Généralisation des opérateurs de dérivation de Galois en recherche d'information basée sur l'analyse formelle de concepts

Yassine Djouadi[1,2]

[1] Université de Tizi-Ouzou, BP 17, 15000 Tizi-Ouzou, Algerie.
[2] IRIT de Toulouse, France. Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex 09, France
djouadi@irit.fr

RÉSUMÉ. *La recherche d'information basée sur l'analyse formelle de concepts repose généralement sur l'utilisation de la structure de treillis des concepts formels. Les noeuds de ce treillis (i.e. les concepts formels) peuvent être interprétés comme des paires ⟨réponse,requête⟩. A ce jour, la quasi-majorité des travaux existant en ce sens utilisent l'opérateur de dérivation de Galois classique (opérateur de suffisance). Il s'avère que cet opérateur se restreint à l'expression de requêtes conjonctives. Pour palier à ces limites, nous proposons de généraliser l'utilisation d'opérateurs de dérivation en s'inspirant de la théorie des possibilités. Nous montrons que ces nouveaux opérateurs permettent de considérer aussi bien des requêtes conjonctives que disjonctives ainsi que la négation.*

ABSTRACT. *The use of formal concept analysis in information retrieval relies on the use of the lattice structure of formal concepts. The nodes of such a lattice (i.e. formal concepts) may be considered as pairs $\langle answer, query \rangle$. Nowadays, almost all related existing approaches consider only the classical Galois derivation operator which expresses a guaranteed possibility. However, this operator is limited to the expression of conjunctive queries. This paper enlarges classical derivation operator to the possibility theory setting which allows for conjunctive and disjunctive queries as well as the negation.*

MOTS-CLÉS : *Analyse formelle de concepts, opérateurs de possibilité, de nécessité, de suffisance et de suffisance duale, réduction linéaire.*

KEYWORDS: *Formal concept analysis, possibility, necessity, sufficiency and dual sufficiency operators, linear reduction.*

1. Introduction

L'analyse formelle de concepts (abr. AFC) consiste à induire, à partir d'une relation binaire *Objets*$\times$*Propriétés*, des paires de sous ensembles $\langle\{$*objets*$\}, \{$*propriétés*$\}\rangle$ telles que $\{$*objets*$\}$ est le sous ensemble maximal d'objets satisfaisant toutes les propriétés de $\{$*propriétés*$\}$ déjà satisfaites par tous les objets de $\{$*objets*$\}$. De telles paires sont appelées concepts formels où $\{$*objets*$\}$ (resp. $\{$*propriétés*$\}$) correspond à l'extension (resp. l'intension) du concept formel. L'ensemble de ces concepts formels forme un treillis complet. Dans la proposition initiale de Wille (Wille, 1982), la relation est Booléenne. Autrement dit, un objet satisfait (vérifie) totalement une propriété ou ne la satisfait pas du tout. Afin de prendre en considération des relations permettant une satisfaction graduelle d'une propriété par un objet, l'AFC floue a été proposée (Pollandt, 1997), (Bělohlávek, 1999). Dans ce cas, la notion de satisfaction peut être exprimée par un degré $\in [0, 1]$. L'approche proposée dans ce papier s'inscrit dans ce cadre.

L'évidente analogie entre la relation binaire *Objets*$\times$*Propriétés* caractérisant l'AFC et une relation binaire de type *Documents*$\times$*Termes* caractérisant la recherche d'information (abr. RI) a rapidemment suscité un engouement certain pour l'utilisation de l'AFC en RI. Dans ce cas, les documents correspondent à des objets formels et les termes d'indexation (descripteurs, éléments de thesaurus, etc.) correspondent aux propriétés formelles (Priss, 2000). Les concepts formels résultant d'une telle relation peuvent être interprétés comme des paires $\langle\{$*réponse*$\}, \{$*requête*$\}\rangle$ où la requête correspond à l'intension du concept formel tandis que la réponse correspond à son extension. La relation de subsomption (relation d'ordre partiel) entre concepts formels peut être considérée comme une relation de spécialisation/généralisation entre requêtes (Messai *et al.,* 2005).

La consolidation de l'AFC (Ganter *et al.,* 1999) sur un modèle mathématique théoriquement fondé a donné lieu à une utilisation effective et efficace de l'AFC en RI. Dans (Carpineto *et al.,* 2004), il est indiqué que la RI utilisant les treillis de concepts formels atteint des performances qui dépassent nettement celles de la recherche Booléenne classique.

L'application la plus naturelle reste le raffinement de requêtes qui repose sur les deux assertions suivantes (Messai *et al.,* 2005) ; i) Un concept formel c d'un treillis quelconque peut être interprété comme une paire $\langle$*réponse,requête*$\rangle$ où la *requête* correspond à l'intension de c et la *réponse* correspond à l'extension de c, ii) Un déplacement à partir d'un noeud du treillis (i.e. une requête), vers le haut (resp. vers le bas) en suivant les arêtes du treillis, produit tous les raffinements (resp. élargissements) minimaux de la requête. L'analyse formelle de concepts a néanmoins été utilisée en RI à d'autres fins. Par exemple, dans (Nauer *et al.,* 2008), elle est appliquée pour de la classification de documents et dans (Dau *et al.,* 2008) pour de l'ordonnancement. Dans ce papier, nous nous limitons au cadre de l'interrogation par navigation. Dans ce contexte précis, nous avons pu constater que la plupart des approches de RI basées sur l'AFC supposent que :

1) La relation *Documents*$\times$*Termes* est Booléenne,

2) Les requêtes sont aussi Booléennes (i.e. il n'y a pas moyen de pondérer les termes de la requête),

3) Les requêtes sont limitées à une forme conjonctive,

4) Il n'y a pas moyen d'obtenir l'ordonnancement des différents résultats car obtenus par navigation ad hoc à travers le treillis.

Dans un papier précédent (Djouadi, 2011), nous avons introduit l'AFC floue qui permet de généraliser une relation Booléenne et permet d'exprimer des correspondances numériques entre *Terme* et *Document* (par exemple obtenu à partir de la fréquence d'apparition du Terme dans le Document). Nous avons aussi proposé d'utiliser de nouveaux opérateurs de dérivation afin d'exprimer en plus la disjonction et la négation. Dans ce papier, nous rappelons d'abord certaines de ces propositions puis présentons une nouvelle approche déterministe permettant l'ordonnancement des résultats d'une requête traditionnellement obtenus par navigation empirique à travers le treillis de concepts. L'approche proposée revient en quelque sorte à "linéariser" un treillis.

Ce papier est organisé comme suit. La section 2 présente d'abord les fondements de l'analyse formelle de concepts dans le cas classique (Booléen) et dans le cas flou. Puis, nous présentons certaines tendances en matière de recherche d'information basée sur l'AFC. Nous mettons en évidence certaines limites des approches existantes. La section suivante présente de nouveaux opérateurs de dérivation que nous proposons d'utiliser afin d'augmenter l'expressivité des requêtes formulées. La section 4, met en évidence cette expressivité par la prise en compte de la disjonction et de la négation. La section 5 présente une proposition permettant l'ordonnancement deterministe des résultats d'une requête.

2. Analyse formelle de concepts et recherche d'information

2.1. *Fondements de l'analyse formelle de concepts*

2.1.1. *Contextes formels Booléens*

L'AFC fournit un cadre théorique pour l'apprentissage de hiérarchies de concepts. Cet apprentissage s'effectue à partir d'un contexte formel $\mathcal{K} := (\mathcal{O}, \mathcal{P}, \mathcal{R})$ où $\mathcal{R}$ est une relation binaire complètement définie entre un ensemble d'objets $\mathcal{O}$ et un ensemble de propriétés Booléennes $\mathcal{P}$ ($\mathcal{R} \subseteq \mathcal{O} \times \mathcal{P}$). Un contexte formel est généralement représenté par une matrice d'adjacence.

Etant donné un objet x et une propriété y, soit $R(x) = \{y \in \mathcal{P} \mid x\mathcal{R}y\}$ l'ensemble des propriétés possédées par l'objet x ($x\mathcal{R}y$ signifie que x possède la propriété y) et, de manière polymorphe, soit $R(y) = \{x \in \mathcal{O} \mid x\mathcal{R}y\}$ l'ensemble des objets possédant la propriété y. On définit en AFC des correspondances entre les ensembles $2^{\mathcal{O}}$ et $2^{\mathcal{P}}$. Ces correspondances sont appelés opérateurs de dérivation de Galois. L'opérateur de Galois proposé par Wille en AFC est aussi appelée *opérateur de suffisance*[1] (Düntsch *et al.,* 2003). Etant donné $X \subseteq \mathcal{O}$ et $Y \subseteq \mathcal{P}$, l'opérateur de suffisance, noté $(.)^{\Delta}$, permet d'exprimer l'ensemble des propriétés satisfaites par tous les objets de X comme :

$$\begin{aligned} X^{\Delta} = &\ \{y \in \mathcal{P} \mid \forall x \in \mathcal{O}\ (x \in X \Rightarrow x\mathcal{R}y)\} \\ = &\ \{y \in \mathcal{P} \mid X \subseteq R(y)\} \\ = &\ \textstyle\bigcap_{x \in X} R(x) \end{aligned}$$

Et permet aussi d'exprimer dualement l'ensemble des objets satisfaisant toutes les propriétés de Y comme :

$$\begin{aligned} Y^{\Delta} = &\ \{x \in \mathcal{O} \mid \forall y \in \mathcal{P}\ (y \in Y \Rightarrow x\mathcal{R}y)\} \\ = &\ \{x \in \mathcal{O} \mid Y \subseteq R(x)\} \\ = &\ \textstyle\bigcap_{y \in Y} R(y) \end{aligned}$$

La paire duale d'opérateurs $((.)^{\Delta}, (.)^{\Delta})$ constitue ainsi une connexion de Galois qui permet d'induire des concepts formels. Un concept formel est une paire $\langle X, Y \rangle$ telle que $X^{\Delta} = Y$ et $Y^{\Delta} = X$. L'ensemble X (resp. Y) est appelé *extension* (resp. *intension*) du concept. L'ensemble de tous les concepts formels est naturellement équipé d'une relation d'ordre (notée $\preccurlyeq$) et définie comme : $\langle X_1, Y_1 \rangle \preceq \langle X_2, Y_2 \rangle$ ssi $X_1 \subseteq X_2$ (ou $Y_2 \subseteq Y_1$). Cet ensemble muni de la relation d'ordre $\preccurlyeq$ forme un treillis complet $\mathfrak{B}(\mathcal{K})$. Les opérateurs *Meet* et *Join* sont décrits par le théorème fondamental de Ganter et Wille (voir (Ganter *et al.,* 1999) pour de plus amples détails).

2.1.2. *Contextes formels flous*

Dans la proposition originale de Wille (Wille, 1982), la relation considérée est Booléenne (bi-valuée). La nécessité d'étendre l'AFC à des relations floues a rapidement été mise en évidence pour prendre en compte des relations incomplètes (vagues, floues, imprécises, incertaines, etc.) (Djouadi *et al.,* 2009) ou même multi-valuées (Messai *et al.,* 2008). Un contexte flou est décrit par $\mathcal{K} := (L, \mathcal{O}, \mathcal{P}, \mathcal{R})$ où $\mathcal{R}$ est définie comme $\mathcal{R} : \mathcal{O} \times \mathcal{P} \rightarrow L$ (généralement $L = [0, 1]$).

La généralisation de l'opérateur de Galois $(.)^{\Delta}$ à des contextes formels flous est basée sur des implications floues (Bělohlávek, 1999). Cette généralisation s'exprime comme une fonction ensembliste entre les parties floues de $\mathcal{O}$ ($L^{\mathcal{O}}$) et de $\mathcal{P}$ ($L^{\mathcal{P}}$). Elle est donnée par :

$$X^{\Delta}(y) = \bigwedge_{x \in \mathcal{O}} (X(x) \rightarrow \mathcal{R}(x, y)) \qquad [1]$$

$$Y^{\Delta}(x) = \bigwedge_{y \in \mathcal{P}} (Y(y) \rightarrow \mathcal{R}(x, y)) \qquad [2]$$

1. Traduction de *sufficiency operator*.

Où $\rightarrow$ dénote l'implication floue. C'est à dire un opérateur décroissant sur le premier opérande et croissant sur le deuxième opérande et qui vérifie les conditions aux limites : i) $0 \rightarrow 0 = 0 \rightarrow 1 = 1 \rightarrow 1 = 1$, et ii) $1 \rightarrow 0 = 0$. Un concept formel flou correspond alors à la paire $\langle X, Y \rangle$ telle que $X^{\Delta} = Y$ et $Y^{\Delta} = X$. On peut remarquer que $X^{\Delta\Delta} = X$ et $Y^{\Delta\Delta} = Y$ ce qui signifie que X et Y sont des points fixes.

2.2. *Recherche d'information basée sur les treillis de concepts formels*

L'utilisation de l'AFC en RI nous amène logiquement à considérer l'ensemble d'objets (resp. de propriétés) comme ensemble de documents (resp. de termes). Aussi, nous noterons dans la suite de ce papier : $\mathcal{D}$ l'ensemble des documents (collection) ; $\mathcal{T}$ l'ensemble des termes ; $\mathcal{R}$ la relation *Documents*$\times$*Termes* ; D un sous ensemble (classique) de documents ($D \subseteq \mathcal{D}$) ; T un sous ensemble (classique) de termes ($T \subseteq \mathcal{T}$) ; $\widetilde{D}$ un sous ensemble flou de documents ; $\widetilde{T}$ un sous ensemble flou de termes.

La plupart des approches de recherche d'information basées sur l'AFC reposent sur un point commun à savoir l'utilisation de la structure de treillis de l'ensemble des concepts formels ((Priss, 2000), (Carpineto *et al.,* 2004), (Nauer *et al.,* 2008), (Dau *et al.,* 2008), (Messai *et al.,* 2008)). A juste titre, nous pouvons remarquer qu'avant même l'avènement de l'AFC, une première formalisation de l'utilisation des treillis en RI avait été proposée début des années 60. Salton lui-même évoqua dans son ouvrage (Salton, 1968) différents approches autres que le classique modèle vectoriel à savoir les modèles topologiques, métriques, graphiques et enfin les treillis.

L'AFC a connu plusieurs applications en recherche d'information. Nous citerons notamment l'interrogation pour la reformulation (raffinement) de requêtes (Messai *et al.,* 2005), l'ordonnancement (Dau *et al.,* 2008) et la classification de documents (Nauer *et al.,* 2008).

Constatant la complexité de la RI basée sur les treillis de concepts, différentes nouvelles approches, dédiées au Web, proposent d'utiliser l'AFC à un méta niveau. C'est-à-dire au dessus d'un moteur de recherche classique (Carpineto *et al.,* 2004), (Dau *et al.,* 2008). Ces approches ont toutes un point commun : la requête est initialement envoyée à un moteur de recherche (Google ou autre). Puis, un treillis de concepts formels est construit à partir des mots du titre et de l'extrait (résumé) des documents retournés. Il est à noter que cette opération étant dynamique, elle doit s'effectuer à chaque requête.

Dans (Nauer *et al.,* 2008), le système CreChainDo proposé permet de coupler un contrôle de pertinence utilisateur à une navigation par treillis. En effet, les auteurs proposent d'intégrer l'utilisateur dans le processus de recherche d'information en lui donnant la possibilité d'apprécier l'intérêt positif ou négatif des concepts et des documents. L'utilisateur explore le treillis et identifie des concepts pertinents et/ou non pertinents. Ces retours utilisateur sont aussi exploités pour faire évoluer le treillis par la même occasion.

A travers les différentes approches existantes et rencontrées, nous avons pu constater les points suivants :

– La notion de *concept requête* est généralement admise dans bon nombre d'approches existantes. C'est-à-dire qu'il est supposé qu'une quelconque requête figure comme l'intension d'un concept formel ce qui n'est pas forcément le cas. Pour qu'une telle supposition soit vérifiée, certains auteurs (Messai *et al.,* 2005) sont par exemple amenés à rajouter un tel concept requête dans le treillis de concepts initial ce qui revient à modifier la relation binaire initiale en rajoutant une ligne correspondant à la requête émise .

$\mathcal{R}_1$	t_1	t_2	t_3
d_1	×		×
d_2	×	×	
d_3		×	×
d_4		×	×
d_5	×	×	×

Tableau 1. *Exemple de contexte formel*

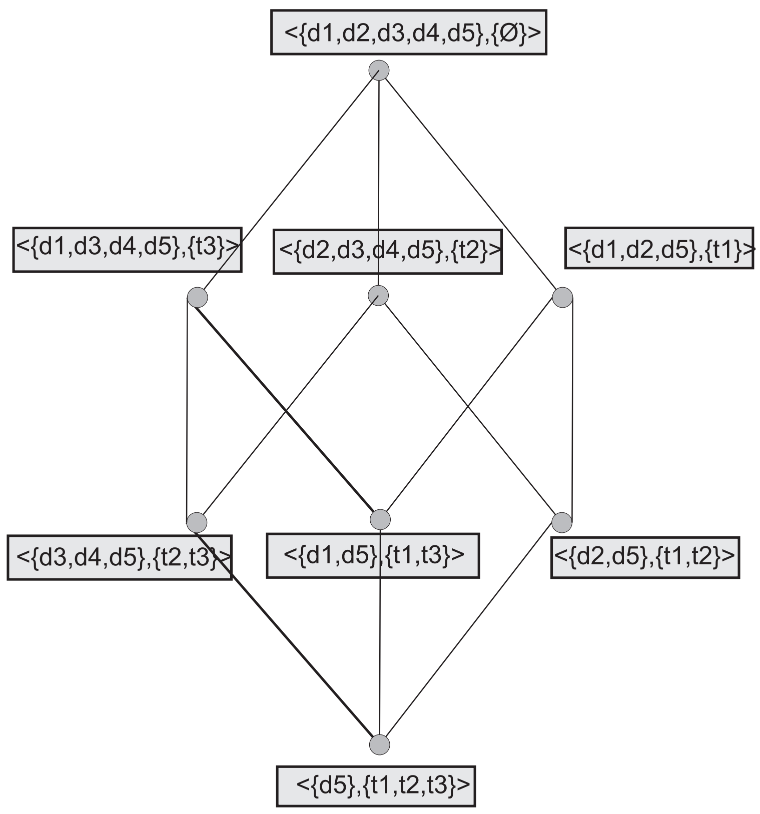

Figure 1. *Treillis de concepts formels.*

– La majorité des approches existantes considèrent une relation Booléenne. Dans pareil cas, aussi bien les requêtes que les réponses sont Booléennes. Une certaine pondération des termes de la requête n'est donc pas permise et en retour, une certaine gradualité des éléments de la réponse n'est aussi pas permise. Pour palier à ce problème une approche intéressante (Latiri *et al.,* 2004) propose l'extension de la connexion de Galois classique (i.e. la paire d'opérateurs $((.)^{\Delta}, (.)^{\Delta})$) à des relations floues. Néanmoins, cette approche considère uniquement des requêtes conjonctives.

– La quasi-majorité[2] des approches existantes utilisent la connexion de Galois classique (Δ, Δ). La sémantique d'un concept formel $\langle D, T\rangle$ induit à partir de cette connexion de Galois peut être énoncée comme suit : *D correspond à l'ensemble maximal de documents possédant (satisfaisant) tous les termes de T qui sont déjà possédés par tous les documents de D*. Illustrons cela à travers l'exemple donné dans le Tableau 1. A la requête $r = \{t_1, t_2\}$, la connexion (Δ, Δ) fournit une seule réponse, à savoir l'ensemble de documents $D = \{d_2, d_5\}$. Les approches existantes proposent ensuite à l'utilisateur un moyen de naviguer dans le treillis de concepts afin d'étendre le résultat aux concepts " voisins ". Une telle navigation peut s'avérer comme un moyen ad hoc au sens où la notion de voisinage ainsi que le choix du chemin de navigation sont laissés à l'initiative de l'utilisateur. A titre d'exemple, dans la Figure 1, l'utilisateur est evidemment confronté à différents chemins de navigation dans le treillis de concepts formels.

3. Extension possibiliste des opérateurs de dérivation de Galois

Dans (Dubois *et al.,* 2007), les auteurs ont mis en évidence de nouveaux opérateurs ensemblistes (opérateurs de dérivation de Galois). Ces opérateurs ont été proposés par les auteurs comme une lecture possibiliste de l'AFC (c'est à dire, inspirée de la théorie des possibilités (Zadeh, 1978)). Il s'agit notamment des quatre opérateurs ensemblistes Π, △, N, and ∇ désignant respectivement la possibilité potentielle, la suffisance, la nécessité potentielle et la suffisance duale. Comme l'opérateur de suffisance est celui qui est déjà utilisé en AFC au sens classique, nous donnons ci-après la définition des trois autres opérateurs dans le cas Booléen.
- Etant donné un ensemble $T \subseteq \mathcal{T}$, l'opérateur de possibilité T^{Π} désigne l'ensemble des documents satisfaisant au moins un terme $t \in T$:

$$\begin{aligned} T^{\Pi} = & \ \{d \in \mathcal{D} \mid T \cap R(d) \neq \emptyset\} \\ = & \ \{d \in \mathcal{D} \mid \exists t \in \mathcal{T},\ d\mathcal{R}t\} \end{aligned}$$

- Etant donné un ensemble $T \subseteq \mathcal{T}$, l'opérateur de nécessité T^{N} désigne l'ensemble des documents satisfaisant uniquement les termes de T :

$$\begin{aligned} T^{N} = & \ \{d \in \mathcal{D} \mid R(d) \subseteq T\} \\ = & \ \{d \in \mathcal{D} \mid \forall t \in \mathcal{T}\big(d\mathcal{R}t \Rightarrow t \in T\big)\} \end{aligned}$$

2. Nous n'avons pas connaissance d'une approche de RI basée sur l'AFC utilisant un opérateur autre que celui de suffisance

- Etant donné un ensemble $T \subseteq \mathcal{T}$, l'opérateur de suffisance duale T^{∇} désigne l'ensemble des documents ne satisfaisant pas au moins un terme de $\overline{T}$.

$$\begin{aligned} T^{\nabla} = & \quad \{d \in \mathcal{D} \mid T \cup R(d) \neq \mathcal{D}\} \\ = & \quad \{d \in \mathcal{D} \mid \exists\, t \in \overline{T},\, d\overline{\mathcal{R}}t\} \end{aligned}$$

Ces opérateurs de dérivation se généralisent également à des relations *Documents*$\times$*Termes* floues. Cette généralisation correspond à des fonctions ensemblistes définies entre l'ensemble des parties floues de $\mathcal{T}$ ($L^{\mathcal{T}}$) et l'ensemble des parties floues de $\mathcal{D}$ ($L^{\mathcal{D}}$) est donnée ci dessous. Le lecteur pourra toutefois se référer au papier (Djouadi *et al.,* 2011) qui établit les propriétés mathématiques nécessaires à l'utilisation des ces trois nouveaux opérateurs.

$$\begin{aligned} \widetilde{T}^{\Pi}(d) = & \quad \bigvee\nolimits_{t \in \mathcal{T}} \big(\widetilde{T}(t) * \mathcal{R}(d,t)\big) \\ \widetilde{T}^{\mathrm{N}}(d) = & \quad \bigwedge\nolimits_{t \in \mathcal{T}} \big(\mathcal{R}(d,t) \rightarrow \widetilde{T}(t)\big) \\ \widetilde{T}^{\nabla}(d) = & \quad \bigvee\nolimits_{t \in \mathcal{T}} \big(\neg\widetilde{T}(t) * \neg\mathcal{R}(d,t)\big) \end{aligned}$$

Où l'opérateur $*$ (exprimant une sémantique de conjonction) est un opérateur décroissant vérifiant : i) $p * 1 = p$ et ii) $0 * 0 = 0 * 1 = 1 * 0 = 0$. Tandis que l'opérateur $\neg$ exprime la négation définie telle que : $\neg\widetilde{T}(t) = 1 - \widetilde{T}(t)$.

4. Généralisation de l'utilisation des opérateurs de dérivation de Galois en recherche d'information

4.1. *Formes conjonctives*

Notons par $\hat{Q}(T) \equiv t_1 \wedge t_2 \wedge \cdots \wedge t_k$ une requête conjonctive sur les termes $t_i, i = 1, k$. En accord avec le modèle Booléen de Salton, nous pouvons définir la satisfaction de la requête $\hat{Q}(T)$ comme suit :

Definition 1 *Un ensemble de documents D satisfait une requête conjontive $\hat{Q}(T)$ (noté $D \models \hat{Q}(T)$) si et seulement si :*

$$D \models \hat{Q}(T) \Leftrightarrow D = \bigcap_{t \in T} \{t\}^{\triangle} \Leftrightarrow D = T^{\triangle} \qquad [3]$$

Cette notion de satisaction peut être généralisée à des requêtes conjonctives floues $\hat{Q}(\widetilde{T})$ de la forme $t_1^{w_1} \wedge t_2^{w_2} \wedge \cdots \wedge t_k^{w_k}$ où $\widetilde{T}(t_i) = w_i$ représente le poids du terme t_i dans la requête $\hat{Q}(\widetilde{T})$. Rappelons qu'une requête floue est ici modélisée par un ensemble flou de termes et fournit en résultat un ensemble flou de documents. Dans ce cas, la satisfaction de $\hat{Q}(\widetilde{T})$ par un ensemble flou de documents est définie comme : $\widetilde{D} \models \hat{Q}(\widetilde{T}) \Leftrightarrow \widetilde{D} = \widetilde{T}^{\triangle}$.

Notons par $Ext(c)$ (resp. $Int(c)$) deux fonctions qui associent à un concept formel c son extension (resp. son intension). Il en résulte qu'un concept formel c satisfait une requête conjonctive $\hat{Q}(\widetilde{T})$ ssi $Ext(c) = \widetilde{T}^{\triangle}$ ou $Int(c) = \widetilde{T}^{\triangle\triangle}$.

Remarque 1 *Il existe un et un seul concept formel satisafisant $\hat{Q}(\widetilde{T})$. Cela est une conséquence de la propriété d'idempotence de l'opérateur $(.)^{\triangle\triangle}$, c'est à dire $(\widetilde{T}^{\triangle\triangle})^{\triangle\triangle} = \widetilde{T}^{\triangle\triangle}$.*

4.2. *Formes disjonctives*

Notons par $\check{Q}(T) \equiv t_1 \vee t_2 \vee \cdots \vee t_k$ une requête (Booléenne) disjonctive sur les termes $t_i, i = 1, k$. Nous donnons ci-dessous la définition de la satisfaction de $\check{Q}(T)$ par un ensemble de documents en s'appuyant sur un nouvel opérateur de dérivation à savoir l'opérateur de possibilité.

Definition 2 *Un ensemble de documents D satisfait une requête disjonctive $\check{Q}(T)$ (noté $D \models \check{Q}(T)$) comme suit :*

$$D \models \check{Q}(T) \Leftrightarrow D \subseteq \bigcup_{t \in T} \{t\}^{\Pi} = T^{\Pi} \qquad [4]$$

Considérons maintenant une requête disjonctive floue $\check{Q}(\widetilde{T})$ de la forme $t_1^{w_1} \vee t_2^{w_2} \vee \cdots \vee t_k^{w_k}$. La définition, ci-dessus enoncée, de la satisfaction d'une requête disjonctive peut être généralisée au cas flou comme suit : $\widetilde{D} \models \check{Q}(\widetilde{T}) \Leftrightarrow \widetilde{D} \subseteq \widetilde{T}^{\Pi}$ où l'opérateur $\subseteq$ désigne l'inclusion ensembliste floue définie par $\widetilde{D} \subseteq \widetilde{T}^{\Pi} \Leftrightarrow \forall d \in \mathcal{D} \ : \ \widetilde{D}(d) \leqslant \widetilde{T}^{\Pi}(d)$.

Remarque 2 *Lorsque l'ensemble des termes est un singleton, nous avons l'égalité : $\{t\}^{\Pi} = \{t\}^{\triangle}$.*

Nous pouvons constater que l'ensemble des concepts formels satisfaisant une requête disjonctive ne se réduit pas nécessairement à un singleton. A ce titre, la proposition suivante établit une importante propriété concernant la structure algébrique de l'ensemble des résultats admissibles d'une requête.

Proposition 1 *Soient $\widetilde{T}$ un ensemble flou de termes, $\check{Q}(\widetilde{T})$ la requête disjonctive pondérée associée et $\mathfrak{B}(\check{Q}(\widetilde{T}))$ l'ensemble des concepts formels satisfaisant $\check{Q}(\widetilde{T})$. L'ensemble $\mathfrak{B}(\check{Q}(\widetilde{T})) \bigcup \{\langle \widetilde{T}^{\Pi}, (\widetilde{T}^{\Pi})^{\triangle} \rangle\}$ est un treillis complet.*

Exemple 1 : Soit la relation binaire $\mathcal{R}_1$ illustrée dans le Tableau 1. Considérons la requête disjonctive $\check{Q}(\{t_1, t_2\}) \equiv t_1 \vee t_2$ (afin de rendre l'exemple plus lisible, nous avons considéré une requête Booléenne). En accord avec la Proposition 1, l'ensemble $\mathfrak{B}(\check{Q}(\{t_1, t_2\})) \bigcup \{\langle \{t_1, t_2\}^{\Pi}, (\{t_1, t_2\}^{\Pi})^{\triangle} \rangle\}$ forme un treillis complet illustré dans la Figure 2.

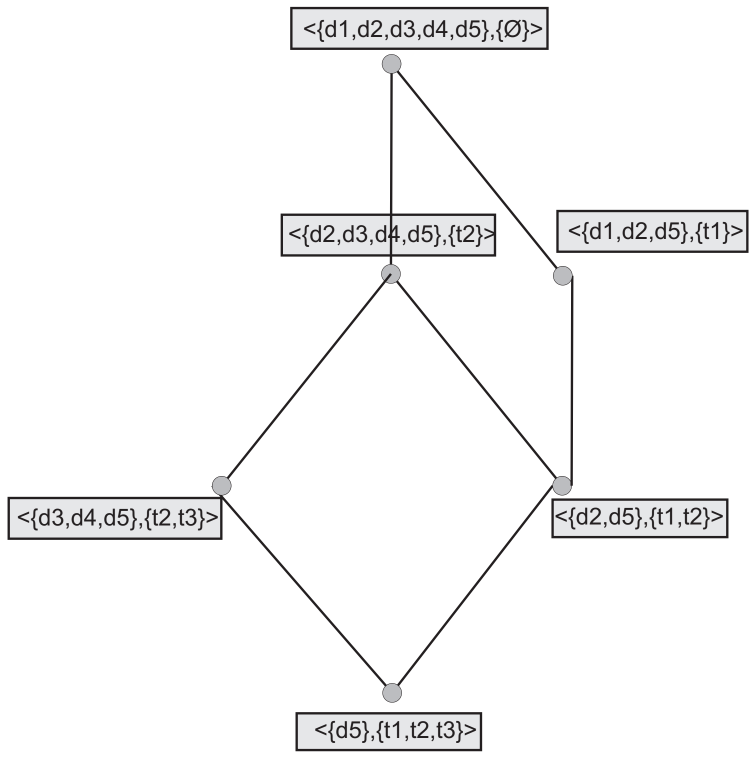

Figure 2. *Treillis des concepts formels satisfaisant la requête disjonctive* $t_1 \vee t_2$.

4.3. *Expression de la négation*

Dans ce qui suit, on considérera que les notations $(.)^{\triangle}_{\overline{\mathcal{R}}}$, $(.)^{\Pi}_{\overline{\mathcal{R}}}$, $(.)^{N}_{\overline{\mathcal{R}}}$, $(.)^{\nabla}_{\overline{\mathcal{R}}}$ font référence à l'utilisation du complémentaire $\overline{\mathcal{R}}$ de la relation $\mathcal{R}$. Rappelons que $\overline{\mathcal{R}}(d,t) = \neg\mathcal{R}(d,t) = 1-\mathcal{R}(d,t)$. Nous avons déjà établi dans (Djouadi *et al.,* 2011) des propriétés entre ces quatre opérateurs. Nous allons mettre à profit certaines de ces propriétés pour exprimer la négation de requêtes conjonctives ou disjonctives.

Considérons la propriété $\overline{\widetilde{T}^{\triangle}} = \widetilde{T}^{\Pi}_{\overline{\mathcal{R}}}$ déjà prouvé dans (Djouadi *et al.,* 2011) (où $\overline{\widetilde{T}^{\triangle}}(d) = 1 - \widetilde{T}^{\triangle}(d)$). Cette propriété exprime naturellement que les ensembles $\widetilde{D}$ (flous ou classiques) de documents satisfaisant la negation d'une requête conjonctive $\overline{\hat{Q}(\widetilde{T})}$ pour une relation binaire donnée $\mathcal{R}$, sont ceux qui satisfont la requête disjonctive $\check{Q}(\widetilde{T})$ pour le complémentaire de $\mathcal{R}$. Puisque la satisfaction d'une requête disjonctive a déjà été traitée dans la Section 4.2, il suffit de construire une unique fois le treillis de concepts formels induit par le complémentaire de la relation $\mathcal{R}$ et appliquer la Proposition 1 pour obtenir l'ensemble des résultats de la requête. Rappelons que cet ensemble sera forcément un treillis complet.

5. Ordonnancement par réduction linéaire

Nous avons mis en évidence l'intérêt d'élargir l'AFC au cas flou pour : i) la prise en compte de relations floues ; ii) la prise en compte de requêtes pondérées (modélisées par des ensembles flous de termes). Un des gros problèmes en AFC floue est que le treillis induit peut comporter un nombre très important de concepts formels et ce, même pour une relation *Documents*×*Termes* assez réduite (entre 10 et 100 documents). Il résulte un treillis relativement dense. Rappelons que l'une des utilisations de l'AFC en RI consiste à naviguer dans le treillis des concepts formels pour explorer les "concepts voisins". Afin de formaliser la notion de "concept voisin", nous proposons de munir l'ensemble $\mathfrak{B}(\mathcal{K})$ des concepts formels d'une distance δ vérifiant les axiomes de symétrie, séparation et inégalité triangulaire. On introduit en outre, la notion de concept cible défini comme suit :

Definition 3 *La paire* $\langle \widetilde{T}^{\triangle}, \widetilde{T}^{\triangle\triangle} \rangle$ (*resp.* $\langle \widetilde{T}^{\Pi}, \widetilde{T}^{\Pi\triangle} \rangle \}$) *est appelée concept cible de la requête* $\hat{Q}(\widetilde{T})$ (*resp.* $\check{Q}(\widetilde{T})$).

Remarque 3 *Un concept cible n'est pas forcément un concept formel.*

Il est à noter que la notion de concept cible a déjà été utilisée (Messai *et al.,* 2008), mais uniquement dans le cas de requêtes conjonctives. Il s'avère que dans ce cas, nous avons montré qu'il existait un et un seul concept formel satisfaisant la requête. En revanche, nous avons vu que l'ensemble des concepts formels satisfaisant une requête disjonctive est un treillis (muni d'un ordre partiel). Nous proposons la réduction linéaire de ce treillis. C'est à dire, la transformation de la structure de treillis en une structure linéaire munie d'un ordre total comme suit.

Soit S un sous-ensemble de concepts formels satisfaisant une requête disjonctive ($S \subseteq \mathfrak{B}(\mathcal{K})$). Soit $\mathfrak{T} = (S, \preccurlyeq)$ le treillis induit par la relation d'ordre partiel $\preccurlyeq$ sur S. Etant donné un concept cible $t \in S$ et une distance δ, on considère la relation d'ordre $\lessdot$ définie telle que $c_1 \lessdot c_2 \Leftrightarrow \delta(Ext(t), Ext(c_2)) \leqslant \delta(Ext(t), Ext(c_1))$. Nous appelons réduction linéaire de $\mathfrak{T}$, le treillis linéaire $\mathfrak{L} = (S, \lessdot)$ tel que :

$$\begin{aligned} \bigvee_{c_i \in S} \mathfrak{L} &= t \\ \bigwedge_{c_i \in S} \mathfrak{L} &= b, \quad \text{tel que} \quad \delta(Ext(b), Ext(t)) = \bigvee_{c_i \in S} \delta(Ext(c_i), Ext(t)) \\ c_1 \wedge c_2 &= c_1 \quad \Longleftrightarrow c_1 \lessdot c_2 \end{aligned}$$

Il est aisé de prouver que $\mathfrak{L}$ est aussi un treillis complet où $\lessdot$ est un ordre total réversible. Nous illustrons l'intérêt de la réduction linéaire à travers l'exemple suivant.

Exemple 2 : Soit la relation floue $\mathcal{R}_2$ illustrée dans le Tableau 2. Considérons la requête pondérée suivante : $t_1^1 \vee t_2^{0.0} \vee t_3^{0.5}$. Le treillis complet des concepts formels satisfaisant cette requête est illustrée dans la Figure 3.

La réduction linéaire du treillis est illustrée dans la Figure 4. Nous avons choisi la distance de Hausdorff qui est fréquemment utilisé pour mesurer l'éloignement entre

$\mathcal{R}_2$	t_1	t_2	t_3
d_1	0.0	0.2	0.3
d_2	1	0.9	0.6
d_3	0.0	0.4	0.1
d_4	0.7	0.9	0.5

Tableau 2. *Relation* Documents×Termes *floue*

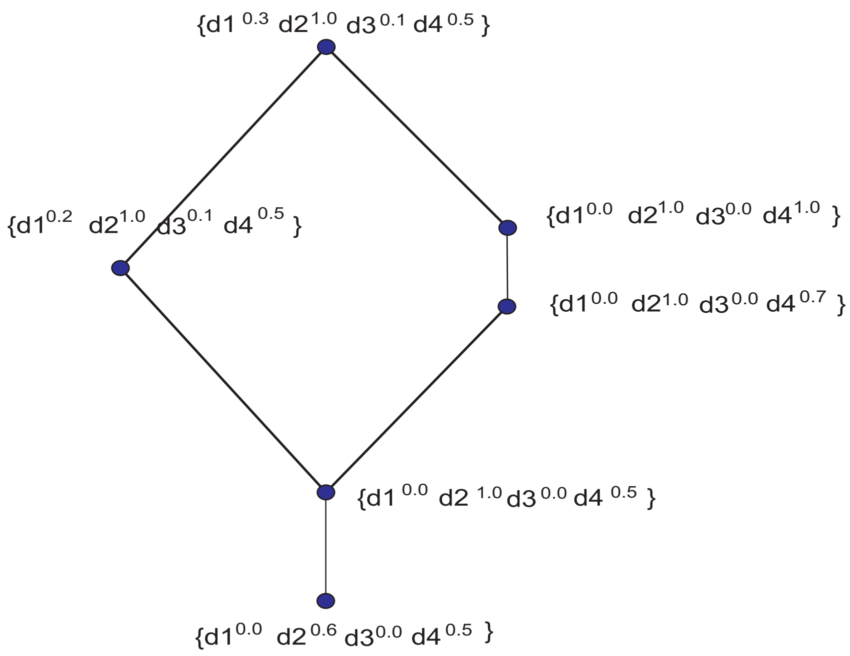

Figure 3. *Concepts formels satisfaisant la requête* $t_1^1 \vee t_2^{0.0} \vee t_3^{0.5}$.

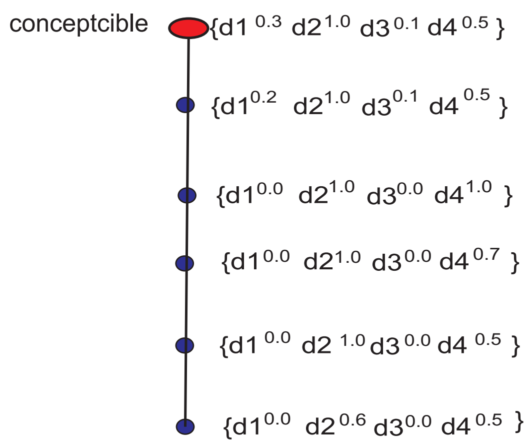

Figure 4. *Ordonnancement des ensembles de documents satisfaisant la requête* $t_1^1 \vee t_2^{0.0} \vee t_3^{0.5}$.

deux ensembles. Etant donné une collection $\mathcal{D}$ de documents, la distance de Hausdorff entre deux ensembles flous de documents $\widetilde{X}$ et $\widetilde{Y}$ est définie comme suit :

$$\delta(\widetilde{X},\widetilde{Y}) = max\Big(\bigwedge_{x\in\mathcal{D}}\bigvee_{y\in\mathcal{D}}|\widetilde{X}(x)-\widetilde{Y}(y)|, \bigwedge_{y\in\mathcal{D}}\bigvee_{x\in\mathcal{D}}|\widetilde{X}(x)-\widetilde{Y}(y)|\Big)$$

Par exemple, le calcul de la distance entre le concept cible $d_1^{0.3}d_2^{1.0}d_3^{0.1}d_4^{0.5}$ et les deux plus proches voisins donne $\delta\big(d_1^{0.3}d_2^{1.0}d_3^{0.1}d_4^{0.5}, d_1^{0.2}d_2^{1.0}d_3^{0.1}d_4^{0.5}\big)$=max$\big($0.5,0.5$\big)$=0.5 tandis que $\delta\big(d_1^{0.3}d_2^{1.0}d_3^{0.1}d_4^{0.5}, d_1^{0.0}d_2^{1.0}d_3^{0.0}d_4^{1.0}\big)$=max$\big($0.5,0.7$\big)$=0.7. De ce fait, $(d_1^{0.0}d_2^{1.0}d_3^{0.0}d_4^{1.0}) \lessdot (d_1^{0.2}d_2^{1.0}d_3^{0.1}d_4^{0.5})$.

6. Conclusion

Ce papier présente une approche originale en recherche d'information basée sur l'analyse formelle de concepts. Cette approche consiste à étendre l'utilisation de l'opérateur de Galois classiquement utilisé à de nouveaux opérateurs ensemblistes inspirés de la théorie des possibilités en l'occurrence les opérateurs de possibilité, nécessité et suffisance duale.

Nous avons aussi constaté que l'ensemble des réponses satisfaisant une requête avait parfois une structure de treillis. La navigation de concept à concept voisin (telle qu'effectuée par les approches existantes) peut s'avérer empirique et complexe (exemple : existence d'anti-chaînes). Pour cela, notre seconde contribution consiste à proposer une réduction linéaire du treillis des réponses ce qui permet un ordonnancement des résultats de la requête.

Comme perspectives, nous comptons investiguer l'utilisation de distances sémantiques. Nous envisageons aussi de mettre en évidence le parallèle entre l'utilisation des opérateurs considérés dans ce papier et les approches de RI basées sur la logique modale.

7. Bibliographie

Bělohlávek R., « Fuzzy Galois connections », Math. Logic Quart, vol. 45, p. 497-504, 1999.

Carpineto C., Romano G., « Exploiting the Potential of Concept Lattices for Information Retrieval with CREDO », *Journal of Universal Computer Science*, vol. 10, n° 8, p. 985-1013, 2004.

Dau F., Ducrou J., Eklund P. W., « Concept Similarity and Related Categories in SearchSleuth », *Proc. 16th International Conference on Conceptual Structures, ICCS 2008, July 7-11, 2008*, Toulouse, France, p. 255-268, 2008.

Djouadi Y., « Extended Galois Derivation Operators for Information Retrieval Based on Fuzzy Formal Concept Lattice », *SUM'2011, Scalable Uncertainty Managment, LNCS 6929*, p. 346-358, 2011.

Djouadi Y., Dubois D., Prade H., « Différentes extensions floues de l'analyse formelle de concepts », *Proc. Rencontres Francophones sur la Logique Floue et ses Applications (LFA), 05/11/09-06/11/09*, Cépaduès Editions, Annecy (France), p. 141-148, 2009.

Djouadi Y., Prade H., « Possibility-theoretic extension of derivation operators in formal concept analysis », *Fuzzy Optimization and Decision Making*, vol. 10, n° 4, p. 287-309, 2011.

Dubois D., Bannay F., Prade H., « A possibilty-theoretic view of formal concept analysis », *Fundamenta Informaticae*, vol. 75, n° 1-4, p. 195-213, 2007.

Düntsch I., Gediga G., « Approximation operators in qualitative data analysis », *In : Theory and application of relational structures as knowledge instruments*, Springer, LNCS 2929, p. 214-230, 2003.

Ganter B., Wille R., *Formal Concept Analysis*, Springer-Verlag, 1999.

Latiri C., , Chevallet J. P., Elloumi S., Jaoua A., « Une Extension de la Connexion de Galois Floue pour la Recherche d'Information », *Revue I3 : Information - Interaction - Intelligence*, vol. 3, n° 2, p. 1-44, 2004.

Messai N., Devignes M.-D., Napoli A., Smaïl-Tabbone M., « Treillis de concepts et ontologies pour l'interrogation d'un annuaire de sources de données biologiques (BioRegistry) », *Actes du XXIIIème Congrès INFORSID, 24-27 mai, 2005*, Grenoble, France, p. 587-602, 2005.

Messai N., Devignes M.-D., Napoli A., Smaïl-Tabbone M., « Many-Valued Concept Lattices for Conceptual Clustering and Information Retrieval », *Proc. ECAI 2008 - 18th European Conference on Artificial Intelligence, July 21-25, 2008*, Patras, Greece, p. 127-131, 2008.

Nauer E., Toussaint Y., « Classification dynamique par treillis de concepts pour la recherche d'information sur le web », *Proc. COnférence en Recherche d'Infomations et Applications - CORIA 2008, March 12-14, 2008*, Trégastel, France, p. 71-86, 2008.

Pollandt S., *Fuzzy Begriffe*, Springer-Verlag, Berlin/Heidelberg, 1997.

Priss U., « Lattice-based information retrieval », *Knowledge Organization*, vol. 27, n° 3, p. 132-142, 2000.

Salton G., *Automatic information organization and retrieval*, McGraw-Hill, New York, 1968.

Wille R., *Restructuring Lattice Theory : an Approach Based on Hierarchies of Concepts*, Rival. I. (Ed) : Ordered Sets. Reidel, Dordrecht. Boston, 1982.

Zadeh L., « Fuzzy sets as a basis for a possibility theory », *Fuzzy Sets and Systems*, vol. 1, n° 1, p. 3-28, 1978.