Building A Fuzzy Knowledge Body for Integrating Domain Ontologies

Konstantin Todorov[1], Peter Geibel[2], and Céline Hudelot[1]

[1]Laboratory MAS, École Centrale Paris
[2]TU Berlin, Sekr. FR 5-8, Fakultät IV

Abstract. This paper deals with the problem of building a common knowledge body for a set of domain ontologies in order to enable their sharing and integration in a collaborative framework. We propose a novel hierarchical algorithm for concept fuzzy set representation mediated by a reference ontology. In contrast to the original concept representations based on instances, this enables the application of methods of fuzzy logical reasoning in order to characterize and measure the degree of the relationships holding between concepts from different ontologies. We presenta an application of the approach in the multimedia domain.

1 Introduction

In collaborative contexts, multiple independently created ontologies often need to be brought together in order to enable their interoperability. These ontologies have an impaired collaborative functionality, due to heterogeneities coming from the decentralized nature of their acquisition, differences in scopes and application purposes and mismatches in syntax and terminology.

We present an approach to building a combined knowledge body for a set of domain ontologies, which captures and exposes various relations holding between the concepts of the domain ontologies, such as their relative generality or specificity, their shared commonality or their complementarity. This can be very useful in a number of real-life scenarioss, especially in collaborative platforms. Let us imagine a project which includes several partners, each of which has its own vocabulary of semantically structured terms that describes its activity. The proposed framework would allow every party to keep its ontology and work with it, but query the combined knowledge body whenever collaboration is necessary. Examples of such queries can be: "which concept of a partner P_1 is closest to my concept A", or "give me those concepts of all of my partners which are equally distant to my concept B", or "find me a concept from partner P_2 which is a strong subsumer of my concept C", or "what are the commonality and specificity between my concept A and my partner's concept D".

We situate our approach in a fuzzy framework, where every domain concept is represented as a fuzzy set of the concepts of a particular *reference* ontology. This can be seen as a projection of all domain source concepts onto a common semantical space, where distances and relations between any two concepts can be expressed under fixed criteria. In contrast to the original instance-representation, we can apply methods of fuzzy logical reasoning in order to characterize the relationship between concepts from different ontologies. In addition, the fuzzy representations allow for quantifying the degree to which a certain relation holds.

The paper is structured as follows. Related work is presented in the next section. Background in the field of fuzzy sets, as well as main definitions and problems from the ontology matching domain are overviewed in Section 3. We present the concept fuzzification algorithm in Section 4, before we discuss how the combined knowledge body can be constructed in Section 5. Experimental results and conclusions are presented in Sections 6 and 7, respectively.

2 Related Work

Fuzzy set theory generalizes classical set theory [19] allowing to deal with imprecise and vague data. A way of handling imprecise information in ontologies is to incorporate fuzzy reasoning into them. Several papers by Sanchez, Calegari and colleagues [4], [5], [13] form an important body of work on fuzzy ontologies where each ontology concept is defined as a fuzzy set on the domain of instances and relations on the domain of instances and concepts are defined as fuzzy relations.

Work on fuzzy ontology matching can be classified in two families: (1) approaches extending crisp ontology matching to deal with fuzzy ontologies and (2) approaches addressing imprecision of the matching of (crisp or fuzzy) concepts. Based on the work on approximate concept mapping by Stuckenschmidt [16] and Akahani *et al.* [1], Xu *et al.* [18] suggested a framework for the mapping of fuzzy concepts between fuzzy ontologies. With a similar idea, Bahri *et al.* [2] propose a framework to define similarity relations among fuzzy ontology components. As an example of the second family of approaches, we refer to [8] where a fuzzy approach to handling matchinging uncertainty is proposed. A matching approach based on fuzzy conceptual graphs and rules is proposed in [3]. To define new intra-ontology concept similarity measures, Cross *et al.* [6] model a concept as a fuzzy set of its ancestor concepts and itself, using as a membership degree function the Information Content (IC) of concept with respect to its ontology.

Crisp instance-based ontology matching, relying on the idea that concept similarity is accounted for by the similarity of their instances, has been overviewed broadly in [7]. We refer particularly to the Caiman approach which relies on estimating concepts similarity by measuring class-means distances [10].

3 Background and Preliminaries

In this section, we introduce basics from fuzzy set theory and discuss aspects of the ontology matching problem.

3.1 Fuzzy Sets

A fuzzy set $\mathcal{A}$ is defined on a given domain of objects X by the function

$$\mu_{\mathcal{A}} : X \longmapsto [0,1], \tag{1}$$

which expresses the degree of membership of every element of X to $\mathcal{A}$ by assigning to each $x \in X$ a value from the interval $[0,1]$ [19]. The fuzzy power set of X, denoted by $\mathcal{F}(X,[0,1])$, is the set of all membership functions $\mu : X \longmapsto [0,1]$.

We recall several fuzzy set operations by giving definitions in terms of Gödel semantics [15]. The *intersection* of two fuzzy sets $\mathcal{A}$ and $\mathcal{B}$ is given by a t-norm function $T(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x)) = \min(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x))$. The *union* of $\mathcal{A}$ and $\mathcal{B}$ is given by $S(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x)) = \max(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x))$ where S is a t-conorm. The *complement* of a fuzzy set $\mathcal{A}$, denoted by $\neg\mathcal{A}$, is defined by the membership function $\mu_{\neg\mathcal{A}}(x) = 1 - \mu_{\mathcal{A}}(x)$. We consider the Gödel definition of a fuzzy *implication*

$$\mu_{\mathcal{A}\to\mathcal{B}}(x) = \begin{cases} 1, & \text{if } \mu_{\mathcal{A}}(x) \leq \mu_{\mathcal{B}}(x), \\ \mu_{\mathcal{B}}(x), & \text{otherwise.} \end{cases} \tag{2}$$

3.2 Ontologies, Heterogeneity and Ontology Matching

An *ontology* consists of a set of semantically related *concepts* and provides in an explicit and formal manner knowledge about a given domain of interest [7]. We are particularly interested in ontologies, whose concepts come equipped with a set of associated instances, defined as it follows.

Definition 1 (Crisp Ontology). *Let C be a set of concepts, $is\_a \subseteq C \times C$, R a set of relations on C, I a set of instances, and $g : C \to 2^I$ a function that assigns a subset of instances from I to each concept in C. We require that is_a and g are compatible, i.e., that $is\_a(A', A) \leftrightarrow g(A') \subseteq g(A)$ holds for all $A', A \in C$. In particular, this entails that is_a has to be a partial order. With these definitions, the quintuple*

$$O = (C, is\_a, R, I, g)$$

forms a crisp ontology.

Above, a concept is modeled *intensionally* by its relations to other concepts, and *extensionally* by a set of instances assigned to it via the function g. By assumption, every instance can be represented as a real-valued vector, defined by a fixed number of variables of some kind (the same for all the instances in I).

Ontology heterogeneity occurs when two or more ontologies are created independently from one another over similar domains. Heterogeneity may be observed on *linguistic or terminological*, on *conceptual* or on *extensional* level [7]. *Ontology matching* is understood as the process of establishing relations between the elements of two or more heterogeneous ontologies. Different matching techniques have been introduced over the past years in order to resolve different types of heterogeneity [9].

Instance-based, or *extensional ontology matching* gathers a set of approaches around the central idea that ontology concepts can be represented as sets of related instances and the similarity measured on these sets reflects the semantic similarity between the concepts that these instances populate.

3.3 Crisp Concept Similarities

Consider the ontologies $O_1 = (C_1, is\_a_1, R_1, I_1, g_1)$ and $O_{ref} = (X, is\_a_{ref}, R_{ref}, I_{ref}, g_{ref})$. We rely on the straightforward idea that determining the similarity $sim(A, x)$ of two concepts $A \in C_1$ and $x \in X$ consists in comparing their instance sets $g_1(A)$ and $g_{ref}(x)$. For doing so, we need a similarity measure for instances $\mathbf{i}^A$ and $\mathbf{i}^x$, where $\mathbf{i}^A \in g_1(A)$ and $\mathbf{i}^x \in g_{ref}(x)$. We have used the scalar product and the cosine $s(\mathbf{i}^A, \mathbf{i}^x) = \frac{\langle \mathbf{i}^A, \mathbf{i}^x \rangle}{\|\mathbf{i}^A\| \|\mathbf{i}^x\|}$. Based on this similarity measure for elements, the similarity measure for the sets can be defined by computing the similarity of the mean vectors corresponding to class prototypes [10]:

$$sim_{proto}(A, x) = s\Big(\frac{1}{|g_1(A)|} \sum_{j=1}^{|g_1(A)|} \mathbf{i}_j^A, \frac{1}{|g_{ref}(x)|} \sum_{k=1}^{|g_{ref}(x)|} \mathbf{i}_k^x \Big) . \quad (3)$$

Note that other approaches of concept similarity can be employed as well, like the variable selection approach in [17]. In the context of our study, we have used the method that both works best and is less complex. A hierarchical application of the similarity measure for the concepts of two ontologies is presented in [17].

4 A Hierarchical Algorithm for Concept Fuzzification

Let $\Omega = \{O_1, ..., O_n\}$ be a set of (crisp) ontologies that will be referred to as *source ontologies* defined as in Def. 1. The set of concepts $C_\Omega = \bigcup_{i=1}^{n} C_i$ will be referred to as the set of *source concepts*. The ontologies from the set Ω are assumed to share similar functionalities and application focuses and to be heterogeneous in the sense of some of the heterogeneity types described in Section 3.2. A certain complementarity of these resources can be assumed: they could be defined with the same application scope, but on different levels, treating different and complementary aspects of the same application problem.

Let $O_{ref} = (X, is\_a_{ref}, R_{ref}, I_{ref}, g_{ref})$ be an ontology, called a *reference* ontology whose concepts will be called *reference concepts*. In contrast to the source ontologies, the ontology O_{ref} is assumed to be a less application dependent, generic knowledge source. As a consequence of Def. 1, the ontologies in Ω and O_{ref} are populated.

The fuzzification procedure that we propose relies on the idea of scoring every source concept by computing its similarities with the reference concepts, using the similarity measure (3). A source concept A will be represented by a function of the kind

$$\mu_{\mathcal{A}}(x) = score_A(x), \forall x \in X, \quad (4)$$

where $score_A(x)$ is the similarity between the concept A and a given reference concept x. Since *score* takes values between 0 and 1, (4) defines a fuzzy set. We will refer to such a fuzzy set as the *fuzzified* concept A denoted by $\mathcal{A}$.

In order to fuzzify the concepts of a source ontology O_1, we propose the following hierarchical algorithm. First, we assign score-vectors, i.e. fuzzy membership functions to all leaf-node concepts of O_1. Every non-leaf node, if it does not contain instances (documents) of its own, is scored as the maximum of the scores of its children for every $x \in X$. If a non-leaf node has directly assigned instances (not inherited from its children), the node is first scored on the basis of these instances with respect to the reference ontology, and then as the maximum of its children and itself. To illustrate, let a concept A have children A' and A'' and let the non-empty function $g^*(A)$ represent the instances assigned directly to the concept A. We compute the following similarity scores for this concept w.r.t. the set X :

$$score_A(x) = max\{score_{A'}(x), score_{A''}(x), score_{g^*(A)}(x)\}, \forall x \in X. \quad (5)$$

Above, $score_{g^*(A)}(x)$ conventionally denotes the similarity obtained for the concept A and a reference concept x by only taking into account the documents assigned directly to A. The algorithm is given in Alg. 1.

It is worth noting that assigning the *max* of all children to the parent for every x leads to a convergence to uniformity of the membership functions for nodes higher up in the hierarchy. Naturally, the functions of the higher level concepts are expected to be less "specific" than those of the lower level concepts. A concept in a hierarchical structure can be seen as the union of its descendants, and a union corresponds to taking the *max* (an approach underlying the single link strategy used in clustering).

The hierarchical scoring procedure has the advantage that every x-score will be larger for a parent node than those for any of its children, and it holds that $\mu_{\mathcal{A}'}(x) \to \mu_{\mathcal{A}}(x) = 1$ for all x and all children $\mathcal{A}'$ of $\mathcal{A}$. From computational viewpoint, the procedure which only scores the populated nodes is less expensive, compared to scoring all nodes one by one.

5 Building a Combined Knowledge Body

The construction of a combined knowledge body for a set of source ontologies aims at making explicit the relations that hold among their concepts, across these ontologies. To these ends, we apply the fuzzy set representations acquired in the previous section. In what follows, we consider two source ontologies O_1 and O_2 but note that all definitions can be extended for multiple ontologies. Let $\mathcal{C}_\Omega = \{\mathcal{A}_1, ..., \mathcal{A}_{|C_1|}, \mathcal{B}_1, ..., \mathcal{B}_{|C_2|}\}$ be the union of the concept sets of O_1 (the $\mathcal{A}$-concepts) and O_2 (the $\mathcal{B}$-concepts). We introduce several relations and operations that can be computed over $\mathcal{C}_\Omega$ and will be used for constructing a combined reduced knowledge body that contains the concepts of interest.

5.1 Fuzzy Concept Relations

The implication $A' \to A$ holds for any A' and A such that $is\_a(A', A)$. We provide a definition for a fuzzy subsumption of two fuzzified concepts $\mathcal{A}'$ and $\mathcal{A}$ based on the fuzzy implication (2).

```
Function score(concept A, ontology O_ref, sim. measure sim)
begin
    for i = 1, ..., |X| do
        sim[i] = sim(A, x_i) // x_i ∈ X
    return sim
end
Procedure hierachicalScoring(ontology O, ontology O_ref, sim. measure sim)
begin
    1. Let C be the list of concepts in O.
    2. Let L be a list of nodes, initially empty
    3. Do until C is empty:
       (a) Let L' be the list of nodes in C that have only children in L
       (b) L = append(L, L')
       (c) C = C − L'
    4. Iterate over L (first to last), with A being the current element:
       if children(A) = ∅ then
           score(A) = score(A, O_ref, sim)
       else
           if g*(A) ≠ ∅ then
               score(A) = max{max_{B∈children(A)} score(B), score(A, O_ref, sim)}
           else
               score(A) = max_{B∈children(A)} score(B)

    return score(A), ∀A ∈ C
end
```
Algorithm 1: An algorithm for hierarchical scoring of the source concepts.

Definition 2 (Fuzzy Subsumption). *The subsumption $\mathcal{A}'$ is a $\mathcal{A}$ is defined and denoted in the following manner:*

$$\mathbf{is\_a}(\mathcal{A}', \mathcal{A}) = \inf_{x \in X} \mu_{\mathcal{A}' \to \mathcal{A}}(x) \tag{6}$$

Equation (6) defines the fuzzy subsumption as a degree between 0 and 1 to which one concept is the subsumer of another. It can be shown that **is_a**, similarly to its crisp version, is reflexive and transitive (i.e. a quasi-order). In addition, the hierarchical procedure for concept fuzzification introduced in the previous section assures that $\mathbf{is\_a}(A', A) = 1$ holds for every child-parent concept pair, i.e. the crisp subsumption relation is preserved by the fuzzification process.

Taking the example of a collaborative platform from the introduction, computing the fuzzy **is_a** between two concepts allows for answering a user query regarding generality and specificity of their partners concepts with respect to a given target concept.

We provide a definition of a fuzzy ontology which follows directly from the fuzzification of the source concepts and their **is_a** relations introduced above.

Definition 3 (Fuzzy Ontology). *Let $\mathcal{C}$ be a set of (fuzzy) concepts,* $\mathbf{is\_a}$: $\mathcal{C} \times \mathcal{C} \to [0, 1]$ *a fuzzy is_a-relationship, $\mathcal{R}$ a set of fuzzy relations on $\mathcal{C}$, i.e., $\mathcal{R}$*

contains relations $r : \mathcal{C}^n \to [0,1]$*, where* n *is the arity of the relation (for the sake of presentation, we only consider binary relations),* $\mathcal{X}$ *a set of objects, and* $\phi : \mathcal{C} \to \mathcal{F}(\mathcal{X}, [0,1])$ *a function that assigns a membership function to every fuzzy concept in* $\mathcal{C}$*. We require that* $\mathbf{is\_a}$ *and* ϕ *are compatible, i.e., that* $\mathbf{is\_a}(\mathcal{A}', \mathcal{A}) = \inf_x \mu_{\mathcal{A}' \to \mathcal{A}}(x)$ *holds for all* $\mathcal{A}', \mathcal{A} \in \mathcal{C}$*. In particular, it can be shown that this entails that* $\mathbf{is\_a}$ *is a fuzzy quasi-order. With these definitions, the quintuple*

$$\mathcal{O} = (\mathcal{C}, \mathbf{is\_a}, \mathcal{R}, \mathcal{X}, \phi)$$

forms a fuzzy ontology.

Above, the set $\mathcal{X}$ is defined as a set of abstract objects. In our setting, these are the concepts of the reference ontology, i.e. $\mathcal{X} = X$. The set $\mathcal{C}$ is any subset of $\mathcal{C}_\Omega$. In case $\mathcal{C} = \mathcal{C}_1$, where $\mathcal{C}_1$ is the set of fuzzified concepts of the ontology O_1, $\mathcal{O}$ defines a fuzzy version of the crisp source ontology O_1. In case $\mathcal{C} = \mathcal{C}_\Omega$, $\mathcal{O}$ defines *a common knowledge body* for the two source ontologies. Note that the membership values of the reference concepts entail fuzzy membership values for the documents populating the reference concepts. However, we will work directly with the concepts scores in what follows.

Based on the subsumption relation defined above, we will define equivalence of two concepts in the following manner.

Definition 4 (Fuzzy θ-Equivalence). *Fuzzy* θ*-equivalence between a concept* $\mathcal{A}$ *and a concept* $\mathcal{B}$*, denoted by* $\mathcal{A} \backsim_\theta \mathcal{B}$ *holds if and only if* $\mathbf{is\_a}(\mathcal{A}, \mathcal{B}) > \theta$ *and* $\mathbf{is\_a}(\mathcal{B}, \mathcal{A}) > \theta$*, where* θ *is a parameter between* 0 *and* 1*.*

The equivalence relation allows to define classes of equivalence on the set $\mathcal{C}_\Omega$. In the collaborative framework described in the introduction, this can be used for querying concepts equivalent (up to a degree defined by the user) to a given user concept from the set of their partners concepts.

5.2 Similarity Measures for Fuzzy Concepts

We propose several measures of closeness of two fuzzy concepts $\mathcal{A}$ and $\mathcal{B}$. We begin by introducing a straightforward measure given by

$$\rho_{\text{base}}(\mu_\mathcal{A}, \mu_\mathcal{B}) = 1 - \max_{x \in X} |\mu_\mathcal{A}(x) - \mu_\mathcal{B}(x)|. \tag{7}$$

We consider a similarity measure based on the Euclidean distance:

$$\rho_{\text{eucl}}(\mu_\mathcal{A}, \mu_\mathcal{B}) = 1 - \|\mu_\mathcal{A} - \mu_\mathcal{B}\|_2 \,, \tag{8}$$

where $\|x\|_2 = \left(\sum_{x \in X} |x|^2\right)^{1/2}$ is the ℓ^2-norm. Several measures of fuzzy set compatibility can be applied, as well. Zadeh's partial matching index between two fuzzy sets is given by

$$\rho_{\text{sup-min}}(\mu_\mathcal{A}, \mu_\mathcal{B}) = \sup_{x \in X} T(\mu_\mathcal{A}(x), \mu_\mathcal{B}(x)). \tag{9}$$

Finally, the Jaccard coefficient is defined by

$$\rho_{\text{jacc}}(\mu_{\mathcal{A}}, \mu_{\mathcal{B}}) = \frac{\sum_x T(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x))}{\sum_x S(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x))} . \qquad (10)$$

It is required that at least one of the functions $\mu_{\mathcal{A}}$ or $\mu_{\mathcal{B}}$ takes a non-zero value for some x. T and S are as defined in Section 3.

The similarity measures listed above provide different information as compared to the relations introduced in the previous subsection. Subsumption and equivalence characterize the structural relation between concepts, whereas similarity measures closeness between set elements. The two types of information are to be used in a complementary manner within the collaboration framework.

5.3 Quantifying Commonality and Relative Specificity

The union of two fuzzy concepts can be decomposed into three components, each quantifying, respectively, the commonality of both concepts, the specificity of the first compared to the second and the specificity of the second compared to the first expressed in the following manner

$$S(\mathcal{A}, \mathcal{B}) = (\mathcal{A}\mathcal{B}) + (\mathcal{A} - \mathcal{B}) + (\mathcal{B} - \mathcal{A}). \qquad (11)$$

Each of these components is defined as follows and, respectively, accounts for:

$$\mathcal{A}\mathcal{B} \ = T(\mathcal{A}, \mathcal{B}) \ \text{// what is common to both concepts;} \qquad (12)$$

$$\mathcal{A} - \mathcal{B} = T(\mathcal{A}, \neg\mathcal{B}) \ \text{// what is characteristic for A;} \qquad (13)$$

$$\mathcal{B} - \mathcal{A} = T(\mathcal{B}, \neg\mathcal{A}) \ \text{// what is characteristic for B.} \qquad (14)$$

Several merge options can be provided to the user with respect to the values of these three components. In case $\mathcal{A}\mathcal{B}$ is significantly larger than each of $\mathcal{A} - \mathcal{B}$ and $\mathcal{B} - \mathcal{A}$, the two concepts can be merged into their union. In case one of $\mathcal{A} - \mathcal{B}$ or $\mathcal{B} - \mathcal{A}$ is larger than the other two components, the concepts can be merged to either $\mathcal{A}$ or $\mathcal{B}$.

6 Experiments

We situate our experiments in the multimedia domain, opposing two complementary heterogeneous ontologies containing annotated pictures. We chose, on one hand, LSCOM [14] initially built in the framework of TRECVID[1] and populated with the development set of TRECVID 2005. Since this set contains images from broadcast news videos, LSCOM is particularly adapted to annotate this kind of content, thus contains abstract and specific concepts (e.g. Science_Technology, Interview_On_Location). On the other hand, we used

[1] http://www-nlpir.nist.gov/projects/tv2005/

WordNet [11] populated with the LabelMe dataset [12], referred to as the LabelMe ontology. Contrarily to LSCOM, this ontology is very general, populated with photographs from daily life and contains concepts such as CAR, COMPUTER, PERSON, etc. The parts of the two multimedia ontologies used in the experiments are shown in Figure 1.

```
natural hazard
        earthquake
        natural disasters
        tornado
        avalanche
        mudslide
conveyance
        airplane
                flying
                landing
                take-off
        ground vehicles
                bus
                truck
        boat
                sailboat
                boat ship
recreational activities
        sports
                baseball
                basketball
                football
                soccer
                tennis
group
single person
        single person male
        single person female
head and shoulders
computers
```

```
animal
        bird
        fish
        mammal
                bear
                monkey
                cat
                dog
group
        people
        social
                couple
                crowd
                audience
body
        arm
        head
        leg
        torso
atmospheric phenomenon
        cloud
        fog
        snow
conveyance
        bus
        vehicle
                airplane
                truck
                vessel
                   boat
                      motorboat
                      gondola
device
        elec. device
                computer
```

Fig. 1. The LSCOM (left) and the LabelMe (right) ontologies.

107_Standing One or more people standing up. 227_Bus Shots of a bus. 224_Outdoor Shots of Outdoor locations. 217_Person Shots depicting a person. The face may be partially visible. 202_Crowd Shots depicting a crowd. 181_Adult Shots showing a person over the age of 18. 104_Male_Person One or more male persons. 290_Daytime_Outdoor shots that take place outdoors during the day. 316_Group We defined a group as 3-10 people. 109_Windows An opening in the wall or roof of a building or vehicle fitted with glass or other transparent material.

Fig. 2. The LSCOM concept Bus: a visual and a textual instance.

A text document has been generated for every image of the two ontologies, by taking the names of all concepts that an image contains in its annotation, as well as the (textual) definitions of these concepts (the LSCOM definitions for

TRECVID images or the WordNet glosses for LabelMe images). An example of a visual instance of a multimedia concept and the constructed textual description is given in Figure 2. Several problems related to this representation are worth noting. The LSCOM keyword descriptions sometimes depend on negation and exclusion which are difficult to handle in a simple bag-of-words approach. Taking the WordNet glosses of the terms in LabelMe introduces problems related to polysemy and synonymy. Additionally, a scene often consists of several objects, which are frequently not related to the object that determines the class of the image. In such cases, the other objects in the image act as noise.

Concept $\mathcal{A}$: Concept $\mathcal{B}$:	LSCM:truck vs. LSCM:gr.vehicle	LSCM:sports vs. LSCM:basketball	LM:computer vs. LM:elec. device	LM:animal vs. LM:bird
$\mathbf{is\_a}(\mathcal{A},\mathcal{B})$	1	0.007	1	0.004
$\mathbf{is\_a}(\mathcal{B},\mathcal{A})$	0.012	1	0.011	1
$\mathbf{is\_a}^{\text{mean}}(\mathcal{A},\mathcal{B})$	1	0.052	1	0.062
$\mathbf{is\_a}^{\text{mean}}(\mathcal{B},\mathcal{A})$	0.326	1	0.07	1
Base Sim.	0.848	0.959	0.915	0.390
Eucl. Sim.	0.835	0.908	0.854	0.350
SupMin Sim.	0.435	0.545	0.359	0.309
Jacc. Sim.	0.870	0.814	0.733	0.399
Cosine Sim.	0.974	0.994	0.975	0.551

Concept $\mathcal{A}$: Concept $\mathcal{B}$:	LM:gondola vs. LSCM:boat_ship	LSCM:group vs. LM:audience	LSCM:truck vs. LM:vehicle	LSCM:truck vs. LM:conveyance
$\mathbf{is\_a}(\mathcal{A},\mathcal{B})$	0.016	0.006	0.022	0.022
$\mathbf{is\_a}(\mathcal{B},\mathcal{A})$	0.009	1	0.012	0.012
$\mathbf{is\_a}^{\text{mean}}(\mathcal{A},\mathcal{B})$	0.86	0.022	0.748	0.769
$\mathbf{is\_a}^{\text{mean}}(\mathcal{B},\mathcal{A})$	0.167	1	0.301	0.281
Base Sim.	0.72	0.78	0.58	0.58
Eucl. Sim.	0.66	0.71	0.40	0.38
SupMin Sim.	0.069	0.082	0.22	0.22
Jacc. Sim.	0.49	0.42	0.54	0.52
Cosine Sim.	0.69	0.82	0.66	0.67

Table 1. Examples of pairs of matched intra-ontology concepts (above) and cross-ontology concepts (below), column-wise.

In order to fuzzify our source concepts, we have applied the hierarchical scoring algorithm from Section 4 independently for each of the source ontologies. As a reference ontology, we have used an extended version of the Wikipedia's so-called main topic classifications (adding approx. 3 additional concepts to every first level class), containing more than 30 categories. For each topic category, we included a set of corresponding documents from the Inex 2007 corpus.

The new combined knowledge body has been constructed by first taking the union of all fuzzified source concepts. For every pair of concepts, we have computed their Gödel subsumptional relations, as well as the degree of their similarities (applying the measures from Section 5.2 and the standard cosine measure). Apart from the classical Gödel subsumption defined in (6), we consider a version of it which takes the average over all x instead of the smallest value, given as $\mathbf{is\_a}^{\text{mean}}(\mathcal{A}', \mathcal{A}) = \text{avg}_{x \in X} \mu_{\mathcal{A}' \to \mathcal{A}}(x)$. The results for several intra-ontology concepts and several cross-ontology concepts are given in Table 1. Fig. 3 shows a fragment of the common fuzzy ontology built for LSCOM and LabelMe. The labels of the edges of the graph correspond to the values of the fuzzy subsumptions between concepts.

We will underline several shortcomings that need to be adressed in future work. Due to data heterogeneity, it appears that the fuzzy **is_a**-structure is reflected better within one single ontology, as compared to cross-ontology relations which are more interesting. Additionally, some part_of relations are expressed as subsumptional (e.g. *torso* **is_a** *person*) which is a natural effect in view of the instance-representations. Indeed, the textual representation of images needs to be improved by accounting for the limitations discussed earlier in this section.

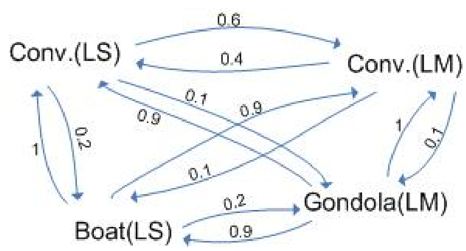

Fig. 3. A fragment of the common fuzzy ontology of LSCOM (LS) and LabelMe (LM).

Note that computing the common fuzzy ontology is inexpensive, once we have in hand the fuzzy representations of the source concepts made available by the hierarchical scoring algiorithm.

7 Conclusion and Open Ends

Whenever collaboration between knowledge resources is required, it is important to provide procedures which make explicit to users the relations that hold between different terms of these resources. In an attempt to solve this problem, we have proposed a fuzzy theoretical approach to build a common ontology for a set of source ontologies which contains these relations, as well as the degrees to which they hold, and can be queried upon need by different parties within a collaborative framework.

In future work, we will investigate the impact of the choice of a reference ontology onto the concept fuzzification and the quality of the constructed fuzzy common ontology. Additionally, the approach will be extended with elements of OWL 2, including relations and axioms between instances which is not covered by the ontology definition used in this work.

References

1. J.-I. Akahani, K. Hiramatsu, and T. Satoh. Approximate query reformulation based on hierarchical ontology mapping. In *In Proc. of Intl Workshop on SWFAT*, pages 43–46, 2003.
2. A. Bahri, R. Bouaziz, and F. Gargouri. Dealing with similarity relations in fuzzy ontologies. In *Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE International*, pages 1–6. IEEE, 2007.
3. P. Buche, J. Dibie-Barthélemy, and L. Ibanescu. Ontology mapping using fuzzy conceptual graphs and rules. In *ICCS Supplement*, pages 17–24, 2008.
4. S. Calegari and D. Ciucci. Fuzzy ontology, fuzzy description logics and fuzzy-owl. In *WILF*, pages 118–126, 2007.
5. S. Calegari and E. Sanchez. A fuzzy ontology-approach to improve semantic information retrieval. In *URSW*, 2007.
6. V. Cross and X. Yu. A fuzzy set framework for ontological similarity measures. In *WCCI 2010, FUZZ-IEEE 2010*, pages 1 – 8. IEEE Compter Society Press, 2010.
7. J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer-Verlag, 1 edition, 2007.
8. A. Ferrara, D. Lorusso, G. Stamou, G. Stoilos, V. Tzouvaras, and T. Venetis. Resolution of conflicts among ontology mappings: a fuzzy approach. OM'08 at ISWC, 2008.
9. A. Gal and P. Shvaiko. Advances in web semantics i. chapter Advances in Ontology Matching, pages 176–198. Springer-Verlag, Berlin, Heidelberg, 2009.
10. M. S. Lacher and G. Groh. Facilitating the exchange of explicit knowledge through ontology mappings. In *Proceedings of the 14th FLAIRS Conf.*, pages 305–309. AAAI Press, 2001.
11. G.A. Miller. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.
12. B.C. Russell, A. Torralba, K.P. Murphy, and W.T. Freeman. Labelme: A database and web-based tool for image annotation. *International Journal of Computer Vision*, 77(1), 2008.
13. E. Sanchez and T. Yamanoi. Fuzzy ontologies for the semantic web. *Flexible Query Answering Systems*, pages 691–699, 2006.
14. J.R. Smith and S.F. Chang. Large-scale concept ontology for multimedia. *IEEE Multimedia*, 13(3):86 – 91, 2006.
15. U. Straccia. Towards a fuzzy description logic for the semantic web (preliminary report). In Asuncin Gomez-Perez and Jrme Euzenat, editors, *The Semantic Web: Research and Applications*, volume 3532 of *Lecture Notes in Computer Science*, pages 73–123. Springer Berlin / Heidelberg, 2005.
16. H. Stuckenschmidt. Approximate information filtering on the semantic web. In M. Jarke, G. Lakemeyer, and J. Koehler, editors, *KI 2002*, volume 2479 of *LNCS*, pages 195–228. Springer Berlin / Heidelberg, 2002.
17. K. Todorov, P. Geibel, and K.-U. Khnberger. Mining concept similarities for heterogeneous ontologies. In P. Perner, editor, *Advances in Data Mining. Applications and Theoretical Aspects*, volume 6171 of *LNCS*, pages 86–100. Springer Berlin / Heidelberg, 2010.
18. B. Xu, D. Kang, J. Lu, Y. Li, and J. Jiang. Mapping fuzzy concepts between fuzzy ontologies. In R. Khosla, R. J. Howlett, and L. C. Jain, editors, *Knowledge-Based Intelligent Information and Engineering Systems*, volume 3683 of *Lecture Notes in Computer Science*, pages 199–205. Springer Berlin / Heidelberg, 2005.
19. L.A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338 – 353, 1965.