# Text retrieval based on dyadic conceptual projection

Sonia Alouane-Ksouri[1], Minyar Sassi Hidri[2], Kamel Barkaoui[3]
[1,2]Université de Tunis El Manar
[1,2]Ecole Nationale d'Ingénieurs de Tunis
LR-SITI-ENIT
BP. 37, Le Belvèdère 1002, Tunis, Tunisia
[3]CEDRIC-CNAM
Rue Saint-Martin Paris 75003, France
{[1,2]sonia.alounae,minyar.sassi}@enit.rnu.tn,[3]kamel.barkaoui@cnam.fr

***Abstract*—This work exploits the use of Triadic Concept Analysis (TCA) for document retrieval to efficiently answer users' queries. The proposed conceptual analysis aims at describing the documents according to three hierarchical levels with a triadic computing model. It is based on normalisation and prototyping, which, by projection, induce a formal dyadic context. This representation has enabled us to visualize triadic concepts associated with the documents, sentences and words through the construction of fuzzy concept lattice. The lattices generated are then used to organize documents in a hierarchical structure to facilitate retrieval process.**

**Keywords-Document Retrieval, Triadic Concepts Analysis, Co-similarities, Galois lattice, Fuzzy clustering.**

## I. Introduction

Document Retrieval (DR) results in a list of documents ordered according to relevance. However, the number of documents is so big that the users often cannot consult them all. Some relevant documents may thus be ignored because they have been badly positioned. This calls for a representation technique that exploits document semantics in order to make the results more explicit and to direct the browsing within the collection of documents.

To overcome this problem, several proposed approaches use the query representation, the similarity measure, or the representation of documents [1]. In [2], the results of a search session is gradually refined using the user's judgments of relevance. The Galois Lattice, the core of Formal Concept Analysis (FCA) [3], can be used as a basis for conceptual data structuring [4]. It is generated from the relation between documents and terms. This structure supports a conceptual analysis and allows the user to widen or to specify his query as concerns the subsets of documents and terms presented in the lattice.

Despite of the efficiency proved by FCA in the field of Information Retrieval (IR) [5], few approaches have been exploited this theory to deal with DR [6], [7], [8], [9].

Classically documents are represented as bags of words to which are assigned weights measuring their importance in the text (binary weight, frequency...). In this case, the search is done on this set of weighted words [10].

Recently, a model based on triadic document co-similarities has been proposed [11]. It extends the classic systems by representing documents as sets of sentences and words. This allows an increase in the level of relevance of the documents as to sentences and words.

The present work is a natural follow-up which leads us to apply a triadic model to DR. This conceptual analysis aims at describing documents according to two levels (sentences and word) corresponding to the computing model of triadic co-similarities. This approach is mainly based on a projection step which leads into a formal dyadic context. The transformation of a fuzzy many-valued concepts associated to documents, sentences and words into a dyadic one allow the visualization and the DR throw fuzzy concept lattice .

This paper is organized as follows: in section II, we present how computes triadic document co-similarities. In section III, we present our conceptual document retrieval model. It gives a formal description of the proposed model based on fuzzy triadic co-similarities of many-valued attributes. Section IV concludes the paper and mentions some future work.

## II. Tri-partite computing of document co-similarities

In [11], a new tripartite computing model for document similarity is proposed. It is used to measure adherence in the context of documents classification and allows to simultaneously compute co-similarity matrices between documents, denoted by $\widetilde{D}_2$, sentences, denoted by $\widetilde{S}_2$, and words, denoted by $\widetilde{W}_2$. Each of the three matrices is built based on the two others. This model stems from the idea that the document clustering should be based not only on text analysis as a set of words, but also as a set of sentences.

However, sentences have a strong descriptive power compared to simple words. Thus, the document can be divided into a set of sentences, each of which is divided into a set of words. This model combines the fuzzy sets [12] in the document pre-processing step in order to determine more relevant memberships.

From $D = \{D_1, D_2, ..., D_i\}$ $(i = 1..I)$ a set of $I$ documents, $S = \{S_1, S_2, ..., S_j\}$ $(j = 1..J)$ a set of $J$ sentences and $W = \{W_1, W_2, ..., W_k\}$ $(k = 1..K)$ a set of $K$ words, we determine $SD = [SD]_{ji}$ $(i = 1..I,\ j = 1..J)$ similarity matrices which represents respectively to the number of occurrences of the $j^{th}$ $(j = 1..J)$ sentence with the $i^{th}$

$(i = 1..I)$ document and the $k^{th}$ $(k = 1..K)$ word by the $j^{th}$ $(j = 1..J)$ sentence.

After the fuzzification step, we obtain the matrices $\widetilde{SD}_i = [\mu]_{ji}$ $(i = 1..I,\ j = 1..J)$ and $\widetilde{WS}_j = [\mu]_{kj}$ $(k = 1..K,\ j = 1..J)$ which correspond respectively to fuzzy similarities of Sentences $\times$ Documents and Words $\times$ Sentences. It is on the basis of these fuzzy matrices that we will compute co-similarity between Documents $\times$ Documents $\widetilde{D}_2 = [\delta]_{lm}$ $(l = 1..I,\ m = 1..I)$ (resp. Sentences $\times$ Sentences $\widetilde{S}_2 = [\alpha]_{lm}$ $(l = 1..J,\ m = 1..J)$ and Words $\times$ Words $\widetilde{W}_2 = [\omega]_{lm}$ $(l = 1..K,\ m = 1..K)$). The elements of the matrices are given as follows [11]:

$$\delta_{l,m}^{\{t\}} = \sum_{i=1}^{J}\sum_{j=1}^{J} \min(\mu_{il}, \mu_{jm}) * \alpha_{ij}^{\tilde{S_2}^{\{t-1\}}} \quad (1)$$

$$\alpha_{l,m}^{\{t\}} = Min[\textstyle\sum_{i=1}^{I}\sum_{j=1}^{I} \min(\mu_{il}, \mu_{jm})\delta_{ij}^{\tilde{D_2}^{\{t-1\}}}$$

$$\sum_{i=1}^{K}\sum_{j=1}^{K} \min(\mu_{il}, \mu_{jm})\omega_{ij}^{\tilde{W_2}^{\{t-1\}}}] \quad (2)$$

$$\omega_{l,m}^{\{t\}} = \sum_{i=1}^{J}\sum_{j=1}^{J} \min(\mu_{il}, \mu_{jm}) * \alpha_{ij}^{\tilde{S_2}^{\{t-1\}}} \quad (3)$$

## III. Document retrieval based on conceptual triadic co-similarities

In this work, we focus on organizing, in a Galois lattice, documents having a common themes. The resulting lattice provides indications of proximity or remoteness of the thematic elements of a corpus.

To study the interaction between documents that address a theme, it is possible to construct the Galois lattice of the relation between the document and the treated theme. This structure is used to organize data in a lattice presented in the form of a table modeling a binary relation between objects and properties.

This problem is also known and studied as an analysis of formal concepts [13]. This analysis lead us to manage a two-dimensional array keywords/documents where concepts are given as pairs $(\{Extension\}, \{Intension\})$. This assumes that a document is decomposed into a set of keywords.

However, it turns out that the choice of keywords is less relevant than that of sentences. A sentence of a document is defined as an ordered sequence of one or more words [14].

In [15], an extraction and identification of significant sentences in the statistical processing of natural language was proposed.

The aim of this work is to exploit the formal representations of concepts, frequently used in a dyadic context, and adapt them to the representation of concepts according to the triadic model proposed in [11]. Researchers have examined the triadic aspect of data representation for classification [4] or for auto-completion of a [16].

The steps of our approach are as follows:

1) Computing triadic document co-similarities : this allows obtaining three co-similarity matrices between documents $\widetilde{D}_2$, sentences $\widetilde{S}_2$ and words $\widetilde{W}_2$;
2) Prototyping and normalization of data : while basing on the results from the computing of fuzzy triadic co-similarities, this step is based on applying a clustering algorithm to documents, sentences and words. Following the clustering, we proceed with standardization and prototyping to obtain a representation that lends itself to the analysis of triadic concepts;
3) Document retrieval: this is the ultimate goal of our approach. Indeed, the two preceding steps have led to the construction of triadic contexts. These are now exploited to perform a document retrieval through incremental algorithms of construction of Galois lattices [4].

Figure 1 shows an overview of the triadic analysis of documents.

### A. Normalization and prototyping

This is a pretreatment preceding the clustering step. Figure 2 presents the various steps of computing the triadic correspondence between documents, sentences and words.

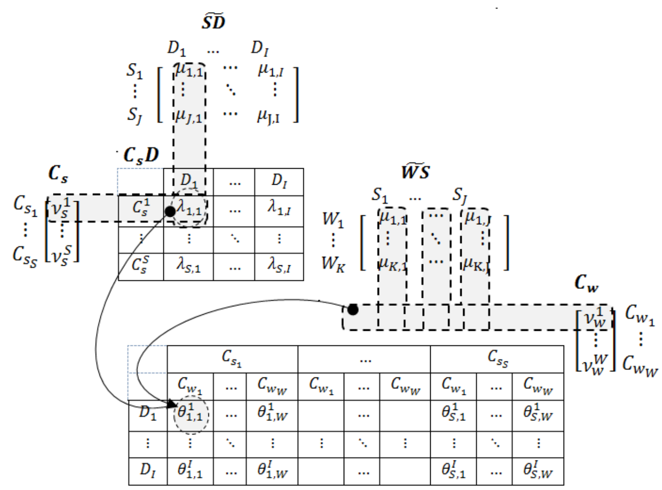


Fig. 2. Normalization and prototyping.

Indeed, the matrices obtained after the application of the triadic similarity computing model, it is possible to apply any clustering algorithm such as Fuzzy C-Means (FCM) [17]. It is based on the optimization of a quadratic criterion which generate, while using the co-similarity matrix $\widetilde{S}_2$, a membership matrix $U = [\tilde{\alpha}_{j,p}]$ $(j = 1..J, p = 1..S)$. Each element of this matrix denotes the membership degree of the $j^{th}$ sentence in the $m^{th}$ cluster.

The centers of the obtained clusters are represented as follows:

$$C_s = \{C_{s_1}, .., C_{s_S}\} = \begin{bmatrix} \nu_s^1 \\ \vdots \\ \nu_s^S \end{bmatrix} \quad (4)$$

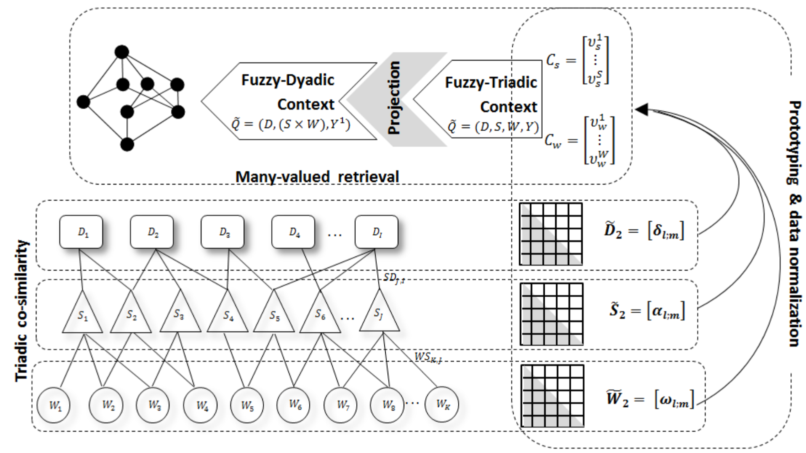


Fig. 1. Overview of the triadic analysis of documents.

The correspondence between the documents and the centers of the obtained clusters from sentences is determined on the basis of the $\widetilde{SD}$ matrix merged by the center $C_s$. The correspondence matrix between Documents and Sentences, denoted by $C_sD = [\lambda_{p,i}]$ with $(p = 1..S, i = 1..I)$, expresses the correspondence degree between the $p^{th}$ sentence and the $i^{th}$ document.

The correspondence between the $p^{th}$ sentence and the $i^{th}$ document is given as follows:

$$\lambda_{p,i} = \sum_{j=1}^{J} \mu_{i,j} * \nu_s^p \tag{5}$$

The clustering performed on sentences will also be performed on words. In effect, FCM is applied to the co-similarity matrix $\widetilde{W}_2$ in order to obtain the membership matrix $U = [\tilde{\omega}_{k,n}]$ $(k = 1..K, n = 1..W)$. It expresses the membership degrees of the $k^{th}$ word in the $n^{th}$ cluster.

The centers of the obtained clusters are represented by :

$$C_w = \{C_{w_1}, .., C_{w_W}\} = \begin{bmatrix} \nu_w^1 \\ \vdots \\ \nu_w^W \end{bmatrix} \tag{6}$$

## B. Fuzzy TCA

Fuzzy clustering methods allow documents to belong simultaneously to several clusters with different membership degrees. According to [18], a triadic context is a triple $K = (E, I, C, Y)$, where $E, I$ and $C$ are sets, and $Y$ is a ternary relationship between $E$, $I$ and $C$.

For the generation of the fuzzy triadic context, we propose to connect the documents with clusters by means of a relation representing the membership degrees of each document to other clusters.

*Definition 1 (Fuzzy triadic context):* A fuzzy triadic context is a quadruplet $\widetilde{Q} = (D, S, W, Y = \theta(D, S, W))$ where $D = \{d_i : i = 1, .., I\}$ is the set of documents, $S = \{\nu_s^p : p = 1, .., S\}$ and $W = \{\nu_w^n : n = 1, .., W\}$ are respectively the sets of obtained clusters of words and sentences. $Y$ is a fuzzy set of the domain $D \times S \times W$. Each fuzzy relation $(d_i, \nu_s^p, \nu_w^n) \in Y$ has a membership degree $(d_i, \nu_s^p, \nu_s^n)$ in $[0, 1]$.

*Definition 2 (Fuzzy triadic concept):* A fuzzy triadic concept of a triadic context $\widetilde{Q} = (D, S, W, Y = \theta(D, S, W))$ is a triple $(U, T, R)$ where $U \subseteq D$, $T \subseteq S$, and $R \subseteq W$ with $U \times T \times R \subseteq Y$. For $U1 \subseteq U$, $T1 \subseteq T$ and $R1 \subseteq R$ with $U1 \times T1 \times R1 \subseteq Y$, the inclusions $U \subseteq U1$, $T \subseteq T1$, and $R \subseteq R1$ imply that $(U, T, R) = (U1, T1, R1)$.

Note that for a fuzzy triadic concept (tri-concept) $(U, T, R)$ $U, R$ and $T$ are respectively called Extension, Intention, and Modus of the tri-concept $(U, T, R)$. In the light of these definitions, $\widetilde{Q} = (D, S, W, Y)$ is the quadruplet describing the fuzzy triadic context with $D$ the set of documents, $S$ the set of sentences' clusters and $W$ the set of words' clusters. $Y$ is a ternary relation between $D$, $S$ and $W$, which by a fuzzy weighting, reflects the fact that words belong to sentences that make up the documents.

Table I describes the triadic context where each $\theta_{p,n}^i$ represents a triadic relationship between a document belonging to $D$, a sentence cluster belonging to $S$ and a word cluster belonging to $W$ .

The fuzzy weight $\theta_{p,n}^i$ is given by :

$$\theta_{p,n}^i = (\sum_{j=1}^{J} (\sum_{k=1}^{K} \mu_{k,j} * \nu_w^n)) * \lambda_{p,i} \tag{7}$$

*Definition 3 (Fuzzy triadic concepts lattice):* The fuzzy triadic concept lattice of a fuzzy triadic context $\widetilde{Q} = (D, S, W, Y = \theta(D, S, W))$ is a set of all fuzzy formal concepts of $\widetilde{Q}$ with a partial order $\leq$.

Several studies have examined the research approaches on triadic analysis concepts. In [19], the input data, represented as a formal triadic context, are transformed into a dyadic one. Then, a Galois lattice is built in correspondence with the generated dyadic concepts. We denote by $Y^1$ the projection of $Y$ on $S \times W$. $(d_1, s_1, w_1)$ is a fuzzy triadic concept of a fuzzy triadic context $\widetilde{Q}$. $(d_1, x)$ is a fuzzy dyadic concept

TABLE I
FUZZY TRIADIC CONTEXT.

| | $C_{s_1}$ | | | $C_{s_2}$ | | | ... | | | $C_{s_S}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $C_{w_1}$ | ... | $C_{w_W}$ | $C_{w_1}$ | ... | $C_{w_W}$ | $C_{w_1}$ | ... | $C_{w_W}$ | $C_{w_1}$ | ... | $C_{w_W}$ |
| $d_1$ | $\theta^1_{1,1}$ | ... | $\theta^1_{1,W}$ | $\theta^1_{2,1}$ | ... | $\theta^1_{2,W}$ | ... | ... | ... | $\theta^1_{S,1}$ | ... | $\theta^1_{S,W}$ |
| $d_2$ | $\theta^2_{1,1}$ | ... | ... | $\theta^2_{2,1}$ | ... | ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| $d_I$ | $\theta^I_{1,1}$ | ... | $\theta^I_{1,W}$ | $\theta^I_{2,1}$ | ... | $\theta^I_{2,W}$ | ... | ... | ... | ... | ... | $\theta^I_{S,W}$ |

where $d_1 \subseteq D$ and $x \subseteq S \times W$. $S \times W$ can be restricted to combinations of $s_1$ and $w_1$.

A projection transforms all of its objects to obtain a formal dyadic context equivalent $\widetilde{Q}^1 = (D, S \times W, Y^1)$. In table II, we present a simplified example of fuzzy triadic context resulting from the projection step. It shows a set of 4 documents according to $C^{w_w}_{s_s} \in \{S \times W\}$ ($p = \{1..3\}$, $n = \{1..4\}$).

To reduce this fuzzy context, an $\alpha$-$Cut$ can be applied to eliminate relationships with low memberships. This threshold may be determined as the reverse of the number of the generated words' clusters. The $\alpha$-$Cut$ is given as follows :

$$\alpha - Cut(U) = C_w^{-1} \tag{8}$$

The reduced fuzzy triadic context is presented in Table III.

It then becomes possible to build a Galois lattice from the reduced fuzzy triadic context. The computing approach adopted is presented in [20]. As presented in Figure 3, a lattice is generated from data shown in Table III.

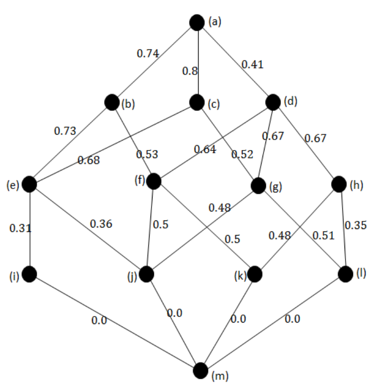


Fig. 3. Fuzzy triadic lattice.

Extensions and intensions of the concepts illustrated in Figure 3 are detailed in Table IV.

### C. Document retrieval

Once the lattice constructed, the search for relevant documents can start according the retrieval process proposed in [21]. For this, we define a query concept $Q = (Q_A, Q_B)$ where $Q_A$ is the name for an extension sought and $Q_B$ the set of clusters describing the properties of sentences' and words' clusters. All $Q_B$ of clusters can be determined by seeking membership of introduced words or their synonyms, via a search engine or other systems, with several clusters of the

TABLE IV
DETAILED FUZZY TRIADIC CONCEPTS.

| **Node** | **Concept** |
|---|---|
| $(a)$ | $(\{d_1^{(0.27)}, d_2^{(0.25)}, d_3^{(0.25)}, d_4^{(0.48)}\}, \{C_{s_1}^{w^2}, C_{s_1}^{w^4}, C_{s_2}^{w^2}, C_{s_2}^{w^4}\})$ |
| $(b)$ | $(\{d_2^{(0.25)}, d_3^{(0.25)}, d_4^{(0.43)}\}, \{C_{s_1}^{w^2}, C_{s_1}^{w^4}, C_{s_2}^{w^2}, C_{s_2}^{w^4}, C_{s_3}^{w^2}\})$ |
| $(c)$ | $(\{d_1^{(0.27)}, d_2^{(0.25)}, d_4^{(0.48)}\}, \{C_{s_1}^{w^2}, C_{s_1}^{w^4}, C_{s_2}^{w^2}, C_{s_2}^{w^4}, C_{s_3}^{w^1}\})$ |
| $(d)$ | $(\{d_1^{(0.27)}, d_2^{(0.25)}, d_3^{(0.25)}\}, \{C_{s_1}^{w^2}, C_{s_1}^{w^4}, C_{s_2}^{w^1}, C_{s_2}^{w^2}, C_{s_2}^{w^4}, C_{s_1}^{w^1}\})$ |
| $(e)$ | $(\{d_2^{(0.25)}, d_4^{(0.43)}\}, \{C_{s_1}^{w^2}, C_{s_1}^{w^4}, C_{s_2}^{w^2}, C_{s_2}^{w^4}, C_{s_3}^{w^1}, C_{s_3}^{w^2}\})$ |
| $(f)$ | $(\{d_2^{(0.25)}, d_3^{(0.25)}\}, \{C_{s_1}^{w^2}, C_{s_1}^{w^4}, C_{s_2}^{w^1}, C_{s_2}^{w^2}, C_{s_2}^{w^4}, C_{s_3}^{w^2}, C_{s_3}^{w^4}, C_{s_1}^{w^1}\})$ |
| $(g)$ | $(\{d_1^{(0.27)}, d_2^{(0.25)}\}, \{C_{s_1}^{w^2}, C_{s_1}^{w^4}, C_{s_2}^{w^1}, C_{s_2}^{w^2}, C_{s_2}^{w^4}, C_{s_3}^{w^1}, C_{s_1}^{w^1}\})$ |
| $(h)$ | $(\{d_1^{(0.27)}, d_3^{(0.25)}\}, \{C_{s_1}^{w^2}, C_{s_1}^{w^4}, C_{s_2}^{w^1}, C_{s_2}^{w^2}, C_{s_2}^{w^4}, C_{s_3}^{w^3}, C_{s_1}^{w^1}\})$ |
| $(i)$ | $(\{d_4^{(0.31)}\}, \{C_{s_1}^{w^2}, C_{s_1}^{w^3}, C_{s_1}^{w^4}, C_{s_2}^{w^2}, C_{s_2}^{w^4}, C_{s_3}^{w^1}, C_{s_3}^{w^2}\})$ |
| $(j)$ | $(\{d_2^{(0.25)}\}, \{C_{s_1}^{w^1}, C_{s_1}^{w^2}, C_{s_1}^{w^4}, C_{s_2}^{w^1}, C_{s_2}^{w^2}, C_{s_2}^{w^4}, C_{s_3}^{w^1}, C_{s_3}^{w^2}, C_{s_3}^{w^4}\})$ |
| $(k)$ | $(\{d_3^{(0.25)}\}, \{C_{s_1}^{w^1}, C_{s_1}^{w^2}, C_{s_1}^{w^4}, C_{s_2}^{w^1}, C_{s_2}^{w^2}, C_{s_2}^{w^4}, C_{s_3}^{w^2}, C_{s_3}^{w^3}, C_{s_3}^{w^4}\})$ |
| $(l)$ | $(\{d_1^{(0.27)}\}, \{C_{s_1}^{w^1}, C_{s_1}^{w^2}, C_{s_1}^{w^4}, C_{s_2}^{w^1}, C_{s_2}^{w^2}, C_{s_2}^{w^4}, C_{s_3}^{w^1}, C_{s_3}^{w^3}\})$ |
| $(m)$ | $(\{\emptyset\}, \{C_{s_1}^{w^1}, C_{s_1}^{w^2}, C_{s_1}^{w^3}, C_{s_1}^{w^4}, C_{s_2}^{w^1}, C_{s_2}^{w^2}, C_{s_2}^{w^3}, C_{s_2}^{w^4}, C_{s_3}^{w^1}, C_{s_3}^{w^2}, C_{s_3}^{w^3}, C_{s_3}^{w^4}\})$ |

triadic context. Once this concept is built, its insertion into the lattice can begin in order to search relevant documents following the incremental Lattice construction of Godin [21].

*Definition 4 (Relevant document):* A document $d$ is relevant to a given query $Q = (Q_A, Q_B)$ if and only if $d$ is characterized by at least one of the clusters $Q_B$ .

This definition of relevance is the basis of the incremental construction process of Godin. The obtained lattice corresponding to the query concept, denoted by $T_{f\oplus} = (C_{\leq f} \oplus Q)$ where $C_{\leq} \oplus Q$, represents the new set of resulting concepts from the insertion of the query in $T_f$ .

We consider all nodes in the initial lattice $T_f$, the insertion of the query concept $Q$ follows the following rules :

1) The generating concept of a new pair is a son concept of this new pair;
2) The sons' concepts of former concepts do not change;
3) The generator is also the only old pair who becomes son of a new pair;
4) A new pair may have a son but that is also a new concept .

TABLE II
EXAMPLE OF A FUZZY TRIADIC CONTEXT.

| | $C_{s_1}^{w_1}$ | $C_{s_1}^{w_2}$ | $C_{s_1}^{w_3}$ | $C_{s_1}^{w_4}$ | $C_{s_2}^{w_1}$ | $C_{s_2}^{w_2}$ | $C_{s_2}^{w_3}$ | $C_{s_2}^{w_4}$ | $C_{s_3}^{w_1}$ | $C_{s_3}^{w_2}$ | $C_{s_3}^{w_3}$ | $C_{s_3}^{w_4}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $d_1$ | 0.66 | 0.31 | 0.09 | 0.41 | 0.61 | 0.27 | 0.10 | 0.29 | 0.33 | 0.05 | 0.42 | 0.04 |
| $d_2$ | 0.68 | 0.29 | 0.08 | 0.44 | 0.63 | 0.25 | 0.12 | 0.38 | 0.29 | 0.67 | 0.07 | 0.33 |
| $d_3$ | 0.71 | 0.27 | 0.10 | 0.50 | 0.66 | 0.25 | 0.12 | 0.45 | 0.12 | 0.39 | 0.86 | 0.46 |
| $d_4$ | 0.19 | 0.58 | 0.31 | 0.48 | 0.12 | 0.64 | 0.20 | 0.58 | 0.63 | 0.43 | 0.21 | 0.17 |

TABLE III
REDUCED FUZZY TRIADIC CONTEXT.

| | $C_{s_1}^{w_1}$ | $C_{s_1}^{w_2}$ | $C_{s_1}^{w_3}$ | $C_{s_1}^{w_4}$ | $C_{s_2}^{w_1}$ | $C_{s_2}^{w_2}$ | $C_{s_2}^{w_3}$ | $C_{s_2}^{w_4}$ | $C_{s_3}^{w_1}$ | $C_{s_3}^{w_2}$ | $C_{s_3}^{w_3}$ | $C_{s_3}^{w_4}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $d_1$ | 0.66 | 0.31 | - | 0.41 | 0.61 | 0.27 | - | 0.29 | 0.33 | - | 0.42 | - |
| $d_2$ | 0.68 | 0.29 | - | 0.44 | 0.63 | 0.25 | - | 0.38 | 0.29 | 0.67 | - | 0.33 |
| $d_3$ | 0.71 | 0.27 | - | 0.50 | 0.66 | 0.25 | - | 0.45 | - | 0.39 | 0.86 | 0.46 |
| $d_4$ | - | 0.58 | 0.31 | 0.48 | - | 0.64 | - | 0.58 | 0.63 | 0.43 | - | - |

The lattice is initialized with the element $(\emptyset, \emptyset)$. Extensions $(D)$ and Intensions $(S \times W)$ are updated as soon as new documents are added. If we assume that $D$ and $S \times W$ contain all elements with an empty relation $Y^1$, the lattice is initialized with the two elements $(D, \emptyset)$ and $(\emptyset, S \times W)$. The goal is to generate $T_{f\oplus} = (C_{\leq f} \oplus Q)$ from $T_{f\oplus} = (C_{\leq f})$.

Given a query $Q = (Q_A, Q_B)$, all relevant documents are in $Q$ and extending its upper bounds in the lattice, since the intent of each of these concepts is included in $Q_B$ (the concept query intension).

## IV. CONCLUSION

In this paper, we proposed a new approach, exploiting the TCA theory and processing textual corpora in a triadic form. Fuzzy logic is used to manage uncertainty. This triadic conceptual analysis enables us to obtain a triadic context that can be used to construct a Galois lattice. A projection has been implemented to transform the triadic context into a dyadic one. The lattices generated are then used to organize documents in a hierarchical structure to facilitate retrieval process.

The purpose of this study is to use formal concept representations, commonly used in a dyadic context, and adapt them to the context of document representation according to the triadic model.

We intend to explore the possibility of representing our fuzzy triadic contexts with nested lattices to exploit them with more precision the membership of different levels of the lattice. This opens up on our future work.

## REFERENCES


[1] M. Ghribi, P. Cuxac, J.-C. Lamirel, and A. Lelu, “Mesures de qualité de clustering de documents : Prise en compte de la distribution des mots clés,” in *Extraction et Gestion des Connaissances*, 2010.

[2] M. Keim, D. Lewis, and D. Madigan, “Bayesian information retrieval: Preliminary evaluation,” in *The Sixth International Workshop on Artificial Intelligence and Statistics*, 1997, pp. 303–310.

[3] R. Wille, “Restructuring lattice theory: An approach based on hierarchies of concepts,” *Ordered sets, Rival Editor*, vol. 23, pp. 445–470, 1982.

[4] R. Godin, G. Mineau, R. Missaoui, and H. Mili, “Méthodes de classification conceptuelle basées sur les treillis de galois et applications,” *Revue d’Intelligence Artificielle*, vol. 9(2), pp. 105–137, 1995.

[5] L. Szathmary and A. Napoli, “Les treillis de galois pour l’organisation et la gestion des connaissances,” in *11èmes Rencontres de la Société Francophone de Classification*, vol. 1, 2004, pp. 153–164.

[6] B. Ganter, J. Stahl, and R. Wille, “Conceptual measurement and many–valued contexts,” in *Classification as a tool of research*, 1986, pp. 169–176.

[7] N. Pernelle, M. Rousset, H. Soldano, and V. Ventos, “Zoom: a nested galois lattices-based system for conceptual clustering,” *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 14, no. 2-3, pp. 157–187, 2002.

[8] C. Carpineto and G. Romano, “Using concept lattices for text retrieval and mining,” in *Formal Concept Analysis*, vol. 3626, 2005, pp. 161–179.

[9] L. Szathmary and A. Napoli, “Knowledge organisation and information retrieval with galois lattices,” in *Engineering Knowledge in the Age of the Semantic Web*, vol. 3257, 2004, pp. 511–512.

[10] G. Salton, A. Wong, and C. Yang, “A vector space model for automatic indexing,” in *Communications of the ACM*, vol. 18, 1975, pp. 613–620.

[11] S. Alouane, M. Sassi, and K. Barkaoui, “A parallel comparator of documents,” in *proceedings of the* $23^{th}$ *International Conference on Database and Expert Systems Applications*, 2013, pp. 48–52.

[12] L. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, pp. 338–353, 1965.

[13] J. Konecny and P. Osicka, “General approach to triadic concept analysis,” in *The International Conference on Concept Latices and Their Applications*, 2010, pp. 116–126.

[14] H. Chim and X. Deng, “Efficient phrase-based document similarity for clustering,” in *Knowledge and Data Engineering, IEEE Transactions*, vol. 20, 2008, pp. 1217–1229.

[15] M. Sven, L. Jorg, and N. Hermann, “Algorithms for bigram and trigram word clustering,” in *Speech Communication*, vol. 24, 1998, pp. 19–37.

[16] C. Trabelsi, N. Jelassi, and S. B. Yahia, “Auto-complétion de requêtes par une base générique de règles d’association triadiques,” in *COnférence en Recherche d’Infomations et Applications*, 2011, pp. 9–24.

[17] J. Bezdek, “Fcm: The fuzzy c-means clustering algorithm,” *Computers et Geosciences*, vol. 2-3, no. 10, pp. 191–203, 1984.

[18] F. Lehmann and R. Wille, “A triadic approach to formal concept analysis,” in *Conceptual Structures: Applications, Implementation and Theory*, 1995, pp. 32–43.

[19] R. Missaoui and L. Kwuida, “Mining triadic association rules from ternary relations,” in *Formal Concept Analysis*, vol. 6628, 2011, pp. 204–218.

[20] T. Quan, S. Hui, and T. Cao, “A fuzzy fca-based approach to conceptual clustering for automatic generation of concept hierarchy on uncertainty data,” in *Concept Lattices and Their Applications Workshop*, 2004, pp. 1–12.

[21] R. Godin, R. Missaoui, and H. Alaoui, “Incremental concept formation algorithms based on galois lattices,” *Computational Intelligence*, vol. 11, no. 2, pp. 246–267, 1995.