A Semi-automatic Method to Fuzzy-Ontology Design by using Clustering and Formal Concept Analysis

Amira Aloui
Université Tunis El Manar
Ecole Nationale d'Ingénieurs de Tunis
LR-SITI
Tunis,Tunisia
aloui_amira@yahoo.fr

Alaa Ayadi
Université Tunis El Manar
Ecole Nationale d'Ingénieurs de Tunis
LR-SITI
Tunis,Tunisia
alaa.ayadi@gmail.com

Amel Grissa-Touzi
Université Tunis El Manar
Ecole Nationale d'Ingénieurs de Tunis,
Faculté des Sciences de Tunis,
LIPAH
Tunis,Tunisia
amel.touzi@enit.rnu.tn

Abstract—**Ontology design is a complex and time-consuming process. It is extremely difficult for experts to discover ontology from given data or texts. This paper presents a semi-automatic method for Fuzzy Ontology extraction and Design (FOD). The method is based on conceptual clustering, fuzzy logic and Formal Concept Analysis (FCA). The FOD approach starts with the organization of the data in homogeneous clusters having common properties which allows to deduce the data's semantic. Then, it models these clusters by an extension of the FCA. This lattice will be used to build a core of ontology that is represented as a set of fuzzy rules. Ontology designer is given this initial ontology expression for further extension by adding concepts and relationships (part-of, related to, etc.). To validate our approach, we used Protégé 4.3, that support the fuzzy concept and generates automatically the script in fuzzy-OWL 2 language.**

Keywords-Data Mining; Clustering; Formal Concept Analysis; Fuzzy Logic; Ontology; Fuzzy OWL2.

I. INTRODUCTION

Manual construction and description of field-specific ontology is a complex and time-consuming procedure. The recent study of ontology design methodologies shows that it is very difficult for a designer to create a precise and consistent ontology [1]. Many researchers in the field of data mining have tried to build an ontology for data mining that intended to solve some specific problems. Most of the developments aimed to automate the planning of data mining workflows [2][3]. Some of them are concerned with the description of the data mining services on the grid [4]. Others explored the possible interactions among FCA and Ontology in the Semantic Web [5] and the text documents [6] fields. The problem of these ontologies is that they are not constructed to describe the complete domain of data mining, but are simply made with a specific task in mind. Accordingly, the limits of these approaches reside in the extraction of this ontology starting from the data or a data variety, which may be huge. The goal of this paper is to present a new semi-automatic approach to extract ontology using clustering and FCA combined with a fuzzy rule-based language[19].

Our approach provides tools for semi-automatic extraction of taxonomy and automatic transformation of initial ontology to fuzzy rules. Validation of ontology is done by using Protégé 4.3 [15].

Thus, we propose a new approach for generating an ontology which takes into consideration another degree of granularity in the process of this generation. Indeed, we propose to define an ontology between classes resulting from a preliminary classification of the data and not from the initial large amount of data. We have proven that this approach optimizes the definition of the ontology, offers a better interpretation of the data and optimizes both the space memory and the time spent on data exploiting.

The remainder of the paper is formed as follows: Section 2 introduces the basic concepts of ontology and FCA. Section 3 presents related work; Section 4 presents our motivation for this work. Section 5 describes our new approach for the semi-automatic generation of Fuzzy Ontology of Data Mining, called FODM. Section 6 validates our approach and represents some applications using the generated fuzzy ontology. Section 7 enumerates the advantages of the proposed approach. We finish this paper with a conclusion and a presentation of some future works.

II. BASIC CONCEPTS

In this section, we present the basic concepts of ontology and FCA.

A. Ontologies

Ontologies [7] are content theories about the classes of individuals, properties of individuals, and relations between individuals that are possible in a specified domain of knowledge. They set the terms for describing our knowledge around the field. An ontology of a domain is beneficial in establishing a common vocabulary describing the domain of interest. This is important for the unification and the sharing of knowledge about the domain and connecting with other domains. In reality, there is no common formal definition of what an ontology is. All the same, most approaches share a few core items, such as: concepts, a hierarchical IS-A-relation, and further relations. For the sake of generality, we do not discuss more specific features like constraints, functions, or axioms in this paper, instead we formalize the core in the following way:

Definition: A (core) ontology is a tuple **O = (C, is_a, R, σ)** where

 ISBN: 978-1-61208-334-6

- ***C*** is a set, whose elements are called *concepts*
- **is_a** is a partial order on ***C*** (I. e., a is a binary relation **is_a (*C, X C*** which is reflexive, transitive, and anti symmetric),
- **R** is a set whose elements is called *relation names* (or *relations* for short),
- **σ :** $R \rightarrow C^{+}$ is a function which assigns to each relation name its arity.

In the last years, several languages have been developed to describe ontologies. For instance, the Ontology Web Language (OWL) [8] and extension of OWL language like OWL 2 [9] or Fuzzy OWL [10]. Likewise, the number of environments and tools for building ontologies has grown exponentially. These tools aimed to provide support for the ontology's development process and for the subsequent ontology usage. Among these tools, the most relevant are: Ontolingua [11], WebODE [12], Protégé-2000 [13], OntoEdit [14] and OilEd [15].

B. Formal concept analysis (FCA)

FCA is a method of data analysis, knowledge representation and information management. It was suggested by Rudolf Wille in 1982 [16]. In late years, FCA has grown into an international research community with applications in many fields, such as linguistics, software technology, psychology, medicine, AI, database, library science, environmental science, information retrieval, ontology building, etc. FCA starts with the concept of a formal context specifying which objects have attributes and thus a formal context may be viewed as a binary relation between the object set and the attribute set. In [17], an ordered lattice extension theory has been proposed: Fuzzy Formal Concept Analysis (FFCA), in which uncertainty information is directly represented by a real number of membership values in the range of [0,1], then the intersection of these membership values should be the minimum of these membership values, according to fuzzy theory [18]. This number is equal to the similarity defined as follows:

Definition. The similarity of a fuzzy formal concept $C_1 = (\varphi(A_1), B_1)$ and its subconcept $C_2 = (\varphi(A_2), B_2)$ is defined as:

$$S(C_1, C_2) = \frac{|\varphi(A_1) \cap \varphi(A_2)|}{|\varphi(A_1) \cup \varphi(A_2)|} \quad (1)$$

In (1), $\cap$ and $\cup$ refer to the intersection and union operators on fuzzy sets [18], respectively. In [19], we showed that these FFCA are very powerful in the interpretation of the results of the Fuzzy Clustering as well as in the optimization of the flexible query.

III. RELATED WORK

Usually, the ontology building is performed manually, but researchers try to build an ontology automatically or semi automatically to save the time and the efforts of building the ontology. We survey in this section the most important approaches that generate ontologies from data.

Clerkin et al. used concept clustering algorithm (COBWEB) to automatically discover and generate ontology. They argued that such an approach is highly appropriate to domains where no expert knowledge exists, and they proposed how they might use software agents to collaborate, as a substitute to human beings, in the construction of shared ontologies [20]. Blaschke et al. presented a methodology that creates structured knowledge for gene-product function directly from the literature. They apply an iterative statistical information extraction method combined with the nearest neighbor clustering to create an ontology structure [21]. FCA is an efficient technique that can formally abstract data as conceptual structures [22]. Quan et al. proposed to incorporate fuzzy logic into FCA to enable FCA to deal with uncertainty in data and interpret the concept hierarchy reasonably, the proposed framework is known as FFCA. They use FFCA for automatic generation of ontology for scholarly Semantic Web [23]. Dahab et al. presented a framework for constructing ontology from natural English text namely TextOntEx. TextOntEx constructs ontology from natural domain text using semantic pattern-based approach, and analyzes natural domain text to extract candidate relations, then maps them into a meaning representation to facilitate ontology representation [24]. Wuermli et al. used different ways to build ontologies automatically, based on data mining outputs represented by rule sets or decision trees. They used the semantic web languages, RDF, RDF-S and DAML+OIL for defining ontologies [25].

IV. MOTIVATION

The motivation for developing an ontology of data mining is multi-fold.

- The area of data mining is rapidly developing and one of the most challenging problems deals with developing a general framework for data mining. By developing an ontology of data mining, we are taking one step towards solving this problem.
- There exist several proposals for ontology of data mining, but all of them are light-weight, aimed at covering a particular use-case in data mining and are of a limited scope and highly use-case dependent.

Accordingly, we would argue that the limits of these approaches are due to the extraction of this ontology departing from the data or a data variety, which may be huge. To solve all these problems, we propose a new approach for generation of the ontology using conceptual clustering, fuzzy logic, and FCA. Indeed, we propose to define an ontology between classes resulting from a preliminary classification of the data. The data classification is to divide a data set into subsets, called classes, so that all data in the same class are similar and data from different classes are dissimilar.

V. PRESENTATION OF THE FUZZY ONTOLOGY DESIGN: FOD

A. Principle of the FOD

In this section, we present the architecture of the Fuzzy Ontology Design (FOD) approach and the process for building fuzzy ontology. Our FOD approach takes the database records and provides the corresponding Fuzzy Ontology Design; Figure 1 shows the proposed approach.

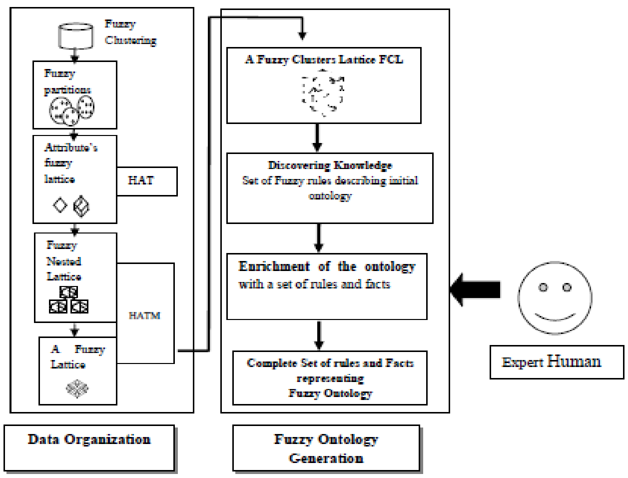

Figure 1. Presentation of the Fuzzy Ontology of Data Mining approach

The FODM approach is organized according to two principal steps: Data Organization step and Fuzzy Ontology Generation step. In the following we describe each step of the method in more detail.

B. *Theoretical Foundation of the FOD model*

In this part, we present the theoretical foundations of the proposed approach, based on the following properties:

Property 1. The number of clusters generated by a clustering algorithm is always lower than the number of starting objects to which one applies the clustering algorithm.

- All objects belonging to one same cluster have the same proprieties. These characteristics can be easily deduced knowing the center and the distance from the cluster.
- The size of the lattice modeling the properties of the clusters is lower than the size of the lattice modeling the properties of the objects.
- The management of the lattice modeling the properties of the clusters is optimum than the management of the lattice modeling the properties of the objects.

Property 2. Let C1, C2 be two clusters, generated by a clustering algorithm and verifying the properties p1 and p2 respectively. Then the following properties are equivalent:

$$C1 \Rightarrow C2 \ \ (CR)$$

$$\Leftrightarrow$$

- $\forall$ object O1 $\in$ C1 => O1 $\in$C2 (CR)
- $\forall$ object O1 $\in$ C1, O1 checks the property p1 of C1 and the property p2 of C2. (CR)

Property 3. Let C1, C2 and C3 are three clusters generated by a classification algorithm and verifying the properties p1, p2 and p3 respectively. Then the following properties are equivalent: C1, C2 = > C3 (CR)

$$\Leftrightarrow$$

- $\forall$ object O1 $\in$ C1 $\cap$ C2 = > O1 $\in$C3 (CR)
- $\forall$ object O1$\in$ C1 $\cap$ C2 then O1 checks the properties p1, p2 and p3 with (CR).

The validation of the two properties rises owing to the fact that all objects which belong to a same cluster check necessarily the same attribute as their cluster.

C. *Data Organization Step*

This step allows us to organize the database records in homogeneous clusters having common properties. This step gives a certain number of clusters for each attribute. Each tuple has values in the interval [0,1] representing these membership degrees according to the formed clusters. We propose to leave the fuzzy formal context, to apply an α-Cut (2) to the set of the degrees of membership, to replace them by values 1 and 0 and to deduce the binary reduced formal context. We define α-Cut as follow:

Definition. *alpha-cut* We define the cut, noted α-Cut, on the fuzzy context as being the reverse of the number of clusters obtained.

$$\alpha^{\text{-Cut}} = (c)^{-1} \qquad (2)$$

Linguistic labels, which are fuzzy partitions, will be assigned to the attribute's domain. This step consists of generating the relieving attributes for the fuzzy concept [19] lattices noted as TAH's and the fuzzy nested lattice noted as MTAH's. This step is very important in the FOD process because it allows us to define and interpret the distribution of objects in the various clusters.

Example: Let a relational database table presented in Table I containing the list of AGE and SALARY of Employee.

TABLE I. A RELATIONAL DATABASE TABLE.

	SALARY	AGE
t1	800	30
t2	600	35
t3	400	26
t4	900	40
t5	1000	27
t6	500	30

TABLE II. FUZZY CONCEPTUAL SCALES FOR AGE AND SALARY ATTRIBUTES

	SALARY			AGE	
	C1	C2	C3	C4	C5
t1	0.1	0.5	0.4	0.5	0.5
t2	0.3	0.6	0.1	0.4	0.6
t3	0.7	0.2	0.1	0.7	0.3
t4	0.1	0.4	0.5	0.2	0.8
t5	-	0.5	0.5	0.6	0.4
t6	0.5	0.5	-	0.5	0.5

Table II shows the results of fuzzy clustering (using Fuzzy C-Means [26]) applied to Age and Salary attributes. For *Salary* attribute, fuzzy clustering generates three clusters (C1, C2 and C3). For *AGE* attribute, two clusters have been generated (C4 and C5).

In our example, α-Cut (Salary) = 0.3 and α-Cut (Age) = 0.5; so, the Table II can be rewritten, as show in Table III.

TABLE III. FUZZY CONCEPTUAL SCALES FOR AGE AND SALARY ATTRIBUTES WITH $\alpha - Cut$.

	SALARY			AGE	
	C1	C2	C3	C4	C5
t1	-	0.5	0.4	0.5	0.5
t2	0.3	0.6	-	-	0.6
t3	0.7	-	-	0.7	-
t4	-	0.4	0.5	-	0.8
t5	-	0.5	0.5	0.6	-
t6	0.5	0.5	-	0.5	0.5

The minimum value (maximal, respectively) of each cluster corresponds to the lower (resp. higher) interval terminal of its values. The corresponding SALARY TAH of fuzzy context presented in Table III are given by the line diagrams presented in Figure 2.

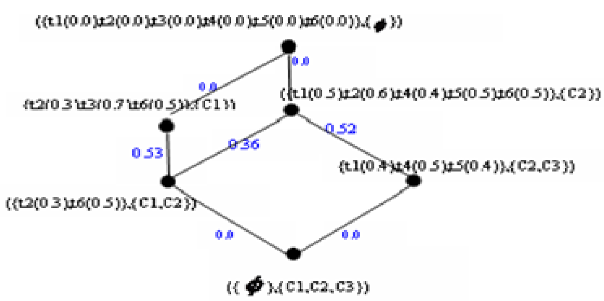

Figure 2. Salary TAH

Each cluster of a partition is labeled with a ***linguistic label*** provided by the user or a domain expert. For example, the fuzzy labels *young* and *adult* could belong to a partition built over the domain of the attribute *AGE*. Besides, the fuzzy labels *low*, *Medium* and *High*, could belong to a partition built over the sphere of the attribute Salary. Table IV presents the correspondence of the linguistic labels and their designations for the attributes Salary and Age. The corresponding fuzzy concept lattices of fuzzy context are shown in Table V.

TABLE IV. CORRESPONDENCE OF THE LINGUISTIC LABELS AND THEIR DESIGNATIONS

Attribute	Linguistic labels	Designation
Salary	Low	C1
Salary	Medium	C2
Salary	High	C3
Age	Young	C4
Age	Adult	C5

TABLE V. FUZZY CONCEPTUAL SCALES FOR AGE AND SALARY ATTRIBUTES WITH $\alpha - Cut$.

	SALARY			AGE	
	Low C1	Medium C2	High C3	Young C4	Adult C5
t1	-	0.5	0.4	0.5	0.5
t2	0.3	0.6	-	-	0.6
t3	0.7	-	-	0.7	-
t4	-	0.4	0.5	-	0.8
t5	-	0.5	0.5	0.6	-
t6	0.5	0.5	-	0.5	0.5

This very simple sorting procedure gives us for each many-valued attribute the distribution of the objects in the line diagram of the chosen fuzzy scale. Usually, we are interested in the interaction between two or more fuzzy many-valued attributes. This interaction can be visualized using the so-called fuzzy nested line diagrams. It is used for visualizing larger fuzzy concept lattices, and combining fuzzy conceptual scales on-line. Figure 3 shows the fuzzy nested lattice constructed from TAH's.

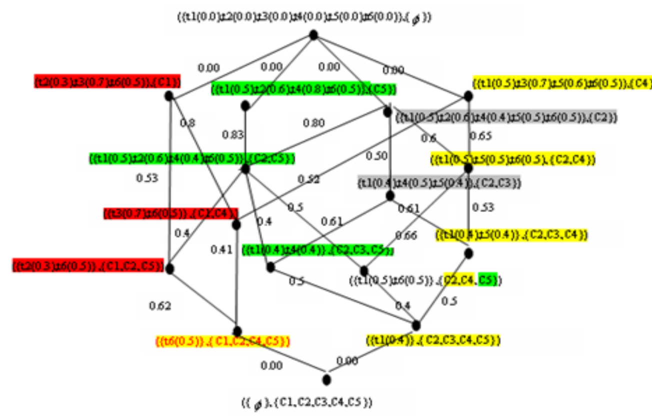

Figure 3. Fuzzy Lattice: MTAH

D. *Fuzzy Ontology Generation step*

This step consists of the construction of a Fuzzy Ontology from the Fuzzy Cluster Lattice generated in the first step.

1) *FCL Generation.*

The goal of this phase is to make a certain abstraction on the list of the objects with their degrees of membership in the clusters. This lattice will be used to build a core of ontology.

Definition. A Fuzzy Clusters Lattice (FCL) of a Fuzzy Formal Concept Lattice, consists on a Fuzzy concept lattice where each equivalence class (a node of the lattice) contains only the intentional description (intent) of the associated fuzzy formal concept.

Definition. A level i of FCL is a is the set of nodes of FCL with cardinality equal to i.

We do a certain abstraction of the list of the objects with their degrees of membership in the clusters. The nodes of FCL are the clusters ordered by the inclusion relation.

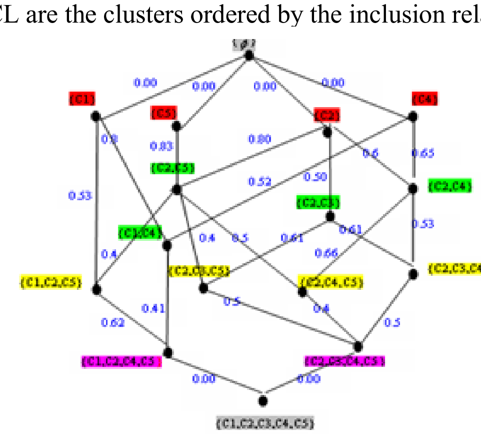

Figure 4. Fuzzy Clusters Lattice FCL

As shown in Figure 5, we obtain a lattice more reduced, simply traversed and stored.

2) Discovering Knowledge

This step consists in the Extraction of knowledge for the fuzzy ontology. To do so, we must build a concept of hierarchy from the conceptual clusters; we need to find the hierarchical relations from the clusters. We first define a concept hierarchy as follows:

Definition (Concept Hierarchy). A concept hierarchy is a poset (partially ordered set) (H,<), where H is a finite set of concepts and < is a partial order on H.

a) Principle of discovering knowledge from FCL.

Taking as input an FCL, the extraction of fuzzy association rules can be performed straightforwardly. Indeed, the rule represents the implications deduced from FCL between two adjacent classes. The confidence factor will be equal to the weight of the link (arc)between the two nodes.

Rule 1. Discovering rule: Let C1= {A1.., A} and C2= {B1.., BM} two nodes of FCL such as C2 is the successors of C1 in the lattice and having as distance d>0 (weight of the arc) the generated rule will be defined as follows:

A1,...,An => B1,...,Bm (*d*)

Notice that, if d=0 this implies that there is no object in common to the two concepts C1, C2. There is no knowledge to discover or to generate.

Rule 2. Discovering rule: Let C1= {A1.., An} and C2= {B1.., BM} two nodes of FCL such as C2 is the successors of C1 in the lattice and having as distance d>0 (weight of the arc). The generated rule will be defined by:

R: A1,.., An => C1,.., Cq (d) such that

{C1,…, Cq} = {B1.., Bm}\{A1.., A} (∀ Ci, Ci ({A1,… An })

Rule 3. Generated rule: Let C1={A1.., An} C2={B1.., Bn} and C3={D1.., Dn} three concepts such as C2 successors of C1 and C2 successor of C3 having respectively as distance d1 and d2. The generated rule will be defined by:

R 3: A1,..,An => D1,..,Dn (d2*d1)

b) Algorithm for Discovering Fuzzy Association rules.

The pseudo-code for this algorithm is as follows:

```
Generating knowledge
Input: Fuzzy Cluster Lattice FCL
Output : FRS: Fuzzy Rules Set
Begin
FRS :=∅
Find_Cluster(FCL, 0, S) ;
 For each subconcept Cj ={y1,▮ym}of
Ci in S
     r.premise= ∅
     r.conclusion={y1,y2,▮ym}
      r.CF=1
      FRS=FRS ∪ {r}
 End For
Nmax :=Niveau_max(FCL) ;
For Niv :=1 to Nmax-1 do
   Find_Cluster(FCL, i, S) ;
For Ci={x1,▮xm} in S do
For each subconcept Cj ={y1,▮ym}of
Ci and having (d>0)
      r.premise= {x1,x2,▮xn}

r.conclusion={y1,y2,▮ym} \ {x1,x2,▮xn}
                  r.CF=d
                  FRS=FRS ∪ {r}
               End For
    End For
End For
Generate_Rule(FRS,FRS1);
FRS:= FRS ∪ FRS1
End.
```

Figure 5. Algorithm for Discovering Fuzzy Association rules

The Algorithm for Discovering Fuzzy Association rules traverses the search space (FCL) by level to square up the Fuzzy Rules Set (FRS). As input it takes the lattice of Clusters FCL and returns, as output, the list of all Fuzzy Rules Set (FRS) generated. It works as follows: For each non empty node ∈ FCL in descending, it generates all rules with one cluster in conclusion (level 1). Then, it generates the set of all rules with two Clusters in conclusion. The same process is applied to generate conclusions with four clusters, and so on until conclusions with *n* clusters are generated.

Proposition 3.

If the system of extraction rules traverses the search space by the level of the lattice of clusters then no rule generated by this system is redundant (all the generated rules are obligatorily distinct).

Proof. This is due to the fact that from a level to another of the lattice the nodes are obligatorily distinct (by definition even of a level of lattice).

3) Ontology Generation.

This step constructs fuzzy ontology from a fuzzy context using the concept hierarchy created by fuzzy conceptual clustering. This is done based on the characteristic that both FCA and ontology support formal definitions of concepts. Thus, we define the fuzzy ontology as follows:

Definition (Fuzzy Ontology). A fuzzy ontology F_o consists of four elements (C, A^C, R, X), where:

- C represents a set of concepts,
- A^C represents a collection of attribute sets, one for each concept,
- R = (R_T; R_N) represents a set of relationships, which consists of two elements:
  R_N is a set of non-taxonomy relationships and
  R_T is a set of taxonomic relationships.
- Each concept c_i in C represents a set of objects, or instances, of the same kind.
- Each object o_{ij} of a concept c_i can be described by a set of attributes values denoted by $A^C(c_i)$.
- Each relationship $\mathbf{r_i(c_p, c_q, \alpha)}$ in R represents a *fuzzy association* between concepts c_p and c_q, and the instances of such a relationship are pairs of (c_p, c_q) concept objects with confidence α; α is in]0..1].
- Each attribute value of an object or the relationship instance is associated with a fuzzy membership value between [0,1] implying the uncertain degree of this attribute value or relationship.
- X is a set of axioms. Each axiom in X is a constraint on the concept's and relationship's attribute values or a constraint on the relationships between concept objects.

In our approach, we consider the Fuzzy Ontology Lattice as a formal domain-specific ontology. This ontology has all lattice properties, which are useful for ontology sharing, and reasoning. The whole process to create a fuzzy ontology was completed. We may consider nodes as concepts. The name of the concept is a concatenation of an attribute and its label linguistics, in accordance with the correspondence in Table IV. Nevertheless, taxonomic relationships between concepts are present in the lattice.

VI. VALIDATION AND APPLICATION OF GENERATING FUZZY ONTOLOGY

The performance of the proposed algorithm for Discovering Fuzzy Association rules can be measured in order to evaluate the generated ontology. To do this, we evaluate the processing time and the number of rules between two approaches: The first one does not apply the clustering concept and the second uses the formal concepts for structuring and building ontology-based classification.

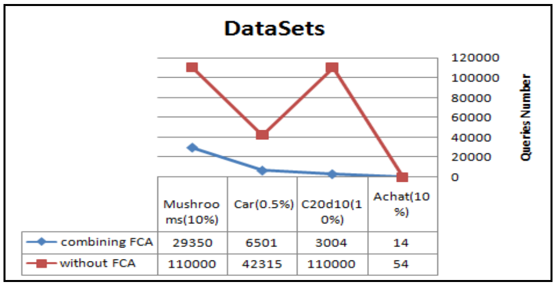

Figure 6. Metrics of the Proposed Approach

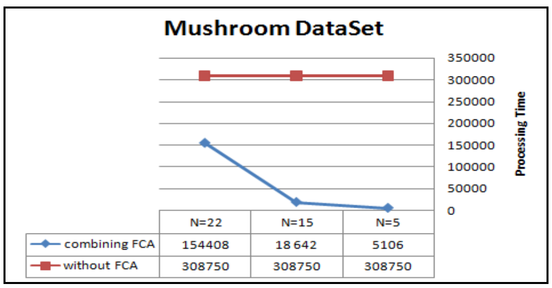

Figure 7. Processing time of the Proposed Approach

We prove that with FCA, we minimize the high time and space complexity of the resulting lattice. We implement, then, the concept lattice (result of fuzzy classification in ClusterFCA) with Protégé 4.3, generate the ontology, test its consistency, and extract the queries. The process of generating ontology is presented in Figure 9.

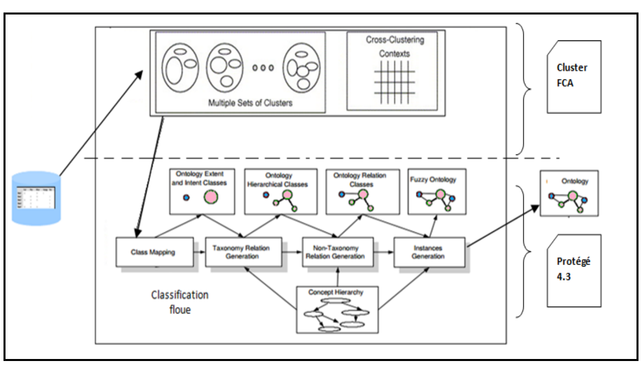

Figure 8. Process of generating/ vaidating Ontology

By taking the abstractions got by FCA as a guideline, the generated ontology in Protégé 4.3 is shown in Figure 10.

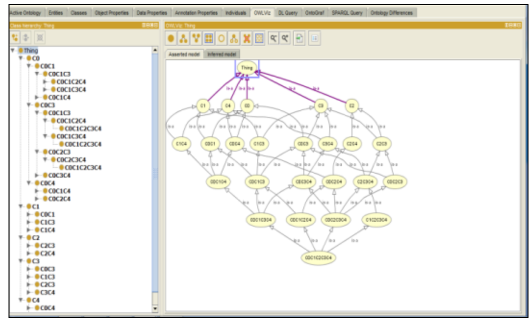

Figure 9. Generated Ontology

Once the queries concepts are defined, we can model the resulting rules deduced from our Fuzzy Ontology using Protégé 4.3 and respond to the user answers. We have also succeeded to generate the description of our ontology with fuzzy-OWL 2 language.

VII. ADVANTAGES OF THE PROPOSED APPROACH

We present in Table VI the advantage of every basic concept used on our approach.

TABLE VI. ADVANTAGES OF THE PROPOSED APPROACH

Operationalization	Advantages	Comments
Using FCL for constructing Ontology	Redundant relation elimination	For each relation concepts, we can have two concept instances which are equivalent. The two concept instances are valid in the sense that concept with a higher membership degree is closer to the concept truth. Eliminating one of the concept relation will not reduce the information conveyed, but will reduce by half the size of the storage. In constructing an ontology we retain the fuzzy relation that has a higher membership degree. This decision strategy will choose a positive concept instance and will choose a stronger relation if the two membership values are close to each other.
	Less meaningful relation elimination	After redundant class relation is removed much potential less meaningful information intact.
	Unrelated concept relation elimination	The relation between two distinct classes cannot be established if both concept never co-occur so that their membership values will be 0. It is obvious that unrelated classes should also not be considered during the ontology creation. These concepts will be automatically excluded by applying alpha-cut as described above.
Using the domain ontology	Less number of generating classes	The number of classes generated is less than the number of objects starting on which we apply the classification algorithm. This improves the quality of the process of information retrieval by considering only a part of the ontology according to a user preference.
	Best answer to the user request	The ontology has been described in OWL2, we took advantage of the progress of this language in terms of expressiveness for greater capacity inference without using a dedicated language to express rules

Nowadays, a few proposals for ontologies of data mining using FCA exist, but all of them start from a data unit, after having done a data cleansing step and an elimination of invalid-value elements. We have come to the conclusion that this idea is very important because it models an abstraction of the data especially in the case of voluminous one.

VIII. CONCLUSION

Motivated by the increased need for formalized representations of the domain of Data Mining, the success of using FCA and Ontology in several Computer Science fields, we presented in this paper a new approach for the semi automatic generation of Fuzzy Ontology Design (*FOD)*, through the fusion of conceptual clustering, fuzzy logic, and FCA. In our approach, we proposed to generate an ontology taking into consideration another degree of granularity in the process of generation. Indeed, we suggest to define an ontology between classes resulting from a preliminary classification of the data. We prove that this approach optimizes the definition of the ontology, offers a better interpretation of the data and optimizes both the space memory and the execution time for exploiting this data. To validate our approach, we used Protégé 4.3, which supports the fuzzy concept, to model our ontology and to generate the script in fuzzy-OWL 2 language.

Knowing that the number of classes has been always lower than the number of starting data, our proposed approach intends to achieve the objectives of offering better interpretation of the data and minimizing both execution time and space memory (by reducing considerably the definition of the ontology). As future perspectives of this work, we intend to test our approach on several large datasets.

REFERENCES

[1] C. Tempich and R. Volz, "Towards a benchmark for Semantic Web reasoners-an analysis of the DAML ontology library," Sure Y (editor) Proceedings of Workshop of Evaluation of Ontology-based Tools (EON 2003) at 2nd Int. Semantic Web Conference (ISWC 2003), USA, (2003) .

[2] A. Bernstein, F. Provost, and S. Hill, "Toward intelligent assistance for a data mining process: An ontology-based approach for cost-sensitive classification", IEEE Trans on Knowl and Data Eng, 2005, pp. 503–518.

[3] M. Zakova, P. Kremen, F. Zelezny, and N. Lavrac, "Planning to learn with a knowledge discovery ontology," In P. Brazdil, A. Bernstein, and L. Hunter, editors, Proceedings of the Second Planning to Learn Workshop (PlanLearn) at the ICML/COLT/UAI, 2008, pp. 29–34.

[4] P. Brezany, I. Janciak, and A. M. Tjoa, "Data Mining with Ontologies Implementations, Findings and Frameworks," chapter Ontology-Based Construction of Grid Data Mining Workflows. IGI Global, 2007.

[5] Q. T. Tho, S. C. Hui, A. C. M. Fong, and Cao, T.H., "Automatic fuzzy ontology generation for semantic web", Knowledge and Data Engineering, IEEE Transactions on, vol. 18, no. 6, 2006, pp. 842-856.

[6] P. Cimiano, A. Hotho, G. Stumme, and J. Tane, "Conceptual knowledge processing with formal concept analysis and ontologies," In ICFCA, 2004, pp. 189-207.

[7] B. Chandrasekaran, J. R. Josephson, and V. R. Benjamins, "What are ontologies, and why do we need them?," IEEE Intelligent Systems, 1999, pp. 20–26.

[8] S. Bechhofer et al. , "OWL Web Ontology Language: Reference". World Wide Web Consertium, February. 2004.

[9] B. Cuenca-Grau, I. Horrocks, B. Motik, B. Parsia, P. F. Patel-Schneider, and U. Sattler, "OWL 2: The next step for OWL", Journal of Web Semantics, 2008, pp. 309-322.

[10] F. Bobillo and U. Straccia, "Representing fuzzy ontologies in OWL 2," in: Proceedings of the 19th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2010), IEEE Press, 2010, pp. 2695–2700.

[11] A. Farquhar, R. Fikes, and J. Rice, "The ontolingua server: a tool for collaborative ontology construction," In the 10th Knowledge Aqcuisition for Knowledge-Based Systems (KAW'96), Canada, 1996.

[12] J. Arpirez, O. Corcho, M. Fernández-López and A. Gómez-Pérez). "WebODE , a Workbench for Ontological Engineering", In First international Conference on Knowledge Capture (K-CAP'01), Victoria, Canada ACM, 2001, pp. 6–13.

[13] N. Noy, R. Fergerson, and M. Musen, "The knowledge model of Protégé2000 : Combining interoperability and flexibility," In R. D IENG & O.CORBY, Eds., 12th International Conference on Knowledge Engineering and Knowledge Management (EKAW'00), volume (1937) of Lecture Notes in Computer Science, Juan-les-Pins, France: Springer Verlag, pp. 17–32.

[14] Y. Sure, M. Erdmann, J. Angele, S. Staab, R. Studer, and D. Wenke, "OntoEdit: Collaborative Ontology Engineering for the Semantic Web", In I. Horrocks & J. Hendler, Eds., First International Semantic Web Conference (ISWC'02), volume (2342) of Lecture Notes in Computer Science, Chia, Sardaigne, Italie: Springer Verlag. 2002, pp. 221–235

[15] S. Bechhofer, I. Horrocks, C. Goble, and R. Stevens, "OilEd: a Reason-able Ontology Editor for the Semantic Web", In Joint German/Austrian conference on Artificial Intelligence (KI'01), volume (2174) of Lecture Notes in Artificial Intelligence, Vienne, Austria: Springer Verlag, 2001, pp. 396–408.

[16] R. Wille, "Restructuring lattice theory: An approach based on hierarchies of concepts", In I. Rival (Ed.), Ordered sets, 1982, pp .445–470.

[17] T. T. Quan, S. C. Hui, and T. H. Cao, "A Fuzzy FCA-based Approach to Conceptual Clustering for Automatic Generation of Concept Hierarchy on Uncertainty Data", Proc. of the 2004 Concept Lattices and Their Applications Workshop (CLA), pp. 1-12, 2004.

[18] L. A. Zadeh, "Fuzzy Logic and Approximate Reasoning," Synthese, vol. 30, 1975, pp. 407-428.

[19] A. Grissa Touzi, M. Sassi, and H. Ounelli, "An innovative contribution to flexible query through the fusion of conceptual clustering, fuzzy logic, and formal concept analysis", International Journal of Computers and Their Applications. Vol. 16, N 4, December. 2009, pp. 220-233.

[20] P. Clerkin, P, P. Cunningham, and C. Hayes, "Ontology Discovery for the Semantic Web Using Hierarchical Clustering" , Proc. European Conf. Machine Learning (ECML) and European Conf. Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD-2001), 2001.

[21] C. Blaschke and A. Valencia, "Automatic Ontology Construction from the Literature", Genome Informatics, vol. 13, 2002, pp. 201–213.

[22] B. Ganter, G. Stumme, and R. Wille, "Formal Concept Analysis", Foundations and Applications. Lecture Notes in Artificial Intelligence, no.3626, Springer-Verlag. ISBN 3-540-27891-5. (Eds.) 2005.

[23] T. T. Quan, S. C. Hui, A. C. M. Fong, and T. H. Cao, "Automatic generation of ontology for scholarly semantic Web", In: Lecture Notes in Computer Science. Vol. 3298, 2004 , pp. 726–740.

[24] M. Y. Dahab, H. Hassan, and A. Rafea, "TextOntoEx: Automatic ontology construction from natural English text", Expert Systems with Applications (2007), doi:10.1016/j.eswa.2007.01.043.

[25] O. Wuermli, A. Wrobel, S. C. Hui, and J. M. Joller ,"Data Mining For Ontology_Building: Semantic Web Overview", Diploma Thesis–Dep. of Computer Science_WS2002/2003, Nanyang Technological University.

[26] H. Sun, S. Wanga, and Q. Jiangb, "FCM-Based Model Selection Algorithms for Determining the Number of Clusters", Pattern Recognition 37, pp. 2027-2037.