Towards Fuzzy Granulation in OWL Ontologies

Francesca A. Lisi and Corrado Mencar

Dipartimento di Informatica, Università degli Studi di Bari "Aldo Moro", Italy
{francesca.lisi,corrado.mencar}@uniba.it

Abstract. The integration of fuzzy sets in ontologies for the Semantic Web can be achieved in different ways. In most cases, fuzzy sets are defined by hand or with some heuristic procedure that does not take into account the distribution of available data. In this paper, we describe a method for introducing a granular view of data within an OWL ontology.

1 Introduction

Endowing OWL ontologies with capabilities of representing and processing imprecise knowledge is a highly desirable feature since the Semantic Web is full of imprecise and uncertain information coming from perceptual data (*i.e.*, data coming from subjective judgments), incomplete data, data with errors, etc. [16]. Moreover, even in the case that precise information is available, imprecise knowledge could be advantageous: tolerance to imprecision may lead to concrete benefits such as compact knowledge representation, and efficient and robust reasoning [20]. Additionally, humans continually acquire, manipulate and communicate imprecise knowledge: therefore any ontology capable of expressing imprecise knowledge, when a precise alternative leads to a complex representation, could be more *interpretable* by human users, i.e. easier to read and understand [1].

A number of mathematical tools are available to deal with imprecision and uncertainty in knowledge representation. The choice of the right tool depends on the type of imprecision. In particular, imprecision due to the lack of boundaries in concepts (such as coldness in the domain of indoor temperatures, interestingness of movies, etc.) are well modeled through fuzzy sets. In essence, fuzzy sets define collections of objects whose membership can be partial. Differently to probability measures, the degree of membership does not measure how likely an object is referred by a concept, but rather it quantifies how much the concept is applicable to the object. Fuzziness pervades human reasoning and allows it to intelligently act in complex environments: since fuzzy sets make possible a computational representation of concepts with no sharp boundaries, they enable machines to carry out human-centered information processing and reasoning [5].

The integration of fuzzy sets in ontologies for the Semantic Web can be achieved in different ways (see [19] for an updated overview). However, in most cases, fuzzy sets are defined by hand or with some heuristic procedure that does not take into account the distribution of available data. In this paper, we propose the adoption of a fuzzy clustering procedure to automatically acquire fuzzy sets from data. Also, we exploit the resulting clusters, together with fuzzy quantifiers, to develop a granular view of the individuals in an OWL ontology.

The paper is structured as follows. Section 2 presents some preliminary information on Description Logics[1] (2.1), Fuzzy Set Theory (2.2), and Fuzzy DLs (2.3). Section 3 describes the proposed granulation method on OWL schemas at increasing levels of complexity. Finally, Section 4 draws some conclusive remarks along with future research directions.

2 Preliminaries

2.1 Description Logics

Description Logics (DLs) are a family of decidable First Order Logic (FOL) fragments that allow for the specification of structured knowledge in terms of classes (*concepts*), instances (*individuals*), and binary relations between instances (*roles*) [2]. Complex concepts (denoted with C) can be defined from atomic concepts (A) and roles (R) by means of the constructors available for the DL in hand. The members of the DL family differ from each other as for the set of constructors, thus for the complexity of concept expressions they can generate. For the sake of illustrative purposes, we present here a salient representative of the DL family, namely $\mathcal{ALC}$ [15], which is often considered to illustrate some new notions related to DLs. A DL *Knowledge Base* (KB) $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ is a pair where $\mathcal{T}$ is the so-called *Terminological Box* (TBox) and $\mathcal{A}$ is the so-called *Assertional Box* (ABox). The TBox is a finite set of *General Concept Inclusion* (GCI) axioms which represent is-a relations between concepts, whereas the ABox is a finite set of *assertions* (or *facts*) that represent instance-of relations between individuals (resp. couples of individuals) and concepts (resp. roles). Thus, when a DL-based ontology language is adopted, an ontology is nothing else than a TBox (*i.e.*, the intensional level of knowledge), and a populated ontology corresponds to a whole KB (*i.e.*, encompassing also an ABox, that is, the extensional level of knowledge).

The semantics of DLs can be defined directly with set-theoretic formalizations or through a mapping to FOL (as shown in [8]). Specifically, an *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ for a DL KB consists of a domain $\Delta^{\mathcal{I}}$ and a mapping function $\cdot^{\mathcal{I}}$. For instance, $\mathcal{I}$ maps a concept C into a set of individuals $C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, *i.e.* $\mathcal{I}$ maps C into a function $C^{\mathcal{I}} : \Delta^{\mathcal{I}} \to \{0, 1\}$ (either an individual belongs to the extension of C or does not belong to it). Under the *Unique Names Assumption* (UNA) [13], individuals are mapped to elements of $\Delta^{\mathcal{I}}$ such that $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ if $a \neq b$. However UNA does not hold by default in DLs. An interpretation $\mathcal{I}$ is a *model* of a KB $\mathcal{K}$ iff it satisfies all axioms and assertions in $\mathcal{T}$ and $\mathcal{A}$. In DLs a KB represents many different interpretations, *i.e.* all its models. This is coherent with the Open World Assumption (OWA) that holds in FOL semantics. A DL KB is *satisfiable* if it has at least one model. We also write $C \sqsubseteq_{\mathcal{K}} D$ if in any

[1] We recap that DLs are the logical foundation of the standard for web ontology languages belonging to the OWL 2 family [12].

Table 1. Syntax and semantics of constructs for $\mathcal{ALC}(\mathbf{D})$.

bottom (resp. top) concept	$\bot$ (resp. $\top$)	$\emptyset$ (resp. $\Delta^{\mathcal{I}}$)
atomic concept	A	$A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
abstract role	R	$R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
concrete role	T	$T^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathbf{D}}$
individual	a	$a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
concrete value	v	$v^{\mathcal{I}} \in \Delta^{\mathbf{D}}$
concept intersection	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
concept union	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
concept negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
universal abstract role restriction	$\forall R.C$	$\{x \in \Delta^{\mathcal{I}} \mid \forall y\ (x,y) \in R^{\mathcal{I}} \to y \in C^{\mathcal{I}}\}$
existential abstract role restriction	$\exists R.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y\ (x,y) \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
universal concrete role restriction	$\forall T.\mathbf{d}$	$\{x \in \Delta^{\mathcal{I}} \mid \forall z\ (x,z) \in T^{\mathcal{I}} \to z \in \mathbf{d}^{\mathbf{D}}\}$
existential concrete role restriction	$\exists T.\mathbf{d}$	$\{x \in \Delta^{\mathcal{I}} \mid \exists z\ (x,z) \in T^{\mathcal{I}} \wedge z \in \mathbf{d}^{\mathbf{D}}\}$
general concept inclusion	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
concept assertion	$a : C$	$a^{\mathcal{I}} \in C^{\mathcal{I}}$
abstract role assertion	$(a,b) : R$	$(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$
concrete role assertion	$(a,v) : T$	$(a^{\mathcal{I}}, v^{\mathcal{I}}) \in T^{\mathcal{I}}$

model $\mathcal{I}$ of $\mathcal{K}$, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ (concept C is subsumed by concept D). Moreover we write $C \sqsubset_{\mathcal{K}} D$ if $C \sqsubseteq_{\mathcal{K}} D$ and $D \not\sqsubseteq_{\mathcal{K}} C$. The *consistency check*, which tries to prove the satisfiability of a DL KB $\mathcal{K}$, is the main reasoning task in DLs. It is performed by applying decision procedures mostly based on tableau calculus. All other reasoning tasks can be reformulated as consistency checks.

In many applications, it is important to equip DLs with expressive means that allow to describe "concrete qualities" of real-world objects such as the length of a car. The standard approach is to augment DLs with a so-called *concrete domain* (or *datatype theory*) $\mathbf{D} = \langle \Delta^{\mathbf{D}}, \cdot^{\mathbf{D}} \rangle$, which consists of a datatype domain $\Delta^{\mathbf{D}}$ (*e.g.*, the set of real numbers in double precision) and a mapping $\cdot^{\mathbf{D}}$ that assigns to each data value an element of $\Delta^{\mathbf{D}}$, and to every n-ary datatype predicate $\mathbf{d}$ an n-ary (typically, $n = 1$) relation over $\Delta^{\mathbf{D}}$ [3]. In DLs extended with concrete domains, each role is therefore either *abstract* (denoted with R) or *concrete* (denoted with T). The set of constructors for $\mathcal{ALC}(\mathbf{D})$ is reported in Table 1.

2.2 Fuzzy Set Theory

A *crisp set* A over a Universe of Discourse X is characterised by a function $A\colon X \to \{0,1\}$, that is, for any $x \in X$ either $x \in A$ (*i.e.*, $A(x) = 1$) or $x \notin A$ (*i.e.*, $A(x) = 0$). A *fuzzy set* F over X is characterised by a membership function $F\colon X \to [0,1]$. For a fuzzy set F, unlike crisp sets, $x \in X$ belongs to F to a degree $F(x)$ in $[0,1]$.

The trapezoidal, the triangular, the left-shoulder, and the right-shoulder functions are frequently used as *membership functions* of fuzzy sets (see Fig. 1 for a graphical representation). In particular, the *trapezoidal function* is defined

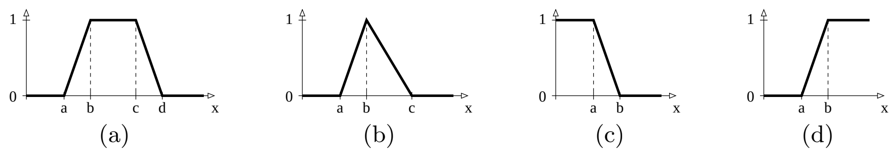

Fig. 1. Four notable membership functions of fuzzy sets: (a) Trapezoidal, (b) triangular, (c) left-shoulder, and (d) right-shoulder.

as follows: let $a < b \leq c < d$ be rational numbers then

$$trz(a,b,c,d)(x) = \begin{cases} 0 & \text{if } x < a \\ (x-a)/(b-a) & \text{if } x \in [a,b) \\ 1 & \text{if } x \in [b,c] \\ (d-x)/(d-c) & \text{if } x \in (c,d] \\ 0 & \text{if } x > d \ . \end{cases} \tag{1}$$

Note that triangular, left-shoulder and right-shoulder fuzzy sets are special cases of trapezoidal fuzzy sets.[2]

Fuzzy sets can be used to represent *information granules*, *i.e.* collections of objects kept together due to their similarity, proximity, etc. [4]. Information granules promote abstraction as far as they can be labeled by symbolic terms. Information granules represented by fuzzy sets are good candidates to represent perceptual information, thus they could be conveniently labeled by linguistic terms coming from natural language. The granularity level (quantifiable as the area of the membership function, for fuzzy sets) assesses the specificity of an information granule: the most specific information granule is a set with a single element (precise information); on the other extreme, an information granule covering the whole universe of discourse is the least specific.

Fuzzy clustering Although fuzzy sets have a greater expressive power than crisp sets, their usefulness depends critically on the capability to construct appropriate membership functions for various given concepts in different contexts. The problem of constructing meaningful membership functions is not a trivial one (see, *e.g.*, [10, Chapter 10]). One easy method is to define a uniform *Strong Fuzzy Partition* (SFP) usually with 5 ± 2 fuzzy sets. A SFP is a collection of fuzzy sets (usually with triangular or trapezoidal fuzzy sets) such that, for each element of the Universe of Discourse, the sum of memberships of all fuzzy sets is always 1. SFPs with trapezoidal fuzzy sets greatly enhance the efficiency of calculations because they guarantee that each element has non-zero membership degree for at most two fuzzy sets. A uniform SFP is based on fuzzy sets with the same granularity. It is very simple to define a uniform SFP, but this approach does not take into account the distribution of available data; in fact, coarse

[2] By convention, whenever the denominator of one of the fractions in (1) is 0, the membership degree is 1.

grained fuzzy sets are more useful to cover regions of the Universe of Discourse where data are more sparse; on the other hand, data crammed in small areas are better represented by more specific (fine grained) fuzzy sets.

The derivation of a SFP with variable granularity, adapted to available data, can be achieved through fuzzy clustering. A widespread algorithm for fuzzy clustering is *Fuzzy C-Means* (FCM) [6], an extension of the well-known K-Means that accommodates partial memberships of data to clusters. FCM, applied to one-dimensional, numerical data, can be used to derive a set of c clusters characterized by prototypes $\pi_1, \pi_2, \ldots, \pi_c$, with $\pi_j \in \mathbb{R}$ and $\pi_j < \pi_{j+1}$. These prototypes, along with the range of data, provide enough information to define a SFP with two trapezoidal fuzzy sets and $c-2$ triangular fuzzy sets according to the following rules:

$$F_j = \begin{cases} trz\,(m, m, \pi_1, \pi_2) & \text{if } j = 1 \\ trz\,(\pi_{j-1}, \pi_j, \pi_j, \pi_{j+1}) & \text{if } 1 < j < c \\ trz\,(\pi_{c-1}, \pi_c, M, M) & \text{if } j = c. \end{cases} \tag{2}$$

where $[m, M]$ is the range of data. In Fig. 2 an example of SFP, consisting of five fuzzy sets with variable granularity, is depicted.

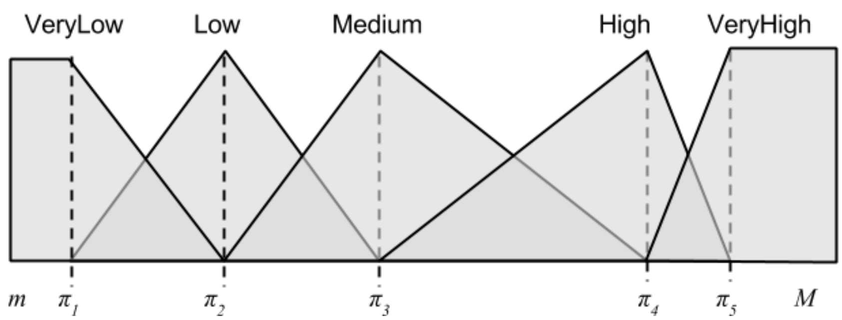

Fig. 2. Example of SFP consisting of $c = 5$ fuzzy sets with variable granularity.

Fuzzy quantifiers Fuzzy sets, like crisp sets, can be quantified in terms of their *cardinality*. Several definitions of cardinality of fuzzy sets are possible [9], although in this paper we consider only relative scalar cardinalities like the *relative σ-count*, defined for a finite Universe of Discourse X as:

$$\sigma(F) = \frac{\sum_{x \in X} F(x)}{|X|} \tag{3}$$

A relative scalar cardinality yields a value within $[0, 1]$ (being $|\emptyset| = 0$ and $|X| = 1$). On this interval, a number of fuzzy sets can be defined to represent granular concepts about cardinalities, such as `Many` (see Fig. 3 for some notable examples). These concepts are called *fuzzy quantifiers*. Note that the usual existential quantifier ($\exists$) and universal quantifier ($\forall$) can be represented as special cases of fuzzy quantifiers: $Q_{\exists}(x) = 1$ iff $x > 0$, 0 otherwise; $Q_{\forall}(x) = 1$ iff $x = 1$, 0 otherwise.

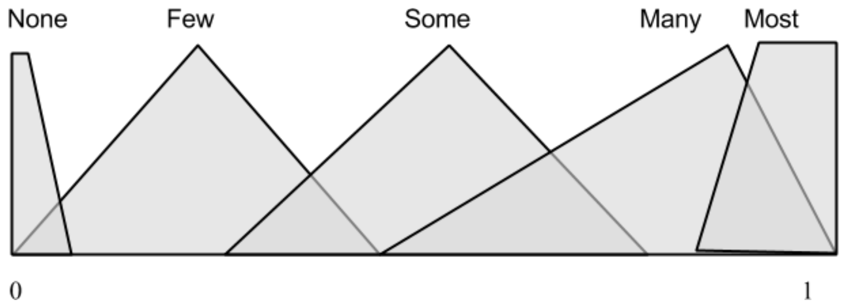

Fig. 3. Some notable fuzzy quantifiers: `None`, `Few`, `Some`, `Many`, and `Most`.

Given a fuzzy quantifier Q and a fuzzy set F, the membership degree $Q(\sigma(F))$ can be intended as the degree of truth of a *fuzzy proposition* of the form "Q`x are` F" (*e.g.*"`Many` x are `Tall`").

2.3 Fuzzy Description Logics

In fuzzy DLs, an *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consists of a nonempty (crisp) set $\Delta^{\mathcal{I}}$ (the *domain*) and of a *fuzzy interpretation function* $\cdot^{\mathcal{I}}$ that, *e.g.*, maps a concept C into a function $C^{\mathcal{I}} : \Delta^{\mathcal{I}} \to [0,1]$ and, thus, an individual belongs to the extension of C to some degree in $[0,1]$, *i.e.* $C^{\mathcal{I}}$ is a fuzzy set. The definition of $\cdot^{\mathcal{I}}$ for $\mathcal{ALC}(\mathbf{D})$ with fuzzy concrete domains is reported in [18]. In particular, $\cdot^{\mathbf{D}}$ maps each concrete role into a function from $\Delta^{\mathbf{D}}$ to $[0,1]$. Typical examples of datatype predicates are

$$\mathbf{d} := ls(a,b) \mid rs(a,b) \mid tri(a,b,c) \mid trz(a,b,c,d) \mid \ \geq_v \ \mid \ \leq_v \ \mid \ =_v \ , \tag{4}$$

where *e.g.* $\geq_v$ corresponds to the crisp set of data values that are greater or equal than the value v.

Axioms in a fuzzy $\mathcal{ALC}(\mathbf{D})$ KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ are graded, *e.g.* a GCI is of the form $\langle C_1 \sqsubseteq C_2, \alpha \rangle$ (*i.e.* C_1 is a sub-concept of C_2 to degree at least α). We may omit the truth degree α of an axiom; in this case $\alpha = 1$ is assumed. An interpretation $\mathcal{I}$ *satisfies* an axiom $\langle \tau, \alpha \rangle$ if $(\tau)^{\mathcal{I}} \geq \alpha$. $\mathcal{I}$ is a *model* of $\mathcal{K}$ iff $\mathcal{I}$ satisfies each axiom in $\mathcal{K}$. We say that $\mathcal{K}$ *entails* an axiom $\langle \tau, \alpha \rangle$, denoted $\mathcal{K} \models \langle \tau, \alpha \rangle$, if any model of $\mathcal{K}$ satisfies $\langle \tau, \alpha \rangle$. Further details of the reasoning procedures for fuzzy DLs can be found in [17].

Fuzzy quantifiers have been also studied in fuzzy DLs. In particular, Sanchez and Tettamanzi [14] define an extension of fuzzy $\mathcal{ALC}(\mathbf{D})$ involving fuzzy quantifiers of the absolute and relative kind, and using qualifiers. They also provide algorithms for performing two important reasoning tasks with their DL: Reasoning about instances, and calculating the fuzzy satisfiability of a fuzzy concept.

3 Fuzzy Granulation of OWL Schemas

In this Section we show our proposal of introducing a granular view within an ontology. We shall proceed incrementally starting from the simplest case. For the sake of simplicity, we shall use the OWL terminology henceforth instead of the DL terminology (We remind the reader that class stands for concept, and property stands for role).

3.1 Case 1

Let C be a class and T a functional datatype property connecting instances of C to values in a numerical range $\mathbf{d}$. See Fig. 4 for a graphical representation of this costruct.

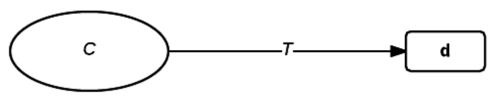

Fig. 4. Graphical representation of a functional datatype property T with domain C and range over a numerical datatype $\mathbf{d}$.

This schema can be directly translated into a table (see Table 2) with two columns and as many rows as the number of individuals of C for which T holds.

Table 2. Tabular representation of the OWL schema depicted in Fig. 4.

C	T
a_1	v_1
a_2	v_2
$\cdots$	$\cdots$
a_n	v_n

The dataset in Table 2 can be easily granulated in a number of fuzzy sets $F_1, F_2, \ldots, F_c$ by applying, *e.g.*, the fuzzy clustering method mentioned in Section 2.2. In essence, the granulation process puts individuals in the same information granule if their respective values are similar. The use of fuzzy sets to define granules ensures a gradual membership degree of individuals to such granules, where the maximal membership is assigned to individuals detected as "prototypes" of each granule. Each fuzzy set represents a fuzzy concept, and can be tagged by a linguistic term, *e.g.* `Low`.

Table 3. Granulated individuals obtained from Table 2.

C	F_1	F_2	$\cdots$	F_c
a_1	μ_{11}	μ_{12}	$\cdots$	μ_{1c}
a_2	μ_{21}	μ_{22}	$\cdots$	μ_{2c}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
a_n	μ_{n1}	μ_{n2}	$\cdots$	μ_{nc}

The result of granulation can be represented in a new table (see Table 3), where each individual a_i is associated to a row of membership values μ_{ij}, being

$$\mu_{ij} = F_j(v_i) \tag{5}$$

For each granule F_j, the relative cardinality $\sigma(F_j)$ can be computed by means of the formula in Eq. (3). Given a fuzzy quantifier Q_k, the membership degree

$$q_{jk} = Q_k(\sigma(F_j)) \tag{6}$$

identifies the degree of truth of the fuzzy proposition "$Q_k x$ are F_j". In this way, a new table can be constructed from a collection $Q_1, Q_2, \ldots, Q_m$ of fuzzy quantifiers, as shown in Table 4. If $cm \ll n$, a sensible reduction of data can be achieved to represent the original property through a granulated view. (To further reduce data, a threshold τ can be set, so that all q_{jk} less than τ are set to zero.)

Table 4. Quantified cardinalities for the granules reported in Table 3.

	Q_1	Q_2	$\cdots$	Q_m
F_1	q_{11}	q_{12}	$\cdots$	q_{1m}
F_2	q_{21}	q_{22}	$\cdots$	q_{2m}
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
F_c	q_{c1}	q_{c2}	$\cdots$	q_{cm}

The new granulated view can be integrated in the ontology as follows. The fuzzy sets F_j are the starting point for the definition of new subclasses of C defined as $D_j \equiv C \sqcap \exists T.F_j$. Also, a new class `Granule` is defined, with individuals $g_1, g_2, \ldots, g_c$, where each individual g_j is an information granule corresponding to F_j. Each individual in D_j is then mapped to g_j by means of an object property `mapsTo`. Also, the cardinality of information granules is modeled by means of a datatype property `hasCardinality` with domain in `Granule` and range in the datatype domain `xsd:double`. Moreover, for each fuzzy quantifier Q_k, a new class is introduced, which models one of the fuzzy sets over the cardinalities of information granules. The connection between the class `Granule` and each class Q_k is established through `hasCardinality`, once fuzzified, with degrees identified as in Table 4. Note that the fuzzy proposition "$Q_k x$ are F_j" is then represented as the fuzzy assertion $g_j : \exists \texttt{hasCardinality}.Q_k$.

Example 1. In the tourism domain, we might consider an OWL ontology which encompasses the datatype property `hasPrice` with the class `Hotel` as domain and range in the datatype domain `xsd:double`. Let us suppose that the room price for Hotel Verdi (instance `verdi` of `Hotel`) is 105, *i.e.* the KB contains the assertion `(verdi, 105):hasPrice`. By applying fuzzy clustering to `hasPrice`, we might obtain three fuzzy sets (Low, `Medium`, `High`) from which the following classes are derived

$\mathtt{LowPriceHotel} \equiv \mathtt{Hotel} \sqcap \exists \mathtt{hasPrice.Low}$
$\mathtt{MidPriceHotel} \equiv \mathtt{Hotel} \sqcap \exists \mathtt{hasPrice.Medium}$
$\mathtt{HighPriceHotel} \equiv \mathtt{Hotel} \sqcap \exists \mathtt{hasPrice.High}$.

With respect to these classes `verdi` shows different degrees of membership, *e.g.* `verdi` is a low-price hotel at degree 0.8 and a mid-price hotel at degree 0.2 (see Fig. 5 for a graphical representation). Subsequently, we might be interested in obtaining aggregated information about hotels. Here the class `Granule` comes into play. Quantified cardinalities allow us, for instance, to represent the fact that "Many hotels are low-price" as the fuzzy assertion $\mathtt{lph} : \exists \mathtt{hasCardinality.Many}$ with truth degree 0.7, where `lph` is an instance of `Granule` (*i.e.*, it is an information granule) which corresponds to `LowPriceHotel`, and `Many` is one of the fuzzy sets obtained from `hasCardinality`. Note that `verdi`, being an instance of `LowPriceHotel`, maps to `lph`, *i.e.* $(\mathtt{verdi}, \mathtt{lph}) : \mathtt{mapsTo}$ holds to some degree.

3.2 Case 2

A natural extension of the proposed granulation method follows when the class C is specialized in subclasses, as in Fig. 6. In this case, there are as many tables with the same structure of Table 2 as the number of subclasses.

Analogously, for each subclass $SubC_j$ a structure of fuzzy information granules $F_{j1}, F_{j2}, \ldots, F_{jc}$ is produced and quantified according to the usual fuzzy quantifiers $Q_1, Q_2, \ldots, Q_m$. (The quantifiers do not depend on the subclass as their definition is fixed for all information granules.)

Example 2. Following Example 1, one may think of having a subsumption hierarchy with the class `Accommodation` as the root and `Hotel` and `B&B` as subclasses (see Fig. 7). Hotels are granulated in three fuzzy subclasses (`LowPriceHotel`, `MidPriceHotel` and `HighPriceHotel`) while B&Bs are granulated in two fuzzy subclasses (`CheapB&B` and `ExpensiveB&B`). These fuzzy classes are related to the classes representing fuzzy quantifiers via `Granule` analogously to Example 1.

3.3 Case 3

A case of particular interest is given by OWL schemas representing ternary relations. A *ternary relation* is a subset of the Cartesian product involving three domains $C \times D \times N$ (for our purposes, we will assume N a numerical domain).

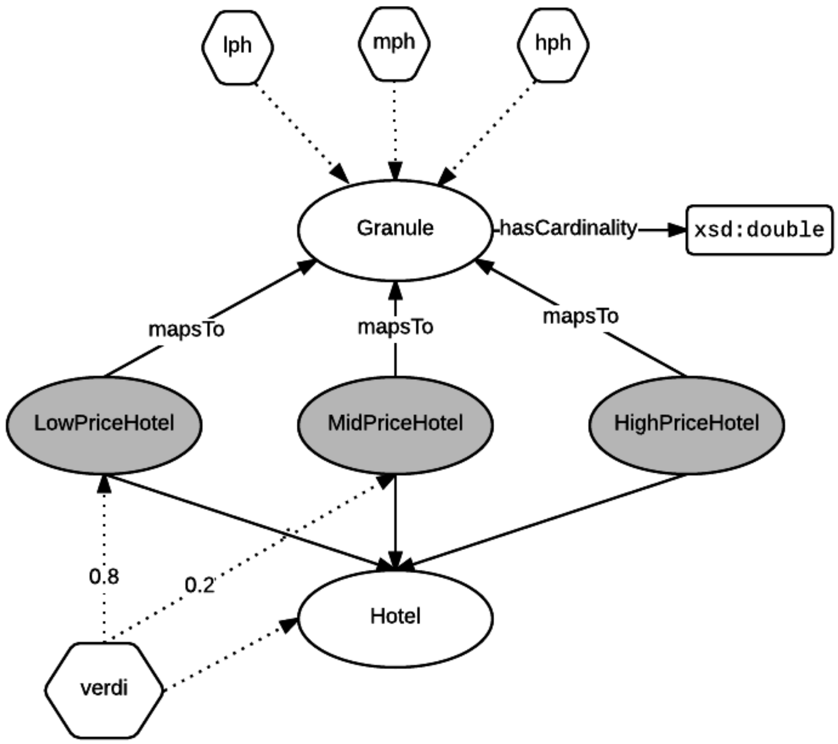

Fig. 5. Graphical representation of the output of the fuzzy granulation process on the OWL schema described in Fig. 4 and instantiated with concepts reported in Example 1. Fuzzy classes are depicted in gray.

Because of DL restrictions, however, ternary relations are not directly representable in OWL, yet they can be indirectly represented through an auxiliary class E, two object properties R_1 and R_2, and one datatype property T, as depicted in Fig. 8.

The structure in Fig. 8 corresponds to a tabular representation with three columns, and as many rows as the number of elements of the relation, as in table 5. By removing one of the two columns in Table 5, the resulting table is in accordance with Table 2, which was the starting point of the granulation process. In particular, as in the previous cases, a number of fuzzy sets $F_1, F_2, \ldots, F_c$ can be derived starting from the dataset represented in Table 5, where one column has been dropped. (We henceforth assume to drop column C.)

In order to connect information granules with classes, we proceed as follows. Given an individual $a \in C$, a subset of Table 5 can be obtained, as in Table 6.

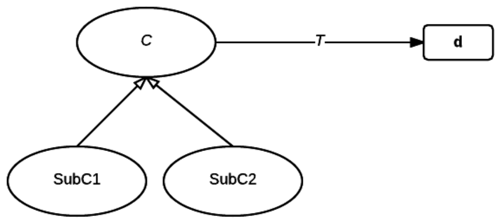

Fig. 6. Variant of the OWL schema shown in Fig. 4 for the case of C having subclasses.

Table 5. Tabular representation of the OWL schema depicted in Fig. 8.

C	D	T
a_1	b_1	v_1
$\cdots$	$\cdots$	$\cdots$
a_i	b_j	v_k
$\cdots$	$\cdots$	$\cdots$
a_n	b_m	v_l

More precisely, for each information granule F_j, it is possible to compute the relative cardinality

$$\sigma_{j_a} = \frac{\sum_{i=1}^{n_a} F_j\left(v_{a_i}\right)}{n_a} \tag{7}$$

Such cardinality can be quantified according to the fuzzy quantifiers $Q_1, \ldots, Q_m$. The result is a table similar to Table 4, but now related to the individual a.

Table 6. A slice of Table 5 obtained by fixing an individual a in C.

C	D	T
a	b_{a_1}	v_{a_1}
a	b_{a_2}	v_{a_2}
$\cdots$	$\cdots$	$\cdots$
a	$b_{a_{n_a}}$	$v_{a_{n_a}}$

Information granules, connected with the individuals in C, are arranged in the ontology in a way that merges the modeling of ternary relations as in Fig. 8 with the granular model illustrated in case 1. The new classes, representing information granules, are defined as $E'_j \equiv E \sqcap \exists R_2.D \sqcap \exists T.F_j$. They are connected to E in order to express a granular view of the relation between C and D. Finally, a natural extension of this case allows the specialization of the class D in subclasses (as in case 2).

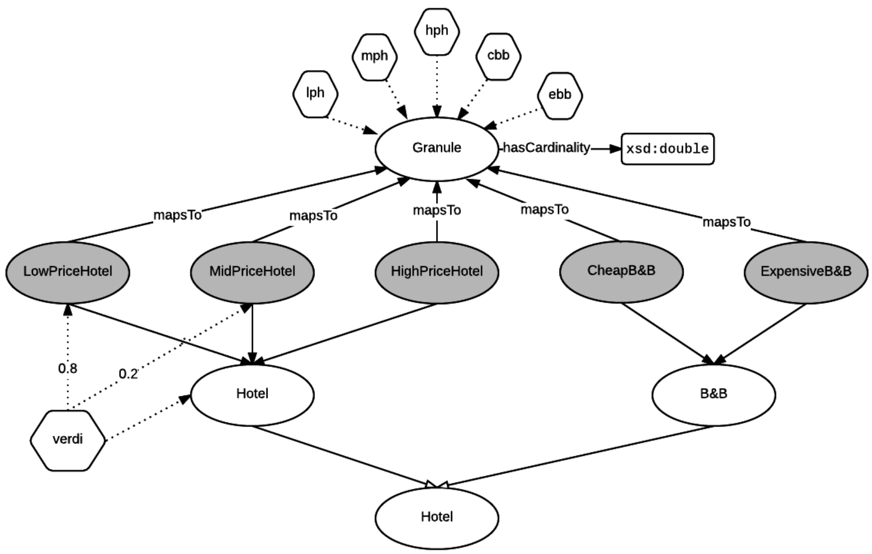

Fig. 7. Graphical representation of the output of the fuzzy granulation process on the OWL schema reported in Fig. 6 and instantiated with the concepts used in Example 2.

Example 3. With reference to the touristic domain, we might also consider the distances between hotels and attractions (see Fig. 9). This is clearly a case of ternary relation which requires to be modeled through an auxiliary class `Distance` which is connected to the classes `Hotel` and `Attraction` by means of the object properties `hasDistance` and `isDistanceFor`, respectively, and plays the role of domain for a datatype property `hasValue` with range `xsd:double`. The knowledge that "Hotel Verdi has a distance of 100 meters from the British Museum" can be therefore modeled as follows:

$(\mathtt{verdi}, \mathtt{d1}) : \mathtt{hasDistance}$
$(\mathtt{d1}, \mathtt{british\_museum}) : \mathtt{isDistanceFor}$
$(\mathtt{d1}, \mathtt{100}) : \mathtt{hasValue}$

After fuzzy granulation, the imprecise sentence "Hotel Verdi has a low distance from many attractions" can be considered as a consequence of the previous and the following axioms and assertions

$\mathtt{LowDistance} \equiv \mathtt{Distance} \sqcap \exists\mathtt{isDistanceFor.Attraction} \sqcap \exists\mathtt{hasValue.Low}$
$\mathtt{d1} : \mathtt{LowDistance}$ (to some degree)
$(\mathtt{d1}, \mathtt{ld}) : \mathtt{mapsTo}$
$(\mathtt{ld}, \mathtt{0.5}) : \mathtt{hasCardinality}$
$\mathtt{ld} : \exists\mathtt{hasCardinality.Many}$ (to some degree)

where `Many` is defined as mentioned in Example 1.

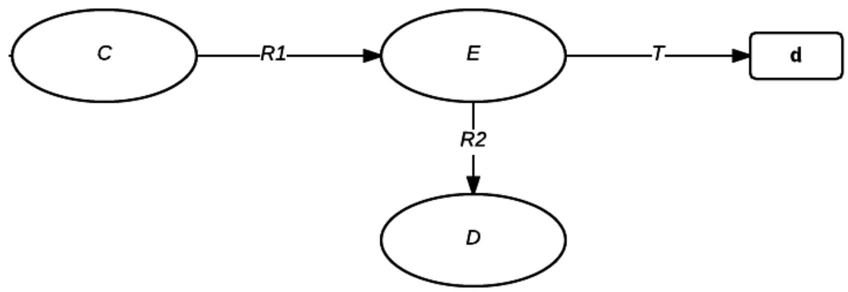

Fig. 8. Graphical representation of the OWL schema modeling a ternary relation.

4 Conclusions

This paper presents a starting point for introducing a granular view of data within an OWL ontology. According to the ideas presented in the paper, a number of individuals belonging to the ontology can be replaced by information granules, represented as fuzzy sets. In particular, the connection between the existing individuals (not involved in granulation) and the granulated view is possible by exploiting the peculiar representation of ternary relations in OWL.

This work is in a preliminary stage. We are currently evaluating the possibility of representing the output of our fuzzy granulation method by using OWL 2, *i.e.* within the current Semantic Web languages as suggested by Bobillo and Straccia in their proposal of Fuzzy OWL 2 [7]. In a certain sense, we are pursuing an alternative direction in comparison with the work of Sanchez and Tettamanzi [14] which, if implemented, could lead to the extension of current Semantic Web languages. However, it should be noted that there are some non-negligible restrictions to make their approach work in current fuzzy DLs. Notably, the approach considers a finite number of individuals, which causes a mismatch with the usual semantics for DLs (*i.e.*, OWA with infinite interpretations).

Future research is aimed at integrating our fuzzy granulation approach within inductive learning algorithms, such as FOIL-$\mathcal{DL}$ [11], with the aim of verifying the benefits of information granulation in terms of efficiency and effectiveness of the learning process, as well as in terms of interpretability of the learning results.

Acknowledgements This work was partially funded by the *Università degli Studi di Bari "Aldo Moro"* under the IDEA Giovani Ricercatori 2011 grant "Dealing with Vague Knowledge in Ontology Refinement".

References

1. Alonso, J.M., Castiello, C., Mencar, C.: Interpretability of Fuzzy Systems: Current Research Trends and Prospects. In: Kacprzyk, J., Pedrycz, W. (eds.) Springer

Handbook of Computational Intelligence. Springer Berlin / Heidelberg (2015)
2. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (eds.): The Description Logic Handbook: Theory, Implementation and Applications (2nd ed.). Cambridge University Press (2007)
3. Baader, F., Hanschke, P.: A scheme for integrating concrete domains into concept languages. In: Mylopoulos, J., Reiter, R. (eds.) Proceedings of the 12th International Joint Conference on Artificial Intelligence. Sydney, Australia, August 24-30, 1991. pp. 452–457. Morgan Kaufmann (1991)
4. Bargiela, A., Pedrycz, W.: Granular computing: an introduction. Springer Science & Business Media (2003)
5. Bargiela, A., Pedrycz, W.: Human-centric information processing through granular modelling, vol. 182. Springer Science & Business Media (2009)
6. Bezdek, J.C.: Fuzzy clustering. In: Ruspini, E.H., Bonissone, P.P., Pedrycz, W. (eds.) Handbook of Fuzzy Computation, p. 2. Institute of Physics Pub. (1998)
7. Bobillo, F., Straccia, U.: Representing fuzzy ontologies in OWL 2. In: FUZZ-IEEE 2010, IEEE International Conference on Fuzzy Systems, Barcelona, Spain, 18-23 July, 2010, Proceedings. pp. 1–6. IEEE (2010)
8. Borgida, A.: On the relative expressiveness of description logics and predicate logics. Artificial Intelligence 82(1–2), 353–367 (1996)
9. Dubois, D., Prade, H.: Fuzzy cardinality and the modeling of imprecise quantification. Fuzzy sets and Systems 16(3), 199–230 (1985)
10. Klir, G.J., Yuan, B.: Fuzzy sets and fuzzy logic: theory and applications. Prentice-Hall, Inc. (1995)
11. Lisi, F.A., Straccia, U.: Dealing with incompleteness and vagueness in inductive logic programming. In: Cantone, D., Nicolosi Asmundo, M. (eds.) Proceedings of the 28th Italian Conference on Computational Logic, Catania, Italy, September 25-27, 2013. CEUR Workshop Proceedings, vol. 1068, pp. 179–193. CEUR-WS.org (2013), http://ceur-ws.org/Vol-1068/paper-l12.pdf
12. OWL 2 Web Ontology Language Document Overview: http://www.w3.org/TR/2009/REC-owl2-overview-20091027/. W3C (2009)
13. Reiter, R.: Equality and domain closure in first order databases. Journal of ACM 27, 235–249 (1980)
14. Sanchez, D., Tettamanzi, A.G.: Fuzzy quantification in fuzzy description logics. In: Sanchez, E. (ed.) Fuzzy Logic and the Semantic Web, Capturing Intelligence, vol. 1, pp. 135 – 159. Elsevier (2006)
15. Schmidt-Schauss, M., Smolka, G.: Attributive concept descriptions with complements. Artificial Intelligence 48(1), 1–26 (1991)
16. Stoilos, G., Simou, N., Stamou, G., Kollias, S.: Uncertainty and the Semantic Web. IEEE Intelligent Systems 21 (2006)
17. Straccia, U.: Reasoning within fuzzy description logics. Journal of Artificial Intelligence Research 14, 137–166 (2001)
18. Straccia, U.: Description logics with fuzzy concrete domains. In: UAI '05, Proceedings of the 21st Conference in Uncertainty in Artificial Intelligence, Edinburgh, Scotland, July 26-29, 2005. pp. 559–567. AUAI Press (2005)
19. Straccia, U.: Foundations of Fuzzy Logic and Semantic Web Languages. CRC Studies in Informatics Series, Chapman & Hall (2013)
20. Zadeh, L.A.: Is there a need for fuzzy logic? Information sciences 178(13), 2751–2779 (2008)

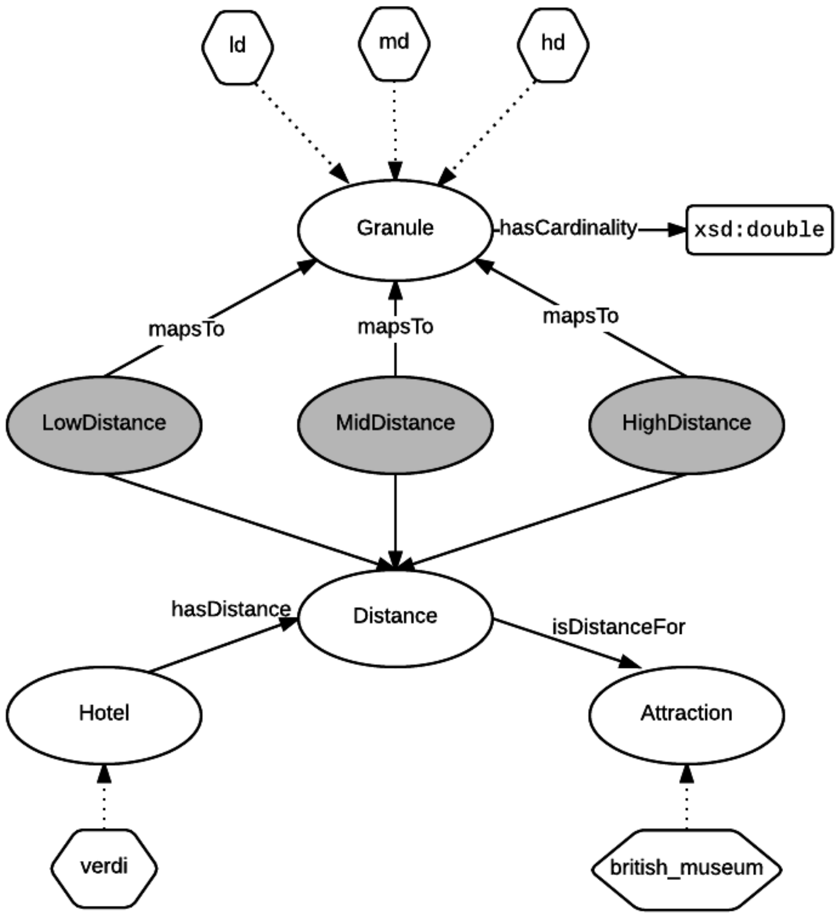

Fig. 9. Graphical representation of the output of the fuzzy granulation process on the OWL schema reported in Fig. 8 and instantiated with the concepts used in Example 3.