

Une Extension de la Connexion de Galois Floue pour la Recherche d'Information

Ch. Latiri Cherif^{*}, J.P. Chevallet[†], S. Elloumi^{*}, A. Jaoua[‡]

^{*} Faculté des sciences de Tunis
Département des Sciences Informatiques
Campus Universitaire, le Belvédère, Tunis, Tunisie
chiraz.latiri@gnet.tn
samir.elloumi@fst.rnu.tn

[†] Équipe Modélisation et Recherche d'Information Multimédia
Laboratoire CLIPS-IMAG
B.P. 53, 38041 Grenoble Cedex 9, France
jean-pierre.chevallet@imag.fr

[‡] University of Qatar
Collège of Science
Department of Computer Sciences
Doha, PO.Box 2713
jaoua@qu.edu.qa

Résumé

En recherche d'information, l'implication logique incertaine $D \longrightarrow RQ$ a fait l'objet de plusieurs pistes de recherche et d'interprétations. Dans cet article nous proposons une nouvelle manière de l'interpréter en se plaçant dans un contexte flou. Nous introduisons ainsi un nouveau schéma de correspondance basé sur une extension de la connexion de Galois floue. Ce schéma tient compte dans sa définition des différentes sémantiques des degrés attribués aux termes de la requête et d'un quantificateur flou relatif.

Mots-clés : recherche d'information, implication floue, quantificateur flou, connexion de Galois floue.

1 INTRODUCTION

Dans [28, 29], nous avons proposé une approche pour l'extraction de règles associatives non redondantes entre termes en se limitant à explorer une relation *terme-document* classique, i.e. non floue. Les seuls degrés figurant dans cette relation sont de type 0 ou 1, traduisant la présence ou l'absence d'un terme d'indexation dans un document du corpus.

Cependant, dans le domaine de la recherche d'information (RI), nous disposons du poids d'un terme dans un document [38], qui peut être une information importante à intégrer dans un processus de découverte de règles associatives entre termes ou encore lors de l'interrogation en RI. Dans la pratique, ce poids n'est pas toujours égal à 0 ou 1, mais appartient plutôt à l'intervalle $[0, 1]$. Ceci nous permet ainsi de nous placer dans un contexte flou, où la relation *terme-document* est considérée comme une relation binaire floue, traduisant des degrés d'appartenance des termes dans les documents du corpus.

C'est ainsi que la théorie des ensembles flous [46] présente une base formelle adéquate pour offrir un instrument de discrimination du rôle des termes dans la représentation des documents.

Aussi, l'objectif de cet article est double :

1. Étendre la connexion de Galois floue afin de pouvoir extraire les règles associatives floues entre termes en considérant une approche par fermeture de la connexion de Galois floue [27] ;
2. Présenter la définition d'un nouveau schéma de correspondance *requête-document* qui est fondé sur un opérateur plus flexible de la connexion de Galois floue, en intégrant le quantificateur flou relatif "*la plupart*". Nous prenons en compte les différentes sémantiques des degrés d'appartenance associés aux termes. La finalité d'un tel schéma est de pouvoir assouplir le mécanisme classique d'appariement *requête-document*.

L'article est organisé comme suit : la deuxième section introduit la représentation floue des données textuelles, retenue pour la nouvelle approche proposée. La troisième section décrit les fondements de l'article. Nous présentons dans la quatrième section le principe d'extension de la connexion de Galois floue. La cinquième section présente la première extension de la connexion de Galois, utilisée pour la génération de règles associatives floues entre termes. Au niveau de la sixième section, nous décrivons et nous détaillons le nouveau schéma de correspondance, basé sur une extension de la connexion de Ga-

lois floue et nous concluons ensuite, dans la dernière section, en introduisant quelques perspectives de l'approche.

2 REPRÉSENTATION FLOUE DES DONNÉES TEXTUELLES

Nous commençons par présenter la définition du poids d'un terme dans un document ainsi que la définition d'un ensemble flou et d'une relation binaire floue.

Définition 1

Le poids P_{ij} d'un terme t_i dans un document d_j , est défini comme la valeur de représentativité du terme pour ce document. Une mesure classiquement utilisée est le produit de la fréquence d'apparition du terme par la fréquence inversée du document, donnée par la formule suivante : $P_{ij} = tf_{ij} \times idf_j$ ¹ [38].

La valeur P_{ij} est souvent la composition d'une mesure locale au document et d'une mesure globale au corpus. Si nous désirons utiliser la théorie des ensembles flous, il est possible de relier cette notion de poids d'un terme à la notion d'appartenance à un ensemble flou, définie comme suit [46] :

Définition 2

Un ensemble flou $\tilde{F}$ ² de l'univers de discours U est caractérisé par une fonction d'appartenance $\mu_{\tilde{F}} : U \rightarrow [0, 1]$, où $\mu_{\tilde{F}}(u)$, pour $u \in U$, représente le degré d'appartenance de u à l'ensemble flou $\tilde{F}$. L'ensemble flou $\tilde{F}$ est noté comme suit : $\tilde{F} = \{(u_1, \mu_{\tilde{F}}(u_1)), (u_2, \mu_{\tilde{F}}(u_2)), \dots, (u_n, \mu_{\tilde{F}}(u_n))\}$, avec $u_1, \dots, u_n \in U$.

La relation entre un terme d'indexation et les documents qu'il indexe peut se modéliser de manière générale par une relation binaire floue, définie comme suit [10] :

¹Rappelons que tf_{ij} (term frequency) est une mesure de la représentativité du terme t_i dans le document d_j . idf (inverted document frequency) mesure par contre le pouvoir de discrimination du terme t_i dans le document d_j .

²Dans la suite de l'article, le symbole $\sim$ sert à désigner une notion définie dans un contexte flou.

Définition 3

Soient U et V deux univers de discours. Une relation binaire floue $\tilde{R}$ est un ensemble flou défini sur le produit cartésien $U \times V$. L'expression $\mu_{\tilde{R}}(u, v)$ avec $u \in U$ et $v \in V$, exprime le degré d'association entre u et v dans la relation $\tilde{R}$.

La définition de la fonction d'appartenance relative à cette relation binaire floue *terme-document* a été discutée dans divers travaux tel que [1]. Ainsi, dans le cadre d'extraction de règles associatives floues entre termes, nous avons proposé dans [27] deux manières possibles pour évaluer de manière pratique le degré d'appartenance $\mu_{\tilde{R}}(t, d)$. Nous avons considéré une première mesure globale basée sur le poids d'un terme dans un document et une deuxième mesure locale au document qui est basée sur la fréquence d'un terme dans un document. Les expérimentations menées dans [27] sur des corpus réels ont montré que la deuxième mesure locale illustre plus fidèlement le degré d'appartenance d'un terme dans un document.

Néanmoins, Une représentation floue des données textuelles a été déjà introduite dans des travaux antérieurs en RI [1, 26, 35, 39], dont le principal objectif est de raffiner et d'assouplir le processus de recherche d'information. La matrice classique *terme-document* est interprétée comme une relation binaire floue, notée par $\tilde{R}$. À chaque couple *terme-document* (t_i, d_j) est associé un degré $\in [0, 1]$ (voir tableau 1).

$\tilde{R}$	t_1	t_2	t_3	t_4
d_1	0.90	0.70	0.60	0.20
d_2	0.70	0.45	0.76	0.80
d_3	0.66	0.88	1.00	0.00
d_4	1.00	0.26	0.60	0.73

TAB. 1 – Relation binaire floue *terme-document*.

Ainsi, partant de la relation binaire floue $\tilde{R}$ illustrée par le tableau 1, nous définissons :

1. Un ensemble fini D de documents ;
2. Un ensemble fini T de termes ;
3. Le couple $(t_i, d_j) \in_{\alpha} \tilde{R}$ représente le fait que le terme $t_i \in T$ est présent dans le document $d_j \in D$ avec un *degré d'appartenance* égal à α appartenant à l'intervalle $[0, 1]$.

Dans les systèmes booléens pondérés [40], la représentation booléenne a été étendue en permettant à la fonction d'indexation F_I de prendre des valeurs dans l'intervalle unité $[0, 1]$, afin de calculer le degré de représentativité du terme t_i dans le document d_j .

Cette valeur peut varier entre une représentativité nulle ($F_I(d_j, t_i) = 0$) et une représentativité complète ($F_I(d_j, t_i) = 1$). En considérant la théorie des ensembles flous, un document $d_j \in D$ est représenté par un ensemble flou $\widetilde{M}$ de termes comme suit :

$$\widetilde{M}(d_j) = \{(t_i, \mu_{d_j}(t_i)) \mid t_i \in T\} \quad (1)$$

L'expression (1) indique que chaque terme t_i dans l'ensemble $\widetilde{M}(d_j)$ a un degré d'appartenance $\mu_{d_j}(t_i) \in [0, 1]$. Dans ce cas $\mu_{d_j}(t_i) = F_I(d_j, t_i)$, c'est-à-dire que la fonction d'indexation F_I est interprétée comme une fonction d'appartenance de l'ensemble flou $\widetilde{M}(d_j)$, exprimant la représentativité du terme t_i dans le document d_j [35]. Une définition possible de la fonction d'appartenance est obtenue à partir d'une variante de la fonction de pondération classique $F_I(d_j, t_i) = tf_{i,j} \times F_N(idf_j)$, où F_N est une fonction de normalisation pour obtenir des valeurs de la fonction F_I dans l'intervalle $[0, 1]$ [40].

Le choix de la fonction d'appartenance floue μ_{d_j} , qu'il convient d'opérer en pratique, est important pour le bon fonctionnement d'un système de recherche d'information. Nous avons discuté de ce choix dans [27], dans le cadre de la génération de règles associatives floues entre termes et leur utilisation dans l'expansion de requêtes en RI. Pour l'instant, nous le laissons en marge de notre étude, dans le but d'énoncer l'ensemble de nos propositions, sans nous restreindre par la fonction d'appartenance qui est mise en place de manière pratique.

3 FONDEMENTS DE L'APPROCHE

Dans cette section, nous présentons les notions de base que nous utiliserons dans la suite de l'article pour énoncer un certain nombre de propositions.

3.1 Connexion de Galois classique et RI

L'objectif de notre travail est d'examiner la problématique de la recherche d'information sous un angle différent, en considérant l'analyse formelle de

concepts comme fondement mathématique [23].

Habituellement, une requête est définie en RI par un ensemble de termes reliés par des opérateurs, i.e. de conjonction ou de disjonction. Dans le cadre de notre approche, nous considérons que la requête est définie uniquement par l'ensemble des termes qu'elle contient.

En effet, l'association *terme-document* peut être considérée comme une relation binaire. Une requête étant un ensemble de termes, si tous ses termes sont d'égale importance pour l'utilisateur. La meilleure réponse à la requête, dans ce cas, est l'ensemble de documents qui sont indexés par tous les termes de la requête. Cet ensemble est donné par l'opérateur h d'une connexion de Galois classique [23]. Nous donnons ici sa définition adaptée au domaine de la RI :

Définition 4

Soit R une relation binaire définie sur $D \times T$ tels que D définit un ensemble de documents et T un ensemble de termes d'indexation. Le triplet (D, T, R) définit un contexte formel. Pour les deux ensembles A et B , tels que $A \subseteq D$, $B \subseteq T$, les opérateurs f et h de la connexion de Galois classique sont définis comme suit [23] :

$$\begin{aligned} f(A) &= \{t \in T \mid \forall d \in A, (t, d) \in R\} \\ h(B) &= \{d \in D \mid \forall t \in B, (t, d) \in R\} \end{aligned}$$

La composition des deux opérateurs f et h , i.e. $f \circ h$, définit la fermeture de la connexion de Galois.

La figure 1 illustre l'utilisation de la connexion de Galois classique dans le contexte de la RI. L'ensemble B désigne les termes composant une requête booléenne et $h(B)$ représente alors l'ensemble de documents contenant tous les termes de cette requête. Inversement, $f(A)$ représente le plus grand ensemble de termes qui sont présents dans tous les documents de l'ensemble A .

Nous allons maintenant introduire quelques notions, que nous utiliserons pour la proposition d'un nouveau schéma de correspondance en RI, basé sur une extension de la connexion de Galois floue.

3.2 Les implications floues

Dans la littérature, différentes implications multivaluées ont été proposées [4, 16]. Elles correspondent principalement à deux classes d'opérations d'im-

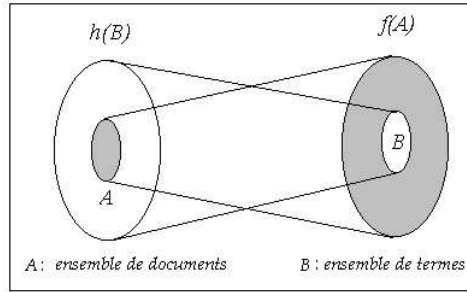


FIG. 1 – Utilisation de la connexion de Galois classique en RI

plication qui sont les *R-implications* et les *S-implications*, définies comme suit [10, 20] :

Définition 5

Pour $a, b \in [0, 1]$, une *R-implication* est définie par :

$$a \xrightarrow{R_{imp}} b = \sup \{x \in [0, 1] \mid a \top x \leq b\} \quad (2)$$

sachant que $\top$ désigne une *t-norme triangulaire* (voir annexe A, page 35).

Définition 6

Pour $a, b \in [0, 1]$, une *S-implication* est définie par :

$$a \xrightarrow{S_{imp}} b = (1 - a) \perp b \quad (3)$$

sachant que $\perp$ désigne une *t-conorme triangulaire* (voir annexe A, page 35).

Les *R-implications* sont souvent utilisées pour l'estimation du niveau de satisfaction, alors que les *S-implications* sont recommandées pour la modélisation de l'importance [19].

Définition 7

Une *implication floue*, ou *implication multivaluée* $I = a \xrightarrow{I_{floue}} b$, est une fonction de $[0, 1] \times [0, 1] \rightarrow [0, 1]$ qui détermine le degré de vérité de la proposition $a \xrightarrow{I_{floue}} b$ [16].

Nom	Abbréviation	Degré de vérité	Type d'implication	Propriété 1	Propriété 2
Lukasiewicz	$I_{Luka}(a, b)$	$\min(1, 1 - a + b)$.	R et S	Vérifiée	Vérifiée
Gödel	$I_{Gö}(a, b)$	1 si $a \leq b$; b sinon.	R	Vérifiée	Vérifiée
Goguen	$I_G(a, b)$	1 si $a = 0$; $\min(1, b/a)$ sinon.	R	Vérifiée	Vérifiée
Rescher-Gaïnes	$I_{RG}(a, b)$	1 si $a \leq b$; 0 sinon.	R	Vérifiée	Vérifiée
Kleene-Dienes	$I_{KD}(a, b)$	$\max(1 - a, b)$.	S	Vérifiée	Non vérifiée
Reichenbach	$I_{RB}(a, b)$	$(1 - a + a \times b)$.	S	Vérifiée	Non vérifiée

TAB. 2 – Implications floues.

Nous présentons deux propriétés des implications floues qui seront utilisées plus loin dans cet article pour élaborer les preuves des propriétés de la connexion de Galois floue que nous définissons pour la RI.

Propriété 1

Soient a_1, a_2, b_1, b_2 des degrés appartenant à l'intervalle $[0, 1]$. Si nous avons $(a_1 \leq a_2)$ et $(b_1 \leq b_2)$ alors $(a_2 \xrightarrow{I_{floue}} b_1 \leq a_1 \xrightarrow{I_{floue}} b_2)$ [34].

Propriété 2

Soient a et b deux degrés appartenant à l'intervalle $[0, 1]$. Nous avons : $((a \xrightarrow{I_{floue}} b) \xrightarrow{I_{floue}} b) \geq a$ [34].

Dans le tableau 2, les principales implications floues sont données [16]. Nous notons le degré de vérité de la prémisse par a et celui de la conclusion par b .

3.3 Les quantificateurs flous

Zadeh a introduit les quantificateurs flous appelés aussi les quantificateurs linguistiques pour décrire des situations intermédiaires entre le quantificateur universel et le quantificateur existentiel [45]. Nous distinguons deux types de quantificateurs flous : les *quantificateurs relatifs* et les *quantificateurs absolus* [45]. Dans le cadre de notre travail, nous nous intéressons uniquement aux quantificateurs relatifs tels que le quantificateur “la plupart”,

dans le but de pouvoir traiter “la plupart” des termes d’une requête, au lieu de considérer “tous” ses termes.

Un quantificateur relatif est défini comme un ensemble flou (noté par Q_{rel}) d’un univers de discours U et il est caractérisé par une fonction d’appartenance $\mu_{Q_{rel}} : U \rightarrow [0, 1]$, où $\mu_{Q_{rel}}(u)$, pour $u \in U$, désigne le degré d’appartenance de la proportion u au quantificateur Q_{rel} [10].

À titre d’exemple, le quantificateur flou relatif “la plupart” défini sur un ensemble fini de termes, peut être représenté par l’ensemble flou “la plupart” comme suit :

$$\mu_{la\ plupart} : T \longrightarrow [0, 1]$$

tels que $\mu_{la\ plupart}(i) \leq \mu_{la\ plupart}(i + 1)$ pour $i = 1..(m - 1)$ termes et $\mu_{la\ plupart}(m) = 1$, sachant que m désigne le nombre de termes qui définit l’univers de discours du quantificateur flou “la plupart”.

Exemple 1

Considérons le quantificateur flou “la plupart” défini sur l’ensemble des termes d’une requête en RI. Supposons que la requête peut contenir au plus cinq termes ($m = 5$). Dans le meilleur des cas où il existe un document qui contient tous les termes de la requête, la valeur $\mu_{la\ plupart}(5)$ est égale à 1. Inversement, dans le cas où le document ne contient aucun terme de la requête, nous avons $\mu_{la\ plupart}(0) = 0$. Les cas intermédiaires sont évalués selon la fonction d’appartenance trapézoïdale du quantificateur flou “la plupart”, illustrée par la figure 2.

Notons que dans cet exemple, la fonction d’appartenance $\mu_{la\ plupart}$ est définie comme suit : $\mu_{la\ plupart}(0) = 0.0$, $\mu_{la\ plupart}(1) = 0.0$, $\mu_{la\ plupart}(2) = 0.0$, $\mu_{la\ plupart}(3) = 0.5$, $\mu_{la\ plupart}(4) = 0.75$ et $\mu_{la\ plupart}(5) = 1.0$.

4 EXTENSION FLOUE DE LA CONNEXION DE GALOIS

Différentes extensions floues de la connexion de Galois et du concept formel ont été proposées dans la littérature selon différents points de vue [5, 25, 34, 44]. Nous distinguons la proposition de Wolff [44] qui consiste à ramener le problème à un contexte non flou, contrairement à Pollandt [34], Belohlavèk [5] et Jaoua et al. [25] qui ont utilisé des opérateurs flous dans les

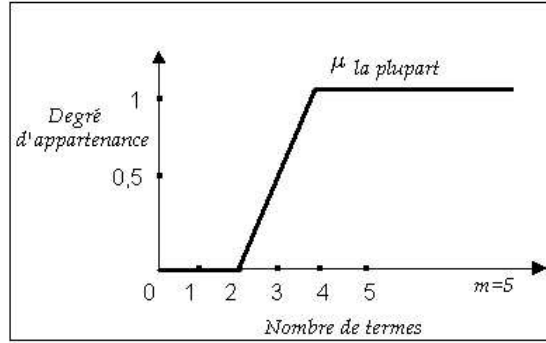


FIG. 2 – L’allure de la fonction d’appartenance du quantificateur flou “la plupart”

nouvelles définitions proposées. Ces différentes propositions sont résumées dans [22]. Nous mettons l’accent sur la proposition de Jaoua et al. [25], puisque elle est utilisée comme base pour définir notre propre extension floue pour la RI.

4.1 Connexion de Galois floue et RI

Jaoua et al. [25] proposent de nouvelles définitions pour les opérateurs $\tilde{f}$ et $\tilde{h}$ de la connexion de Galois floue et une nouvelle forme pour le concept formel floue. Ils considèrent que seule l’intention du concept formel floue est représentée par un ensemble floue.

Afin d’adapter ces définitions à un contexte d’extraction de connaissances ou de RI, nous allons redéfinir les opérateurs de la connexion de Galois floue $\tilde{f}$ et $\tilde{h}$, en considérant le contexte formel textuel floue $\widetilde{\mathcal{TF}} = (\mathcal{C}, \mathcal{T}, \tilde{R})$, défini comme suit :

Définition 8

Le triplet $\widetilde{\mathcal{TF}} = (\mathcal{C}, \mathcal{T}, \tilde{R})$ définit un contexte formel textuel floue, tel que $\mathcal{C}$ décrit l’ensemble de documents du corpus, $\mathcal{T}$ l’ensemble des termes du corpus et $\tilde{R}$ une relation binaire floue, définie sur $\mathcal{C} \times \mathcal{T}$.

Pour un ensemble de documents $D \subseteq \mathcal{C}$ et un ensemble floue de termes $\widetilde{RQ}$

défini sur $\mathcal{T}$, les opérateurs de la connexion de Galois floue $\tilde{f}$ et $\tilde{h}$ [25] sont reformulés comme suit :

$$\tilde{f}(D) = \{t \mid \alpha = \min_{d \in D} (\mu_{\tilde{R}}(t, d)), t \in \mathcal{T}\} \quad (4)$$

$$\tilde{h}(\widetilde{RQ}) = \{d \in \mathcal{C} \mid \forall t \in \mathcal{T}, (\mu_{\widetilde{RQ}}(t) \xrightarrow{I_{RG}} \mu_{\tilde{R}}(t, d)) = 1\} \quad (5)$$

sachant que :

- I_{RG} désigne l’implication floue de *Rescher-Gaïnes* (voir tableau 2) ;
- $\tilde{t}$ désigne le terme t appartenant à tous les documents d_j de l’ensemble D avec un degré d’appartenance α qui n’est autre que le minimum des $\mu_{\tilde{R}}(t, d_j)$. À titre d’exemple, considérons la relation binaire floue terme-document illustrée par le tableau 1, si $D = \{d_2, d_3, d_4\}$ alors $\tilde{f}(D) = \begin{matrix} 0.66 & 0.26 & 0.60 & 0.00 \\ \{ t_1, & t_2, & t_3, & t_4 \}. \end{matrix}$

Nous remarquons ainsi, que l’ensemble $\widetilde{RQ}$ est un sous-ensemble flou de termes, où à chaque terme est associé un degré de l’intervalle $[0, 1]$, qui exprime le seuil de représentativité minimal exigé par l’utilisateur. L’ensemble $\widetilde{RQ}$ peut donc faire office de requête dans un modèle de RI flou.

Appliquer l’opérateur $\tilde{h}$ à l’ensemble flou de termes $\widetilde{RQ}$, i.e. $\tilde{h}(\widetilde{RQ})$, permet d’obtenir un ensemble classique, i.e. non flou, de documents qui sont annotés par les termes de $\widetilde{RQ}$, en respectant le degré de représentativité exigé pour chaque terme de la requête $\widetilde{RQ}$. Inversement, appliquer l’opérateur $\tilde{f}$ à l’ensemble classique de documents D , i.e. $\tilde{f}(D)$, permet d’obtenir un ensemble flou de termes qui sont pondérés par leurs degrés d’appartenance respectifs.

Selon l’expression (5), l’opérateur $\tilde{h}$ de la connexion de Galois floue peut être défini comme la fonction de correspondance qui va apparier les documents pertinents avec “tous” les termes d’une requête floue donnée, sans pour autant évaluer leurs degrés de pertinence respectifs. Ceci constitue une limite de l’extension de Galois floue proposée dans [25]. Néanmoins, elle reste intéressante dans un contexte de datamining, étant donné qu’une approche d’extraction de règles associatives floues, basée sur la fermeture de cette connexion de Galois floue, a été proposée dans [6].

Cependant, en associant aux termes de la requête des degrés, nous spécifions plus de restrictions sur les documents cherchés. Deux types de restriction sont citées dans la littérature [42], à savoir des *restrictions qualitatives* et des

restrictions quantitatives. Dans le cadre de notre étude, nous nous intéresserons uniquement aux restrictions qualitatives qui illustrent le cas où les degrés associés aux termes de la requête expriment un critère qui affecte la *qualité* des documents cherchés. Ainsi, les degrés traduisent des contraintes qui doivent être satisfaites par les degrés d'appartenance des termes au niveau de la relation *terme-document*. L'utilisateur peut par exemple fixer un seuil de représentativité minimal pour les termes de la requête.

4.2 Sémantique RI de l'implication

Dans le modèle vectoriel de recherche d'information, la requête est interprétée comme la description d'un document idéal [38]. La recherche consiste alors à sélectionner les documents qui sont les plus proches de la requête. Nous remarquons que le schéma de correspondance basé sur l'opérateur $\tilde{h}$ de la connexion de Galois floue proposée dans [25], ne correspond pas à cette sémantique. Cet opérateur utilise l'implication de *Rescher-Gaïnes* (voir expression (5)) qui présente un certain nombre de limites, liées à l'utilisation de l'inégalité " $\leq$ ".

En effet, considérons par exemple la relation *terme-document* illustrée par le tableau 1 (voir page 3) et supposons que $\mu_{\widetilde{RQ}}(t_2) = 0.10$. Dans une interprétation où la requête représente le document idéal, un terme de la requête avec un faible degré signifie explicitement que le terme ne doit pas apparaître dans les documents. Or, d'après la définition de l'opérateur $\tilde{h}$, donnée par l'expression (5), le degré de vérité de l'implication $\mu_{\widetilde{RQ}}(t_2) \xrightarrow{IRG} \mu_{\widetilde{R}}(t_2, d_j)$ est égal à 1.00 pour tous les documents de la relation, puisque $\mu_{\widetilde{RQ}}(t_2) = 0.10 \leq \mu_{\widetilde{R}}(t_2, d_j)$, avec $j = 1..4$ (voir tableau 1, page 3). Nous remarquons que les documents d_3 et d_4 sont inclus dans le résultat de $\tilde{h}(\widetilde{RQ})$ bien qu'ils ne répondent pas à la requête interprétée avec la sémantique du modèle vectoriel.

Inversement, supposons que $\mu_{\widetilde{RQ}}(t_1) = 0.68$. D'après la définition de l'opérateur $\tilde{h}$ de la connexion de Galois, donnée par l'expression (5), l'implication $\mu_{\widetilde{RQ}}(t_1) \xrightarrow{IRG} \mu_{\widetilde{R}}(t_1, d_3)$ donne 0 car $\mu_{\widetilde{RQ}}(t_1) = 0.68 > \mu_{\widetilde{R}}(t_1, d_3) = 0.66$. Ainsi, le document d_3 ne sera pas inclus dans le résultat de $\tilde{h}(\widetilde{RQ})$ bien qu'il soit assez pertinent puisque les deux valeurs $\mu_{\widetilde{RQ}}(t_1) = 0.68$ et $\mu_{\widetilde{R}}(t_1, d_3) = 0.66$ sont très proches.

Nous signalons que l'implication de *Rescher-Gaïnes* demeure un cas par-

ticulier des implications floues car elle rejoint le cas binaire avec un seuil de vérification binaire égal à 0 ou 1. Pour éviter cette sélection binaire, il semble plus approprié de considérer d'autres implications floues plus flexibles que l'implication de *Rescher-Gaïnes* et d'étudier les critères de choix de ces dernières dans le contexte de la définition d'une connexion de Galois floue. De plus, l'utilisation de l'implication de *Rescher-Gaïnes* signifie que le degré de vérité de l'implication $\mu_{\widetilde{RQ}}(t) \xrightarrow{I_{RG}} \mu_{\widetilde{R}}(t, d)$ n'est pris en compte que si l'égalité $\mu_{\widetilde{RQ}}(t) \xrightarrow{I_{RG}} \mu_{\widetilde{R}}(t, d) = 1.00$ est vérifiée. Il serait intéressant d'associer un seuil de vérification θ compris entre 0 et 1 relatif au degré de vérité de l'implication $\mu_{\widetilde{RQ}}(t) \xrightarrow{I_{RG}} \mu_{\widetilde{R}}(t, d)$, en considérant une autre implication floue. Ceci peut réduire le problème de silence, i.e. réponse vide, en considérant un seuil de vérification minimal.

Nous développons ainsi l'opérateur $\widetilde{h}$ de la connexion de Galois floue dans le contexte de la RI afin de pouvoir trouver, par le biais de cet opérateur, non seulement l'ensemble de documents pertinents, mais aussi leurs degrés de pertinence respectifs, évalués par l'implication $\mu_{\widetilde{RQ}}(t) \xrightarrow{I_{RG}} \mu_{\widetilde{R}}(t, d)$, pour chaque terme t de la requête $\widetilde{RQ}$.

Dans un premier temps, nous introduisons une nouvelle définition de la connexion de Galois floue, adaptée à l'extraction de règles associatives floues à partir de texte. Nous avons détaillé dans [27] l'extraction de règles associatives floues à partir de texte basée sur la fermeture de la connexion de Galois floue que nous définissons dans la section qui suit ainsi que leur utilisation pour l'expansion de requêtes en recherche d'information. Cette approche est validée par une série d'expérimentations [27] effectuées sur la collection test OFIL du projet AMARYLLIS³.

5 CONNEXION DE GALOIS FLOUE POUR L'EXTRACTION DE RÈGLES ASSOCIATIVES FLOUES À PARTIR DE TEXTES

Nous proposons une nouvelle définition de la connexion de Galois pour le support de relation binaire floue *terme-document*. En partant des propositions présentées dans [22, 25], nous suggérons les deux améliorations suivantes :

³<http://amaryllis.inist.fr/>

1. L'utilisation d'une implication floue moins stricte que celle de *Rescher-Gaines*. Dans le cadre de notre étude, nous proposons d'utiliser l'implication de Lukasiewicz qui est de type R-implication et S-implication à la fois (voir tableau 2, page 7) et qui vérifie les propriétés 1 et 2 (voir page 7). Cette implication supporte ainsi les deux types de sémantiques à savoir : la satisfaction et l'importance.
2. Étant donné que le degré de vérité d'une implication floue, autre que celle de *Rescher-Gaines*, appartient à l'intervalle $[0, 1]$, nous proposons de fixer un seuil minimal de vérification $\theta \in [0, 1]$ de l'implication $(\mu_{\tilde{T}}(t) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t, d) \geq \theta)$, pour chaque terme de l'ensemble flou de termes $\tilde{T}$.

5.1 Définition de la connexion de Galois floue

Nous énonçons la nouvelle définition de la connexion de Galois floue [30] :

Définition 9

Soit $\tilde{\mathcal{TF}} = (\mathcal{C}, \mathcal{T}, \tilde{R})$ un contexte formel textuel flou. Pour un ensemble de documents $D \subseteq \mathcal{C}$, un ensemble flou de termes $\tilde{T}$ défini sur $\mathcal{T}$ et un seuil de vérification $\theta \in [0, 1]$, les opérateurs $\tilde{f}$ et $\tilde{h}_\theta$ sont définis comme suit :

$$\tilde{f}(D) = \{t \mid \alpha = \min_{d \in D} (\mu_{\tilde{R}}(t, d)), t \in \mathcal{T}\}$$

$$\tilde{h}_\theta(\tilde{T}) = \{d \in \mathcal{C} \mid \forall t \in \mathcal{T}, (\mu_{\tilde{T}}(t) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t, d)) \geq \theta\}$$

Sachant que $\xrightarrow{I_{floue}}$ est une implication floue de type R-implication ou S-implication, vérifiant les propriétés 1 et 2 (voir page 7).

L'image d'un sous-ensemble de documents D du corpus par l'opérateur $\tilde{f}$ de la connexion de Galois floue est un sous-ensemble flou de termes, tel que le degré d'appartenance pour un terme représente le minimum de ses degrés d'appartenance dans la relation binaire floue *terme-document*, respectivement, pour tous les documents du sous-ensemble D . Lorsque le degré d'appartenance est binaire, nous retrouvons exactement la définition de l'opérateur f de la connexion de Galois classique [23].

Inversement, l'image d'un sous-ensemble flou de terme $\tilde{T}$ défini sur $\mathcal{T}$, par l'opérateur $\tilde{h}$ de la connexion de Galois floue, est un sous-ensemble classique

de documents tel que le degré d'appartenance pour un terme appartenant à $\tilde{T}$ est atteint jusqu'à un seuil θ dans le sous-ensemble de documents de la relation binaire floue *terme-document* $\tilde{R}$. Cette sémantique est donnée par le calcul du degré de vérité de l'implication $\mu_{\tilde{T}}(t) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t, d) \geq \theta$. Lorsque le degré de vérité de cette implication est binaire, comme dans le cas de l'utilisation de l'implication floue de *Rescher-Gaines* (voir tableau 2, page 7), nous retrouvons exactement la définition de l'opérateur h de la connexion de Galois classique [23].

La fermeture d'un ensemble flou de termes $\tilde{T}$, i.e. $\tilde{f} \circ \tilde{h}_\theta(\tilde{T})$, représente un sous-ensemble flou de termes, tel que cet ensemble est fortement lié à un ensemble de documents. Nous désignerons le sous-ensemble flou de termes $\tilde{f} \circ \tilde{h}_\theta(\tilde{T})$ par *concept réduit flou*. Partant de la définition de concept flou donnée dans [25], nous énonçons la définition suivante [31] :

Définition 10

Soit $\tilde{T}$ un ensemble flou de termes défini sur $\mathcal{T}$. Il est dit *concept réduit flou*, si et seulement si il est égal à sa fermeture, i.e. $\tilde{f} \circ \tilde{h}_\theta(\tilde{T}) = \tilde{T}$.

Notons que la relaxation apportée par le seuil de vérification minimum θ , au niveau de la définition de l'opérateur $\tilde{h}_\theta$ de la connexion de Galois floue, permet de ne pas extraire de concepts réduits flous vides à partir du contexte formel flou $\tilde{\mathcal{F}} = (\mathcal{C}, \mathcal{T}, \tilde{R})$.

Les concepts réduits flous traduisent une connaissance additionnelle qui peut être représentée formellement par des règles associatives floues entre termes. Ces règles peuvent être utilisées pour l'expansion de requêtes en RI [27].

5.2 Propriétés de la connexion de Galois floue

Proposition 1

Les opérateurs $\tilde{f}$ et $\tilde{h}_\theta$ définissent une connexion de Galois floue. Les propriétés suivantes sont vérifiées pour tout ensemble classique de documents $D \subseteq \mathcal{C}$ et tout ensemble flou de termes $\tilde{T}$ défini sur $\mathcal{T}$, en considérant une implication floue I_{floue} , vérifiant les propriétés 1 et 2 (voir page 7).

- (P1) $D_1 \subseteq D_2 \Rightarrow \tilde{f}(D_1) \supseteq \tilde{f}(D_2)$
- (Q1) $\tilde{T}_1 \subseteq \tilde{T}_2 \Rightarrow \tilde{h}_\theta(\tilde{T}_1) \supseteq \tilde{h}_\theta(\tilde{T}_2)$
- (P2) $D \subseteq \tilde{h}_\theta \circ \tilde{f}(D)$
- (Q2) $\tilde{T} \subseteq \tilde{f} \circ \tilde{h}_\theta(\tilde{T})$

- (P3) $\tilde{f}(D) = \tilde{f} \circ \tilde{h}_\theta \circ \tilde{f}(D)$
- (Q3) $\tilde{h}_\theta(\tilde{T}) = \tilde{h}_\theta \circ \tilde{f} \circ \tilde{h}_\theta(\tilde{T})$
- (P4) $D_1 \subseteq D_2 \Rightarrow \tilde{h}_\theta \circ \tilde{f}(D_1) \subseteq \tilde{h}_\theta \circ \tilde{f}(D_2)$
- (Q4) $\tilde{T}_1 \subseteq \tilde{T}_2 \Rightarrow \tilde{f} \circ \tilde{h}_\theta(\tilde{T}_1) \subseteq \tilde{f} \circ \tilde{h}_\theta(\tilde{T}_2)$
- (P5) $\tilde{h}_\theta \circ \tilde{f}(\tilde{h}_\theta \circ \tilde{f}(D_1)) = \tilde{h}_\theta \circ \tilde{f}(D_1)$
- (Q5) $\tilde{f} \circ \tilde{h}_\theta(\tilde{f} \circ \tilde{h}_\theta(\tilde{T}_1)) = \tilde{f} \circ \tilde{h}_\theta(\tilde{T}_1)$
- (Q6) $D \subseteq \tilde{h}_\theta(\tilde{T}) \Leftrightarrow \tilde{T} \subseteq \tilde{f}(D)$

Les preuves de ces différentes propriétés sont données dans [22].

6 EXTENSION DE LA CONNEXION DE GALOIS FLOUE POUR LA RI

6.1 Motivations

Dans ses articles de 1986, Van Rijsbergen [36, 37] a suggéré une modélisation de la recherche d'information à l'aide d'une logique. Cette idée a été la base d'un courant de recherche basé sur l'utilisation de la logique en tant que modèle de description pour la RI. Il modélise ainsi la correspondance entre un document et une requête par une implication d'un document d vers une requête RQ , i.e. $d \longrightarrow RQ$. Cette suggestion a été reprise par Nie [32] et étendue à une double implication. Mais différentes autres interprétations logiques de cette implication ont été étudiées par Sebastiani [41].

Dans sa proposition initiale, Van Rijsbergen [37] ne traite pas l'implication $d \longrightarrow RQ$ en tant qu'implication matérielle de la logique classique, mais il considère que l'opération primitive pour la RI est une implication incertaine. D'où notre intérêt pour l'utilisation de la logique floue.

Dans le reste de l'article, nous proposons une nouvelle interprétation de l'implication directe $d \longrightarrow RQ$ en considérant l'analyse formelle de concepts et la logique floue comme fondements mathématiques. Notre objectif est en fait d'évaluer cette implication par un opérateur de la connexion de Galois floue, qui peut faire office d'un schéma de correspondance entre une requête et un document.

Nous nous sommes inspirés de la proposition de Pasi [33] qui a repris la structure de l'évaluation du modèle booléen pondéré [40] dans un contexte logique en intégrant la possibilité d'utiliser un formalisme flou. Nous reprenons ces travaux à partir de la vision du modèle logique décrite dans [13].

Par ailleurs, de nombreux travaux [23, 43] ont traité la découverte de concepts formels et la génération de règles à partir du treillis de Galois [24]. Ces travaux ont eu des applications en recherche d'information [2, 3]. Le recours à une connexion de Galois dans le contexte de la RI est intéressant dans la mesure où il permet de :

- Considérer directement le treillis de concepts, qui demeure une représentation plus structurée de la relation *terme-document* lors de l'interrogation, de l'indexation et de la navigation dans le treillis.
- Exploiter la connaissance véhiculée par le concept qui peut être représentée sous la forme de règles associatives entre termes dans l'expansion de requêtes en RI.
- Exploiter l'ordre partiel qui est défini sur les concepts du treillis de Galois où il existe une relation duale entre les concepts et leurs extensions dans le treillis de concepts. Autrement dit, plus les intentions des concepts (i.e. l'ensemble des termes) s'élargissent au niveau du treillis et plus leurs extensions (i.e. l'ensemble des documents) respectives se réduisent. Cette propriété s'avère intéressante pour la RI car le résultat d'une requête peut être traduit par un concept réduit au niveau du treillis.

Dans cette section, nous allons donc proposer une nouvelle extension de la connexion de Galois floue. Notre objectif est d'aboutir à un nouveau schéma de correspondance basé sur un opérateur de cette connexion de Galois floue et assoupli par le quantificateur flou relatif "*la plupart*".

L'approche proposée est basée sur les hypothèses suivantes :

1. Nous considérons que les termes représentent les contraintes élémentaires au niveau de la requête et l'estimation de la pertinence d'un document est calculée avec "*la plupart*" des termes de la requête, en utilisant un quantificateur flou relatif "*la plupart*";
2. Nous nous basons sur les diverses sémantiques véhiculées par les degrés d'appartenance associés aux termes et par les implications floues ;
3. Le concept de pertinence d'un document d peut être modélisé comme un concept graduel lié au degré de vérité de l'implication floue $t^{RQ} \rightarrow t^d$ pour chaque terme t appartenant à la requête RQ .

6.2 Sémantiques des degrés associés aux termes de la requête

Il nous paraît important d'étudier de manière explicite, les différentes sémantiques possibles des degrés associés aux termes ainsi que des implications

floues et de discuter de leur intérêt dans le cadre d'un schéma de correspondance flou en RI.

En effet, l'interrogation de la relation *terme-document* floue, moyennant une requête floue, correspond bien au principe de la division relationnelle floue (DRF), où l'ensemble $\widetilde{RQ}$ est la requête floue à évaluer et $\widetilde{R}$ est la relation binaire floue *terme-document*. Ici, nous faisons référence aux travaux de Dubois et al. [19, 21] et des travaux de Bosc et al. [8], dans lesquels les auteurs discutent du choix de l'implication floue, en tenant compte des différentes sémantiques des degrés.

En d'autres termes, dans un contexte de RI, nous cherchons des documents de $\widetilde{R}$ qui sont représentés par tous les termes pondérés de la requête $\widetilde{RQ}$. Formellement, nous avons [19, 21] :

$$\mu_{\widetilde{Réponse}}(d) = \min_{t \in T} \mu_{\widetilde{RQ}}(t) \xrightarrow{I_{floue}} \mu_{\widetilde{R}}(t, d) \quad (6)$$

sachant que :

- $\widetilde{Réponse}$ est l'ensemble flou des documents satisfaisant la requête $\widetilde{RQ}$;
- $\xrightarrow{I_{floue}}$ dénote une implication floue ;
- $\mu_{\widetilde{RQ}}(t) \xrightarrow{I_{floue}} \mu_{\widetilde{R}}(d, t)$ désigne l'implication floue $t^{RQ} \xrightarrow{I_{floue}} t^d$ que nous désirons évaluer pour chaque terme t de la requête.

L'implication $\mu_{\widetilde{RQ}}(t) \xrightarrow{I_{floue}} \mu_{\widetilde{R}}(d, t)$ mesure le degré d'appartenance avec lequel un terme t de la requête $\widetilde{RQ}$ appartient au document d de la relation $\widetilde{R}$, sachant que t possède le niveau d'exigence $\mu_{\widetilde{RQ}}(t)$. Ce niveau d'exigence traduit le degré de représentativité du terme que l'utilisateur souhaiterait avoir dans les documents trouvés. La requête désigne en fait le prototype d'un document qui lui répond parfaitement. La pertinence d'un document par rapport à la requête est mesurée par $\mu_{\widetilde{Réponse}}(d)$, qui n'est autre que le minimum des degrés de vérité des implications floues $\mu_{\widetilde{RQ}}(t) \xrightarrow{I_{floue}} \mu_{\widetilde{R}}(d, t)$ pour chaque terme t de la requête $\widetilde{RQ}$.

Selon Dubois et al. [19, 21], le choix de l'implication floue est étroitement lié à la sémantique des degrés d'appartenance associés aux termes (en occurrence la sémantique assignée à $\mu_{\widetilde{R}}(t_i, d_j)$ et $\mu_{\widetilde{RQ}}(t_i)$). Les différentes sémantiques possibles ont été détaillées et discutées dans [19]. Nous retenons deux sémantiques dans le contexte de la RI :

1. *Sémantique de la satisfaction* : Les degrés définissent des seuils minimaux de représentativité relatifs aux termes de la requête, qui doivent être satisfaits au moment de la correspondance avec le document. Nous cherchons ainsi les documents suffisamment représentés par tous les termes de la requête.
2. *Sémantique de l'importance* : Les degrés traduisent des niveaux d'importance relatifs aux termes de la requête. La sémantique de l'importance exprime la priorité que l'utilisateur accorde à un terme au niveau global de la requête (c'est une importance relative). Dans la pratique, ceci veut dire que l'utilisateur exige que l'évaluation de la pertinence du document soit donnée par rapport aux termes ayant les degrés d'appartenance les plus forts dans la requête.

Cependant, dans le schéma de correspondance que nous proposons, nous considérons que seule la sémantique de type *satisfaction* peut être adoptée pour les degrés d'appartenance de la relation binaire floue *terme-document*, car ils décrivent le degré de représentativité de chaque terme dans chaque document de la relation floue *terme-document*. Par contre, les deux sémantiques de type *satisfaction* et *importance* peuvent être considérées pour les degrés des termes de la requête.

6.3 Intérêt de l'utilisation d'un quantificateur flou dans le contexte de la RI

L'expression (6) montre que la pertinence d'un document par rapport à la requête est mesurée par $\mu_{\widetilde{Réponse}}(d)$, qui n'est autre que le *minimum* des degrés de vérité des implications floues $\mu_{\widetilde{RQ}}(t) \xrightarrow{I_{floue}} \mu_{\widetilde{R}}(d, t)$ pour chaque terme t d'une requête $\widetilde{RQ}$. L'utilisation de la t-norme *MIN* de Zadeh [10] (voir annexe A, page 35), qui illustre la conjonction floue, est justifiée par le fait que l'utilisateur désire trouver les documents contenant “tous” les termes de la requête $\widetilde{RQ}$. Ceci paraît restrictif dans un contexte de RI.

Nous proposons de relaxer l'évaluation du degré de pertinence du document en utilisant le quantificateur relatif flou “la plupart” [45] avec l'idée de trouver les documents contenant “la plupart” des termes d'une requête $\widetilde{RQ}$. À ce sujet, signalons qu'une forme de relaxation des requêtes floues a été proposée par Bosc et al. [9] en utilisant les α -coupures sur la requête floue $\widetilde{RQ}$.

En effet, l'utilisation d'un quantificateur flou “la plupart” correspond à la situation où il existe quelques termes de la requête $\widetilde{RQ}$ dont les seuils de satisfaction ou d'importance exigés ne sont pas atteints pour certains documents de la relation $\widetilde{R}$. Nous modélisons la notion de “la plupart des termes” en associant au quantificateur flou relatif “la plupart”, un ensemble flou $I_{la\ plupart}$, défini par [19] :

$$\mu_{I_{la\ plupart}}(i) = 1 - \mu_{la\ plupart}(i - 1) \text{ pour } i = 1..m \text{ et } \mu_{la\ plupart}(0) = 0.0$$

Le sous-ensemble flou $I_{la\ plupart}$ définit un ordre de l'importance des m termes dans le quantificateur flou “la plupart”, selon la propriété suivante :

$$\mu_{I_{la\ plupart}}(1) = 1.0 \text{ et } \mu_{I_{la\ plupart}}(i) \geq \mu_{I_{la\ plupart}}(i + 1), \text{ pour } i = 1..(m - 1)$$

Ainsi, un terme ayant le degré d'appartenance le plus élevé dans la relation binaire floue $\widetilde{R}$, est considéré comme le terme le plus important selon la définition du quantificateur flou “la plupart”, i.e. $\mu_{I_{la\ plupart}}(1)$.

En supposant que la requête $\widetilde{RQ}$ contient m termes, la fonction σ définit une permutation au niveau des termes de $\widetilde{RQ}$ en respectant l'ordre suivant [19] :

$$\mu_{\widetilde{R}}(t_{\sigma(1)}, d) \geq \mu_{\widetilde{R}}(t_{\sigma(2)}, d) \geq \dots \geq \mu_{\widetilde{R}}(t_{\sigma(m)}, d)$$

En se basant sur la conjonction pondérée [15], la pertinence du document par rapport à la requête $\widetilde{RQ}$ est calculée comme suit [19] :

$$\mu_{\widetilde{R\acute{e}ponse}}(d) = \min_{i=1..m} (\max(\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{I_{floue}} \mu_{\widetilde{R}}(t_{\sigma(i)}, d), 1 - \mu_{I_{la\ plupart}}(i))), \forall t \in \mathcal{T}$$

L'utilisation de la fonction σ est justifiée par le fait que lors de l'évaluation de la pertinence d'un document, nous donnons d'abord la priorité au terme t de la requête $\widetilde{RQ}$ ayant le degré d'appartenance le plus élevé dans la relation $\widetilde{R}$. Ceci permet de trouver le degré de vérité de l'implication $\mu_{\widetilde{RQ}}(t) \xrightarrow{I_{floue}} \mu_{\widetilde{R}}(d, t)$ le plus élevé pour ce terme t . Plus le degré de vérité de l'implication diminue, plus le quantificateur flou “la plupart” intervient dans l'évaluation

de la pertinence d'un document. Autrement dit, pour le terme $t_{\sigma(1)}$, i.e. le terme ayant le degré d'appartenance le plus fort dans la relation, le degré $1 - \mu_{I_{la\ plupart}}(1)$ est égal à 0 et l'expression suivante [19] :

$$\max(\mu_{\widetilde{RQ}}(t_{\sigma(1)}) \xrightarrow{I_{f\ l\ o\ u\ e}} \mu_{\widetilde{R}}(t_{\sigma(1)}, d), 1 - \mu_{I_{la\ plupart}}(1))$$

sera réduite au degré de vérité de l'implication $\mu_{\widetilde{RQ}}(t_{\sigma(1)}) \xrightarrow{I_{f\ l\ o\ u\ e}} \mu_{\widetilde{R}}(t_{\sigma(1)}, d)$. Ceci veut dire que le quantificateur flou n'intervient pas dans ce cas de figure.

La particularité de l'utilisation du quantificateur flou "*la plupart*" réside dans le fait que ce quantificateur s'applique progressivement sur une requête de $m + 1$ termes (i.e. de 0 à m termes). Ainsi, pour le rang $i = 2$, un seul terme de la requête vérifie le quantificateur flou "*la plupart*", pour $i = 3$, deux termes de la requête vérifient le quantificateur "*la plupart*", et ainsi de suite jusqu'à arriver à $i = (m + 1)$, où m termes de la requête vérifient le quantificateur flou "*la plupart*".

Nous pouvons donc déduire que le degré de vérité de l'implication traitée dans le cadre de la RI $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{I_{f\ l\ o\ u\ e}} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$ est pondéré par $\mu_{I_{la\ plupart}}(i)$, de manière à ce que le degré $\mu_{I_{la\ plupart}}(i)$ soit aussi grand que le degré de l'implication $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{I_{f\ l\ o\ u\ e}} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$ est grand. Ceci explique l'utilisation de $1 - \mu_{I_{la\ plupart}}(i)$.

6.4 Schéma de correspondance en RI basé sur une extension de la connexion de Galois floue

Dans le cadre de notre travail, pour chaque terme de la requête, nous associons un degré d'appartenance, exprimant une sémantique de type *satisfaction* relative au seuil de représentativité exigé ou une sémantique de type *d'importance* relative à son importance dans la requête. Nous proposons une nouvelle extension de la connexion de Galois floue adaptée à la RI, où l'opérateur $\widetilde{h}$ de la connexion représente le nouveau schéma de correspondance flou *requête-document*, qui permet d'évaluer, en plus des documents pertinents, leurs degrés de pertinence respectifs.

Nous proposons d'utiliser le quantificateur relatif "*la plupart*" tel qu'il a été défini dans la section 6.3 (voir page 18).

En supposant que la requête $\widetilde{RQ}$ contient m termes et que l'ensemble $\widetilde{D}$ de documents contient l documents, la fonction γ définit une permutation

au niveau de $\widetilde{D}$ et la fonction σ définit une permutation au niveau de $\widetilde{RQ}$ comme suit :

1. Permutation des m termes qui figurent dans $\widetilde{RQ}$ en respectant l'ordre suivant :

$$\mu_{\widetilde{R}}(t_{\sigma(1)}, d) \geq \mu_{\widetilde{R}}(t_{\sigma(2)}, d) \geq \dots \geq \mu_{\widetilde{R}}(t_{\sigma(m)}, d);$$

2. Permutation des l documents de l'ensemble $\widetilde{D}$ en respectant l'ordre suivant :

$$\mu_{\widetilde{R}}(t, d_{\gamma(1)}) \geq \mu_{\widetilde{R}}(t, d_{\gamma(2)}) \geq \dots \geq \mu_{\widetilde{R}}(t, d_{\gamma(l)}).$$

Nous énonçons ainsi la nouvelle définition de la connexion de Galois pour la RI [31].

Définition 11

Soit $\widetilde{\mathcal{TF}} = (\mathcal{C}, \mathcal{T}, \widetilde{R})$ un contexte formel textuel flou. Pour un ensemble flou de documents $\widetilde{D}$ défini sur $\mathcal{C}$ et un ensemble flou de termes $\widetilde{RQ}$ défini sur $\mathcal{T}$, les opérateurs $\widetilde{f}$ et $\widetilde{h}$ de la connexion de Galois floue sont définis comme suit :

$$\begin{aligned} \widetilde{f}(\widetilde{D}) = \{t \mid \mu_{\widetilde{f}(\widetilde{D})}(t) = \min_j (\max(\mu_{\widetilde{D}}(d_{\gamma(j)}) \xrightarrow{I_{floue}} \mu_{\widetilde{R}}(t, d_{\gamma(j)}), \\ 1 - \mu_{I_{la\ plupart}}(j))), t \in \mathcal{T}\} \end{aligned}$$

$$\begin{aligned} \widetilde{h}(\widetilde{RQ}) = \{d \mid \mu_{\widetilde{h}(\widetilde{RQ})}(d) = \min_i (\max(\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{I_{floue}} \mu_{\widetilde{R}}(t_{\sigma(i)}, d), \\ 1 - \mu_{I_{la\ plupart}}(i))), d \in \mathcal{C}\} \quad (7) \end{aligned}$$

sachant que $\xrightarrow{I_{floue}}$ désigne une implication floue, vérifiant les propriétés 1 et 2 (voir page 7) et $I_{la\ plupart}$ représente le sous-ensemble flou, définissant le quantificateur flou “la plupart”.

Dans l'expression (7), chaque terme de la requête ($\widetilde{RQ}$) est traité individuellement afin de pouvoir calculer par le biais de l'implication floue $\xrightarrow{I_{floue}}$ et le quantificateur flou “la plupart”, le degré de vérification de sa représentativité ou de son importance dans le document d en question. Ainsi, la valeur

de δ , calculée par la fonction $\tilde{h}$ pour un document d représente le degré de pertinence de ce document par rapport à “la plupart” des termes pondérés de la requête floue $\widetilde{RQ}$. Ceci est obtenu en utilisant la conjonction pondérée [14].

6.4.1 Sémantique de la fermeture de la connexion de Galois floue en RI

Nous définissons dans cette section la nouvelle sémantique de la fermeture de la connexion de Galois floue, présentée dans la section précédente, i.e. $\tilde{f} \circ \tilde{h}$. Ainsi, en appliquant l’opérateur $\tilde{h}$ de la connexion de Galois floue sur un ensemble flou de termes $\widetilde{RQ}$ qui désigne la requête floue dans un contexte de RI, nous trouvons l’ensemble flou $\tilde{D}$ de documents qui sont pondérés par leurs degrés de pertinence respectifs. Ces documents contiennent “la plupart” des termes pondérés de la requête $\widetilde{RQ}$. De la même manière, si nous appliquons l’opérateur $\tilde{f}$ de la connexion de Galois floue sur un ensemble flou de documents $\tilde{D}$, nous trouvons l’ensemble flou $\tilde{T}$ de termes pondérés par leurs degrés d’appartenance respectifs à cet ensemble et qui annotent “la plupart” des documents de l’ensemble $\tilde{D}$. Ainsi la fermeture d’un ensemble flou de termes $\widetilde{RQ}$, i.e. $\tilde{f} \circ \tilde{h}(\widetilde{RQ})$, définit un *concept réduit flou* (voir définition 10, page 14).

En effet, le concept réduit flou correspond bien à un premier filtrage significatif, permettant de traiter plus précisément les associations existantes entre les termes de cet ensemble.

Dans le contexte de la RI, en considérant la matrice *terme-terme*, Bruandet constate dans [11, 12] qu’il est difficile d’interpréter le graphe associé à une telle matrice. Ainsi, afin d’extraire plus facilement des connaissances utiles à partir de cette matrice *terme-terme*, Bruandet propose de construire, à partir du graphe associé, les sous-graphes maximaux complets appelés *cliques* [7]. Ces cliques fournissent ainsi une représentation de la matrice *terme-terme* sans perte d’informations. À titre d’exemple, une clique peut constituer un bon résumé d’une phrase. Par rapport à la matrice initiale, les cliques apportent plus d’informations quant aux liaisons entre les termes. Chaque clique donne un ensemble de termes complètement connectés entre eux au sens d’une certaine mesure d’association entre deux termes définie dans [12].

Nous constatons que, par rapport aux *cliques* [11, 12] utilisées pour l’extraction de concepts⁴ à partir de texte, les concepts réduits de termes, qu’ils

⁴Ici, ce terme ne désigne pas un concept réduit, mais plutôt le concept traité par les termes

soient classiques ou flous, sont plus riches en information et illustrent formellement les liens qui existent entre les termes et les documents dans une relation binaire *terme-document*. Cet avantage provient du fait que l'extraction de concepts réduits se fait sur la relation binaire *terme-document*, alors que la construction de cliques se fait à partir du graphe associé à la relation *terme-terme* [12].

Pour conclure, nous pouvons dire que l'extraction de cliques, tout comme celle de concepts réduits, permet d'accéder à des connaissances plus aisément interprétables et correspondant mieux à l'expression d'une connaissance additionnelle. Les cliques nous apportent plus d'informations quant aux liaisons entre les termes alors que les concepts réduits décrivent les liaisons étroites existantes entre les termes et les documents.

6.4.2 Propriétés de la connexion de Galois floue

Proposition 2

Soit la connexion de Galois floue $(\tilde{f}, \tilde{h})$ (voir définition 11, page 21). Les propriétés énoncées dans la proposition 1 (voir page 14) sont vérifiées pour tout ensemble flou de documents $\tilde{D}$ défini sur $\mathcal{C}$ et tout ensemble flou de termes $\tilde{T}$ défini sur $\mathcal{T}$, en considérant le quantificateur flou relatif "la plupart" et une implication floue I_{floue} de type *R-implication* ou *S-implication* et vérifiant les propriétés 1 et 2 (voir page 7).

La propriété **(Q1)** (voir page 14) peut être interprétée en RI comme suit : en considérant une requête floue initiale, le fait d'étendre cette requête par d'autres termes permettrait de la rendre plus précise et de cibler la recherche, en trouvant un nombre plus réduit de documents. L'objectif serait ainsi de minimiser le bruit lors de l'interrogation, tout en conservant un bon taux de rappel.

Les preuves des différentes propriétés de la connexion de Galois floue définie pour la RI sont données dans l'annexe B.

6.5 Traitement sémantique d'une requête

Une requête traduit un besoin en information exprimé par l'utilisateur. Ce besoin peut être défini en mettant l'accent soit sur le niveau de satisfaction

du texte.

exigé, soit sur l'importance relative de ce besoin ou encore sur les deux à la fois.

En considérant l'opérateur $\tilde{h}$ de la connexion de Galois floue comme la fonction d'appariement *requête-document*, nous allons discuter les trois cas possibles relatifs au choix de l'implication floue, utilisée dans la nouvelle connexion de Galois floue définie pour la RI. Notons que le résultat implicite de l'interrogation dans chaque cas, dépend fortement de la sémantique donnée à $\mu_{\widetilde{RQ}}(t)$ et $\mu_{\widetilde{R}}(t, d)$ [19, 21].

Cas 1) Les degrés attribués aux termes de la requête expriment une sémantique de type *satisfaction*

Dans ce cas, pour chaque terme de la requête, l'utilisateur attribue un niveau de représentativité minimal, qu'il désire trouver dans les documents, i.e. $\mu_{\widetilde{RQ}}(t_{\sigma(i)})$. Notons que tous les termes de la requête sont considérés à importance égale, i.e. qu'aucune priorité entre les termes n'est fixée, et tous les termes ont un niveau d'importance maximal égal à 1.0. La requête est ainsi représentée par un ensemble flou $\widetilde{RQ}_S$ (S pour satisfaction), qui modélise la représentativité du terme exigée dans le document. L'idée est de trouver les documents qui contiennent "*la plupart*" des termes de la requête $\widetilde{RQ}_S$. Formellement, le degré de pertinence d'un document d dans le cadre de cette sémantique est défini comme suit :

$$\mu_{\widetilde{Réponse}}(d) = \min_{i=1..m} (\max(\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{R-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d), 1 - \mu_{I_{la\ plupart}}(i))) \quad (8)$$

Pour évaluer le degré de vérité de cette implication, Dubois et al. [14] proposent d'utiliser une R-implication. L'expression (8) exprime avec quel degré de représentativité un terme d'une requête $\widetilde{RQ}_S$ doit figurer dans la relation floue *terme-document* $\widetilde{R}$. Ainsi, il est naturel de considérer que le document satisfait complètement la requête si le degré d'appartenance exigé pour un terme $t_{\sigma(i)}$, i.e. $\mu_{\widetilde{RQ}}(t_{\sigma(i)})$, est inférieur ou égal au degré d'appartenance existant dans la relation $\mu_{\widetilde{R}}(t_{\sigma(i)}, d_j)$. Ceci veut dire que le degré exigé est largement atteint. Autrement dit, ce document appartient à $\tilde{h}(\widetilde{RQ})$ avec un niveau de pertinence égal à 1. Formellement, l'implication floue $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{R-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$ est évaluée dans ce cas, comme suit :

$$(\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{R-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d) = 1) \iff \mu_{\widetilde{RQ}}(t_{\sigma(i)}) \leq \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$$

Évaluation de l'implication	Commentaires
$\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{R-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$ Utilisation de l'implication de Gödel : $(\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{G\ddot{o}} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)) = \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$	Préserver, pour le degré de vérité de l'implication le degré d'appartenance du terme en question dans la relation floue.
Utilisation de l'implication de Goguen : $(\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{G} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)) = \frac{\mu_{\widetilde{R}}(t_{\sigma(i)}, d)}{\mu_{\widetilde{RQ}}(t_{\sigma(i)})}$	Garder un degré de pertinence relatif, qui est le degré d'appartenance du terme dans la relation divisé par le degré de représentativité exigé dans la requête.
Utilisation de l'implication de Lukasiewicz : $(\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{Luka} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)) = 1 - \mu_{\widetilde{RQ}}(t_{\sigma(i)}) + \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$	Basée sur l'écart entre le degré de représentativité exigé et le degré d'appartenance existant dans la relation.

TAB. 3 – Propositions avec les R-implications floues.

Les R-implications les plus communes telles que celles de Gödel, Goguen et Lukasiewicz (voir tableau 2, page 7) vérifient cette propriété. Dans le cas contraire, où le document d ne satisfait pas complètement la requête, i.e. $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) > \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$, le terme $t_{\sigma(i)}$ appartient au document d dans la relation floue *terme-document*, avec un degré d'appartenance inférieur au degré exigé dans la requête. Le degré de satisfaction exigé n'est pas donc atteint. Dans ce cas, l'implication $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{R-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$ est évaluée en utilisant les R-implications de Gödel, de Goguen et de Lukasiewicz. Le tableau 3 résume les différentes propositions avec les R-implications floues.

Cas 2) Les degrés attribués aux termes de la requête expriment une sémantique de type *importance*

Dans ce cas, à chaque terme de la requête est associé un degré d'importance, i.e. son degré d'appartenance à la requête floue $\mu_{\widetilde{RQ}}(t_{\sigma(i)})$, qui n'est pas toujours égal à 1. Par contre, tous les degrés de satisfaction qui expriment les niveaux de représentativité respectifs des termes sont fixés à 1. La requête est ainsi représentée par un ensemble flou $\widetilde{RQ}_I$ (I pour importance), qui modélise l'importance d'un terme dans la requête par rapport aux autres termes. Nous utilisons le quantificateur flou relatif “*la plupart*” afin

de trouver les documents qui contiennent “la plupart” des termes importants dans la requête $\widetilde{RQ}_I$. Formellement, le degré de pertinence d’un document dans le contexte de cette sémantique est défini comme suit :

$$\mu_{\widetilde{Réponse}}(d) = \min_{i=1..m} (\max(\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{S-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d), 1 - \mu_{I_{la\ plupart}}(i))) \quad (9)$$

Pour évaluer le degré de vérité de l’implication $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{S-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$ pour un terme donné de la requête $\widetilde{RQ}_I$, Dubois et al. [14] proposent d’utiliser une S-implication. L’expression (9) exprime avec quel degré d’appartenance un terme important de la requête $\widetilde{RQ}_I$ doit figurer dans la relation floue *terme-document* $\widetilde{R}$. Nous supposons que $\widetilde{RQ}_I$ est un ensemble flou normalisé (i.e. il existe au moins un terme de la requête ayant un degré d’importance maximal égal à 1). Ainsi, il est naturel de considérer dans le cadre de la sémantique *importance* que le document satisfait complètement la requête si tous les termes de cette requête, quels que soient leurs niveaux d’importance, sont inclus dans la relation *terme-document* avec un degré de représentativité maximal, i.e. égal à 1. Formellement, nous avons :

$$\mu_{\widetilde{Réponse}}(d) = 1 \iff (\forall t \in \mathcal{T}, \mu_{\widetilde{RQ}_I}(t_{\sigma(i)}) > 0 \Rightarrow \mu_{\widetilde{R}}(t_{\sigma(i)}, d) = 1) \quad (10)$$

Lorsque $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) = 1$ (i.e. le terme a un degré d’importance maximal), le degré de vérité de l’implication $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{S-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$, pour un terme donné $t_{\sigma(i)}$ de la requête $\widetilde{RQ}_I$, est alors égal au degré d’appartenance du terme existant dans la relation, i.e. $\mu_{\widetilde{R}}(t_{\sigma(i)}, d)$.

Le degré de vérité de l’implication $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{S-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d_j)$ n’est nul, que lorsque le degré d’importance est maximal, i.e. $\mu_{\widetilde{RQ}_I}(t_{\sigma(i)}) = 1$, et que le degré d’appartenance du terme dans la relation est nul (i.e. $\mu_{\widetilde{R}}(t_{\sigma(i)}, d_j) = 0$). Formellement, nous avons :

$$(\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{S-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d) = 0) \iff \mu_{\widetilde{RQ}_I}(t_{\sigma(i)}) = 1 \text{ et } \mu_{\widetilde{R}}(t_{\sigma(i)}, d) = 0 \quad (11)$$

Ce type de comportement nécessite l’utilisation des S-implications comme celles de Kleene-Dienes et Reichenbach [19] qui vérifient les deux expressions (10) et (11).

Évaluation de l'implication	Commentaires
$\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{S-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d_j)$	
Utilisation de l'implication de Kleene-Dienes : $(\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{KD} \mu_{\widetilde{R}}(t_{\sigma(i)}, d_j))$ $= \max(1 - \mu_{\widetilde{RQ}_I}(t_{\sigma(i)}), \mu_{\widetilde{R}}(t_{\sigma(i)}, d_j))$	Le terme $t_{\sigma(i)}$ de la requête peut être ignoré jusqu'à un certain degré.
$(\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{KD} \mu_{\widetilde{R}}(t_{\sigma(i)}, d_j)) = 1$ $\forall$ la valeur de $\mu_{\widetilde{R}}(t_{\sigma(i)}, d_j)$.	Le terme est complètement ignoré, i.e. $\mu_{\widetilde{RQ}_I}(t_{\sigma(i)}) = 0$.
Utilisation de l'implication de Reichenbach : $(\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{RB} \mu_{\widetilde{R}}(t_{\sigma(i)}, d_j))$ $= 1 - \mu_{\widetilde{RQ}_I}(t_{\sigma(i)}) + \mu_{\widetilde{RQ}_I}(t_{\sigma(i)}) \times \mu_{\widetilde{R}}(t_{\sigma(i)}, d_j)$	L'importance d'un terme de la requête peut être réduite jusqu'à un certain degré (qui est le degré $\mu_{\widetilde{RQ}_I}(t_{\sigma(i)})$).

TAB. 4 – Propositions avec les S-implications floues.

Le cas où $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) < 1$ représente le fait que le terme $t_{\sigma(i)}$ n'est pas très important dans la requête, et il n'est pas considéré comme prioritaire dans la recherche. Si de plus $\mu_{\widetilde{R}}(t_{\sigma(i)}, d) = 0$, alors le degré de vérité de l'implication $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{S-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$ doit être strictement positif. Le tableau 4 résume les différentes propositions avec les S-implications [19].

Nous remarquons que le degré de vérité de l'implication $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{KD} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$, donné par l'utilisation de l'implication de Kleene-Dienes (voir tableau 2), s'approche de 1 quand $\mu_{\widetilde{RQ}_I}(t_{\sigma(i)})$ est faible (sans tenir compte de la valeur de $\mu_{\widetilde{R}}(t_{\sigma(i)}, d)$). De même, ce degré de vérité est égal à $\mu_{\widetilde{R}}(t_{\sigma(i)}, d)$ quand le degré d'importance du terme $\mu_{\widetilde{RQ}_I}(t_{\sigma(i)})$ est élevé (sans tenir compte de la valeur de $\mu_{\widetilde{RQ}_I}(t_{\sigma(i)})$).

Notons également que le degré de vérité de l'implication $\mu_{\widetilde{RQ}}(t_{\sigma(i)}) \xrightarrow{RB} \mu_{\widetilde{R}}(t_{\sigma(i)}, d)$, obtenu par l'implication de Reichenbach (voir tableau 2, page 7), croît d'une manière continue de la valeur de $\mu_{\widetilde{R}}(t_{\sigma(i)}, d)$ vers 1, quand $\mu_{\widetilde{RQ}_I}(t_{\sigma(i)})$ décroît.

Cas 3) Combinaison des deux sémantiques au niveau de la requête

Nous considérons dans ce qui suit le cas le plus général où les degrés d'importance exprimés au niveau de la requête sont liés aux degrés de représen-

tativité exigés pour les termes, dans la relation *terme-document*. La combinaison des deux sémantiques illustre une vision plus globale de la requête. Dans ce cas de figure, le degré d'importance exprime, par exemple, qu'il est plus important d'avoir dans le document un terme (avec un degré de représentativité spécifié) que d'avoir un autre terme avec son propre degré d'appartenance dans la relation floue *terme-document*. En utilisant le quantificateur flou “la plupart”, la pertinence d'un document est définie comme suit :

$$\mu_{\widetilde{Réponse}}(d) = \min_{i=1..m} (\max(\mu_{\widetilde{RQ}_I}(t_{\sigma(i)}) \xrightarrow{S-imp} (\mu_{\widetilde{RQ}_S}(t_{\sigma(i)}) \xrightarrow{R-imp} \mu_{\widetilde{R}}(t_{\sigma(i)}, d))), 1 - \mu_{I_{la\ plupart}}(i)) \quad (12)$$

L'expression (12) estime avec quel degré l'ensemble flou $\widetilde{RQ}_I$ des termes importants est inclus dans l'ensemble flou $(\widetilde{RQ}_S \xrightarrow{R-imp} \widetilde{R})$ des termes qui appartiennent au document, en satisfaisant le degré de représentativité exigé au niveau de $\widetilde{RQ}_S$. Remarquons que lorsque $\forall t \in \mathcal{T}, \mu_{\widetilde{RQ}_I}(t_{\sigma(i)}) = 1$ ou $\forall t \in \mathcal{T}, \mu_{\widetilde{RQ}_S}(t_{\sigma(i)}) = 1$, l'expression (12) se réduit respectivement aux deux expressions (8) et (9).

Pour évaluer la pertinence d'un document dans ce cas, il serait judicieux d'utiliser deux implications, i.e. une R-implication et une S-implication, vérifiant les propriétés 1 et 2 (voir page 7) et qui donnent les meilleures valeurs de pertinence, respectivement, pour les deux types de sémantique, i.e. satisfaction et importance,

Exemple 2

Nous considérons la relation floue *terme-document* illustrée par le tableau 1 (voir page 3) et les deux ensembles flous $\widetilde{RQ}_S$ et $\widetilde{RQ}_I$ définis sur $\mathcal{T}$ et qui expriment les besoins de l'utilisateur. Ceux-ci sont définis en considérant les deux types de sémantique, i.e. satisfaction et importance. La requête $\widetilde{RQ}_S$ définit le niveau de représentativité minimal exigé pour chaque terme de la requête alors que $\widetilde{RQ}_I$ décrit l'importance relative de chaque terme dans la requête. Nous supposons que $\widetilde{RQ}_S = \{t_1, t_2, t_3\}$ et $\widetilde{RQ}_I = \{t_1, t_2, t_3\}$.

Dans cet exemple, le quantificateur relatif “la plupart” possède une cardinalité maximale égale à 3 termes et il est défini comme suit :

$$\mu_{I_{la\ plupart}}(0) = 0.0, \mu_{I_{la\ plupart}}(1) = 0.50, \mu_{I_{la\ plupart}}(2) = 0.75 \text{ et } \mu_{I_{la\ plupart}}(3) = 1.00.$$

Nous lui associons un sous-ensemble flou $I_{la\ plupart}$ défini par :

$\mu_{I_{la\ plupart}}(1) = 1.00$, $\mu_{I_{la\ plupart}}(2) = 0.50$, $\mu_{I_{la\ plupart}}(3) = 0.25$ et $\mu_{I_{la\ plupart}}(4) = 0.00$

Nous proposons d'utiliser l'implication floue de Lukasiewicz (voir tableau 2, page 7), qui est à la fois une *R-implication* et une *S-implication* et qui vérifie les deux propriétés 1 et 2 (voir page 7). Nous allons discuter les résultats de la fonction de correspondance $\tilde{h}(\widetilde{RQ})$ en utilisant les différentes versions illustrées par le tableau 5.

Sémantique	Requête	Version	Commentaires
Satisfaction	$\widetilde{RQ}_S$	Version 1	Utilisation de la t-norme MIN de Zadeh et l'implication de Lukasiewicz.
	$\widetilde{RQ}_S$	Version 2	Utilisation du quantificateur flou "la plupart" et l'implication de Lukasiewicz.
Importance	$\widetilde{RQ}_I$	Version 3	Utilisation de la t-norme MIN de Zadeh et l'implication de Lukasiewicz.
	$\widetilde{RQ}_I$	Version 4	Utilisation du quantificateur flou "la plupart" et l'implication de Lukasiewicz.
Combinaison des deux sémantiques	$\widetilde{RQ}_S$ et $\widetilde{RQ}_I$	Version 5	Utilisation de la t-norme MIN de Zadeh, et l'implication de Lukasiewicz pour les deux types de sémantique.
	$\widetilde{RQ}_S$ et $\widetilde{RQ}_I$	Version 6	Utilisation du quantificateur flou "la plupart" et l'implication de Lukasiewicz pour les deux types de sémantique.

TAB. 5 – Versions retenues.

À la lecture du tableau 6, nous remarquons que le document d_3 est pertinent avec un degré égal à 1, et ce, quelle que soit la version retenue relativement à une sémantique de type satisfaction (i.e. version 1 et 2). Ceci est expliqué par le fait que les degrés de représentativité exigés pour tous les termes de la requête $\widetilde{RQ}_S$ sont largement atteints au niveau de la relation floue terme-document $\widetilde{R}$ (voir tableau 1, page 3). Nous constatons également que les degrés de pertinence des documents d_2 et d_4 sont nettement meilleurs, pour les deux types de sémantique, dans le cas où nous utilisons le quantificateur flou "la plupart" (i.e. version 2 et 4). Ceci est justifié par la relaxation apportée par ce quantificateur.

Dans le cas d'une sémantique de type importance, l'utilisateur se montre plus intéressé par des documents contenant le terme t_1 , ensuite par les autres termes de la requête. Nous remarquons que le document d_1 est pertinent avec un degré égal à 0.90 (version 3). Ceci s'explique par le fait qu'il soit

	Degré de pertinence du document			
	d_1	d_2	d_3	d_4
Version 1	0.90	0.65	1.00	0.46
Version 2	0.90	0.75	1.00	0.75
Version 3	0.90	0.65	0.66	0.46
Version 4	0.90	0.70	0.75	0.75
Version 5	1.00	0.85	1.00	0.66
Version 6	1.00	0.85	1.00	0.75

TAB. 6 – Degrés de pertinence des documents relatifs aux différentes versions.

représenté d'une manière assez forte par les trois termes de la requête $\widetilde{RQ}_I$. Soulignons que la pertinence du document d_4 est affaiblie par le fait que le terme t_2 , qui apparaît comme un terme relativement important dans la requête $\widetilde{RQ}_I$, appartient faiblement à ce document. Cependant, en utilisant le quantificateur flou “la plupart”, sa pertinence s'améliore nettement puisqu'il est représenté par les deux autres termes qui sont moins importants.

Pour la combinaison des deux sémantiques, nous obtenons les valeurs suivantes : $\{ \overset{1.00}{d_1}, \overset{0.85}{d_2}, \overset{1.00}{d_3}, \overset{0.66}{d_4} \}$ en considérant la cinquième version. Les documents d_1 et d_3 apparaissent comme les documents les plus pertinents pour la requête. Nous remarquons que l'utilisation du quantificateur flou “la plupart” améliore toutes les valeurs de pertinence, notamment pour le document d_4 .

7 CONCLUSION

Nous avons proposé dans cet article une nouvelle interprétation de l'implication logique en recherche d'information $d \longrightarrow RQ$ dans un contexte flou. L'intérêt de cette approche réside dans un nouveau schéma de correspondance qui est basé sur une extension de la connexion de Galois floue. En intégrant les différentes sémantiques que peut avoir les degrés attribués aux termes de la requête et un quantificateur flou, dans la nouvelle définition de la connexion de Galois floue, nous avons discuté du choix de l'implication floue afin de sélectionner celle qui est la plus appropriée au domaine de la RI. Nous avons également montré que les propriétés de la connexion restent valides sur cette extension. Nous pensons que l'exploration de ce point de vue doit se poursuivre pour l'intégrer dans la modélisation d'autres éléments du

domaine de la RI, comme le degré de certitude d'un terme d'indexation. De manière pratique, cette étude théorique permet d'envisager l'expérimentation d'un moteur de recherche, qui ne sera plus basé sur un calcul de distance entre vecteurs de termes, mais plutôt fondé sur un algorithme d'extraction de concepts formels flous, par fermeture de la connexion de Galois floue telle qu'elle est définie dans cet article. Ce travail peut être également approfondi dans deux autres directions à savoir : l'étude dans le même contexte flou de l'implication logique réciproque en RI $RQ \longrightarrow D$ et l'utilisation de la connexion de Galois floue pour l'extraction de règles associatives floue entre termes, comme nous l'avons déjà proposée dans [27].

RÉFÉRENCES

- [1] R. Baeza-Yates et B. Ribeiro-Neto. *Modern Information Retrieval*. ACM press, 1999.
- [2] N. Belkhiter, C. Bourhfir, M.M. Gammoudi, A. Jaoua, N. Le Thanh et M. Reguig. Décomposition Rectangulaire Optimale d'une Relation Binaire : Application aux bases de Données Documentaires. *INformation Systems and Operational Research Journal*, 1 :33–54, 1994.
- [3] N. Belkhiter, J. Desharnais, A. Jaoua et T. Moukam. Providing Relevant Additional Information to Users Asking Queries Using a Galois Lattice Structure. In *Proceedings of the IEEE 8th International Symposium on Computer and Information Sciences, ISCIS VIII, Istanbul, Turkey*, pages 594–604, 3–5 November 1993.
- [4] R. Bellman, L. Kalaba et L.A. Zadeh. Abstraction and pattern Classification. *J. of Mathematical Analysis and Applications*, (13) :1–7, 1966.
- [5] R. Belohlavek. Fuzzy Galois connections. *Math. Logic*, 45(4) :497–504, 1999.
- [6] S. Benyahia et A. Jaoua. Discovering knowledge from fuzzy concept lattice. In *Data mining and Computational Intelligence, Studies in Fuzziness and Soft Computing*, Vol. 68, chapitre 7, pages 167–190. Physica Verlag, Heidelberg, March 2001.
- [7] C. Berge. *Théorie des graphes et ses applications*. Edition Dunod, Paris, 1967.

- [8] P. Bosc, L. Liétard et O. Pivert. Bases de données et flexibilité : les requêtes graduelles. *Techniques et Sciences Informatiques*, 17(3) :355–378, 1998.
- [9] P. Bosc, L. Liétard et H. Prade. On fuzzy queries involving fuzzy quantifiers. In *Proceedings of the ECAI Workshop on Uncertainty in Information Systems. Budapest, Hungary*, pages 6–10, August 1996.
- [10] B. Bouchon-Meunier. *La logique floue et ses applications*. Edition Addison Wesley, 1995.
- [11] M.F. Bruandet. Outline of a knowledge-base model for an intelligent information retrieval system. *Information Processing & Management*, 25(1) :89–115, 1989.
- [12] M.F. Bruandet. Construction automatique d’une base de connaissances du domaine dans un système de recherche d’informations. Habilitation à Diriger des Recherches (HDR), Université Joseph Fourier, Grenoble I, 1990.
- [13] Y. Chiamarella et J.P. Chevallet. About Retrieval Models and Logic. *The Computer Journal*, 35(3) :233–242, 1992.
- [14] D. Dubois, M. Nakata et H. Prade. Extended division for flexible queries in relational databases. In O. Pons M. A. Villa et J. Kacprzyk, éditeurs, *Knowledge management in fuzzy databases*, pages 105–121. Physica Verlag, Heidelberg, 1999.
- [15] D. Dubois et H. Prade. Weighted minimum and maximum operations in fuzzy set theory. *Information Sciences*, (39) :205–210, 1986.
- [16] D. Dubois et H. Prade. Fuzzy sets in approximate reasoning, Part2 : Logical approaches. *Fuzzy Sets and Systems Journal*, 40 :203–244, 1991.
- [17] D. Dubois et H. Prade. Bases de règles floues en commande : Une discussion critique. Technical report irit/59-48-r, IRIT/Institut de Recherche en Informatique de Toulouse, France, 1995.
- [18] D. Dubois et H. Prade. Logique floue, interpolation et commande. *Journal Européen des systèmes Automatisés*, 30 :607–644, 1996.
- [19] D. Dubois et H. Prade. Semantics of quotient operators in fuzzy relational databases. *Fuzzy sets and systems*, 78 :89–93, 1996.
- [20] D. Dubois et H. Prade. What are fuzzy rules and how to use them ? *Fuzzy Sets and Systems*, (84) :169–185, 1996.

- [21] D. Dubois et H. Prade. The three semantics of fuzzy sets. *Fuzzy Sets and Systems*, (90) :141–150, 1997.
- [22] S. Elloumi. Apprentissage Supervisé par Localisation de Concepts dans les Contextes Flous ou Réels. Thèse de Doctorat en Informatique, Faculté des Sciences de Tunis, Département Informatique, Université d'El Manar, Juillet 2002.
- [23] B. Ganter et R. Wille. *Formal Concept Analysis*. Edition Springer-Verlag, Heidelberg, 1999.
- [24] R. Godin et R. Missaoui. An incremental concept formation approach for learning from databases. *Theoretical Computer Science*, (133) :387–419, 1994.
- [25] A. Jaoua, S. Elloumi, S. BenYahia et F. Alvi. Galois connection in fuzzy binary relations : applications for discovering association rules and decision making. Methodos Publisher, 2002.
- [26] D. Kraft et D. Buel. Fuzzy sets and generalized boolean retrieval systems. *International Journal Man-Machine Studies*, pages 45–56, 1983.
- [27] C. Cherif Latiri, S. Benyahia et A. Jaoua. Query expansion using fuzzy association rules. In *Actes des Journées Informatiques de Metz, Knowledge Discovery and Discrete Mathematics*, September 2003.
- [28] C. Cherif Latiri, S. Benyahia et G. Mineau. Conceptual non-redundant association rules discovery : Application to query expansion. In *Proceedings of the First International Conference on Formal Concept Analysis : The State of the Art (ICFCA03)*, Damstadt, Allemagne, February-March 2003.
- [29] C. Cherif Latiri, S. Benyahia, G. Mineau et A. Jaoua. Découverte des règles associatives non redondantes : Application aux corpus textuels. In *Actes des Journées francophones EGC'2003, Lyon. Publié dans la revue d'Intelligence Artificielle, Vol 17, No. 1, 2003*, pages 131–144, Janvier 2003.
- [30] C. Cherif Latiri, J.P. Chevallet et S. Elloumi. Fuzzy Galois Connection based Information Retrieval Model. In *Proceedings of the Workshop on Information Systems, Villa Monastero, Varenna, Italy*, September 2002.
- [31] C. Cherif Latiri, J.P. Chevallet, S. Elloumi et A. Jaoua. Extension of fuzzy galois connection for information retrieval using a fuzzy quantifier. In *Proceedings of the ACS/IEEE International Conference on*

Computer Systems and Applications, AICCSA'03, Tunis, Tunisia, July 2003.

- [32] J. Nie. Un modèle logique général pour les Systèmes de Recherche d'Informations. Application au prototype RIMRE. Thèse de Doctorat en Informatique, Laboratoire de Génie Informatique - IMAG, Université Joseph Fourier, Grenoble I, Juillet 1990.
- [33] G. Pasi. A logical formulation of the boolean model and of weighted boolean model. In *Proceedings of the Workshop on Logical and Uncertainty Models for Information Systems, London, UK*, pages 1–11, 1999.
- [34] S. Pollard. *Fuzzy-Concepts. Formal Concept Analysis for imprecise data*. Edition Springer Verlag, Berlin, 1996.
- [35] T. Radecki. Fuzzy Set Theoretical Approach to Document Retrieval. *Information Processing and Management*, 15 :247–259, 1979.
- [36] C.J. Van Rijsbergen. A non-classical logic for information retrieval. *The Computer Journal*, 29(6) :481–485, 1986.
- [37] C.J. Van Rijsbergen. A new theoretical framework for information retrieval. In *Proceedings of the 1986-ACM Conference on Research and Development in Information Retrieval*, pages 194–200, 1986.
- [38] G. Salton et C. Buckley. Term weighting approaches in automatic text retrieval. In *Information Processing and Management*, pages 513–523, May 1988.
- [39] G. Salton, C. Buckley et C.T. Yu. An Evaluation of Term Dependence Models in Information Retrieval. *Research and Development in Information Retrieval*, pages 151–173, 1982.
- [40] G. Salton, E. Fox et H. WU. Extended Boolean Information Retrieval. *Communications of the ACM*, 26(11) :1022–1036, November 1983.
- [41] F. Sebastiani. On the role of logic in Information Retrieval. *Information Processing and Management*, 38(1) :1–18, 1998.
- [42] E.H. Viedma. An information retrieval system with ordinal linguistic queries based on the weighting semantics. In *Proceedings of the 7th International Conference on Information Processing and Management of Uncertainty in knowledge based System, Madrid, Spain, 2000*.
- [43] R. Wille. *Knowledge acquisition by methods of formal concept analysis*. Edition Nova Science, New York, 1989.

- [44] K. E. Wolff. Concepts in fuzzy scaling theory : Order and granularity. *Fuzzy Sets and Systems*, 132(1) :63–75, November 2002.
- [45] L. A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. *Computing and Mathematics with Applications*, (9) :149–184, 1983.
- [46] L.A. Zadeh. Fuzzy sets. *Information and Control*, (69) :338–353, June 1965.

A NORMES ET CONORMES TRIANGULAIRES

Les opérations d’intersection, d’union et de complémentation de sous-ensembles flous, habituellement employées, peuvent être remplacées par des opérations construites à l’aide d’opérateurs différents du minimum, du maximum et de la complémentation à 1 [10]. Ces opérateurs sont souvent utilisés lorsque les opérations habituelles ne s’avèrent pas satisfaisantes, par exemple en logique floue ou dans la commande floue [17, 18]. Ils sont définis à partir d’une *norme triangulaire* et d’une *conorme triangulaire* définies comme suit [10] :

Définition 12

Une *norme triangulaire* “t-norme” est une fonction $\top : [0, 1] \times [0, 1] \mapsto [0, 1]$ qui vérifie, pour tout x, y, z dans $[0, 1]$, les propriétés suivantes :

- $\top(x, y) = \top(y, x)$
- $\top(x, \top(y, z)) = \top(\top(x, y), z)$
- $\top(x, y) \leq \top(z, t)$ si $x \leq z$ et $y \leq t$
- $\top(x, 1) = x$

L’opérateur “*min*” vérifie toutes les propriétés de la définition d’une t-norme. À toute t-norme, nous pouvons associer un opérateur d’intersection entre deux ensembles flous $\tilde{A}$ et $\tilde{B}$ d’un univers de discours U , défini par :

$$\mu_{\tilde{A} \cap_{\top} \tilde{B}}(x) = \top(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)). \quad (13)$$

Définition 13

Une *conorme triangulaire* “t-conorme” est une fonction $\perp : [0, 1] \times [0, 1] \mapsto [0, 1]$, qui vérifie, pour tout x, y, z dans $[0, 1]$, les propriétés suivantes :

- $\perp(x, y) = \perp(y, x)$
- $\perp(x, \perp(y, z)) = \perp(\perp(x, y), z)$

t-norme	t-conorme	nom
$\min(x,y)$	$\max(x,y)$	Zadeh
$\max(x + y - 1, 0)$	$\min(x + y, 1)$	Lukasiewicz
$\begin{cases} x \text{ si } x = 1 \\ y \text{ si } y = 1 \\ 0 \text{ sinon} \end{cases}$	$\begin{cases} x \text{ si } y = 0 \\ y \text{ si } x = 0 \\ 1 \text{ sinon} \end{cases}$	Weber

TAB. 7 – Normes et t-conormes duales

- $\perp(x, y) \leq \perp(z, t)$ si $x \leq z$ et $y \leq t$
- $\perp(x, 0) = x$

L'opérateur "max" vérifie les propriétés de la définition d'une t-conorme. À toute t-conorme, nous pouvons associer un opérateur d'union entre deux ensembles flous $\tilde{A}$ et $\tilde{B}$ d'un univers de discours U , défini par :

$$\mu_{\tilde{A} \cup \perp \tilde{B}}(x) = \perp(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)). \quad (14)$$

Les t-normes et les t-conormes les plus utilisées sont indiquées dans le tableau A.1. Elles sont ordonnées de la manière suivante :

$$\forall x, y \in [0, 1] \quad \top_{Weber} \leq \top(x, y) \leq \min(x, y) \quad (15)$$

$$\max(x, y) \leq \perp(x, y) \leq \perp_{Weber}(x, y) \quad (16)$$

B LES PREUVES DES PROPRIÉTÉS DE LA CONNEXION DE GALOIS FLOUE DÉFINIE POUR LA RI

Nous prouvons les propriétés de la connexion de Galois floue définie pour la RI. Ces propriétés restent valides avec n'importe quelle implication floue I_{floue} , vérifiant les deux propriétés 1 et 2 (voir page 7).

$$\text{Propriété (P1)} : \tilde{D}_1 \subseteq \tilde{D}_2 \Rightarrow \tilde{f}(\tilde{D}_1) \supseteq \tilde{f}(\tilde{D}_2)$$

Preuve : Étant donné deux ensembles flous de documents $\tilde{D}_1$ et $\tilde{D}_2$ définis sur $\mathcal{C}$, nous avons :

$$\text{Si } \tilde{D}_1 \subseteq \tilde{D}_2 \text{ alors } \forall d \in \mathcal{C}, \mu_{\tilde{D}_2}(d) \geq \mu_{\tilde{D}_1}(d).$$

Pour un terme $t_k \in \mathcal{T}$, nous avons :

$$\mu_{\tilde{f}(\tilde{D}_1)}(t_k) = \min_j (\max(\mu_{\tilde{D}_1}(d_{\sigma(j)}) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\sigma(j)}), 1 - \mu_{I_{la\ plupart}}(j))), \forall d \in \mathcal{C};$$

$$\mu_{\tilde{f}(\tilde{D}_2)}(t_k) = \min_j (\max(\mu_{\tilde{D}_2}(d_{\gamma(j)}) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(j))), \forall d \in \mathcal{C};$$

En respectant la fonction de permutation γ sur les l documents de la relation $\tilde{R}$ stipulant que :

$$\mu_{\tilde{R}}(t, d_{\gamma(1)}) \geq \mu_{\tilde{R}}(t, d_{\gamma(2)}) \geq \mu_{\tilde{R}}(t, d_{\gamma(3)}) \geq \dots \geq \mu_{\tilde{R}}(t, d_{\gamma(l)}), \quad (17)$$

nous avons :

$$\mu_{\tilde{D}_2}(d_{\gamma(j)}) \geq \mu_{\tilde{D}_1}(d_{\gamma(j)}), \quad (18)$$

D'après la propriété 1 (voir page 7), l'équation (18) devient :

$$\mu_{\tilde{D}_2}(d_{\gamma(j)}) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}) \leq \mu_{\tilde{D}_1}(d_{\gamma(j)}) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}). \quad (19)$$

D'après la propriété de la monotonie des opérateurs min et max, l'équation (19) devient :

$$\begin{aligned} & \min_j (\max(\mu_{\tilde{D}_2}(d_{\gamma(j)}) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(j))) \leq \\ & \min_j (\max(\{\mu_{\tilde{D}_1}(d_{\gamma(j)}) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(j)\})); \end{aligned} \quad (20)$$

Il en découle de l'équation (20) que $\mu_{\tilde{f}(\tilde{D}_2)}(t_k) \leq \mu_{\tilde{f}(\tilde{D}_1)}(t_k)$.

Ainsi, nous pouvons conclure que $\tilde{f}(\tilde{D}_1) \supseteq \tilde{f}(\tilde{D}_2)$. Ceci prouve la propriété **(P1)**.

Propriété (Q1) : $\tilde{T}_1 \subseteq \tilde{T}_2 \Rightarrow \tilde{h}(\tilde{T}_1) \supseteq \tilde{h}(\tilde{T}_2)$

Preuve : Soient $\tilde{T}_1$ et $\tilde{T}_2$ deux ensembles flous de termes définis sur $\mathcal{T}$.

Si $\tilde{T}_1 \subseteq \tilde{T}_2$ alors $\forall t \in \mathcal{T}$, $\mu_{\tilde{T}_2}(t) \geq \mu_{\tilde{T}_1}(t)$.

Pour un document $d_k \in \mathcal{C}$, nous avons :

$$\begin{aligned} \mu_{\tilde{h}(\tilde{T}_1)}(d_k) &= \min_i (\max(\mu_{\tilde{T}_1}(t_{\sigma(i)}) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k), \\ & 1 - \mu_{I_{la\ plupart}}(i))), \forall t \in \mathcal{T}; \end{aligned}$$

$$\mu_{\tilde{h}(\tilde{T}_2)}(d_k) = \min_i (\max(\mu_{\tilde{T}_2}(t_{\sigma(i)}) \xrightarrow{I_{f_{loue}}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k), 1 - \mu_{I_{la\ plupart}}(i))), \forall t \in \mathcal{T};$$

En respectant la fonction de permutation σ sur les m termes de la relation $\tilde{R}$ tel que :

$$\mu_{\tilde{R}}(t_{\sigma(1)}, d) \geq \mu_{\tilde{R}}(t_{\sigma(2)}, d) \geq \mu_{\tilde{R}}(t_{\sigma(3)}, d) \geq \dots \geq \mu_{\tilde{R}}(t_{\sigma(m)}, d), \quad (21)$$

nous avons :

$$\mu_{\tilde{T}_2}(t_{\sigma(i)}) \geq \mu_{\tilde{T}_1}(t_{\sigma(i)}). \quad (22)$$

D'après la propriété 1 (voir page 7), l'équation (22) devient :

$$\mu_{\tilde{T}_2}(t_{\sigma(i)}) \xrightarrow{I_{f_{loue}}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k) \leq \mu_{\tilde{T}_1}(t_{\sigma(i)}) \xrightarrow{I_{f_{loue}}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k). \quad (23)$$

Par la propriété de la monotonie des opérateurs min et max, l'équation (23) peut s'écrire comme suit :

$$\begin{aligned} \min_i (\max(\mu_{\tilde{T}_2}(t_{\sigma(i)}) \xrightarrow{I_{f_{loue}}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k), 1 - \mu_{I_{la\ plupart}}(i))) \leq \\ \min_i (\max(\{\mu_{\tilde{T}_1}(t_{\sigma(i)}) \xrightarrow{I_{f_{loue}}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k), 1 - \mu_{I_{la\ plupart}}(i)\})); \end{aligned} \quad (24)$$

Ainsi, nous déduisons de l'équation (24) que :

$$\mu_{\tilde{h}(\tilde{T}_2)}(d_k) \leq \mu_{\tilde{h}(\tilde{T}_1)}(d_k).$$

Nous pouvons ainsi conclure que $\tilde{h}(\tilde{T}_1) \supseteq \tilde{h}(\tilde{T}_2)$. Ceci prouve la propriété **(Q1)**.

Propriété (P2) : $\tilde{D} \subseteq \tilde{h} \circ \tilde{f}(\tilde{D})$

Preuve : Nous avons :

$$\begin{aligned} \tilde{f}(\tilde{D}) = \{t \mid \mu_{\tilde{f}(\tilde{D})}(t) = \min_j (\max(\mu_{\tilde{D}}(d_{\gamma(j)}) \xrightarrow{I_{f_{loue}}} \mu_{\tilde{R}}(t, d_{\gamma(j)}), \\ 1 - \mu_{I_{la\ plupart}}(j))), t \in \mathcal{T}\} \end{aligned}$$

$$\tilde{h} \circ \tilde{f}(\tilde{D}) = \{d_{\gamma(j)} \mid \mu_{\tilde{h} \circ \tilde{f}(\tilde{D})}(d_{\gamma(j)}) = \min_i (\max(\mu_{\tilde{f}(\tilde{D})}(t_{\sigma(i)}) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(i))), d \in \mathcal{C}\}$$

Nous voulons prouver que $\tilde{D} \subseteq \tilde{h} \circ \tilde{f}(\tilde{D})$. Ceci revient à prouver que pour un document d_k , $\mu_{\tilde{D}}(d_k) \leq \mu_{\tilde{h} \circ \tilde{f}(\tilde{D})}(d_k)$.

Pour un document $d_k \in \mathcal{C}$, nous avons :

$$\mu_{\tilde{h} \circ \tilde{f}(\tilde{D})}(d_k) \geq \min_i (\max(\mu_{\tilde{f}(\tilde{D})}(t_{\sigma(i)}) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(d_k, t_{\sigma(i)}), 1 - \mu_{I_{la\ plupart}}(i))), \forall t \in \mathcal{T}; \quad (25)$$

sachant que :

$$\mu_{\tilde{f}(\tilde{D})}(t_{\sigma(i)}) = (\min_j (\max(\mu_{\tilde{D}}(d_{\gamma(j)}) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(j)))), \forall d \in \mathcal{C} \quad (26)$$

D'après l'équation (26), pour un document d_k de rang k , nous avons :

$$\mu_{\tilde{f}(\tilde{D})}(t_{\sigma(i)}) \leq \max(\mu_{\tilde{D}}(d_k) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k), 1 - \mu_{I_{la\ plupart}}(k)). \quad (27)$$

D'après les équations (25) et (27) et la propriété 1 (voir page 7), nous avons :

$$\begin{aligned} \max(\mu_{\tilde{D}}(d_k) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k), 1 - \mu_{I_{la\ plupart}}(k)) &\xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k) \\ &\leq \mu_{\tilde{f}(\tilde{D})}(t_{\sigma(i)}) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k), \forall d \in \mathcal{C}; \end{aligned} \quad (28)$$

En maximisant l'équation (28) par rapport à $1 - \mu_{I_{la\ plupart}}(i)$ et en minimisant sur i , nous déduisons que :

$$\begin{aligned} \mu_{\tilde{h} \circ \tilde{f}(\tilde{D})}(d_k) &\geq (\min_i (\max((\max(\mu_{\tilde{D}}(d_k) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k), 1 - \mu_{I_{la\ plupart}}(k)) \\ &\xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k)), 1 - \mu_{I_{la\ plupart}}(i))), \forall t \in \mathcal{T}. \end{aligned} \quad (29)$$

À partir de l'équation (29), nous pouvons conclure que :

$$\begin{aligned} \mu_{\tilde{h} \circ \tilde{f}(\tilde{D})}(d_k) &\geq \max((\max(\mu_{\tilde{D}}(d_k) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k), 1 - \mu_{I_{la\ plupart}}(k)) \\ &\xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k)), 1 - \mu_{I_{la\ plupart}}(i)) \end{aligned} \quad (30)$$

Puisque $\max(a, b) \geq a$, alors l'équation (30) devient :

$$\mu_{\tilde{h} \circ \tilde{f}}(\tilde{D})(d_k) \geq (\max(\mu_{\tilde{D}}(d_k) \xrightarrow{I_{f l o u e}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k), 1 - \mu_{I_{la\ plupart}}(k)) \xrightarrow{I_{f l o u e}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k)).$$

De la même manière, nous obtenons :

$$\mu_{\tilde{h} \circ \tilde{f}}(\tilde{D})(d_k) \geq \mu_{\tilde{D}}(d_k) \xrightarrow{I_{f l o u e}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k) \xrightarrow{I_{f l o u e}} \mu_{\tilde{R}}(t_{\sigma(i)}, d_k).$$

D'après la propriété 2 (voir page 7), il en découle :

$$\mu_{\tilde{h} \circ \tilde{f}}(\tilde{D})(d_k) \geq \mu_{\tilde{D}}(d_k).$$

Nous pouvons ainsi conclure que $\tilde{D} \subseteq \tilde{h} \circ \tilde{f}(\tilde{D})$. Ceci prouve la propriété **(P2)**.

Propriété (Q2) : $\tilde{T} \subseteq \tilde{f} \circ \tilde{h}(\tilde{T})$

Preuve : Nous avons :

$$\tilde{h}(\tilde{T}) = \{d \mid \mu_{\tilde{h}(\tilde{T})}(d) = \min_i (\max(\mu_{\tilde{T}}(t_{\sigma(i)}) \xrightarrow{I_{f l o u e}} \mu_{\tilde{R}}(t_{\sigma(i)}, d), 1 - \mu_{I_{la\ plupart}}(i))), d \in \mathcal{C}\}$$

$$\tilde{f} \circ \tilde{h}(\tilde{T}) = \{t \mid \mu_{\tilde{f} \circ \tilde{h}(\tilde{T})}(t) = \min_j (\max(\mu_{\tilde{h}(\tilde{T})}(d_{\gamma(j)}) \xrightarrow{I_{f l o u e}} \mu_{\tilde{R}}(d_{\gamma(j)}, t), 1 - \mu_{I_{la\ plupart}}(j))), t \in \mathcal{T}\};$$

Nous voulons prouver que $\tilde{T} \subseteq \tilde{f} \circ \tilde{h}(\tilde{T})$. Ceci revient à prouver que pour un terme t_k , $\mu_{\tilde{T}}(t_k) \leq \mu_{\tilde{f} \circ \tilde{h}(\tilde{T})}(t_k)$.

Pour un terme $t_k \in \mathcal{T}$, nous avons :

$$\mu_{\tilde{f} \circ \tilde{h}(\tilde{T})}(t_k) = \min_j (\max(\mu_{\tilde{h}(\tilde{T})}(d_{\gamma(j)}) \xrightarrow{I_{f l o u e}} \mu_{\tilde{R}}(d_{\gamma(j)}, t_k), 1 - \mu_{I_{la\ plupart}}(j))), \forall d \in \mathcal{C}; \quad (31)$$

sachant que :

$$\mu_{\tilde{h}(\tilde{T})}(d_{\gamma(j)}) = (\min_i (\max(\mu_{\tilde{T}}(t_k) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(j))), \forall t \in \mathcal{T}; \quad (32)$$

D'après l'équation (32), pour un terme de rang k , nous avons :

$$\mu_{\tilde{h}(\tilde{T})}(d_{\gamma(j)}) \leq \max(\mu_{\tilde{T}}(t_k) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(k)) \quad (33)$$

D'après les équations (31) et (33) et la propriété 1 (voir page 7), nous avons :

$$\max(\mu_{\tilde{T}}(t_k) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(k)) \xrightarrow{I_{floue}} \tilde{R}(t_k, d_{\gamma(j)}) \leq \mu_{\tilde{h}(\tilde{T})}(d_{\gamma(j)}) \xrightarrow{I_{floue}} \tilde{R}(t_k, d_{\gamma(j)}) \quad (34)$$

En maximisant l'équation (34) par rapport à $1 - \mu_{I_{la\ plupart}}(j)$ et en minimisant sur j , nous déduisons que :

$$\mu_{\tilde{f}_{oh}(\tilde{T})}(t_k) \geq (\min_j (\max(\max(\mu_{\tilde{T}}(t_k) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(k)) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(j))), \forall d \in \mathcal{C}. \quad (35)$$

À partir de l'équation (35), nous pouvons conclure :

$$\mu_{\tilde{f}_{oh}(\tilde{T})}(t_k) \geq \max((\max(\mu_{\tilde{T}}(t_k) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(k)) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(j)). \quad (36)$$

Puisque $\max(a, b) \geq a$, l'équation (36) devient :

$$\mu_{\tilde{f}_{oh}(\tilde{T})}(t_k) \geq (\max(\mu_{\tilde{T}}(t_k) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}), 1 - \mu_{I_{la\ plupart}}(k)) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)})).$$

De la même manière nous obtenons :

$$\mu_{\tilde{f}_{oh}(\tilde{T})}(t_k) \geq \mu_{\tilde{T}}(t_k) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}) \xrightarrow{I_{floue}} \mu_{\tilde{R}}(t_k, d_{\gamma(j)}). \quad (37)$$

D'après la propriété 2 (voir page 7) et l'équation (37), nous obtenons :

$$\mu_{\tilde{f} \circ \tilde{h}(\tilde{T})}(t_k) \geq \mu_{\tilde{T}}(t_k).$$

Nous pouvons ainsi conclure que $\tilde{T} \subseteq \tilde{f} \circ \tilde{h}(\tilde{T})$. Ceci prouve la propriété **(Q2)**.

Propriété (P3) : $\tilde{f}(\tilde{D}) = \tilde{f} \circ \tilde{h} \circ \tilde{f}(\tilde{D})$

Preuve : Pour $\tilde{T} = \tilde{f}(\tilde{D})$ et par la propriété **(P2)**, nous avons :

$$\tilde{f}(\tilde{D}) \subseteq \tilde{f} \circ \tilde{h} \circ \tilde{f}(\tilde{D}) \quad (38)$$

et à partir de $\tilde{D} \subseteq \tilde{h} \circ \tilde{f}(\tilde{D})$, par la propriété **(P1)**, nous avons :

$$\tilde{f}(\tilde{D}) \supseteq \tilde{f} \circ \tilde{h} \circ \tilde{f}(\tilde{D}) \quad (39)$$

Les résultats (38) et (39) impliquent que $\tilde{f}(\tilde{D}) = \tilde{f} \circ \tilde{h} \circ \tilde{f}(\tilde{D})$.

La propriété **(Q3)** peut être prouvée de la même manière.

Propriété (P4) : $\tilde{D}_1 \subseteq \tilde{D}_2 \Rightarrow \tilde{h} \circ \tilde{f}(\tilde{D}_1) \subseteq \tilde{h} \circ \tilde{f}(\tilde{D}_2)$

Preuve : Soient deux ensembles flous $\tilde{D}_i$ et $\tilde{D}_j$ définis sur $\mathcal{C}$ tel que $\tilde{D}_i \subseteq \tilde{D}_j$.

Par la propriété **(P1)**, nous obtenons $\tilde{f}(\tilde{D}_i) \supseteq \tilde{f}(\tilde{D}_j)$.

Ceci implique par la propriété **(Q1)** que $\tilde{h} \circ \tilde{f}(\tilde{D}_i) \subseteq \tilde{h} \circ \tilde{f}(\tilde{D}_j)$.

Propriété (Q4) : $\tilde{T}_1 \subseteq \tilde{T}_2 \Rightarrow \tilde{f} \circ \tilde{h}(\tilde{T}_1) \subseteq \tilde{f} \circ \tilde{h}(\tilde{T}_2)$

Preuve : Soient deux ensembles flous $\tilde{T}_i$ et $\tilde{T}_j$ définis sur $\mathcal{T}$ tel que $\tilde{T}_i \subseteq \tilde{T}_j$.

Par la propriété **(Q1)**, nous obtenons $\tilde{h}(\tilde{T}_i) \supseteq \tilde{h}(\tilde{T}_j)$.

Ceci implique par la propriété **(P1)** que $\tilde{f} \circ \tilde{h}(\tilde{T}_i) \subseteq \tilde{f} \circ \tilde{h}(\tilde{T}_j)$.

Propriété (P5) : $\tilde{h} \circ \tilde{f}(\tilde{h} \circ \tilde{f}(\tilde{D}_1)) = \tilde{h} \circ \tilde{f}(\tilde{D}_1)$

Preuve : À partir de la propriété **(P3)**, nous avons :

$$\tilde{f} \circ \tilde{h}(\tilde{f}(\tilde{D})) = \tilde{f}(\tilde{D}).$$

$$\text{D'où } \tilde{f}(\tilde{h} \circ \tilde{f}(\tilde{D})) = \tilde{f}(\tilde{D})$$

$$\text{Donc } \tilde{h} \circ \tilde{f}(\tilde{h} \circ \tilde{f}(\tilde{D})) = \tilde{h} \circ \tilde{f}(\tilde{D}).$$

Propriété (Q5) : $\tilde{f} \circ \tilde{h}(\tilde{f} \circ \tilde{h}(\tilde{T}_1)) = \tilde{f} \circ \tilde{h}(\tilde{T}_1)$

Preuve : A partir de la propriété **(Q3)**, nous avons :

$$\tilde{h} \circ \tilde{f}(\tilde{h}(\tilde{T})) = \tilde{h}(\tilde{T})$$

$$\text{D'où } \tilde{h}(\tilde{f} \circ \tilde{h}(\tilde{T})) = \tilde{h}(\tilde{T})$$

$$\text{Donc } \tilde{f} \circ \tilde{h}(\tilde{f} \circ \tilde{h}(\tilde{T})) = \tilde{f} \circ \tilde{h}(\tilde{T}).$$

Propriété (PQ6) : $\tilde{D} \subseteq \tilde{h}(\tilde{T}) \Leftrightarrow \tilde{T} \subseteq \tilde{f}(\tilde{D})$

Preuve : Soit $\tilde{D} \subseteq \tilde{h}(\tilde{T})$. Par la propriété **(P1)**, nous avons $\tilde{f}(\tilde{D}) \supseteq \tilde{f} \circ \tilde{h}(\tilde{T})$ et par la propriété **(Q2)**, nous avons $\tilde{f}(\tilde{D}) \supseteq \tilde{T}$. Ceci prouve que si $\tilde{D} \subseteq \tilde{h}(\tilde{T})$ alors $\tilde{T} \subseteq \tilde{f}(\tilde{D})$. L'autre direction suit d'une manière symétrique.