Fuzzy logic framework for ontology instance alignment

Bogumiła Hnatkowska[1[0000−0003−1706−0205]], Adrianna Kozierkiewicz[1[0000−0001−8445−3979]], and Marcin Pietranik[1[0000−0003−4255−889X]]

Faculty of Information and Communication Technology
Wroclaw University of Science and Technology
Wybrzeze Wyspianskiego 27
50-370 Wroclaw, Poland
{bogumila.hnatkowska,adrianna.kozierkiewicz,marcin.pietranik}@pwr.edu.pl

Abstract. The widely addressed topic of ontology alignment to this day contains several open research questions that remain either unanswered or only vaguely tackled. One of them is designating alignments of concept instances, which according to the literature are addressed in a handful of publications. Therefore, in this paper we propose a formal framework based on fuzzy logic that can be used to determine such mappings. We provide several similarity functions and a set of inference rules for combining them. The approach has been experimentally verified using widely accepted datasets provided by the Ontology Alignment Evaluation Initiative, yielding promising results.

Keywords: ontology alignment · instance mappings · fuzzy logic · knowledge management

1 Introduction

Ontology alignment, a frequently researched topic, is the task of asserting a sound communication between separate computer systems which utilize different ontologies as their knowledge bases. To deliver such means of communication, a specific ontology alignment tool needs to select which elements from two ontologies refer to the same (or sufficiently similar) elements of the real world. When such selection is done, the final ontology alignment is a set of pairs of such elements along with a confidence degree to which these elements can be aligned and a relationship that holds between them.

Many different ontology alignment tools are based on computing a variety of different similarity measures between ontology elements, which are eventually combined into a single value used to judge whether or not the considered pair of elements can form a valid mapping. However, a plethora of the aforementioned ontology alignment tools focus mainly on designating alignments of concepts (also referred to as TBox alignment), while two remaining levels of ontology

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

abstractions, relations, and instances, are frequently treated neglectfully ([4], [16]). There is very little research available in the literature that addresses these issues.

In our previous publication [11] we addressed the level of relations. The main idea came from analyzing earlier approaches to the task, which (as aforementioned) we based on calculating different similarity measures. A naive approach to combining them would include calculating their average value. However, we wanted to include another layer of experts' knowledge concerning ontology alignment in the process. We achieved this goal by successfully applying the fuzzy logic to the task, in the form of fuzzy inference rules. This approach has been proved useful, rendering good results obtained from experimental verification we conducted using a state of the art datasets provided by the Ontology Alignment Evaluation Initiative ([16]). The obtained results became a straight inspiration for the next research.

The following paper is devoted to the level of instances. This level can be treated as the actual reflection of the objects taken from the real world, expressed using concepts definitions. In other words, while the level concepts contain abstract descriptions of the world (e.g *a Person* or *a Book*), the instance level contains materializations of the real objects (e.g stating that "John" is *a Person*). Therefore, the level of instances expresses the real knowledge about the assumed universe of discourse, and not generic properties.

Thus, the main contribution of the following paper is twofold. Firstly, we provide a set of functions that can be used to calculate a similarity between two instances from independent ontologies. Secondly, we formulated a set of fuzzy inference rules that could be used to reason about how close two instances describe the same elements from the real world.

The article is structured as follows. In the next section, an overview of the related research done in the field is given. Section 3 contains basic notions used throughout the paper, while Section 4 describes our approach to ontology alignments on the instance level. Section 5 is split into two subsections - the first walk the reader through the experimental procedure we designed to verify our framework. The second contains the experimental results gathered during the process. The last section is a summary and a brief overview of our upcoming research plans.

2 Related Works

The increasing number of methods available for schema or ontology matching mandate consensus for evaluation of these methods. The Ontology Alignment Evaluation Initiative (OAEI) is a coordinated international initiative to forge this consensus OAEI [16]. Since 2004, OAEI organises evaluation campaigns aiming at evaluating ontology matching technologies. Organizers provide benchmark datasets consisting of a set of pairs of ontologies with their corresponding alignment, which is supposed to be treated as the correct one. Most of these datasets are devoted to schema matching, while instance matching is treated

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

more neglectfully. Nevertheless, it is possible to find datasets such as: IIMB, Sabine, Doremus, SPIMBENCH, Sandbox ([7]). In our work we especially focus on IIMB dataset because it contains 80 ontologies. Each ontology has been created by systematically applying a set of transformations to the reference ontology such as: data value transformation, data structure transformation and data semantics transformation. IIMB dataset has been used in OAEI campaigns held in 2009, 2010, 2011, 2012 and 2018.

In the last years the interest in participating in the OAEI instance matching competition has not been prominent. Among the systems which participated the one which stand out are out LogMap [14], AML [8], Codi [12], SBUEI [15] and semsim. Performance of those systems have been verified in 2011-2018 years based on IIMB datasets.

AML (AgreementMakerLight) [8] approach to instance matching is build on Data Property values of individuals and the relations between individuals. In the newer version, the AML added to its instance matching arsenal the same lexical-based strategy it was already using for class and property matching based on ontologies annotations. However, the efficiency of the system has not been satisfying because the authors were unable to properly configure this matching strategy and ensure its efficiency.

The better results were achieved by LogMap [14]. LogMap since 2011 has evolved from a logic-based tool to an advanced system that applies additional features like lexical indexation, propositional horn reasoning, axiom tracking, local repair, and semantic indexation. LogMap performance is very high, in particular for schema matching tasks. However, its performance for instance mapping leaves space for improvement, especially in the case of the Recall measure.

CODI (CombinatorialOptimization for Data Integration) [12] uses terminological structure for ontology matching. The current implementation produces mappings between concepts, properties, and individuals. The system combines lexical similarity measures with schema information. Authors assume that they have one common TBox and two different ABoxes, where both TBoxes have been integrated beforehand. The efficiency of CODI is not high, because in benchmark ontologies some individuals (instances) are not assigned to any concepts, hence no TBoxes are available.

SBUEI addresses two issues - instance matching and schema matching [15]. It utilizes schema matching results in instance matching tasks in order to track direct matching on the schema level. Similar to CODI, SBUEI is not efficient if the instances are not associated with any concepts.

To the best of our knowledge there is no formal information about semsim system. The system participated in OAEI 2018 competition, however no documentation has been provided.

The multitude of alignment systems forces the improvement of their effectiveness by applying different techniques. The most popular are string- and language-based methods (i.e LogMap, RIMOM, FALCON, SAMBO, AML, etc.) [1], [2]. Some systems incorporate external sources like WordNet (i.e. LogMap, YAM, SAMBO, etc.). Others, like i.e SEMINT, ProbaMap, LSD, MoTo apply the

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

newest achievements from the machine learning and artificial intelligence fields. It is possible to find an application of Naive Bayes classifiers, neural networks, SVM, or clustering techniques for correspondence determination. To the best of our knowledge, the fuzzy logic-based approach has not been widely investigated.

The papers [6], [9] present a fuzzy-based approach of concept alignment. However, both works are in the preliminary stage and focus only on the concept level of the ontology. Our previous works [11] partially fills this gap. We incorporated fuzzy rules for designating ontology alignments on the relation level. We claim that this is the very first research that shows the usefulness of the fuzzy logic-based framework for ontology instance alignment.

3 Basic notions

Before presenting our fuzzy based approach to instance alignment we will introduce some basic notions important to understand our ideas. Our ontology model is defined as a quintuple. Let (A, V) be a pair, where A is a set of attributes describing objects and V is a set of valuations of such attributes (their domains) such that $V = \bigcup_{a \in A} V_a$, where V_a is a domain of a particular attribute. The (A, V)-based ontology is represented as follow:

$$O = (C, H, R^C, I, R^I) \tag{1}$$

where:

- C is a finite set of concepts,
- H is a concepts' hierarchy, that may be treated as a distinguished relation between concepts,
- R^C is a finite set of binary relations between concepts $R^C = \{r_1^C, r_2^C, ..., r_n^C\}$, $n \in N$, such that every $r_i^C \in R^C$ $(i \in [1, n])$ is a subset of a cartesian product, $r_i^C \subset C \times C$,
- I denotes a finite set of instances' identifiers,
- $R^I = \{r_1^I, r_2^I, ..., r_n^I\}$ is used to denote a finite set of binary relations between concepts' instances.

In our previous works like [10] or [11] we describe in details the ontology model and their components. In this work, we will focus only on concepts and instances level.

Instances are understood as a specific materialisation of concepts. Instances can not exists without belonging to concepts, Thus, firstly, we will introduce the concepts level of an ontology:

$$c = (id^c, A^c, I^c) \tag{2}$$

where:

- id^c is an identifier of the concept c,
- A^c is a set of its attributes,

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

– I^c is a set of concepts' c instances.

By $a \in c$ we denote the fact, that the attribute a belongs to the concept's c set of attributes A^c. The limitations of many ontology model is the lack of information about attributes semantic. For example, the same attribute *address* may carry different meanings while included in the *Home* concept and completely different when incorporated in the *Personal Website* concept. Our ontology model is based on a notion of attributes' semantics, which gives explicit meanings to attributes when they are included in different concepts. Thus, we need to define a sub-language of the sentence calculus denotes as L_S^A. The set L_S^A consists of an atomic description of attributes from the set D_A and logical operators of conjunction, disjunction, and negation. A partial function:

$$S_A : A \times C \to L_S^A \tag{3}$$

allows us to assign a logical sentence from L_S^A to attributes within a specific concept. The context of concept c is defined as a conjunction of semantics of each of its attributes: $ctx(c) = S_A(a_1, c) \wedge S_A(a_2, c) \wedge ... \wedge S_A(a_n, c)$.

The function S_A allows us to formally define relation: *equivalency* (denoted by $\equiv$), *generalization* (denoted by $\leftarrow$) and *contradiction* (denoted by $\sim$) between attributes:

– Two attributes $a \in A^{c_1}, b \in A^{c_2}$ are semantically equivalent $a \equiv b$ if the formula $S_A(a, c_1) \Leftrightarrow S_A(b, c_2)$ is a tautology for any two $c_1 \in C_1, c_2 \in C_2$.
– The attribute $a \in A^{c_1}$ in concept $c_1 \in C_1$ is more general than the attribute $b \in A^{c_2}$ in concept $c_2 \in C_2$ (denoted by $a \leftarrow b$) if the formula $S_A(b, c_2) \Rightarrow S_A(a, c_1)$ is a tautology for any two $c_1 \in C_1, c_2 \in C_2$.
– Two attributes $a \in A^{c_1}, b \in A^{c_2}$ are in semantical contradiction $a \sim b$ if the formula $\neq (S_A(a, c_1) \wedge S_A(b, c_2))$ is a tautology for any two $c_1 \in C_1, c_2 \in C_2$.

The paper is devoted to designating an alignment between two ontologies on the level of instances. Thus, we broadly describe this ontology level. For given a concept c, its instances from the set I^c are defined as a tuple:

$$i = (id^i, v_c^i) \tag{4}$$

where

– id^i is an instance identifier,
– v_c^i is a function with a signature:

$$v_c^i : A^c \to 2^V \tag{5}$$

According to the equation above, the valuation of a particular attribute within an instance can be multivalued. In other words, it can be represented as a set with repetitions of atomic values taken from the domain. Such approach is cohesive with variety of ontology representation formats (e.g. OWL).

For simplicity, we write $i \in c$ which can be understood that the instance i belongs to the concept c. By $I = \bigcup_{c \in C} \{id^i | (id^i, v_c^i) \in I^c\}$ we denote the set of

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

all instances' identifiers. In the further part of this paper we used two auxiliary functions. First of them, returns a set containing identifiers of instances assigned to a given concept c: $Ins(c) = \{id^i | (id^i, v^i_c) \in I^c\}$. The second one, gives a set of concepts to which an instance with some identifier belongs: $Ins^{-1}(i) = \{c | c \in C \wedge i \in c\}$.

To simplify a notation by $i \in r^I$ we will denote a situation in which an instance i participates in a relation $r^I \in R^I$ with some other unspecified instance. Formally, we can define this as $i \in r^I \implies \exists_{i' \in I} : (i, i') \in r^I \vee (i', i) \in r^I$. Additionally, we introduce a helper function rng used during processing of instances and their relations. It is used to designate sets of instances with which some instance is connected through certain relation:

$$rng(i, r^I) = \{i' \in I | (i, i') \in r^I\} \tag{6}$$

Assuming the existences of two ontologies, the integration of them is possible only in case of existing alignment. Such alignment, can allows us to "translate" content of one (source) ontology to the content of some other ontology (target). Formally speaking, between two independent (A, V)-based ontologies $O_1 = (C_1, H_1, R^{C_1}, I_1, R^{I_1})$ and $O_2 = (C_2, H_2, R^{C_2}, I_2, R^{I_2})$ there exist a set of correspondences, called alignment, defined in the following way:

$$Align(O_1, O_2) = \{Align_C(O_1, O_2), Align_I(O_1, O_2), Align_R(O_1, O_2)\} \tag{7}$$

The main aim of this work is determination of $Align_I(O_1, O_2)$. Due the fact, that instances are tightly connected with concepts, two instances can be mapped if belong to the mapped concepts. Thus, to determine an instance alignment we need as an input a concepts alignment $Align_C(O_1, O_2)$ defined in the following way:

$$Align_C(O_1, O_2) = \{(c_1, c_2, \lambda_C(c_1, c_2)) | c_1 \in C_1 \wedge c_2 \in C_2 \wedge \lambda_C(c_1, c_2) \geq T_C\} \tag{8}$$

where:

- c_1, c_2 are concepts from O_1 and O_2 respectively,
- $\lambda_C(c_1, c_2)$ is a value of a degree to which concept c_1 can be aligned into the concept c_2, a vast majority of alignments between two ontologies include only mappings of concepts that are equivalent with 100% certainty. The value of $\lambda_C(c_1, c_2)$ can be calculated in different way i.e like in our previous works [11] or taken directly from other alignment systems.
- T_C represents an assumed threshold

The ontology alignment on the instance level is a set of sets of alignments of instances belonging to two already aligned concepts and can be formulated as:

$$Align_I(O_1, O_2) = \{(i_1, i_2) | i_1 \in I^{C_1} \wedge i_2 \in I^{C_2}\} \tag{9}$$

We consider only relation pairs that have been processed and eventually selected by fuzzy-based alignment algorithm described in the further sections.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

4 Fuzzy based approach to instance alignment

The main aim of our work is to determine the mappings between two ontologies on the instance level. It is performed with the use of a fuzzy system, which has one output variable connection (*con*) with two possible values: *independent* and *equivalent*. We distinguish seven input variables presented on Table 2 and eight inference rules to decide if two instances represent the same (equivalent) or different (independent) phenomenon. The rules are presented in Table 1.

In our fuzzy framework, we use the Mamdani type rule inference, the centroid of gravity method to defuzzify the output variable, and the maximum operator to accumulate the activated terms.

Table 1. Fuzzy inference rules

ID	Rule
1	IF PRSIM IS veryhigh AND CPR IS high AND CCO IS high THEN con IS equivalent
2	IF PRSIM IS veryhigh AND CCO IS high AND (PARTSIM IS high OR MAXSIM IS high) THEN con IS equivalent
3	IF PRSIM IS high AND CCO IS high AND MAXSIM IS high THEN con IS equivalent
4	IF MAXSIM IS high AND ASIM IS high THEN con is equivalent
5	IF ASIM IS high AND RSIM IS high THEN con is equivalent
6	IF CCO IS high AND (MAXSIM IS high AND ASIM IS high) THEN con is equivalent
7	IF PRSIM IS medium AND (NOT MAXSIM IS high AND NOT CPR IS high) THEN con IS independent
8	IF PRSIM IS medium AND (NOT CCO IS high AND (NOT RSIM IS high OR NOT ASIM IS high)) THEN con IS independent

Property similarity ($PRSIM$) is the most important input variable. How it is calculated is shown below. The definition is long because it requires the provision of auxiliary elements. The computation of the other input variables either relies on $PRSIM$ or is easily explained in natural language. The details are shown in Table 2.

The function λ_{value} takes as an input two sets (x and y) of elementary values (e.g. strings) and uses a similarity function $sim\_function$ for comparing atomic values. It calculates an overall similarity between the given sets:

$$\lambda_{value}(x,y) = \frac{\sum\limits_{v \in x} \max\limits_{v' \in y} sim\_function(v,v') + \sum\limits_{v \in y} \max\limits_{v' \in x} sim\_function(v,v')}{|x| + |y|} \tag{10}$$

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

The $sim\_function$ can be any arbitrary given similarity function for particular datatypes. Since all of the values found in ontologies, during the experiment described in the next section, were cast on text type, in further parts of the article the widely known Longest Common Subsequence similarity is used as the $sim\_function$.

By $to\_value\_representation$ we denote a function which takes as an input an instance and converts it in a set of values of its attributes. If an instance does not contain any attributes, then the function returns a single-element set containing this instance identifier:

$$to\_value\_representation(i) = \begin{cases} \bigcup_{c \in Ins^{-1}(i)} \bigcup_{a \in c} \{v_c^i(a)\} & \text{, if } \exists c \in Ins^{-1}(i) : A^c \neq \phi \\ \{id^i\} & \text{, otherwise} \end{cases} \tag{11}$$

We introduce an auxiliary set A_{AR} containing four kinds of explicitly given alignments of two ontologies O_1 and O_2. Its elements represent different types of connections established between two ontologies on a level of concepts and relations, thus they include attribute-attribute mappings, attribute-relation mappings, relation-attribute mappings and relation-relation mappings. The set A_{AR} is formally defined below:

$$\begin{aligned} A_{AR}(O_1, O_2) = & \{(a_1, r^{C_2}) | a_1 \in A, r^{C_2} \in R^{C_2}\} \cup \{(r^{C_1}, a_2) | a_2 \in A, r^{C_1} \in R^{C_1}\} \cup \\ & \{(a_1, a_2) | a_1, a_2 \in A\} \cup \{(r^{C_1}, r^{C_2}) | r^{C_1} \in R^{C_1}, r^{C_2} \in R^{C_2}\} \end{aligned} \tag{12}$$

In order to process two instances in the context of the alignment of their mutual characterising properties (attributes and relations) we use a helper function which for given two instances i_1 and i_2 from two ontologies O_1 and O_2 returns a subset of $A_{AR}(O_1, O_2)$ with alignments somehow connected with the given instances $a \in \bigcup_{c \in Ins^{-1}(i_1)} A^c$.

$$\begin{aligned} & \tilde{A}_{AR}(i_1, i_2, O_1, O_2) = \\ & \quad \{(a_1, r^{C_2}) | (a_1, r^{C_2}) \in A_{AR}(O_1, O_2), a_1 \in \bigcup_{c_1 \in Ins^{-1}(i_1)} A^{c_1}, i_2 \in r^{C_2}\} \cup \\ & \quad \{(r^{C_1}, a_2) | (r^{C_1}, a_2) \in A_{AR}(O_1, O_2), a_2 \in \bigcup_{c_2 \in Ins^{-1}(i_2)} A^{c_2}, i_1 \in r^{C_1}\} \cup \\ & \quad \{(a_1, a_2) | (a_1, a_2) \in A_{AR}(O_1, O_2), a_1 \in \bigcup_{c_1 \in Ins^{-1}(i_1)} A^{c_1}, a_2 \in \bigcup_{c_2 \in Ins^{-1}(i_2)} A^{c_2}\} \cup \\ & \quad \{(r_1, r_2) | (r_1, r_2) \in A_{AR}(O_1, O_2), i_1 \in r_1, i_2 \in r_2\}\} \end{aligned} \tag{13}$$

Final version of the property relatedness ($PRSIM$) calculation is presented on Algorithm 1.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

Algorithm 1: Calculate property relatedness

Input : $i_1 \in I_1,\ i_2 \in I_2, O_1, O_2$
Output: $PRSIM(i_1, i_2) \in [0, 1]$

$partial := 0$
foreach $(e_1, e_2) \in \tilde{A}_{AR}(i_1, i_2, O_1, O_2)$ **do**
  foreach $(c_1, c_2) \in Ins^{-1}(i_1) \times Ins^{-1}(i_2)$ **do**
    if $e_1 \in A \wedge e_2 \in A$ **then**
      $v = v^i_{c_1}(e_1)$
      $v' = v^i_{c_2}(e_2)$
    end
    else if $e_1 \in R^{C_1} \wedge e_2 \in R^{C_2}$ **then**
      $v = \bigcup_{i \in rng(i_1, e_1)} to\_value\_representation(i)$
      $v' = \bigcup_{i' \in rng(i_2, e_2)} to\_value\_representation(i')$
    end
    else if $e_1 \in A \wedge e_2 \in R^{C_2}$ **then**
      $v = v^i_{c_1}(e_1)$
      $v' = \bigcup_{i' \in rng(i_2, e_2)} to\_value\_representation(i')$
    end
    else if $e_1 \in R^{C_1} \wedge e_2 \in A$ **then**
      $v = \bigcup_{i \in rng(i_1, e_1)} to\_value\_representation(i)$
      $v' = v^i_{c_2}(e_2)$
    end
    $partial = partial + \lambda_{value}(v, v')$
  end
end
if $|\tilde{A}_{AR}(i_1, i_2, O_1, O_2)| > 0$ **then**
  return $\frac{partial}{|\tilde{A}_{AR}(i_1, i_2, O_1, O_2)|}$
end
else
  return 0
end

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

Table 2. Input variables with calculation description

Variable	Comment	Calculation description
PRSIM	Similarity of properties (attributes and relations)	PRSIM (Algorithm 1)
MAXSIM	Max similarity of properties	Ratio of mapped properties defined in two instances i_1 and i_2 with maximal (1.0) similarity to the number of all their properties
PARTSIM	Partial similarity of attributes	Ratio of mapped attributes defined in two instances i_1 and i_2 with maximal (1.0) similarity to the number of all their attributes
ASIM	Similarity of attributes	PRSIM limitted to attributes
RSIM	Similarity of relations	PRSIM limitted to relations
CPR	Consistency of properties	$p * \lambda_c(c_1, c_2)$ where p – Ratio of mapped properties defined in two instances i_1 and i_2 to the number of all of their properties for such $c_1 \in Ins^{-1}(i_1)$ and $c_2 \in Ins^{-1}(i_2)$ with the highest $\lambda_c(c_1, c_2)$
CCO	Consistency of concepts	Ratio of mapped concepts two instances i_1 and i_2 belong to calculated as: $\frac{\lvert(Ins^{-1}(i_1) \times Ins^{-1}(i_2)) \cap Align_C(O_1, O_2)\rvert}{\lvert Ins^{-1}(i_1) \times Ins^{-1}(i_2)\rvert}$

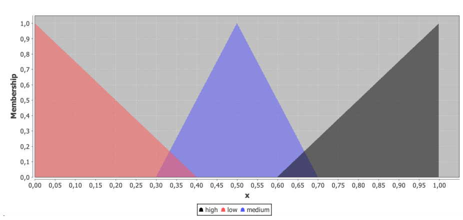

Fig. 1. ASIM fuzzy variable

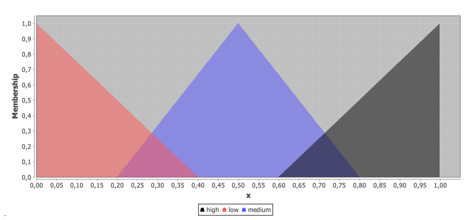

Fig. 2. CCO fuzzy variable

5 Experimental verification

Our fuzzy based approach to instance alignment, described in Section 4, has been implemented and verified against a widely accepted and incorporated benchmark dataset provided by the Ontology Alignment Evaluation Initiative (OAEI), aforementioned in Section 2. We have developed a special dedicated java tool that is

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

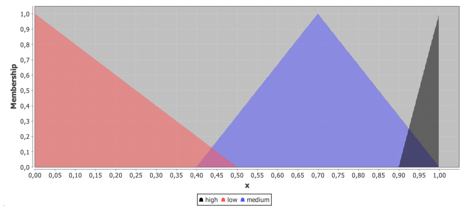

Fig. 3. CPR fuzzy variable

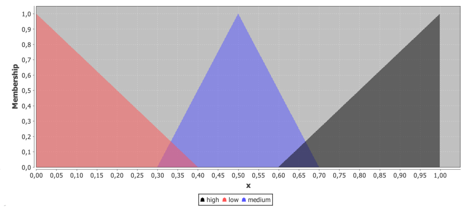

Fig. 4. MAXSIM fuzzy variable

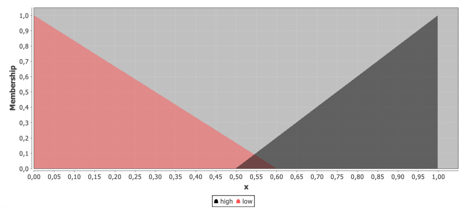

Fig. 5. PARTSIM fuzzy variable

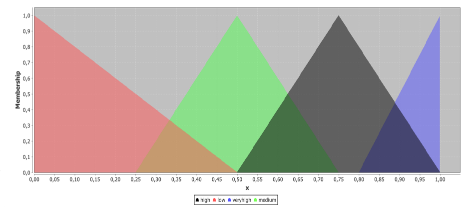

Fig. 6. PRSIM fuzzy variable

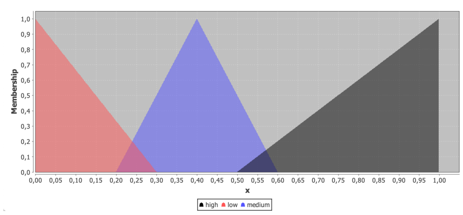

Fig. 7. RSIM fuzzy variable

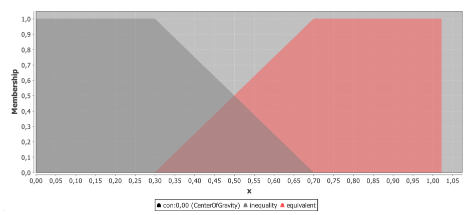

Fig. 8. CON fuzzy variable

able to parse the ontologies expressed in OWL and mappings between concepts expressed in RDF. Our tool (which is online available [18]) uses the jfuzzylogic library [5] to handle fuzzy logic computations.

The results of the instance alignment calculated by our framework were then compared with the expert alignment provided by OAEI along with ontology datasets. Basic measures like Precision, Recall and F-measure were used for this purpose. The IIMB benchmark dataset was divided into four separate test suits – one with data value transformations, one with data structure transformations, one with data semantics transformations and the final one with mixed transformations. According to this division, our experiment was also divided into four stages.

We wanted to verify a hypothesis that the performance of our approach is better or at least not worse than the existing alignment system. All of the collected data can be found in [18]. They have been compared with results presented in the website [17]. A summary of the results is presented in Table 3.

Thus we obtain nine samples – Precision, Recall, and F-measure for our framework, AML, and LogMap, respectively. All the analysis was made with a

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

Table 3. The avarage values of measures from different experiment stages

		Precision	**Recall**	**F-Measure**
Data value transformation	Our framework	0.9630	0.5280	0.6319
	AML	0.8933	0.7890	0.8280
	LogMap	0.8964	0.8928	0.8894
Data structure transformation	Our framework	0.9241	0.4585	0.4811
	AML	0.4194	0.4329	0.4241
	LogMap	0.9343	0.9856	0.9592
Data semantics transformation	Our framework	0.9488	0.9469	0.9388
	AML	0.7466	0.8888	0.7964
	LogMap	0.8546	0.9466	0.8926
Mixed transformations	Our framework	0.9243	0.0680	0.1095
	AML	0.3342	0.2943	0.2953
	LogMap	0.9199	0.7581	0.8195

significance level $\alpha = 0.05$. Before selecting the appropriate test, we analyzed the distribution of all samples using the Shapiro-Wilk test. None of the samples have a normal distribution, thus we used the non-parametric Friedman ANOVA test for further analysis.

For Precision samples, we obtained the Friedman value test equal to 43.075. The *p-value* is less than 0.000001. The Dunn Benferroni's post-hoc test allows us to conclude that our approach achieves the best results. Log Map (statistical value equal to 3.499, *p-value* equal to 0.002), as well as AML (statistical value equal to 6.56, *p-value* less than 0.000001), generate smaller number o correct correspondences than our framework.

The result of the Friedman test for Recall samples is equal to 59.299 and the *p-value* is less than 0.000001, which means that at least one system works differently from another. Dunn Benferroni's test chosen as post-hoc points out that our approach is not worse than the AML system (statistic value equal to 2.25, *p-value* equal to 0.073) and there is a statistical difference between Recall value obtained by our approach and the LogMap system (statistic value equal to 5.23, *p-value* less than 0.000001).

A similar conclusion is a result of statistical analysis for F-measure samples. The statistical value of the Friedman test is equal to 28.825 and *p-value* less than 0.000001. The Dunn Bonferroni confirm that there is no significant difference for the result obtained by our and AML system (statistic value equal to 0.533, *p-value* equal to 1). In terms of F-measure LogMap is better than both (our and AML approach) - statistic value equal to 4.348, *p-value* less than 0.000041.

The results of experiments are promising. We have noticed that our framework deals with data semantics transformation perfectly. In other test cases, the performance of our approach is also very good - statistically not worse than AML, and in some cases even better than AML or LogMap. The proposed frame-

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

work copes with incoherence, which entails reducing the number of false-positive results.

In our experiments, for all test cases, we obtained a very high value of Precision. This means that almost all correspondences found were correct. The lower value of Recall is caused by the non-uniform quality of expected, automatically-generated correspondences. The expected correspondences were created by applying a sequence of transformations of various lengths (i.e., number of transformations) and complexity (i.e., strength of data manipulations applied) [2]. Correspondences in benchmark alignments are often not intuitive and more difficult to agree with than to detect.

6 Future works and summary

The following paper is devoted to finding ontology alignment on the instance level. The proposed solution is a fuzzy-logic-based framework built on a set of functions that calculate similarities between concept instances. The values of these functions are eventually treated as fuzzy variables, which are then incorporated into a set of inference rules used for determining the final instance mappings.

The entire fuzzy framework has been experimentally verified utilizing a state-of-the-art dataset provided by the Ontology Alignment Evaluation Initiative. The results obtained are promising and in many cases outperform competing ontology alignment solutions in terms of assumed quality measures. We claim that adjusting fuzzy variables and inference rules can further improve the quality of alignments collected by our framework.

In the upcoming future, we plan to conduct more extensive experiments using different datasets created by the Ontology Alignment Evaluation Initiative, focusing on the scalability of our framework. Furthermore, we will extend the framework to work on the the level of concepts, a recent area of interest that has not beed addressed in our previous research.

References

1. Aguirre J.L., Grau B.C., Eckert K., Euzenat J., Ferrara A., Hague R.W., Hollink L., Jimenez-Ruiz E., Meilicke CH., Nikolov A, Ritze D., Scharffe F., Shvaiko P., Svab-Zamazal O., Trojahn C., Zapilko B. (2012),Results of the ontology alignment evaluation initiative 2012. In: Proceedings of the 7th International Ontology matching workshop, Boston (MA, US), pages 73–115, 2012
2. Algergawy, A., et all (2018). Results of the ontology alignment evaluation initiative 2018. In Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th ISWC (OM 2018) (Vol. 2288, pp. 76-116).
3. Ardjani F., Bouchiha D., Malki M. (2015). Ontology-Alignment Techniques: Survey and Analysis. International Journal of Modern Education, Computer Science, 7(11)
4. Cheatham, M., Pesquita, C., Oliveira, D., McCurdy, H. B. (2018). The properties of property alignment on the semantic web. International Journal of Metadata, Semantics and Ontologies, 13(1), 42-56.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68

5. Cingolani P., Alcalá-Fdez J. (2013). jFuzzyLogic: a java library to design fuzzy logic controllers according to the standard for fuzzy control programming. International Journal of Computational Intelligence Systems, 6(sup1), 61-75.
6. de Lourdes Martínez-Villaseñor M., González-Mendoza M. (2017, November). Fuzzy-Based Approach of Concept Alignment. In International Conference on Ubiquitous Computing and Ambient Intelligence (pp. 172-180). Springer, Cham.
7. Daskalaki E., Flouris G., Fundulaki I., Saveta T. (2016). Instance matching benchmarks in the era of linked data. Journal of Web Semantics, 39, 1-14.
8. Faria, D., Pesquita, C., Balasubramani, B., Tervo, T., Carriço, D., Garrilha, R., Couto F.M., Cruz, I. F. (2018). Results of AML participation in OAEI 2018. In Proceedings of the 13th International Workshop on Ontology Matching co-located with the 17th International Semantic Web Conference (Vol. 2288)
9. Fernandez S., Velasco J. R., Lopez-Carmona M. A. (2009).A fuzzy rule-based system for ontology mapping. In International Conference on Principles and Practice of Multi-Agent Systems (pp. 500-507). Springer, Berlin, Heidelberg
10. Hnatkowska B., Kozierkiewicz A., Pietranik M. (2020). Semi-Automatic Definition of Attribute Semantics for the Purpose of Ontology Integration. IEEE Access, vol. 8, pp. 107272-107284, 2020, doi: 10.1109/ACCESS.2020.3000035.
11. Hnatkowska, B., Kozierkiewicz, A., Pietranik, M. (2021). Fuzzy based approach to ontology relations alignment. In 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-7). IEEE
12. Huber, J., Sztyler, T., Noessner, J., Meilicke, C. (2011). Codi: Combinatorial optimization for data integration–results for oaei 2011. Ontology Matching, 134.
13. Pietranik M., Nguyen N. T. (2011). Semantic Distance Measure between Ontology Concept's Attributes. In International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (pp. 210-219). Springer, Berlin, Heidelberg.
14. Ruiz E.J, Grau B.C., Zhou Y., Horrocks I. (2012). Large-scale Interactive Ontology Matching: Algorithms and Implementation. In the 20th European Conference on Artificial Intelligence (ECAI 2012)
15. Taheri, A., Shamsfard, M. (2012). SBUEI: results for OAEI 2012. In Ontology Matching.
16. http://oaei.ontologymatching.org/
17. http://islab.di.unimi.it/content/im_oaei/2018/
18. https://github.com/bhnatkowska/FuzzyLogicInstanceAlignment

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_68