_{\neg} \right) \\ \text{(PF4)} \quad & \forall \tilde{D} \forall T \neq \emptyset, \left(\langle \tilde{D}, T \rangle \in \mathcal{B}(\bar{\tilde{\mathcal{K}}}) \Rightarrow (\tilde{D}, \neg \check{Q}(T)) \in \check{F}_{\neg} \right) \end{aligned}$$

Preuve.

(PF1) :

$$\begin{aligned} \langle \tilde{D}, T \rangle \in \mathcal{B}(\tilde{\mathcal{K}}) & \implies \forall d \in \mathcal{D}, \tilde{D}(d) = T^{\Delta}(d) \\ & \implies \tilde{D} = \{d_i^{\alpha_i}, i = 1 \dots m \mid \alpha_i = \bigwedge_{t_j \in T} \mathcal{I}(d_i, t_j)\} \\ & \implies \tilde{D} = \bigcap_{t_j \in T} \{d_i^{\alpha'_i}, i = 1 \dots m \mid \alpha'_i = \mathcal{I}(d_i, t_j)\} \\ & \implies \tilde{D} = \bigcap_{t_j \in T} \tilde{f}(t_j) \\ & \implies (\tilde{D}, \hat{Q}(T)) \in \hat{F} \end{aligned}$$

(PF2) :

$$\begin{aligned} \langle \tilde{D}, T \rangle \in \mathcal{B}_{\text{NII}}(\tilde{\mathcal{K}}) & \implies \forall d \in \mathcal{D}, \tilde{D}(d) = T^{\Pi}(d) \\ & \implies \tilde{D} = \{d_i^{\alpha_i}, i = 1 \dots m \mid \alpha_i = \bigvee_{t_j \in T} \mathcal{I}(d_i, t_j)\} \\ & \implies \tilde{D} = \bigcup_{t_j \in T} \{d_i^{\alpha'_i}, i = 1 \dots m \mid \alpha'_i = \mathcal{I}(d_i, t_j)\} \\ & \implies \tilde{D} = \bigcup_{t_j \in T} \tilde{f}(t_j) \\ & \implies (\tilde{D}, \check{Q}(T)) \in \check{F} \end{aligned}$$

(PF3) :

$$\begin{aligned}
\langle \tilde{D}, T \rangle \in \mathcal{B}_{\text{NII}}(\tilde{\mathcal{K}}) &\implies \forall d \in \mathcal{D}, \tilde{D}(d) = T^{\frac{\text{II}}{\tilde{\mathcal{K}}}}(d) \\
&\implies \tilde{D} = \{d_i^{\alpha_i}, i = 1 \dots m \mid \alpha_i = \bigvee_{t_j \in T} \bar{\mathcal{I}}(d_i, t_j)\} \\
&\implies \tilde{D} = \{d_i^{\alpha_i}, i = 1 \dots m \mid \alpha_i = \bigvee_{t_j \in T} (1 - \mathcal{I}(d_i, t_j))\} \\
&\implies \tilde{D} = \bigcup_{t_j \in T} \{d_i^{\alpha'_i}, i = 1 \dots m \mid \alpha'_i = 1 - \mathcal{I}(d_i, t_j)\} \\
&\implies \tilde{D} = \bigcup_{t_j \in T} \mathcal{D} \setminus \tilde{f}(t_j) \quad (\text{Complément de } \tilde{f}(t)) \\
&\implies \tilde{D} = \mathcal{D} \setminus \bigcap_{t_j \in T} \tilde{f}(t_j) \quad (\text{Lois de De Morgan}) \\
&\implies (\tilde{D}, \neg \tilde{Q}(T)) \in \hat{F}_-
\end{aligned}$$

Où $(\cdot)^{\frac{\text{II}}{\tilde{\mathcal{K}}}}$ correspond à l'application de l'opérateur $(\cdot)^{\text{II}}$ dans le L -contexte $\tilde{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \bar{\mathcal{I}})$.

(PF4) :

$$\begin{aligned}
\langle \tilde{D}, T \rangle \in \mathcal{B}(\tilde{\mathcal{K}}) &\implies \forall d \in \mathcal{D}, \tilde{D}(d) = T^{\hat{\Delta}}_{\tilde{\mathcal{K}}}(d) \\
&\implies \tilde{D} = \{d_i^{\alpha_i}, i = 1 \dots m \mid \alpha_i = \bigwedge_{t_j \in T} (1 - \mathcal{I}(d_i, t_j))\} \\
&\implies \tilde{D} = \bigcap_{t_j \in T} \{d_i^{\alpha'_i}, i = 1 \dots m \mid \alpha'_i = 1 - \mathcal{I}(d_i, t_j)\} \\
&\implies \tilde{D} = \bigcap_{t_j \in T} \mathcal{D} \setminus \tilde{f}(t_j) \\
&\implies \tilde{D} = \mathcal{D} \setminus \bigcup_{t_j \in T} \tilde{f}(t_j) \\
&\implies (\tilde{D}, \neg \hat{Q}(T)) \in \check{F}_-
\end{aligned}$$

□

En utilisant la notion de paires distinctes introduite dans le Chapitre. 4, nous définissons les ensembles de paires distinctes $\hat{\mathbb{F}}, \check{\mathbb{F}}, \hat{\mathbb{F}}_-, \check{\mathbb{F}}_-$ correspondant respectivement aux ensembles $\hat{F}, \check{F}, \hat{F}_-$ et \check{F}_- . La proposition suivante établit des propriétés utiles entre les ensembles de paires distinctes définies ci-dessus d'un coté, les treillis de concepts formels et de NII-paires des L -contextes $\tilde{\mathcal{K}}$ et $\hat{\mathcal{K}}$ de l'autre coté.

Proposition 5.2.4. Soit le L -contexte $\tilde{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \mathcal{I})$. Les propriétés suivantes sont vérifiées $\forall T \in 2^{\mathcal{T}}, T \neq \emptyset$,

$$(PF1') \quad \forall (\tilde{D}, \hat{Q}(T)) \in \hat{\mathbb{F}}, \exists! \langle \tilde{D}', T' \rangle \in \mathcal{B}(\tilde{\mathcal{K}}) \text{ tel que } \tilde{D}' = \tilde{D}.$$

$$(PF2') \quad \forall (\tilde{D}, \check{Q}(T)) \in \check{\mathbb{F}}, \exists! \langle \tilde{D}', T' \rangle \in \mathcal{B}_{\text{NII}}(\tilde{\mathcal{K}}) \text{ tel que } \tilde{D}' = \tilde{D}.$$

$$(PF3') \quad \forall (\tilde{D}, -\hat{Q}(T)) \in \hat{\mathbb{F}}_-, \exists! \langle \tilde{D}', T' \rangle \in \mathcal{B}_{\text{NII}}(\tilde{\mathcal{K}}) \text{ tel que } \tilde{D}' = \tilde{D}.$$

$$(PF4') \quad \forall (\tilde{D}, -\check{Q}(T)) \in \check{\mathbb{F}}_-, \exists! \langle \tilde{D}', T' \rangle \in \mathcal{B}(\tilde{\mathcal{K}}) \text{ tel que } \tilde{D}' = \tilde{D}.$$

Preuve.

$$\begin{aligned} (PF1') : (\tilde{D}, \hat{Q}(T)) \in \hat{\mathbb{F}} &\implies \tilde{D} = \bigcap_{t \in T} \tilde{f}(t) && \text{(à partir de (IF1))} \\ &\implies \tilde{D} = \{d^\alpha, d \in \mathcal{D} \mid \alpha = \bigwedge_{t \in T} \mathcal{I}(d, t)\} \\ &&& \text{(de la définition de } \tilde{f} \text{ et de } \cap \text{)} \\ &\implies \tilde{D} = T^\Delta && \text{(la définition de } \Delta \text{)} \end{aligned}$$

En prenant $T' = (T^\Delta)^{\Delta^c}$ nous obtenons l'unique concept formel flou $\langle T^\Delta, (T^\Delta)^{\Delta^c} \rangle \in \mathcal{B}(\tilde{\mathcal{K}})$

$$\begin{aligned} (PF2') : (\tilde{D}, \check{Q}(T)) \in \check{\mathbb{F}} &\implies \tilde{D} = \bigcup_{t \in T} \tilde{f}(t) && \text{(à partir de (IF2))} \\ &\implies \tilde{D} = \{d^\alpha, d \in \mathcal{D} \mid \alpha = \bigvee_{t \in T} \mathcal{I}(d, t)\} \\ &&& \text{(De la définition de } \tilde{f} \text{ et de } \cup \text{)} \\ &\implies \tilde{D} = T^\Pi && \text{(La définition de } \Pi \text{)} \end{aligned}$$

En supposant $T' = (T^\Pi)^{\text{Nc}}$ nous obtenons l'unique NII-paire floue $\langle T^\Pi, (T^\Pi)^{\text{Nc}} \rangle \in \mathcal{B}_{\text{NII}}(\tilde{\mathcal{K}})$

$$\begin{aligned} (PF3') : (\tilde{D}, -\hat{Q}(T)) \in \hat{\mathbb{F}}_- &\implies \tilde{D} = \mathcal{D} \setminus \bigcap_{t \in T} \tilde{f}(t) && \text{(à partir de (IF3))} \\ &\implies \tilde{D} = \bigcup_{t \in T} \mathcal{D} \setminus \tilde{f}(t) && \text{(à partir de (IF3))} \\ &\implies \tilde{D} = \{d^\alpha, d \in \mathcal{D} \mid \alpha = \bigvee_{t \in T} (1 - \mathcal{I}(d, t))\} \\ &&& \text{(De la définition de } \tilde{f} \text{ et de } \cup \text{)} \\ &\implies \tilde{D} = T^{\frac{\Pi}{\kappa}} \end{aligned}$$

En considérant $T' = (T^{\frac{\Pi}{\kappa}})^{\frac{N_c}{\kappa}}$ nous obtenons l'unique NII-paire floue $\langle T^{\frac{\Pi}{\kappa}}, (T^{\frac{\Pi}{\kappa}})^{\frac{N_c}{\kappa}} \rangle \in \mathcal{B}_{\text{NII}}(\widetilde{\mathcal{K}})$

$$\begin{aligned}
(\text{PF4}') : (\tilde{D}, \neg\tilde{Q}(T)) \in \check{\mathbb{F}}_{\neg} &\implies \tilde{D} = \mathcal{D} \setminus \bigcup_{t \in T} \tilde{f}(t) && \text{(à partir de (IF4))} \\
&\implies \tilde{D} = \bigcap_{t \in T} \mathcal{D} \setminus \tilde{f}(t) && \text{(lois de De Morgan)} \\
&\implies \tilde{D} = \{d^\alpha, d \in \mathcal{D} \mid \alpha = \bigwedge t \in T(1 - \mathcal{I}(d, t))\} \\
&&& \text{(de la définition de } \tilde{f} \text{ et de } \cap \text{)} \\
&\implies \tilde{D} = T^{\frac{\Delta}{\kappa}} && \text{(la définition de } \Delta \text{)}
\end{aligned}$$

En prenant $T' = (T^{\frac{\Delta}{\kappa}})^{\frac{\Delta_c}{\kappa}}$ nous obtenons l'unique concept formel flou $\langle T^{\frac{\Delta}{\kappa}}, (T^{\frac{\Delta}{\kappa}})^{\frac{\Delta_c}{\kappa}} \rangle \in \mathcal{B}(\widetilde{\mathcal{K}})$ □

En conséquence aux deux Propositions. 5.2.3 et 5.2.4, le théorème suivant donne des propriétés algébriques significatives en caractérisant les ensembles de requêtes élémentaires.

Théorème 5.2.2. *Soit le L -contexte $\tilde{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \mathcal{I})$ (et $\bar{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \bar{\mathcal{I}})$) alors,*

$$(\text{TF1}) \hat{\mathbb{F}} \text{ est isomorphe à } \mathcal{B}(\tilde{\mathcal{K}}) \setminus \{\langle 1_{\mathcal{D}}, \emptyset \rangle\}$$

$$(\text{TF2}) \check{\mathbb{F}} \text{ est isomorphe à } \mathcal{B}_{\text{NII}}(\tilde{\mathcal{K}}) \setminus \{\langle 0_{\mathcal{D}}, \emptyset \rangle\}$$

$$(\text{TF3}) \hat{\mathbb{F}}_{\neg} \text{ est isomorphe à } \mathcal{B}_{\text{NII}}(\bar{\mathcal{K}}) \setminus \{\langle 0_{\mathcal{D}}, \emptyset \rangle\}$$

$$(\text{TF4}) \check{\mathbb{F}}_{\neg} \text{ est isomorphe à } \mathcal{B}(\bar{\mathcal{K}}) \setminus \{\langle 1_{\mathcal{D}}, \emptyset \rangle\}$$

Où, $0_{\mathcal{D}} = \{d_1^0, d_2^0, \dots, d_m^0\}$ et $1_{\mathcal{D}} = \{d_1^1, d_2^1, \dots, d_m^1\}$

Preuve.

(TF1) Considérons les applications g et h données par :

$$\begin{aligned}
g : \hat{\mathbb{F}} &\longrightarrow \mathcal{B}(\tilde{\mathcal{K}}) \setminus \{\langle 1_{\mathcal{D}}, \emptyset \rangle\} \\
(\tilde{D}, \hat{Q}(T)) &\longmapsto \langle \tilde{D}, (T^{\Delta})^{\Delta_c} \rangle
\end{aligned}$$

$$\begin{aligned}
h : \mathcal{B}(\tilde{\mathcal{K}}) \setminus \{\langle 1_{\mathcal{D}}, \emptyset \rangle\} &\longrightarrow \hat{\mathbb{F}} \\
\langle \tilde{D}, (T^{\Delta})^{\Delta_c} \rangle &\longmapsto (\tilde{D}, \hat{Q}(T))
\end{aligned}$$

i) À partir de (PF1'), chaque élément de $\hat{\mathbb{F}}$ est associé (en utilisant g) exactement à un et un seul concept formel (une seule image) de $\mathcal{B}(\tilde{\mathcal{K}}) \setminus \{\langle 1_{\mathcal{D}}, \emptyset \rangle\}$,
ii) Inversement, à partir de la propriété (PF1), il s'ensuit que h associe à chaque élément de $\mathcal{B}(\tilde{\mathcal{K}}) \setminus \{\langle 1_{\mathcal{D}}, \emptyset \rangle\}$ un élément de $\hat{\mathbb{F}}$,
de *i)* et *ii)* nous déduisons que $\hat{\mathbb{F}}$ est isomorphe à $\mathcal{B}(\tilde{\mathcal{K}}) \setminus \{\langle 1_{\mathcal{D}}, \emptyset \rangle\}$
(TF2), (TF3) et (TF4) sont prouvées de la même manière. \square

Tableau 5.1 – Exemple de Contexte Formel Flou

$\tilde{\mathcal{R}}_e$	t_1	t_2	t_3	t_4
d_1	0.4	0.1	0.4	0.4
d_2	0.6	0.1	0.5	0.1
d_3	0.3	0.2	0.2	0.2
d_4	0.4	0.1	0.5	0.1
d_5	0.6	0.2	0.4	0.4

Afin d'illustrer les isomorphismes du Théorème. 5.2.2, on considère le contexte formel flou $\tilde{\mathcal{K}}_e$ illustré dans le Tableau. 5.1. La Figure. 5.1 illustre l'isomorphisme entre l'ensemble des concepts formels flous de $\mathcal{B}(\tilde{\mathcal{K}}_e) \setminus \{\langle \{d_1^1, d_2^1, \dots, d_5^1\}, \emptyset \rangle\}$ (resp. l'ensemble de NII-paires $\mathcal{B}_{\text{NII}}(\tilde{\mathcal{K}}_e) \setminus \{\langle \{d_1^0, d_2^0, \dots, d_6^0\}, \emptyset \rangle\}$) et l'ensemble $\hat{\mathbb{F}}$ (resp. $\check{\mathbb{F}}$).

5.3 Pondération de Requêtes

La contribution, présentée dans la section précédente, a mis en évidence l'intérêt de considérer les treillis de concepts formels semi-flous et les treillis de NII-paires semi-floues pour enrichir le modèle \mathcal{M}_f . Afin de prendre en charge la pondération de requêtes, nous proposons dans cette section d'étendre l'approche de RI basée sur l'ACF floue [Djouadi, 2011, 2012] à l'utilisation des treillis de NII-paires floues. Ces treillis nous permettent d'effectuer une recherche par navigation pour les requêtes pondérées disjonctives et les négations des requêtes conjonctives. Rappelons que la pondération des requêtes conjonctives a été traitée dans [Djouadi, 2011, 2012].

Considérons l'algèbre distributive $\mathbf{L} = (L, \wedge, \vee, 0, 1, \rightarrow, *, \sim)$, où L est un ensemble totalement ordonné, 0 (resp. 1) est la borne inférieure (resp. supérieure) de L , \wedge et \vee dénotent l'infimum et le supremum, \sim est un opérateur de négation floue vérifiant $\sim 0 = 1$ et $\sim 1 = 0$, \rightarrow est une implication floue vérifiant les conditions de limites ($0 \rightarrow 0 = 0 \rightarrow 1 = 1 \rightarrow 1 = 1$ et $1 \rightarrow 0 = 0$), $*$ est une conjonction floue qui n'est pas nécessairement une t-norme et la paire $(\rightarrow, *)$ n'est pas nécessairement une paire résiduelle.

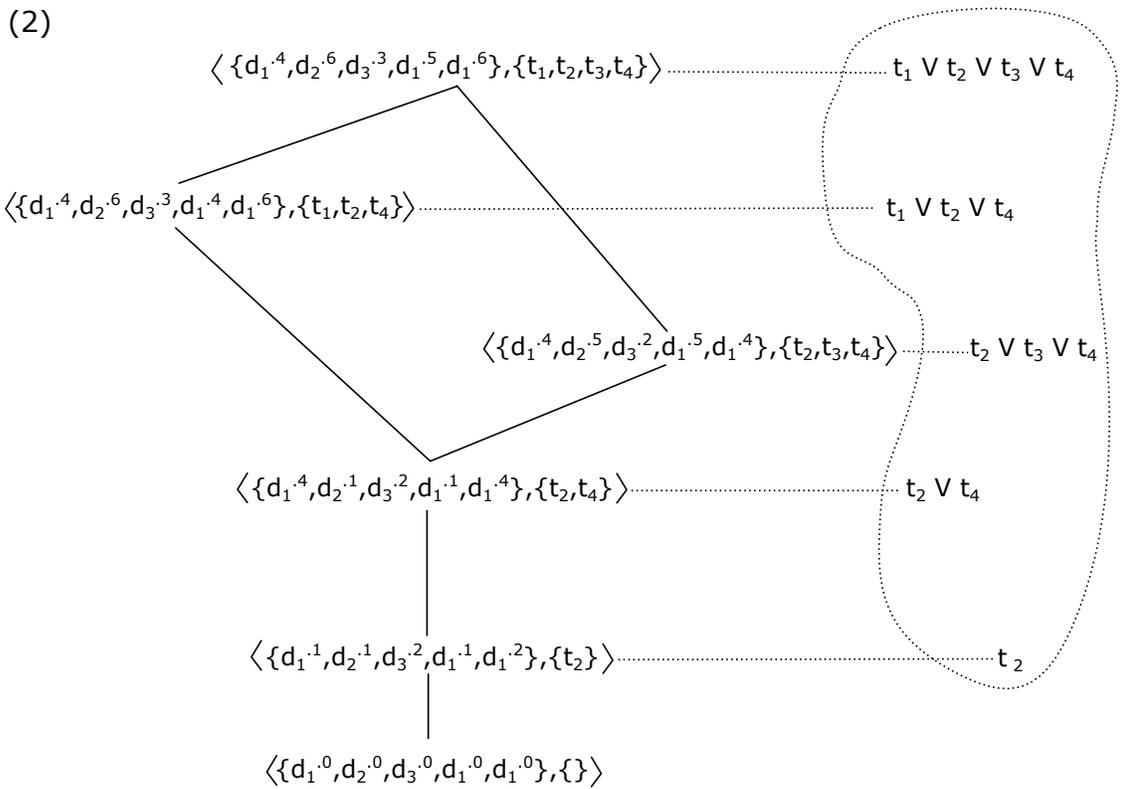
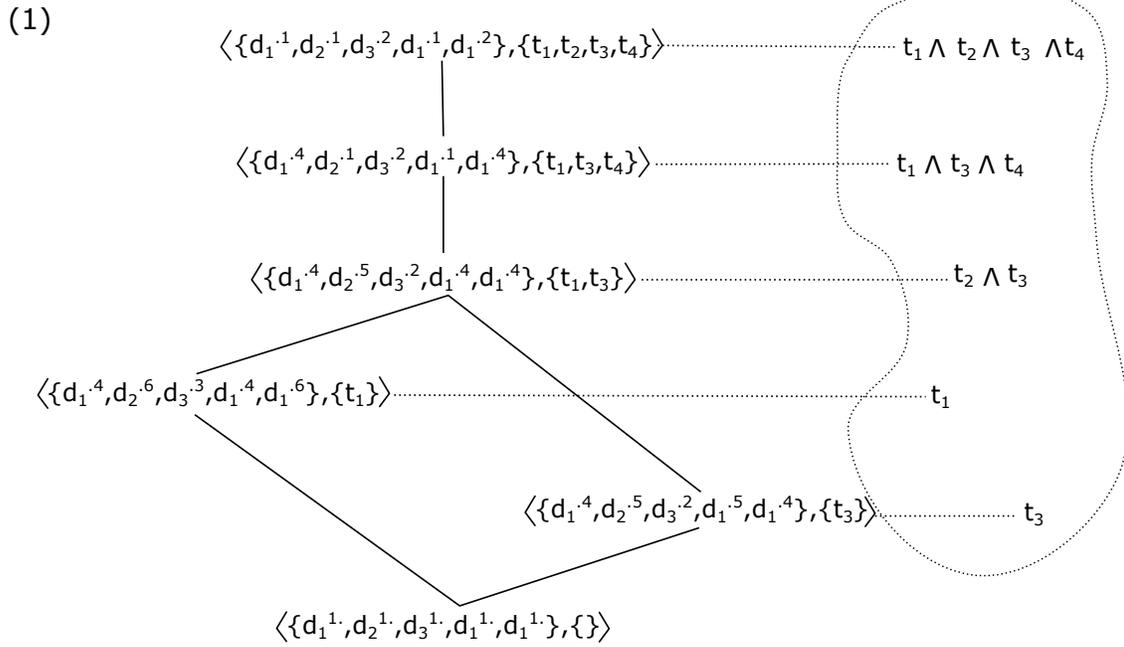


FIGURE 5.1 – (1) Isomorphisme entre $\hat{\mathbb{F}}$ et $\mathcal{B}(\tilde{\mathcal{K}}_e) \setminus \{\langle \{d_1^1, d_2^1, \dots, d_5^1\}, \emptyset \rangle\}$ (2) Isomorphisme entre $\tilde{\mathbb{F}}$ et $\mathcal{B}_{\text{NII}}(\tilde{\mathcal{K}}_e) \setminus \{\langle \{d_1^0, d_2^0, \dots, d_5^0\}, \emptyset \rangle\}$

5.3.1 Requêtes Disjonctives

Pour un ensemble flou de termes $\tilde{T} \in L^{\mathcal{T}}$ ($\tilde{T} = \{t_1^{\alpha_1}, t_2^{\alpha_2}, \dots, t_n^{\alpha_n}\}$), la requête disjonctive floue $\check{Q}(\tilde{T})$ construite sur cet ensemble et de la forme $t_1^{\alpha_1} \vee t_2^{\alpha_2} \vee \dots \vee t_n^{\alpha_n}$. En généralisant le cas de requêtes Booléennes, l'ensemble flou de documents \tilde{D} satisfaisant la requête $\check{Q}(\tilde{T})$ est exprimé par l'opérateur de possibilité $(.)^{\Pi}$ appliqué à l'ensemble \tilde{T} . Formellement, \tilde{D} est donné par :

$$\tilde{D}(d) = \tilde{T}^{\Pi}(d) = \bigvee_{t \in \mathcal{T}} (\tilde{T}(t) * \mathcal{I}(d, t)) \quad (5.8)$$

Où $*$ est une T-norme (un opérateur croissant) avec une sémantique de conjonction floue, vérifiant pour tout $a \in L$: *i*) la condition de l'identité $a * 1 = a$ et *ii*) les conditions de bornes $0 * 0 = 0 * 1 = 1 * 0 = 0$.

$\tilde{T}^{\Pi}(d)$ correspond alors au score maximal associé au document d pour la requête floue disjonctive $\check{Q}(\tilde{T})$.

Remarque. Pour $L = \{0, 1\}$ le score associé à un document d correspond exactement au RSV du modèle Booléen.

Différemment de l'approche proposée dans [Djouadi, 2011, 2012] où la réponse à une requête disjonctive floue est fournie par l'intension d'un concept cible (qui n'est pas forcément un concept formel), nous proposons, dans cette thèse, de calculer la NII-paire relative à cette forme de requêtes. Ainsi, pour une requête $\check{Q}(\tilde{T})$, nous associons la paire $\langle \tilde{T}^{\mathcal{N}}, (\tilde{T}^{\mathcal{N}})^{\Pi} \rangle \in L^{\mathcal{D}} \times L^{\mathcal{T}}$, où \mathcal{N} est donnée par :

$$\tilde{D}^{\mathcal{N}}(t) = \bigwedge_{d \in \mathcal{D}} (\mathcal{I}(d, t) \rightarrow \tilde{D}(d)) \quad (5.9)$$

Proposition 5.3.1. Soit L -contexte formel $\tilde{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \mathcal{I})$ alors, $\langle \tilde{T}^{\mathcal{N}}, (\tilde{T}^{\mathcal{N}})^{\Pi} \rangle$ est une NII-paire si et seulement si le couple $(\rightarrow, *)$ vérifie la propriété suivante :

$$\forall a, b \in L, \quad a \rightarrow (a * b) \geq b$$

Preuve. Pour prouver cette proposition, il suffit de prouver la fermeture de la composition $\mathcal{N} \circ \Pi$ [Djouadi et Prade, 2011] \square

Dans [Djouadi et Prade, 2011], les auteurs ont montré que l'ensemble de toutes les NII-paires induites à partir du L -contexte $\tilde{\mathcal{K}}$ (noté $\mathcal{P}(\tilde{\mathcal{K}})$) est un treillis complet isomorphe à $\tilde{\mathcal{K}}$. Le treillis de concepts formels flous $\mathcal{L}(\tilde{\mathcal{K}})$ (resp. des NII-paires floues $\mathcal{P}(\tilde{\mathcal{K}})$) fournit un moyen de recherche par navigation pour les requêtes pondérées conjonctives (resp. disjonctives). Dans la pratique les treillis $\mathcal{L}(\tilde{\mathcal{K}})$ et $\mathcal{P}(\tilde{\mathcal{K}})$ peuvent contenir un nombre important (voir infini) d'éléments pour certaines paires $(\rightarrow, *)$. Cependant, pour d'autres implications et t-normes ces treillis sont finis. À titre d'exemple, si on considère l'implication de Gödel ($a \rightarrow b = 1$ si $a \leq b$; b sinon) et la t-norme de Zadeh (*min*) nous obtenons un treillis fini.

5.3.2 Formes Négatives de Requêtes Pondérées

Le traitement de la négation (involutive $\forall a \in L, \neg\neg a = a$) des formes conjonctives ($\neg\hat{Q}(\tilde{T})$) et disjonctives ($\neg\check{Q}(\tilde{T})$) de requêtes pondérées revient à considérer les compléments des ensembles de documents flous satisfaisant les requêtes en forme positive. La proposition suivante donne des relations entre les opérateurs de dérivation de nécessité et de possibilité.

Proposition 5.3.2. [Djouadi, 2011] *Étant donné le L -contexte $\tilde{\mathcal{K}} = (\mathcal{O}, \mathcal{A}, \mathcal{I})$ les égalités suivantes sont vérifiées si la condition $\forall a, b \in L : a \rightarrow b = \neg(a * \neg b)$ est satisfaite.*

$$\overline{\tilde{T}^\Pi} = \tilde{T}^{\Delta_{\tilde{\mathcal{K}}}} \quad (5.10)$$

$$\overline{\tilde{T}^\Delta} = \tilde{T}^{\Pi_{\tilde{\mathcal{K}}}} \quad (5.11)$$

La Proposition. 5.3.2 nous permet de traiter la négation des requêtes conjonctives et disjonctives pondérées. En effet, en utilisant l'Expression. 5.10, pour retrouver l'ensemble flou de documents satisfaisant la requête $\neg\check{Q}(\tilde{T})$ revient à retrouver l'ensemble flou de documents satisfaisant la requête $\hat{Q}(\tilde{T})$ dans le L -contexte $\tilde{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \bar{\mathcal{I}})$. De manière similaire, en se basant sur l'Expression 5.11, le traitement de la requête $\neg\hat{Q}(\tilde{T})$ revient à retrouver l'ensemble flou de documents satisfaisant la requête $\check{Q}(\tilde{T})$ dans le L -contexte $\tilde{\mathcal{K}} = (\mathcal{D}, \mathcal{T}, \bar{\mathcal{I}}(d, t))$. En conséquence, le treillis $\mathcal{L}(\tilde{\mathcal{K}})$ (resp. $\mathcal{P}(\tilde{\mathcal{K}})$) caractérise toutes les requêtes négatives disjonctive (resp. conjonctives) construites sur l'ensemble L^T .

Exemple 5.3.1. *En considérant le L -contexte $\tilde{\mathcal{K}}_e$ du Tableau. 5.1, nous illustrons dans la Figure.5.2 (resp. Figure.5.3) le treillis flou $\mathcal{L}(\tilde{\mathcal{K}}_e)$ (resp. $\mathcal{P}(\tilde{\mathcal{K}}_e)$) caractérisant les requêtes pondérées disjonctives négatives (resp. disjonctives).*

5.4 Ignorance Partielle/Totale dans la Matrice d'Incidence

Habituellement, il est plus commode d'attribuer un intervalle pour évaluer la pertinence² d'un terme donné dans un document. Cela permet d'exprimer l'incertitude, l'imprécision et même la gradualité dans la matrice d'incidence *Documents* \times *Termes*. Pour exprimer l'ignorance totale sur certaines relations *document* – *terme*, nous proposons d'assigner l'intervalle $[0, 1]$ aux entrées manquantes (les informations absentes). Notons que cette représentation est également proposée dans [Alcalde

2. Il est plus facile à un expert d'attribuer un intervalle qu'une valeur précise.

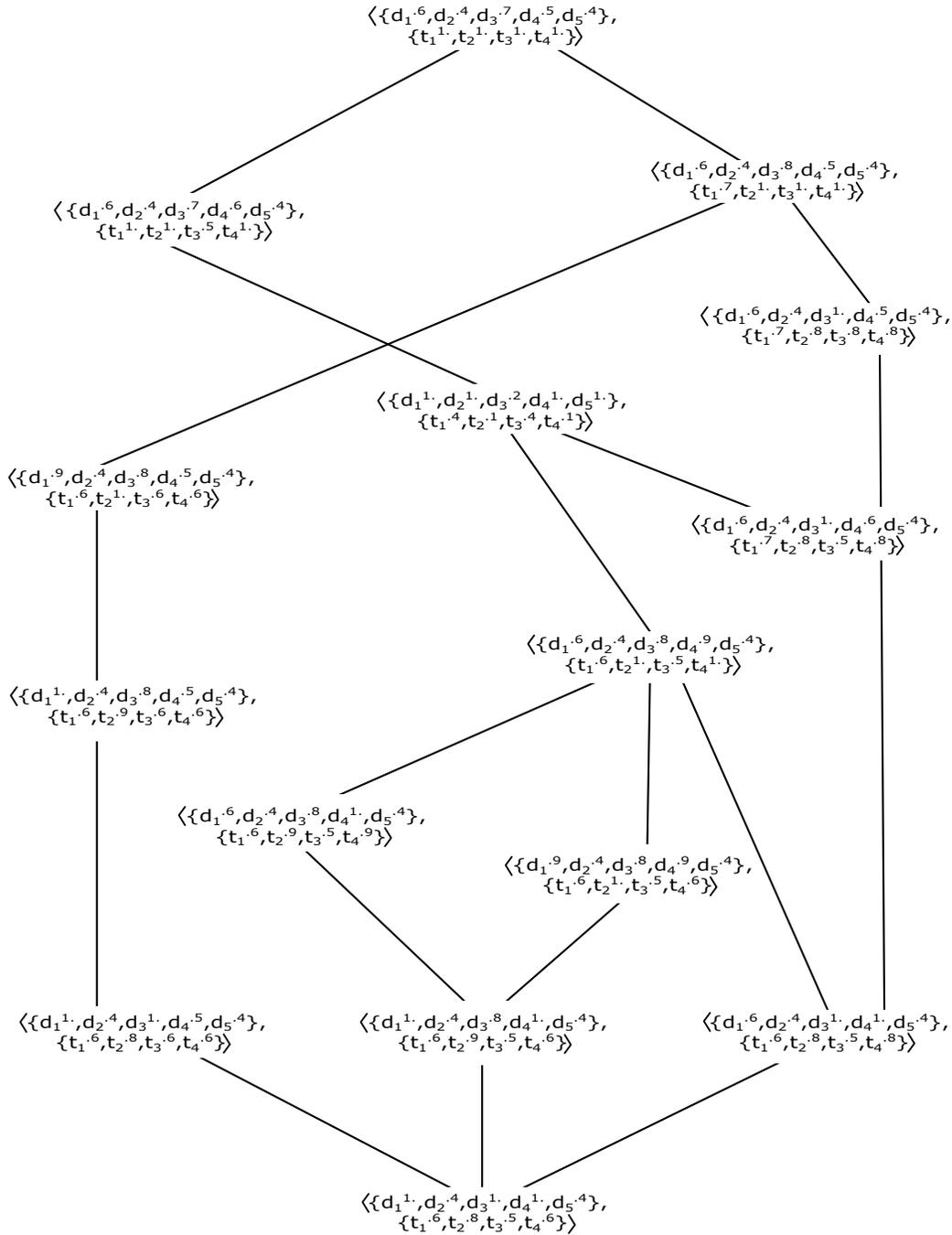


FIGURE 5.2 – Le Treillis de Concepts Formels Flous $\mathcal{L}(\widetilde{K}_e)$

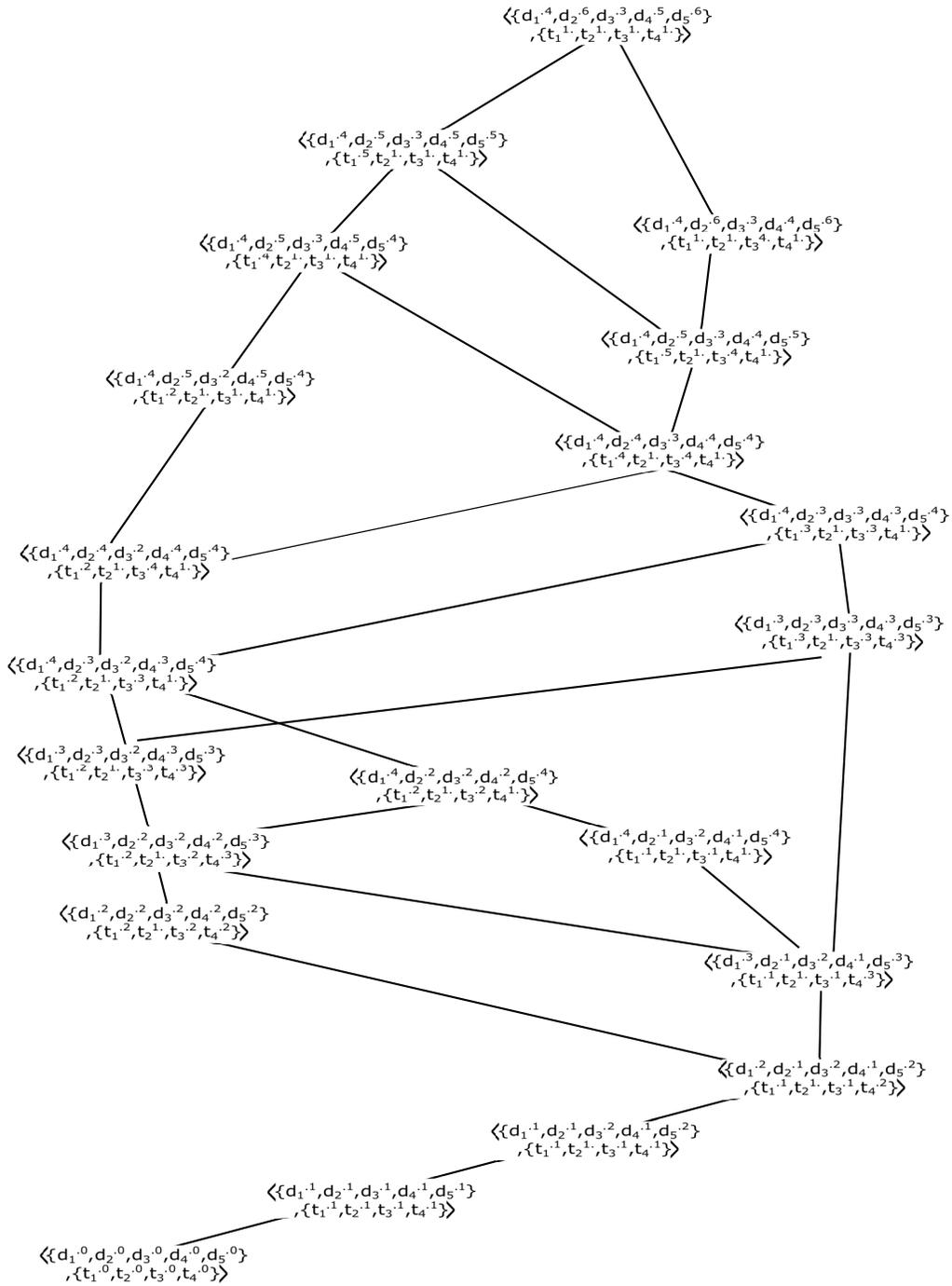


FIGURE 5.3 – Le Treillis de NII-paires $\mathcal{P}(\tilde{\mathcal{K}}_e)$

et al., 2009] où les auteurs ont souligné l'intérêt de substituer les entrées manquantes de la relation flou du $\mathcal{L}[L]$ -contexte par l'intervalle $[0, 1]$.

Soit $L = [0, 1]$ et $\mathcal{L}[L]$ l'ensemble de tous les intervalles fermés donnés par $\mathcal{L}[L] = \{x = [\underline{x}, \bar{x}] \mid \underline{x}, \bar{x} \in L \text{ et } \underline{x} \leq \bar{x}\}$. Considérons l'algèbre $(\mathcal{L}[L], \wedge, \vee, 0_{\mathcal{I}}, 1_{\mathcal{I}}, \preceq, \hookrightarrow, \otimes, \sim)$ où :

- $(\mathcal{L}[L], \preceq)$ est un treillis, où $0_{\mathcal{I}} = [0, 0]$ est le plus petit élément, $1_{\mathcal{I}}$ est le plus grand élément et \preceq est définie par : $x \preceq y \Leftrightarrow \underline{x} \leq \underline{y} \text{ et } \bar{x} \leq \bar{y}$. \wedge et \vee correspond respectivement à l'infimum et le supremum,
- \hookrightarrow est une implication floue d'intervalle de vérité vérifiant les propriétés données dans la Définition.1.8.1,
- \otimes est une t-norme floue d'intervalle de vérité vérifiant les propriétés données dans la Définition.1.7.2, avec $1_{\mathcal{I}}$ est l'élément neutre,
- \sim est une négation involutive $\sim \sim x = x$

En considérant le $\mathcal{L}[L]$ -contexte $\tilde{\mathcal{K}}_{\mathcal{I}} = (\mathcal{D}, \mathcal{T}, \mathcal{I}_{\mathcal{I}})$ où $\mathcal{I}_{\mathcal{I}}$ est une relation binaire qui associe à chaque couple $(d, t) \in \mathcal{D} \times \mathcal{T}$ un intervalle de vérité correspondant au poids du terme t dans le document d (i.e. $\mathcal{I}_{\mathcal{I}}(d, t) \in \mathcal{L}[L]$). Dans la suite de cette section, nous nous intéressons aux requêtes pondérées par des valeurs d'intervalle de vérité dans $\mathcal{L}[L]$.

Étant donné un ensemble flou d'intervalle de vérité de termes $\tilde{T}_{\mathcal{I}}$, où à chaque terme t_i est associé l'intervalle $w_i = [\underline{w}_i, \bar{w}_i]$ d'appartenance à $\tilde{T}_{\mathcal{I}}$. En généralisant les solutions proposées, dans la Section. 5.3, pour les requêtes floues au cas du $\mathcal{L}[L]$ -contexte $\tilde{\mathcal{K}}_{\mathcal{I}}$, nous obtenons :

- 1) Pour la forme conjonctive $\hat{Q}(\tilde{T}_{\mathcal{I}})$, l'ensemble de documents satisfaisant cette requête est donné par $\tilde{D}_{\mathcal{I}} = \tilde{T}_{\mathcal{I}}^{\Delta}$. Et le concept formel relatif à cette requête induit à partir de $\tilde{\mathcal{K}}_{\mathcal{I}}$ est $\langle \tilde{T}_{\mathcal{I}}^{\Delta}, (\tilde{T}_{\mathcal{I}}^{\Delta})^{\Delta} \rangle$. L'opérateur Δ est obtenu en considérant l'implication \hookrightarrow dans l'Expression. 1.6 (Page. 26).
- 2) Pour la forme disjonctive $\check{Q}(\tilde{T}_{\mathcal{I}})$, l'ensemble de documents satisfaisant cette requête est donné par $\tilde{D}_{\mathcal{I}} = \tilde{T}_{\mathcal{I}}^{\Pi}$. Le NII-paire correspondant est donné par $\langle \tilde{T}_{\mathcal{I}}^{\Pi}, (\tilde{T}_{\mathcal{I}}^{\Pi})^N \rangle$. Les opérateurs Π et N sont obtenus en considérant \otimes et \hookrightarrow dans les Expressions. 1.7 et 1.8.
- 3) Pour les formes négatives des requêtes, nous considérons le $\mathcal{L}[L]$ -contexte $\bar{\mathcal{K}}_{\mathcal{I}} = (\mathcal{D}, \mathcal{T}, \bar{\mathcal{I}}_{\mathcal{I}})$ complémentaire de $\tilde{\mathcal{K}}_{\mathcal{I}}$, où $\bar{\mathcal{I}}_{\mathcal{I}}(d, t) = \sim \mathcal{I}_{\mathcal{I}}(d, t)$.

Pour le mode de recherche par navigation nous considérons le treillis de concepts formels flous à valeur d'intervalle $\mathcal{L}(\tilde{\mathcal{K}}_{\mathcal{I}})$ pour les requêtes conjonctives et le treillis de NII-paires $\mathcal{P}(\tilde{\mathcal{K}}_{\mathcal{I}})$ pour les requêtes disjonctives. De même pour les formes négatives, on utilise $\mathcal{L}(\bar{\mathcal{K}}_{\mathcal{I}})$ pour les requête disjonctives et $\mathcal{P}(\bar{\mathcal{K}}_{\mathcal{I}})$ pour les requête conjonctives.

Exemple 5.4.1. Nous illustrons dans le Tableau. 5.2 un exemple de $\mathcal{L}[L]$ -contexte, avec des entrées imprécises ($\mathcal{I}_J(d_2, t_1) = [0.7, 0.7]$, $\mathcal{I}_J(d_2, t_2) = [0.8, 0.8]$), graduel ($\mathcal{I}_J(d_1, t_3) = [0, 0.5]$). Les valeurs associée aux entrées $\mathcal{I}_J(d_2, t_1)$ et $\mathcal{I}_J(d_2, t_1)$ permettent de représenté l'ignorance totale. Le Tableau. 5.3 dénombre les concepts formels flous à valeur d'intervalle induits à partir de ce $\mathcal{L}[L]$ -contexte.

Tableau 5.2 – Exemple de Contexte Formel Flou à Intervalle de Vérité

\mathcal{I}_J	t_1	t_2	t_3	t_4
d_1	[0.0, 1.0]	[1.0, 1.0]	[0.0, 0.5]	[0.0, 1.0]
d_2	[0.7, 0.7]	[0.8, 0.8]	[0.6, 0.6]	[0.7, 0.8]
d_3	[0.8, 0.8]	[0.8, 0.8]	[0.8, 0.8]	[0.8, 0.8]
d_4	[0.8, 0.8]	[1.0, 1.0]	[0.6, 0.6]	[1.0, 1.0]

5.5 Réduction Linéaire des Treillis

Une des difficultés dans les approches de RI basées sur l'AFC floue est que les treillis induits peuvent comporter un nombre très important de paires (concepts formels et NII-paires). Ils résultent des diagrammes de Hasse relativement denses. Par conséquent, parcourir le voisinage d'un concept formel ou d'une NII-paire dans le diagramme est assez difficile. D'où l'intérêt d'ordonner ces paires de façon à obtenir des représentations linéaires de ces diagrammes. L'idée est de munir ces treillis par des distances permettant d'ordonner les paires floues $\langle Documents, Termes \rangle$ par rapport à une paire de référence.

Soit $\mathbb{C} = \{c_0, c_1, \dots, c_n\}$ l'ensemble de paires quelconques³, où c_0 correspond à l'infimum de \mathbb{C} . Nous considérons la fonction $Ext(c_i)$ qui désigne l'extension de c_i pour toute paire $c_i \in \mathbb{C}$ (i.e. $Ext(\langle \tilde{D}, \tilde{T} \rangle) = \tilde{D}$). Nous définissons la distance δ_P entre les deux paires c_1 et c_2 par la distance d_P entre leurs extensions respectives, avec $p = 1 \dots + \infty$. Rappelons que les distances entre les ensembles floues permettent de définir des relations de type "à peu près égal à" ou "supérieur à". Dans notre cas d'utilisation, pour toutes paires $c_1 = \langle \tilde{D}_1, \tilde{T}_1 \rangle$ et $c_2 = \langle \tilde{D}_2, \tilde{T}_2 \rangle$, la distance $\delta_P(c_1, c_2)$ est définie par :

$$d_P(\tilde{D}_1, \tilde{D}_2) = \left(\sum_{d \in \mathcal{D}} | \tilde{D}_1(d) - \tilde{D}_2(d) |^P \right)^{\frac{1}{P}} \quad (5.12)$$

Rappelons également que cette distance vérifie les axiomes suivant :

3. \mathbb{C} peut être l'ensemble de concepts formels, ou de NII-paires, induit à partir de $\tilde{\mathcal{K}}$ ou $\tilde{\mathcal{K}}$.

1. Séparation : $\forall \tilde{D}_1, \tilde{D}_2 \in L^{\mathcal{P}}, d_P(\tilde{D}_1, \tilde{D}_2) = 0 \Leftrightarrow \tilde{D}_1 = \tilde{D}_2,$
2. Symétrie : $\forall \tilde{D}_1, \tilde{D}_2 \in L^{\mathcal{P}}, d_P(\tilde{D}_1, \tilde{D}_2) = d_P(\tilde{D}_2, \tilde{D}_1)$
3. Inégalité triangulaire : $\forall \tilde{D}_1, \tilde{D}_2, \tilde{D}_3 \in L^{\mathcal{P}}, d_P(\tilde{D}_1, \tilde{D}_2) \leq d_P(\tilde{D}_1, \tilde{D}_3) + d_P(\tilde{D}_3, \tilde{D}_1)$

Nous définissons sur \mathbb{C} la relation \ll par : $c_1 \ll c_2 \Leftrightarrow \delta_P(c_1, c_0) \leq \delta_P(c_2, c_0)$. La sémantique associée à l'inégalité $c_1 \ll c_2$ est que c_1 est plus proche que c_2 à l'infimum c_0 . Le treillis linéaire (\mathbb{C}, \ll) est obtenu comme suit :

- $\bigwedge_{c_i \in \mathbb{C}} (\mathbb{C}, \ll) = c_0$
- $\bigvee_{c_i \in \mathbb{C}} (\mathbb{C}, \ll) = c_t$, où $\delta_P(c_t, c_0) = \max_{c_i \in \mathbb{C}} \delta_P(c_i, c_0)$
- $c_1 \vee c_2 = c_2$ si $c_1 \ll c_2$
- $c_1 \wedge c_2 = c_1$ si $c_1 \ll c_2$

Proposition 5.5.1. \ll est une relation de préordre (i.e. réflexive et transitive).

Preuve.

Réflexivité :

$$\forall c \in \mathbb{C}, \delta_P(c_0, c) = \delta_P(c_0, c) \Rightarrow c \ll c$$

Transitivité : Admettant que $c_i \ll c_j$ et $c_j \ll c_k$,

$$\begin{aligned} & c_i \ll c_j \text{ et } c_j \ll c_k \\ \Rightarrow & \delta_P(c_0, c_i) \leq \delta_P(c_0, c_j) \text{ et } \delta_P(c_0, c_j) \leq \delta_P(c_0, c_k) \\ \Rightarrow & \delta_P(c_0, c_i) \leq \delta_P(c_0, c_k) && \text{(par la transitivité de } \leq \text{)} \\ \Rightarrow & c_i \ll c_k \end{aligned}$$

□

Pour illustrer l'intérêt de la réduction linéaire des treillis de concepts formels et des treillis de NII-paires, nous illustrons dans la Figure.5.4 les treillis linéaires correspondants aux treillis $\mathcal{L}(\tilde{\mathcal{K}}_e)$ et $\mathcal{P}(\tilde{\mathcal{K}}_e)$ des Figures. 5.2 et 5.3. Pour cette exemple nous avons utilisé la distance de Hamming donnée par l'Expression. 5.12 avec $P = 1$.

En ce qui concerne les treillis $\mathcal{L}(\tilde{\mathcal{K}}_{\mathcal{J}})$, $\mathcal{P}(\tilde{\mathcal{K}}_{\mathcal{J}})$, $\mathcal{L}(\tilde{\mathcal{K}}_{\mathcal{J}})$ et $\mathcal{P}(\tilde{\mathcal{K}}_{\mathcal{J}})$, la même démarche de linéarisation est réalisable en utilisant une distance entre les ensembles flous à valeur d'intervalle. À titre d'exemple, la distance de Hamming entre ces types d'ensembles est définie par :

$$d_h(\tilde{D}_{\mathcal{J}1}, \tilde{D}_{\mathcal{J}2}) = \sum_{d \in \mathcal{D}_{\mathcal{J}}} \max\{|\underline{\mu}_{\tilde{D}_{\mathcal{J}1}}(d) - \underline{\mu}_{\tilde{D}_{\mathcal{J}2}}(d)|, |\bar{\mu}_{\tilde{D}_{\mathcal{J}1}}(d) - \bar{\mu}_{\tilde{D}_{\mathcal{J}2}}(d)|\} \quad (5.13)$$

Nous illustrons dans le Tableau. 5.3 les distances associées aux concepts formels induits à partir du $\mathcal{L}[L]$ -contexte de l'Exemple. 5.4.1.

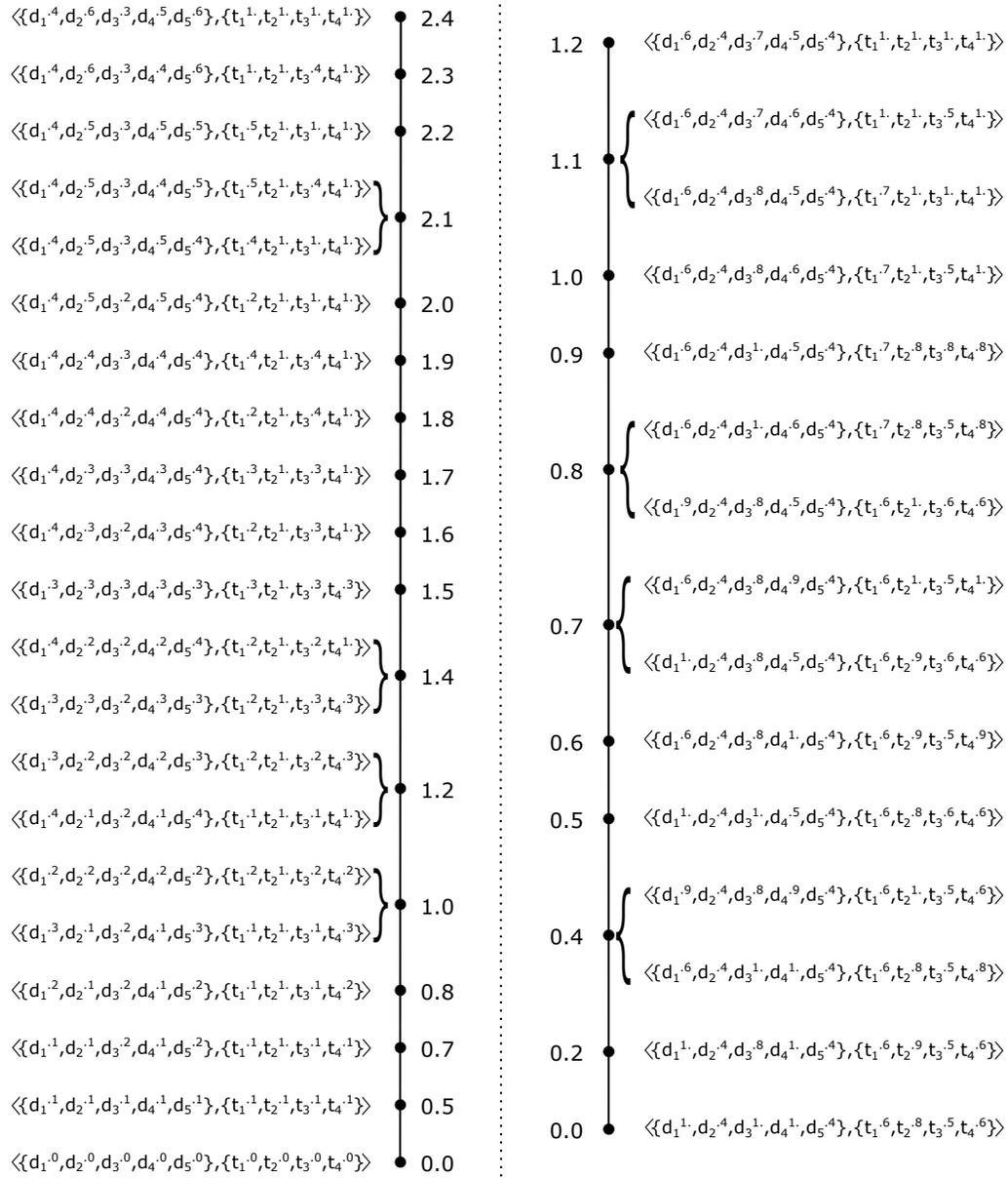


FIGURE 5.4 – Les Treillis Linéaires Correspondants aux Treillis $\mathcal{P}(\widetilde{\mathcal{K}}_e)$ (à gauche) et $\mathcal{L}(\widetilde{\mathcal{K}}_e)$ (à droite)

5.6 Conclusion

Les contributions présentées dans ce chapitre répondent principalement aux deux problématiques suivantes : 1) la pondération des termes dans les documents et/ou dans les requêtes, 2) le classement de résultats de recherche. D'abord, en utilisant la théorie axiomatique \mathfrak{T} et le modèle \mathcal{M}_f qui correspond au modèle flou standard de Bookstein [1980], nous avons proposé des caractérisations algébriques des formes de requêtes élémentaires de ce modèle par des treillis de concepts formels semi-flous et de NII-paires semi-floues. Pour cela, nous avons défini des extensions des opérateurs de dérivation de possibilité et de nécessité pour générer des treillis de NII-paires semi-floues. Puis, nous avons étendu l'approche de RI basée sur l'ACF floue (de Djouadi) à l'utilisation des treillis de NII-paires floues. Ces treillis nous permettent d'effectuer des recherches par navigation pour les requêtes pondérées disjonctives et les négations des requêtes conjonctives.

Nous avons ensuite proposé l'utilisation de l'extension floue à valeur d'intervalle de vérité de l'ACF dans la recherche d'information. Cela permet d'exprimer l'incertitude, l'imprécision et la gradualité des poids des termes dans les documents et dans les requêtes, et permet, également, l'expression de l'ignorance totale sur certains termes d'indexation en considérant l'intervalle $[0, 1]$.

Enfin, pour faciliter le mode recherche par navigation sur les treillis, de concepts formels flous et de NII-paires floues, nous avons proposé une méthode de linéarisation de ces treillis en se basant sur les distances entre les ensembles flous.

De manière générale, dans toutes ces contributions, pour une requête Q donnée (Booléenne ou pondérée), les méthodes utilisées retournent un ensemble flou \tilde{D} de documents satisfaisant cette requête. Le degré d'appartenance d'un document d à l'ensemble \tilde{D} (i.e. $\tilde{D}(d)$) représente le score de pertinence de ce document par rapport à la requête Q . Ces degrés offrent alors un moyen de classer les documents selon leur pertinence.

Tableau 5.3 – Les Distances Associée Aux Concepts Formels à Valeur d'Intervalle

#	Concepts Formels Flous	Distance
C_0	$\langle \{d_1^{[1.0,1.0]}, d_2^{[1.0,1.0]}, d_3^{[1.0,1.0]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.0,0.7]}, t_2^{[0.8,0.8]}, t_3^{[0.0,0.5]}, t_4^{[0.0,0.8]}\} \rangle$	0.0
C_1	$\langle \{d_1^{[1.0,1.0]}, d_2^{[0.7,0.7]}, d_3^{[1.0,1.0]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.0,0.8]}, t_2^{[0.8,0.8]}, t_3^{[0.0,0.5]}, t_4^{[0.0,0.8]}\} \rangle$	0.3
C_2	$\langle \{d_1^{[1.0,1.0]}, d_2^{[0.8,0.8]}, d_3^{[0.8,0.8]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.0,0.7]}, t_2^{[1.0,1.0]}, t_3^{[0.0,0.5]}, t_4^{[0.0,0.8]}\} \rangle$	0.4
C_3	$\langle \{d_1^{[1.0,1.0]}, d_2^{[0.7,0.8]}, d_3^{[0.8,0.8]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.0,0.7]}, t_2^{[1.0,1.0]}, t_3^{[0.0,0.5]}, t_4^{[0.0,1.0]}\} \rangle$	0.5
C_4	$\langle \{d_1^{[1.0,1.0]}, d_2^{[0.7,0.7]}, d_3^{[0.8,0.8]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.0,0.8]}, t_2^{[1.0,1.0]}, t_3^{[0.0,0.5]}, t_4^{[0.0,1.0]}\} \rangle$	0.5
C_5	$\langle \{d_1^{[1.0,1.0]}, d_2^{[0.7,0.7]}, d_3^{[0.8,0.8]}, d_4^{[0.8,0.8]}\}, \{t_1^{[0.0,1.0]}, t_2^{[1.0,1.0]}, t_3^{[0.0,0.5]}, t_4^{[0.0,1.0]}\} \rangle$	0.7
C_6	$\langle \{d_1^{[0.0,1.0]}, d_2^{[1.0,1.0]}, d_3^{[1.0,1.0]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.7,0.7]}, t_2^{[0.8,0.8]}, t_3^{[0.0,0.5]}, t_4^{[0.7,0.8]}\} \rangle$	1.0
C_7	$\langle \{d_1^{[0.0,0.5]}, d_2^{[1.0,1.0]}, d_3^{[1.0,1.0]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.7,0.7]}, t_2^{[0.8,0.8]}, t_3^{[0.6,0.6]}, t_4^{[0.7,0.8]}\} \rangle$	1.0
C_8	$\langle \{d_1^{[0.0,1.0]}, d_2^{[0.7,0.8]}, d_3^{[1.0,1.0]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.7,0.7]}, t_2^{[0.8,0.8]}, t_3^{[0.0,0.5]}, t_4^{[0.8,0.8]}\} \rangle$	1.3
C_9	$\langle \{d_1^{[0.0,0.5]}, d_2^{[0.7,0.8]}, d_3^{[1.0,1.0]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.7,0.7]}, t_2^{[0.8,0.8]}, t_3^{[0.6,0.6]}, t_4^{[0.8,0.8]}\} \rangle$	1.3
C_{10}	$\langle \{d_1^{[0.0,1.0]}, d_2^{[0.7,0.7]}, d_3^{[1.0,1.0]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.8,0.8]}, t_2^{[0.8,0.8]}, t_3^{[0.0,0.5]}, t_4^{[0.8,0.8]}\} \rangle$	1.3
C_{11}	$\langle \{d_1^{[0.0,0.5]}, d_2^{[0.7,0.7]}, d_3^{[1.0,1.0]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.8,0.8]}, t_2^{[0.8,0.8]}, t_3^{[0.6,0.6]}, t_4^{[0.8,0.8]}\} \rangle$	1.3
C_{12}	$\langle \{d_1^{[0.0,1.0]}, d_2^{[0.8,0.8]}, d_3^{[0.8,0.8]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.7,0.7]}, t_2^{[1.0,1.0]}, t_3^{[0.0,0.5]}, t_4^{[0.7,0.8]}\} \rangle$	1.4
C_{13}	$\langle \{d_1^{[0.0,0.5]}, d_2^{[0.8,0.8]}, d_3^{[0.8,0.8]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.7,0.7]}, t_2^{[1.0,1.0]}, t_3^{[0.6,0.6]}, t_4^{[0.7,0.8]}\} \rangle$	1.4
C_{14}	$\langle \{d_1^{[0.0,1.0]}, d_2^{[0.7,0.8]}, d_3^{[0.8,0.8]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.7,0.7]}, t_2^{[1.0,1.0]}, t_3^{[0.0,0.5]}, t_4^{[1.0,1.0]}\} \rangle$	1.5
C_{15}	$\langle \{d_1^{[0.0,0.5]}, d_2^{[0.7,0.8]}, d_3^{[0.8,0.8]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.7,0.7]}, t_2^{[1.0,1.0]}, t_3^{[0.6,0.6]}, t_4^{[1.0,1.0]}\} \rangle$	1.5
C_{16}	$\langle \{d_1^{[0.0,1.0]}, d_2^{[0.7,0.7]}, d_3^{[0.8,0.8]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.8,0.8]}, t_2^{[1.0,1.0]}, t_3^{[0.0,0.5]}, t_4^{[1.0,1.0]}\} \rangle$	1.5
C_{17}	$\langle \{d_1^{[0.0,0.5]}, d_2^{[0.7,0.7]}, d_3^{[0.8,0.8]}, d_4^{[1.0,1.0]}\}, \{t_1^{[0.8,0.8]}, t_2^{[1.0,1.0]}, t_3^{[0.6,0.6]}, t_4^{[1.0,1.0]}\} \rangle$	1.5
C_{18}	$\langle \{d_1^{[0.0,1.0]}, d_2^{[0.7,0.7]}, d_3^{[0.8,0.8]}, d_4^{[0.8,0.8]}\}, \{t_1^{[1.0,1.0]}, t_2^{[1.0,1.0]}, t_3^{[0.0,0.5]}, t_4^{[1.0,1.0]}\} \rangle$	1.7
C_{19}	$\langle \{d_1^{[0.0,0.5]}, d_2^{[0.7,0.7]}, d_3^{[0.8,0.8]}, d_4^{[0.8,0.8]}\}, \{t_1^{[1.0,1.0]}, t_2^{[1.0,1.0]}, t_3^{[0.6,0.6]}, t_4^{[1.0,1.0]}\} \rangle$	1.7
C_{20}	$\langle \{d_1^{[0.0,0.5]}, d_2^{[0.6,0.6]}, d_3^{[1.0,1.0]}, d_4^{[0.6,0.6]}\}, \{t_1^{[0.8,0.8]}, t_2^{[0.8,0.8]}, t_3^{[0.8,0.8]}, t_4^{[0.8,0.8]}\} \rangle$	1.8
C_{21}	$\langle \{d_1^{[0.0,0.5]}, d_2^{[0.6,0.6]}, d_3^{[0.8,0.8]}, d_4^{[0.6,0.6]}\}, \{t_1^{[1.0,1.0]}, t_2^{[1.0,1.0]}, t_3^{[1.0,1.0]}, t_4^{[1.0,1.0]}\} \rangle$	2.0

Conclusion Générale et Perspectives

Les propositions et contributions présentées dans cette thèse reposent sur l'idée d'utiliser le cadre théorique offert par l'ACF dans le domaine de la RI, domaine fort souvent abordé sous un volet expérimental. Dans ce contexte, nous avons élaboré une théorie axiomatique multi-sorte (\mathfrak{T}) qui formalise le processus de recherche et les objets liés à la RI. La théorie ainsi proposée présente une nouvelle vision par rapport aux approches logiques de RI existantes, et offre un cadre générique capable de reproduire et de représenter les modèles de RI existants. Cette généralité a été prouvée par le biais de trois interprétations différentes, satisfaisant chacune cette théorie ; à savoir le modèle Booléen (\mathcal{M}_B), le modèle flou standard (\mathcal{M}_f) et le modèle vectoriel (\mathcal{M}_v).

En se basant sur la théorie \mathfrak{T} et le modèle \mathcal{M}_B , nous avons démontré des propriétés algébriques caractérisant les différentes formes de requêtes de ce modèle par les treillis de concepts formels et les treillis de NII-paires relatifs à l'ACF classique et sa lecture possibiliste. Dans une démarche similaire, nous avons caractérisé les requêtes du modèle \mathcal{M}_f par les treillis de concepts formels semi-flous et les treillis de NII-paires semi-flous. Ces propriétés algébriques approuvent d'une part l'utilisation de la théorie de l'ACF (classique et floue) et sa lecture possibiliste dans le domaine de la RI, et définissent d'autre part des fondements à de nouvelles approches de RI, notamment la classification de documents. Toutefois, ces deux contributions ne tiennent pas compte de la pondération des termes dans les requêtes. Pour cela, nous avons étendu l'approche de RI basée sur l'ACF floue [Djouadi, 2011, 2012] à l'utilisation des treillis de NII-paires floues. Ces treillis nous permettent d'effectuer des recherches par navigation pour les requêtes pondérées disjonctives et les requêtes négatives.

Dans une autre contribution, nous avons proposé une approche de RI structurée

Booléenne basée sur l'extension triadique de l'ACF. L'approche permet de traiter les différentes formes de requêtes orientées contenu et de requêtes orientées contenu et structure.

Nous avons proposé également une approche de RI basée sur l'extension floue à valeur d'intervalle de l'ACF, qui permet d'exprimer l'incertitude, l'imprécision et la gradualité des poids des termes dans les documents et dans les requêtes, et permet, également, l'expression de l'ignorance totale sur certains poids d'indexation en considérant l'intervalle $[0, 1]$.

Dans une dernière contribution, nous avons développé une méthode de linéarisation des treillis de concepts formels et de treillis de NII-paires. Les treillis obtenus définissent une relation de préordre sur les ensembles de concepts formels et les ensembles de NII-paires ce qui facilite le mode de recherche par navigation.

Les travaux effectués dans le cadre de cette thèse ouvrent de nombreuses perspectives aussi bien sur le plan expérimental que théorique. Les perspectives expérimentales concernent, de manière générale, la validation des différentes approches développées dans cette thèse. De manière assez opportune, notre approche offre un cadre unifié de recherche d'information, à partir de la théorie \mathfrak{T} , permettant de comparer les modèles existants. Cette comparaison passe notamment par des tests et des évaluations sur des benchmarks de référence (INEX et TREC). Dans le cadre de RI basée sur l'ACF floue, il est également envisagé d'évaluer de manière expérimentale l'impact du choix des paires $(\rightarrow, *)$ sur la performance en terme de rappel et de précision.

En ce qui concerne les perspectives théoriques, un premier besoin est de démontrer que d'autres modèles de RI, notamment les modèles logiques, satisfont la théorie \mathfrak{T} . Il sera également intéressant de se focaliser sur les modèles qui prennent en considération les liens entre les termes d'indexation.

Nos différentes contributions se limitent au cas de requête élémentaires, or dans la RI nous traitons des requête complexes, d'où la demande de définition d'opérateurs de dérivation généralisés qui permettent de traiter des requête complexes, données par exemple sous forme normale conjonctive ou disjonctive.

Dans nos contributions pour la RI structurée, nous avons considéré des contextes formels dyadiques dérivés du contexte formel triadique. Toutefois nous estimons qu'il sera bénéfique de considérer les concepts formels triadiques et voir des triplets générés par les opérateurs Π et N . Pour cela, une lecture possibiliste de l'ACF triadique est bien envisageable.

Du côté de l'ACF floue, nous avons défini des algèbres minimales qui préservent la fermeture ou l'ouvertures des opérateurs de dérivation flous, toutefois un besoin particulier a surgi, il concerne la caractérisation des familles d'implications floues et de paires $(\rightarrow, *)$ pour générer des treillis flous finis de concepts formels et de NII-paires.

Bibliographie

- I. Ahamd et Taek-Sueng Jang. Old fashion text-based image retrieval using fca. In *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, volume 3, pages III–33–6 vol.2, Sept 2003.
- C. Alcalde, A. Burusco, et R. Fuentes-González. A constructive method for the definition of interval-valued fuzzy implication operators. *Fuzzy Sets and Systems*, 153(2) :211 – 227, 2005.
- Cristina Alcalde, Ana Burusco, et Ramón Fuentes-González. Interval-valued fuzzy propositions. an application to the l-fuzzy contexts with absent values. In João Paulo Carvalho, Didier Dubois, Uzay Kaymak, et João Miguel da Costa Sousa, editors, *Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference, Lisbon, Portugal, July 20-24, 2009*, pages 58–63, 2009. ISBN 978-989-95079-6-8.
- G. Amato et C. Meghini. Faceted content-based image retrieval. In *2008 19th International Workshop on Database and Expert Systems Applications*, pages 402–406, Sept 2008.
- Pauline V. Angione. On the equivalence of boolean and weighted searching based on the convertibility of query forms. *Journal of the American Society for Information Science*, 26(2) :112–124, 1975. ISSN 1097-4571.
- Cherukuri Aswani Kumar, Subramanian Chandra Mouliswaran, Pandey Amriteya, et S. R. Arun. *Fuzzy Formal Concept Analysis Approach for Information Retrieval*, pages 255–271. Springer International Publishing, Cham, 2015.
- Ricardo Baeza-Yates et Berthier Ribeiro-Neto. *Modern Information Retrieval*. ACM Press. Addison-Wesley, 1999.

- Jon Barwise et John Perry. *Situations and Attitudes*. Bradford Books. Massachusetts Institute of Technology, 1983.
- Radim Bělohlávek, Vladimír Sklenář, et Jiří Zacpal. *Crisply Generated Fuzzy Concepts*, pages 269–284. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- Dominik Benz, Andreas Hotho, Robert Jäschke, Beate Krause, Folke Mitzlaff, Christoph Schmitz, et Gerd Stumme. The social bookmark and publication management system bibsonomy. *The VLDB Journal*, 19(6) :849–875, Dec 2010.
- G. Birkhoff. *Lattice Theory*. American Mathematical Society, 3rd edition, 1967.
- R Bělohlávek. Lattices generated by binary fuzzy relations (extended abstract). In *Abstracts of the Fourth International Conference on Fuzzy Sets Theory and Its Applications*, page 11. Liptovsky Jan, Slovakia, 1998.
- R. Bělohlávek et V. Vychodil. What is a fuzzy concept lattice? In *CLA 2005, Proceedings of the 3rd International Workshop*, volume 162, pages 34–45, Olomouc, 2005. CEUR-WS.org.
- Radim Bělohlávek. Concept lattices and order in fuzzy logic. *Annals of Pure and Applied Logic*, 128(1) :277 – 298, 2004.
- Radim Bělohlávek. Lattices of fixed points of fuzzy galois connections. *Mathematical Logic Quarterly*, 47(1) :111–116, 2001.
- Abraham Bookstein. Fuzzy requests : An approach to weighted boolean searches. *Journal of the American Society for Information Science*, 31(4) :240–247, 1980.
- Abraham Bookstein. Outline of a general probabilistic retrieval model. *Journal of Documentation*, 39(2) :63–72, 1983.
- J. P. Bordat. Calcul pratique du treillis de galois d’une correspondance. *Mathématiques et Sciences Humaines*, 96 :31–47, 1986.
- Mohand Boughanem, Yannick Loiseau, et Henri Prade. *Rank-Ordering Documents According to Their Relevance in Information Retrieval Using Refinements of Ordered-Weighted Aggregations*, pages 44–54. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- M. Bruno, G. Canfora, M. Di Penta, et R. Scognamiglio. An approach to support web service classification and annotation. In *2005 IEEE International Conference on e-Technology, e-Commerce and e-Service*, pages 138–143, March 2005.

- T. W. C. Bruza, P. D. and Huibers. Investigating aboutness axioms using information fields. In Bruce W. Croft et C. J. van Rijsbergen, editors, *SIGIR '94 : Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, organised by Dublin City University*, pages 112–121. Springer London, London, 1994. ISBN 978-1-4471-2099-5. doi : 10.1007/978-1-4471-2099-5{_}12.
- Duncan A. Buell et Donald H. Kraft. A model for a weighted retrieval system. *Journal of the American Society for Information Science*, 32(3) :211–216, 1981.
- Duncan A. Buell et Donald H. Kraft. Liars : A software environment for testing query processing strategies. In *Proceedings of the 5th Annual ACM Conference on Research and Development in Information Retrieval, SIGIR '82*, pages 20–27, New York, NY, USA, 1982. Springer-Verlag New York, Inc.
- Ana Juandeaburre Burusco et R. Fuentes-González. The study of the L-fuzzy concept lattice. *Mathware & Soft Computings*, 3 :209 – 218, 1994.
- R. Bělohlávek et V. Vychodil. Reducing the size of fuzzy concept lattices by hedges. In *The 14th IEEE International Conference on Fuzzy Systems, 2005. FUZZ '05.*, pages 663–668, May 2005. doi : 10.1109/FUZZY.2005.1452473.
- C. Carpineto et G. Romano. GALOIS : An order-theoretic approach to conceptual clustering. In *Proceedings of the 10th International Conference on Machine Learning*, pages 33–40, July 1993.
- Claudio Carpineto et Giovanni Romano. A lattice conceptual clustering system and its application to browsing retrieval. *Machine Learning*, 24(2) :95–122, 1996.
- Claudio Carpineto et Giovanni Romano. Effective reformulation of boolean queries with concept lattices. In Troels Andreasen, Henning Christiansen, et Henrik Legind Larsen, editors, *Flexible Query Answering Systems, Third International Conference, FQAS'98, Roskilde, Denmark, May 13-15, 1998, Proceedings*, volume 1495 of *Lecture Notes in Computer Science*, pages 83–94. Springer, 1998.
- Claudio Carpineto et Giovanni Romano. *Using Concept Lattices for Text Retrieval and Mining*, pages 161–179. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- Claudio Carpineto, Giovanni Romano, et Fondazione Ugo Bordoni. Exploiting the potential of concept lattices for information retrieval with credo. *Journal of Universal Computer Science*, 10 :985–1013, 2004.
- P. Ceravolo, A. Gusmini, M. Leida, et Zhan Cui. An fca-based mapping generator. In *2007 IEEE Conference on Emerging Technologies and Factory Automation (EFTA 2007)*, pages 796–803, Sept 2007.

- Michel Chein. Algorithme de recherche des sous-matrices premières d'une matrice. *Bull. Math. Soc. Sc. Math. de Roumanie*, 1(13) :21–25, 1969.
- Jean-Pierre Chevallet et Yves Chiaramella. Extending a Logic-based Model with Algebraic Knowledge. In *MIRO Multimedia Information Retrieval, final workshop*, page 9, Glasgow, UK, 1995.
- Juan M. Cigarrán, Julio Gonzalo, Anselmo Peñas, et Felisa Verdejo. *Browsing Search Results via Formal Concept Analysis : Automatic Selection of Attributes*, pages 74–87. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
- Juan M. Cigarrán, Anselmo Peñas, Julio Gonzalo, et Felisa Verdejo. *Automatic Selection of Noun Phrases as Document Descriptors in an FCA-Based Information Retrieval System*, pages 49–63. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- Victor Codocedo et Amedeo Napoli. Formal concept analysis and information retrieval – a survey. In Jaume Baixeries, Christian Sacarea, et Manuel Ojeda-Aciego, editors, *Formal Concept Analysis*, pages 61–77, Cham, 2015. Springer International Publishing.
- Victor Codocedo, Ioanna Lykourantzou, et Amedeo Napoli. A semantic approach to concept lattice-based information retrieval. *Annals of Mathematics and Artificial Intelligence*, 72(1) :169–195, Oct 2014.
- Sara Cohen, Jonathan Mamou, Yaron Kanza, et Yehoshua Sagiv. Xsearch : A semantic search engine for xml. In *Proceedings of the 29th International Conference on Very Large Data Bases - Volume 29, VLDB '03*, pages 45–56. VLDB Endowment, 2003. ISBN 0-12-722442-4.
- Richard Cole, Peter Eklund, et Gerd Stumme. Document retrieval for e-mail search and discovery using formal concept analysis. *Applied Artificial Intelligence*, 17(3) : 257–280, 2003.
- F. Crestani et C.J. Van Rijsbergen. Information retrieval by logical imaging. *Journal of Documentation*, 51(1) :3–17, 1995.
- Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, et Richard Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6) :391–407, 1990. ISSN 1097-4571.
- Yassine Djouadi. Généralisation des opérateurs de dérivation de galois en recherche d'information basée sur l'analyse formelle de concepts. In Michel Beigbeder, Véronique Eglin, Nicolas Ragot, et Mathias Géry, editors, *CORIA*, pages 373–386, 2012.

- Yassine Djouadi. Extended galois derivation operators for information retrieval based on fuzzy formal concept lattice. In Salem Benferhat et John Grant, editors, *Scalable Uncertainty Management : 5th International Conference, SUM 2011, Dayton, OH, USA, October 10-13, 2011. Proceedings*, pages 346–358. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- Yassine Djouadi et Henri Prade. *Interval-Valued Fuzzy Formal Concept Analysis*, pages 592–601. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- Yassine Djouadi et Henri Prade. Possibility-theoretic extension of derivation operators in formal concept analysis over fuzzy lattices. *Fuzzy Optimization and Decision Making*, 10(4) :287–309, 2011.
- Didier Dubois, Florence Dupin de Saint Cyr Bannay, et Henri Prade. A possibility-theoretic view of formal concept analysis. *Fundamenta Informaticae*, 75(1-4) : 195–213, 2007.
- Jon Ducrou. *DVDSleuth : A Case Study in Applied Formal Concept Analysis for Navigating Web Catalogs*, pages 496–500. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.
- Jon Ducrou et Peter W. Eklund. Searchsleuth : The conceptual neighbourhood of an web query. In Peter W. Eklund, Jean Diatta, et Michel Liquiere, editors, *Proceedings of the Fifth International Conference on Concept Lattices and Their Applications, Montpellier, France, 2007*.
- Jon Ducrou, Bastian Wormuth, et Peter Eklund. *Dynamic Schema Navigation Using Formal Concept Analysis*, pages 398–407. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- Jon Ducrou, Björn Vormbrock, et Peter Eklund. *FCA-Based Browsing and Searching of a Collection of Images*, pages 203–214. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- Ivo Düntsch et Günther Gediga. Approximation operators in qualitative data analysis. In Harrie de Swart, Ewa Orłowska, Gunther Schmidt, et Marc Roubens, editors, *Theory and Applications of Relational Structures as Knowledge Instruments : COST Action 274, TARSKI. Revised Papers*, pages 214–230. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003. ISBN 978-3-540-24615-2. doi : 10.1007/978-3-540-24615-2{-}10.
- Peter Eklund et Jon Ducrou. *Navigation and Annotation with Formal Concept Analysis*, pages 118–121. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

- Peter Eklund et Bastian Wormuth. *Restructuring Help Systems Using Formal Concept Analysis*, pages 129–144. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- Peter W. Eklund, Jean Diatta, et Michel Liquiere, editors. *Proceedings of the Fifth International Conference on Concept Lattices and Their Applications, CLA 2007, Montpellier, France, October 24-26, 2007*, volume 331 of *CEUR Workshop Proceedings*, 2008. CEUR-WS.org.
- Sébastien Ferré. Camelis : Organizing and browsing a personal photo collection with a logical information system. In [Eklund et al. \[2008\]](#).
- Norbert Fuhr. Models for retrieval with probabilistic indexing. *Information Processing & Management*, 25(1) :55 – 72, 1989. Special Issue : Modeling Data, Information and Knowledge.
- Norbert Fuhr. Probabilistic datalog- a logic for powerful retrieval methods. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '95*, pages 282–290, New York, NY, USA, 1995. ACM. ISBN 0-89791-714-6.
- Norbert Fuhr. Probabilistic datalog : Implementing logical information retrieval for advanced applications. *Journal of the American Society for Information Science*, 51(2) :95–110, 2000. ISSN 1097-4571.
- Bernhard Ganter. Two basic algorithms in concept analysis. FB4–Preprint 831, TH Darmstadt, 1984.
- Bernhard Ganter et Klaus Reuter. Finding all closed sets : A general approach. *Order*, 8(3) :283–290, 1991.
- Bernhard Ganter et Rudolf Wille. *Formal concept analysis - mathematical foundations*. Springer, 1999. ISBN 978-3-540-62771-5.
- B. Van Gasse, C. Cornelis, G. Deschrijver, et E.E. Kerre. Triangle algebras : A formal logic approach to interval-valued residuated lattices. *Fuzzy Sets and Systems*, 159 (9) :1042 – 1060, 2008.
- George Georgescu et Andrei Popescu. Non-dual fuzzy connections. *Archive for Mathematical Logic*, 43(8) :1009–1039, 2004. doi : 10.1007/s00153-004-0240-4.
- R. Godin, C. Pichet, et J. Gecsei. Design of a browsing interface for information retrieval. In *Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '89*, pages 32–39, New York, NY, USA, 1989. ACM. ISBN 0-89791-321-3.

- Robert Godin, Eugene Saunders, et Jan Gecsei. Lattice model of browsable data spaces. *Information Sciences*, 40(2) :89 – 116, 1986.
- Robert Godin, Rokia Missaoui, et Alain April. Experimental comparison of navigation in a galois lattice with conventional information retrieval methods. *International Journal of Man-Machine Studies*, 38(5) :747 – 767, 1993.
- Robert Godin, Rokia Missaoui, et Hassan Alaoui. Incremental concept formation algorithms based on galois (concept) lattices. *Computational Intelligence*, 11(2) : 246–267, 1995. ISSN 1467-8640.
- F. A. Grootjen et T. P. van der Weide. Conceptual relevance feedback. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 2, pages 471–476, Oct 2002.
- F.A. Grootjen et Th.P. van der Weide. Conceptual query expansion. *Data & Knowledge Engineering*, 56(2) :174 – 193, 2006.
- Alain Guénoche et Iven Van Mechelen. Galois approach to the induction of concepts. In I. Van Mechelen, J. Hampton, R. Michalski, et P. Theuns, editors, *Categories and Concepts : Theoretical Views and Inductive Data Analysis*. Academic Press, 1993.
- Narjes Hachani, Mohamed Ali Ben Hassine, Hanène Chettaoui, et Habib Ounelli. Cooperative answering of fuzzy queries. *Journal of Computer Science and Technology*, 24(4) :675–686, Jul 2009.
- T. W. C. Huibers. *An Axiomatic Theory for Information Retrieval*. Phd thesis, Utrecht University, The Netherlands, 1996.
- Dmitry I. Ignatov et Sergei O. Kuznetsov. *Frequent Itemset Mining for Clustering Near Duplicate Web Documents*, pages 185–200. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- Peter Ingwersen. Cognitive perspectives of information retrieval interaction : Elements of a cognitive ir theory. *Journal of Documentation*, 52(1) :3–50, 1996.
- Mihye Kim et Paul Compton. Evolutionary document management and retrieval for specialized domains on the web. *International Journal of Human-Computer Studie*, 60(2) :201–241, February 2004. ISSN 1071-5819.
- Bjoern Koester. Conceptual knowledge retrieval with focca : Improving web search engine results with contexts and concept hierarchies. In Petra Perner, editor, *Advances in Data Mining. Applications in Medicine, Web Mining, Marketing, Image*

- and *Signal Mining*, volume 4065 of *Lecture Notes in Computer Science*, pages 176–190. Springer, 2006.
- Matthew B. Koll. Weird : An approach to concept-based information retrieval. In *Proceedings of the 1st Annual International ACM SIGIR Conference on Information Storage and Retrieval*, SIGIR '78, pages 180–, New York, NY, USA, 1978. ACM.
- Donald H. Kraft et Duncan A. Buell. Fuzzy sets and generalized boolean retrieval systems. *International Journal of Man-Machine Studies*, 19(1) :45 – 56, 1983.
- S Krajčí. Cluster based efficient generation of fuzzy concepts. *Neural Network World*, 13 :521–530, 2003.
- Stanislav Krajčí. A generalized concept lattice. *Logic Journal of the IGPL*, 13(5) : 543–550, 2005.
- Saul A. Kripke. Semantical analysis of modal logic i normal modal propositional calculi. *Mathematical Logic Quarterly*, 9(5-6) :67–96, 1963.
- Sergei O Kuznetsov. A fast algorithm for computing all intersections of objects in a finite semi-lattice. *Automatic Documentation and Mathematical Linguistics*, 27 (5) :11–21, 1993.
- Mounia Lalmas. *XML Retrieval*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2009.
- Mounia Lalmas et Peter D. Bruza. The use of logic in information retrieval modelling. *The Knowledge Engineering Review*, 13(3) :263–295, 1998.
- Mounia Lalmas et Keith van Rijsbergen. *A Logical Model of Information Retrieval Based on Situation Theory*, pages 1–13. Springer London, London, 1993.
- Joon Ho Lee. Properties of extended boolean models in information retrieval. In Bruce W. Croft et C. J. van Rijsbergen, editors, *SIGIR '94 : Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, organised by Dublin City University*, pages 182–190. Springer London, London, 1994.
- Fritz Lehmann et Rudolf Wille. A triadic approach to formal concept analysis. In *Proceedings of the Third International Conference on Conceptual Structures : Applications, Implementation and Theory*, ICCS '95, pages 32–43, London, UK, UK, 1995. Springer-Verlag. ISBN 3-540-60161-9.

- Min Liu, Mingwen Shao, Wenxiu Zhang, et Cheng Wu. Reduction method for concept lattices based on rough set theory and its application. *Computers & Mathematics with Applications*, 53(9) :1390 – 1410, 2007.
- David E. Losada et Alvaro Barreiro. A logical model for information retrieval based on propositional logic and belief revision. *The Computer Journal*, 44(5) :410–424, 2001.
- C. De Maio, G. Fenza, V. Loia, et M. Parente. Online query-focused twitter summarizer through fuzzy lattice. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8, Aug 2015.
- Carmen De Maio, Giuseppe Fenza, Vincenzo Loia, et Sabrina Senatore. Hierarchical web resources retrieval by exploiting fuzzy formal concept analysis. *Information Processing & Management*, 48(3) :399 – 418, 2012.
- M. E. Maron et J. L. Kuhns. On relevance, probabilistic indexing and information retrieval. *J. ACM*, 7(3) :216–244, July 1960. ISSN 0004-5411.
- Jose Martinez et Erwan Loisant. Browsing image databases with galois’ lattices. In *Proceedings of the 2002 ACM Symposium on Applied Computing, SAC ’02*, pages 791–795, New York, NY, USA, 2002. ACM. ISBN 1-58113-445-2.
- Jan Martinovic et Petr Gajdos. Vector model improvement by fca and topic evolution. In *DATESO*, volume 129, pages 46–57, 2005.
- Carlo Meghini, Fabrizio Sebastiani, Umberto Straccia, et Costantino Thanos. A model of information retrieval based on a terminological logic. In *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’93*, pages 298–307, New York, NY, USA, 1993. ACM. ISBN 0-89791-605-0.
- Nizar Messai, Marie-Dominique Devignes, Malika Smaïl-Tabbone, et Amedeo Napoli. Treillis de concepts et ontologies pour l’interrogation d’un annuaire de sources de données biologiques (bioregistry). *CoRR*, abs/cs/0506017, 2005.
- Nizar Messai, Marie-Dominique Devignes, Amedeo Napoli, et Malika Smaïl-Tabbone. Treillis de concepts et ontologies pour interroger l’annuaire de sources de données biologiques bioregistry. *Ingénierie des Systèmes d’Information*, 11(1) :39–60, 2006a.
- Nizar Messai, Marie-Dominique Devignes, Amedeo Napoli, et Malika Smaïl-Tabbone. BR-Explorer : A sound and complete FCA-based retrieval algorithm (Poster). In *4th International Conference on Formal Concept Analysis - ICFCA 2006*, Dresden/Germany, February 2006b.

- Nizar Messai, Marie-Dominique Devignes, Amedeo Napoli, et Malika Smaïl-Tabbone. Many-valued concept lattices for conceptual clustering and information retrieval. In Malik Ghallab, Constantine D. Spyropoulos, Nikos Fakotakis, et Nikolaos M. Avouris, editors, *ECAI 2008 - 18th European Conference on Artificial Intelligence, Patras, Greece, July 21-25, 2008, Proceedings*, volume 178 of *Frontiers in Artificial Intelligence and Applications*, pages 127–131. IOS Press, 2008.
- W. Muangon et S. Intakosum. Retrieving design patterns by case-based reasoning and formal concept analysis. In *2nd IEEE International Conference on Computer Science and Information Technology*, pages 424–428, Aug 2009.
- Ibtissem Nafkha et Ali Jaoua. Using formal concept analysis for heterogeneous information retrieval. 162, 01 2005.
- Emmanuel Nauer et Yannick Toussaint. Crechaindo : an iterative and interactive web information retrieval system based on lattices. *International Journal of General Systems*, 38(4) :363–378, 2009.
- Emmanuel Nauer et Yannick Toussaint. Dynamical modification of context for an iterative and interactive information retrieval process on the web. In [Eklund et al. \[2008\]](#).
- Jianyun Nie. An information retrieval model based on modal logic. *Information Processing & Management*, 25(5) :477 – 491, 1989. ISSN 0306-4573.
- Terry Noreault, Matthew Koll, et Michael J. McGill. Automatic ranked output from boolean searches in sire. *Journal of the American Society for Information Science*, 28(6) :333–339, 1977.
- E. M. Norris. An algorithm for computing the maximal rectangles in a binary relation. *Revue Roumaine de Mathématiques Pures et Appliquées*, 23(2) :243–250, 1978.
- Lhouari Nourine et Olivier Raynaud. A fast algorithm for building lattices. *Information Processing Letters*, 71(5-6) :199–204, September 1999.
- Dunlu Peng, Sheng Huang, Xiaoling Wang, et Aoying Zhou. *Concept-Based Retrieval of Alternate Web Services*, pages 359–371. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- X. Peng et W. Zhao. An incremental and fca-based ontology construction method for semantics-based component retrieval. In *Seventh International Conference on Quality Software (QSIC 2007)*, pages 309–315, Oct 2007.

- Justin Picard et Jacques Savoy. *A Logical Information Retrieval Model Based on a Combination of Propositional Logic and Probability Theory*, pages 225–258. Physica-Verlag HD, Heidelberg, 2000.
- Jonas Poelmans, Dmitry I. Ignatov, Stijn Viaene, Guido Dedene, et Sergei O. Kuznetsov. Text mining scientific papers : A survey on fca-based information retrieval research. In Petra Perner, editor, *Advances in Data Mining. Applications and Theoretical Aspects*, pages 273–287, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
- G. Polaillon, M. A. Aufaure, B. L. Grand, et M. Soto. Fca for contextual semantic navigation and information retrieval in heterogeneous information systems. In *18th International Workshop on Database and Expert Systems Applications (DEXA 2007)*, pages 534–539, Sept 2007.
- Silke Pollandt. *Fuzzy-Begriffe*. Springer-Verlag, Berlin/Heidelberg, 1997.
- Jay M. Ponte et W. Bruce Croft. A language modeling approach to information retrieval. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '98*, pages 275–281, New York, NY, USA, 1998. ACM. ISBN 1-58113-015-5.
- D. Poshyvanyk et A. Marcus. Combining formal concept analysis with information retrieval for concept location in source code. In *15th IEEE International Conference on Program Comprehension (ICPC '07)*, pages 37–48, June 2007.
- Uta Priss. Lattice-based information retrieval. *Knowledge Organization*, 27 :132–142, 2000.
- Juan A. Recio-García, Marco A. Gómez-Martín, Belén Díaz-Agudo, et Pedro A. González-Calero. *Improving Annotation in the Semantic Web and Case Authoring in Textual CBR*, pages 226–240. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- R. Reiter. In Matthew L. Ginsberg, editor, *Readings in Nonmonotonic Reasoning*, chapter On Closed World Data Bases, pages 300–310. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987. ISBN 0-934613-45-1.
- Raymond Reiter. On closed world data bases. In Hervé Gallaire et Jack Minker, editors, *Logic and Data Bases*, pages 55–76. Springer US, Boston, MA, 1978.
- S. E. Robertson et S. Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information*

- Retrieval*, SIGIR '94, pages 232–241, New York, NY, USA, 1994. Springer-Verlag New York, Inc. ISBN 0-387-19889-X.
- S. E. Robertson et S. Walker. On relevance weights with little relevance information. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '97, pages 16–24, New York, NY, USA, 1997. ACM.
- S. E. Robertson, M. E. Maron, et W. S. Cooper. *The unified probabilistic model for IR*, pages 108–117. Springer Berlin Heidelberg, Berlin, Heidelberg, 1982.
- J. J. Rocchio. Relevance feedback in information retrieval. In Gerard Salton, editor, *The SMART Retrieval System - Experiments in Automatic Document Processing*. Prentice Hall, Englewood, Cliffs, New Jersey, 1971.
- G. Salton et C. S. Yang. On the specification of term values in automatic indexing. *Journal of Documentation.*, 29(4) :351–372, 1973.
- G. Salton, A. Wong, et C. S. Yang. A vector space model for automatic indexing. *Commun. ACM*, 18(11) :613–620, November 1975. ISSN 0001-0782. doi : 10.1145/361219.361220.
- Gerard Salton. *Dynamic Information and Library Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1975. ISBN 0132213257.
- Gerard Salton. *Automatic information organization and retrieval*. McGraw-Hill, New York, 1968.
- Gerard Salton, Edward A. Fox, et Harry Wu. Extended boolean information retrieval. *Commun. ACM*, 26(11) :1022–1036, November 1983. ISSN 0001-0782.
- Kleber Xavier Sampaio de Souza et Joseph Davis. *Using an Aligned Ontology to Process User Queries*, pages 44–53. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
- Fabrizio Sebastiani. *A Probabilistic Terminological Logic for Modelling Information Retrieval*, pages 122–130. Springer London, London, 1994.
- J. F. Sowa. *Conceptual Structures : Information Processing in Mind and Machine*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1984. ISBN 0-201-14472-7.
- Nenad Stojanovic. On the query refinement in the ontology-based searching for information. *Information Systems*, 30(7) :543–563, November 2005. ISSN 0306-4379.

- G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, et L. Lakhal. Fast computation of concept lattices using data mining techniques. In M. Bouzeghoub, M. Klusch, W. Nutt, et U. Sattler, editors, *Proc. 7th Intl. Workshop on Knowledge Representation Meets Databases*, 2000.
- Gerd Stumme, Rafik Taouil, Yves Bastide, Nicolas Pasquier, et Lotfi Lakhal. Computing iceberg concept lattices with titanic. *Data & Knowledge Engineering*, 42(2) :189 – 222, 2002.
- Julien Tane, Philipp Cimiano, et Pascal Hitzler. Query-based multicontexts for knowledge base browsing : An evaluation. In Henrik Schärfe, Pascal Hitzler, et Peter Øhrstrøm, editors, *Conceptual Structures : Inspiration and Application, 14th International Conference on Conceptual Structures, ICCS 2006, Aalborg, Denmark, July 16-21, 2006, Proceedings*, volume 4068 of *Lecture Notes in Computer Science*, pages 413–426. Springer, 2006.
- P. Valtchev, R. Missaoui, et P. Lebrun. A partition-based approach towards constructing galois (concept) lattices. *Discrete Mathematics*, 256(3) :801 – 829, 2002.
- Petko Valtchev et Rokia Missaoui. *Building Concept (Galois) Lattices from Parts : Generalizing the Incremental Methods*, pages 290–303. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.
- Dean van der Merwe, Sergei Obiedkov, et Derrick Kourie. *AddIntent : A New Incremental Algorithm for Constructing Concept Lattices*, pages 372–385. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
- C. J. van Rijsbergen. A non-classical logic for information retrieval. *The Computer Journal*, 29(6) :481–485, June 1986.
- D Ďuráková, S Krajčí, V. Snášel, et P Vojtáš. Conceptual structures. *Arbeitstagung Allgemeine Algebra*, 2002.
- Hao Wang. Logic of many-sorted theories. *Journal of Symbolic Logic*, 17(2) :105–116, 1952. doi : 10.2307/2266241.
- R. Wille. Restructuring lattice theory : an approach based on hierarchies of concepts. In *Rival, I. (ed.) : Ordered Sets*, pages 445–470. Boston, 1982.
- Rudolf Wille. The basic theorem of triadic concept analysis. *Order, Springer*, 12 : 149–158, 1995. ISSN 0167-8094.
- Karl Erich Wolff. A first course in formal concept analysis how to understand line diagrams. 2003.

-
- Sadok Ben Yahia et Ali Jaoua. *Discovering Knowledge from Fuzzy Concept Lattice*, pages 167–190. Physica-Verlag HD, Heidelberg, 2001.
- L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3) :338 – 353, 1965.
- L.A Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1) :3 – 28, 1978. ISSN 0165-0114.
- L. Zerarga et Y. Djouadi. Content-only querying structured contexts using formal concept analysis. In *11th International Symposium on Programming and Systems (ISPS)*, pages 58–64, 2013.
- Loutfi Zerarga et Yassine Djouadi. *Interval-Valued Fuzzy Extension of Formal Concept Analysis for Information Retrieval*, pages 608–615. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. doi : 10.1007/978-3-642-34475-6_73.
- Loutfi Zerarga et Yassine Djouadi. A many-sorted theory proposal for information retrieval : axiomatization and semantics. *Knowledge and Information Systems*, Jul 2017. doi : 10.1007/s10115-017-1074-9.
- Y. Zhang, B. Feng, et Y. Xue. A new search results clustering algorithm based on formal concept analysis. In *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, volume 2, pages 356–360, Oct 2008.
- Yan Zhao, Dong Xu, Wei Liu, Ying Zhu, Wen Zhou, Zongtian Liu, et Yu Zheng. Fuzzy concept lattice based classifier. *2007 International Conference on Fuzzy Systems and Knowledge Discovery*, 01 :598–602, 2007. doi : doi.ieeecomputersociety.org/10.1109/FSKD.2007.322.

ملخص

البحث المقدم في هذه الأطروحة يدخل في نطاق استخدام التحليل الضبابي للمفاهيم الأساسية في مجال استرجاع المعلومات.

في هذه الأطروحة نقتراح نظرية بديهية لاسترجاع المعلومات. هذه النظرية تعتمد على مجموعة من البديهيات التي تقود عملية الاسترجاع مثل براهين النظريات. وفي هذا السياق تقدم أيضا ثلاثة نماذج لاسترجاع المعلومات المرتبطة بهذه النظرية و هي البولاني، الضبابي و الشعاعي. نستعمل هذه النظرية و النموذجين البولاني و الضبابي لتمييز مختلف صيغ الطلبات بخصائص جبرية. نقتراح أيضا استعمال التمديد الضبابي لتحليل المفاهيم الأساسية في معالجة الطلبات الضبابية و كذا استعمال تحليل المفاهيم الأساسية الثلاثي الأبعاد في مجال استرجاع المعلومات المنظمة. للتعبير عن عدم اليقين و عدم الدقة والتدرج في مصفوفة استرجاع المعلومات نقتراح استعمال المجالات في تحليل المفاهيم الأساسية. اعتمادا على القراءة الاحتمالية لتحليل المفاهيم الأساسية نقتراح طرق معالجة الطلبات المنفصلة و الطلبات ذات الصيغة النكرة. و في آخر اقتراح، نقدم طريقة لإنشاء بنية خطية للمفاهيم المتعلقة بتحليل المفاهيم الأساسية و قراءتها الإحتالية.

الكلمات الرئيسية : تحليل المفاهيم الأساسية، استرجاع المعلومات، نظرية بديهية، نظرية الإحتالات، نظرية المجموعات الضبابية.

