# Comparing Argumentation Frameworks for Composite Ontology Matching

Cássia Trojahn[12], Paulo Quaresma[12], Renata Vieira[3], Antoine Isaac[4]

[1]Departamento de Informática, Universidade de Évora, Portugal
[2]CENTRIA - Centro de Inteligencia Artificial, Universidade Nova de Lisboa, Lisboa, Portugal
[3]Faculdade de Informática, Pontifícia Universidade Católica do Rio Grande do Sul, Brazil
[4]Vrije Universiteit, Department of Computer Science, Amsterdam, Netherlands
cassia@di.uevora.pt, pq@di.uevora.pt, renata.vieira@pucrs.br, aisaac@few.vu.nl

**Abstract.** Resolving the semantic heterogeneity problem is crucial to allow interoperability between ontology-based systems. Ontology matching based on argumentation is an innovative research area that aims at solving this issue, where agents encapsulate different matching techniques and the distinct mapping results are shared, compared, chosen and agreed. In this paper, we compare three argumentation frameworks, which consider different notions of acceptability: based on values and preferences between audiences promoting these values, based on the confidence level of the arguments, and based on voting on the arguments. We evaluate these frameworks using realistic ontologies from an established ontology matching evaluation test set. The best matcher varies depending on specific characteristics of each set, while considering voting on arguments the results are similar to the best matchers for all sets.

## 1 Introduction

Ontologies have proven to be an essential element in a range of applications in which knowledge plays a key role. Resolving the semantic heterogeneity problem by means of *ontology matching* is crucial to allow the interoperability between ontology-based systems. Ontology matching is the process of linking corresponding entities (classes or proprieties) from different ontologies.

Many different approaches to the matching problem are found in the literature. The distinction between them is accentuated by the manner in which they exploit the features within an ontology. Whereas syntactic approaches consider measures of string similarity; semantic ones consider semantic relations usually on the basis of lexical oriented linguistic resources. Other approaches consider term positions in the ontology hierarchy or instances of the ontologies. However, each category of approaches offers a wide diversity of options. Such approaches have been surveyed from different perspectives in [6].

The matching systems are usually aware that a combination of different techniques are required for dealing with the problem. The different techniques are then aggregated in a unified process, which involves parallel or sequential execution of matchers, where the results can be combined varying from a single weighted sum of individual results to automatically learning the best matcher from preliminary results.

Some techniques will perform better than others for specific cases, depending on how well the technique fits the material available. Also, approaches that perform well for a specific case may not be successful in other ones.

An important issue in ontology matching is therefore to find effective ways of choosing among many techniques and their variations, and then combining their results. An innovative approach to solve this problem is to use frameworks of argumentation, where different matchers may work on the basis of alternative approaches arriving to distinct matching results (arguments) that must be shared, compared, chosen and agreed. The matchers exchange the arguments and construct their frameworks of argumentation. The preferred extensions of each matcher are then compared in order to identify set of globally acceptable arguments.

In this paper, we compare three different frameworks of argumentation, Value-based Argumentation Framework (VAF), Strength-based Argumentation Framework (S-VAF), and Voting-based Argumentation Framework (V-VAF). The VAF [2] is based on the classical framework of Dung, aggregating notions of audiences and preferences. The idea of VAF is to relate the arguments in the dispute to the social values represented by their acceptability for given audiences. Both S-VAF and V-VAF frameworks are based on the VAF, in order to effectively combine different audiences. The S-VAF [23] allows to associate to each argument, a confidence value that represents the strength of the argument. Using V-VAF it is possible to manage consensus, i.e., showing that the more often an argument is agreed on, the more chances for it to be valid. This paper extends the V-VAF presented in [9], in which we had compared V-VAF and S-VAF by evaluating their application to a range of individual mappers, in the context of a real-world library case.

The comparison and evaluation of the frameworks is carried out using real ontologies, commonly used in the evaluation of state-of-the-art matching systems. Particularly, the results of each individual matcher is compared with the results of the different frameworks.

The paper is structured as follows. Section 2 introduces the three argumentation frameworks. In Section 3, the argumentation process for ontology matching is described. In Section 4, the experiments are detailed and the discussion on the results is presented. Section 5 comments on related works. Finally, in Section 6, concluding remarks and future work are presented.

## 2 Argumentation Frameworks

Both S-VAF and V-VAF are development of the VAF, which is based on the classical system of Dung [4]. This section presents the basic notions upon these frameworks rely.

### 2.1 Classical Argumentation Framework

Dung, observing that the core notion of argumentation lies in the opposition between arguments and counter-arguments, defines an argumentation framework (AF) as follows:

**Definition 2.1.1** [4] An Argumentation Framework is a pair $AF = (AR, attacks)$, $AR$ is a set of arguments and $attacks$ is a binary relation on $AR$.

$attacks(a, b)$ means that the argument $a$ attacks the argument $b$. A set of arguments $S$ attacks an argument $b$ if $b$ is attacked by an argument in $S$. The key question about the framework is whether a given argument $a \in AR$ should be accepted or not. Dung proposes that an argument should be accepted only if every attack on it is rebutted by an accepted argument. This notion then leads to the definition of acceptability (for an argument), admissibility (for a set of arguments) and preferred extension:

**Definition 2.2.2** [4] An argument $a \in AR$ is *acceptable* with respect to set arguments $S$, noted $acceptable(a, S)$, if $\forall x \in AR\ (attacks(x, a) \longrightarrow \exists y \in S, attacks(y, x))$

**Definition 2.1.3** [4] A set $S$ of arguments is *conflict-free* if $\neg\exists x, y \in S, attacks(x, y)$. A conflict-free set of arguments $S$ is *admissible* if $\forall x \in S$, $acceptable(x, S)$. A set of arguments $S$ is a *preferred extension* if it is a maximal (with respect to set inclusion) admissible set of $AR$.

A preferred extension represents a consistent position within $AF$, which defends itself against all attacks and cannot be extended without raising conflicts.

### 2.2 Value-based Argumentation Framework (VAF)

In Dung's framework, all arguments have equal strength, and an attacks always succeed, except if the attacking argument is otherwise defeated. However, as noted in [14], in many domains, including ontology alignment, arguments may provide reasons which may be more or less persuasive. Moreover, their persuasiveness may vary according to their audience. Bench-Capon has extended the notion of AF so as to associate arguments with the social values they advance:

**Definition 2.2.1** [2] A Value-based Argumentation Framework (VAF) is a 5-tuple $VAF = (AR, attacks, V, val, P)$ where $(AR, attacks)$ is an argumentation framework, $V$ is a nonempty set of values, $val$ is a function which maps elements of $AR$ to elements of $V$ and $P$ is a set of possible audiences.

Practically, in [13], the role of value is played by the types of ontology match that ground the arguments, covering general categories of matching approaches: semantic, structural, terminological and extensional. We argue further — and will use later — that any kind of matching ground identified during a mapping process or any specific matching tools may give rise to a value. The only limitations are (i) a value can be identified and shared by a source of mapping arguments and the audience considering this information (ii) audiences can give preferences to the values. An extension to this framework, required for deploying argumentation processes, indeed allows to represent how audiences with different interests can grant preferences to specific values:

**Definition 2.2.2** [2] An Audience-specific Value-based Argumentation Framework (AVAF) is a 5-tuple $VAF_p = (AR, attacks, V, val, valpref_{aud})$ where $AR$, $attacks$, $V$ and $val$ are as for a VAF, $aud$ is an audience and $valpref_{aud}$ is a preference relation (transitive, irreflexive and asymmetric), $valpref_{aud} \subseteq V \times V$.

$valpref_{aud}(v_1, v_2)$ means that audience $aud$ prefers $v_1$ over $v_2$. Attacks are then deemed successful based on the preference ordering on the arguments' values. This leads to re-defining the notions seen previously:

**Definition 2.2.3** [2] An argument $a \in AR$ defeats an argument $b \in AR$ for audience *aud*, noted $defeats_{aud}(a, b)$, if and only if both $attacks(a, b)$ and not $valpref_{aud}(val(b), val(a))$. An argument $a \in AR$ is *acceptable* to audience $aud$ with respect to a set of arguments $S$, noted $acceptable_{aud}(a, S)$, if $\forall x \in AR,\ defeats_{aud}(x, a) \longrightarrow \exists y \in S,\ defeats_{aud}(y, x)$.

**Definition 2.2.4** [2] A set $S$ of arguments is *conflict-free* for audience $aud$ if $\forall x, y \in S,\ \neg attacks(x, y) \lor valpref_{aud}(val(y), val(x))$. A *conflict-free* set of arguments $S$ for audience $aud$ is *admissible* for $aud$ if $\forall x \in S, acceptable_{aud}(x, S)$. A set of arguments $S$ in the VAF is a *preferred extension* for audience $aud$ if it is a maximal admissible set (with respect to set inclusion) for *aud*.

In order to determine preferred extensions with respect to a value ordering promoted by distinct audiences, *objective* and *subjective* acceptance are defined:

**Definition 2.2.5** [2, 13] An argument $a \in AR$ is *subjectively acceptable* if and only if $a$ appears in **some** preferred extension for some specific audiences. An argument $a \in AR$ is *objectively acceptable* if and only if $a$ appears in **all** preferred extension for every specific audience.

### 2.3 Strength-based Argumentation Framework (S-VAF)

Value-based argumentation acknowledges the importance of preferences when considering arguments. However, in the specific context of ontology alignment, an objection can still be raised about the lack of complete mechanisms for handling persuasiveness. Indeed, many mapping tools actually output mappings with a strength that reflects the confidence they have in the similarity between the two entities. These confidence levels are usually derived from similarity assessments made during the alignment process, *e.g.* from edit distance measure between labels, or overlap measure between instance sets, as in [10]. They are therefore often based on objective grounds.

It is one of our goals to investigate whether considering confidence levels gives better results or not.[1] To this end, we adapt a formulation introduced in [24, 25] to consider the strength granted to mappings for determining attacks' success:

**Definition 2.3.1** A Strength and value-based Argumentation Framework (S-VAF) is a 6-tuple $(AR, attacks, V, val, P, str)$ where $(AR, attacks, V, val, P)$ is a value-based argumentation framework, and $str$ is a function which maps elements of $AR$ to real values from the interval $[0, 1]$, representing the *strength* of the argument. An audience-specific S-VAF is an S-VAF where the generic set of audiences is replaced by the definition of a specific $valpref_{aud}$ preference relation over V.

**Definition 2.3.2** In an audience-specific S-VAF, an argument $a \in AR$ defeats an argument $b \in AR$ for audience $aud$ if and only if $attacks(a, b) \wedge (\, str(a) > str(b) \vee (str(a) = str(b) \wedge valpref_{aud}(val(a), val(b)))\, )$

In other words, for a given audience, an attack succeeds if the strength of the attacking argument is greater than the strength of the attacked one; or, if both arguments have equal strength, the attacked argument is not preferred over the attacking argument by the concerned audience. Similarly to what is done for VAFs, an argument is acceptable for a given audience *w.r.t* a set of arguments if every argument defeating it is defeated by other members of the set. A set of arguments is conflict-free if no two members can defeat each other. Such a set is admissible for an audience if all its members are acceptable for this audience *w.r.t* itself. A set of arguments is a preferred extension for an audience if it is a maximal admissible set for this audience.

### 2.4 Argumentation Frameworks with Voting (V-VAF)

The previously described frameworks capture the possible conflicts between mappers, and find a way to solve them. However, they still fail at rendering the fact that sources of mappings often agree on their results, and that this agreement can be meaningful. Some large-scale experiments involving several alignment tools — as the OAEI 2006 Food track campaign [5] — have indeed shown that the more often a mapping is agreed on, the more chances for it to be valid.

We have adapted the S-VAF presented above to consider the level of consensus between the sources of the mappings, by introducing the notions of support and voting into the definition of successful attacks. Support enables arguments to be counted as defenders or co-attackers during an attack: **(i.e., support, as well as attack, is an abstract notion, that depends of the framework instantiation)**:

[1] Note that as opposed to what is done [24, 25] this paper aims at experimenting with matchers that were developed prior to the experiment, and hence more likely to present strength mismatches.

**Definition 2.4.1** A *Voting Argumentation Framework* (V-VAF) is a 7-tuple *(AR, attacks, supports, V,val,P,str)* where $(AR, attacks, V, val, P, str)$ is a S-VAF, and *supports* is a (reflexive) binary relation over $AR$. *supports* and *attacks* are disjoint relations.

Voting it then used to determine whether an attack is successful or not. For this paper, we have chosen to test further the most simple voting scheme – the plurality voting system – where the number of supporters decides for success of attacks.

**Definition 2.4.2** In a *Voting Argumentation Framework (V-VAF)* an argument $a \in AR$ $defeats_{aud}$ an argument $b \in AR$ for audience *aud* if and only if $attacks(a, b) \quad \wedge \quad ( \quad |\{x|supports(x, a)\}| > |\{y|supports(y, b)\}| \quad \vee$ $(|\{x|supports(x, a)\}| = |\{y|supports(y, b)\}| \wedge valpref_{aud}(val(a), val(b))) \quad )$.

This voting mechanism is based on simple counting. In fact, as we have seen previously, matchers sometimes return mappings together with a confidence value. There are voting mechanisms which address this confidence information. The first and most elementary one would be to sum up the confidence values of supporting arguments. However, as for the S-VAF, this would rely on the assumption that the strengths assigned by different mappers are similarly scaled, which as we have seen is debatable in practice[2].

## 3 Argumentation Process

The argumentation process has two main steps: *argument generation* and *preferred extension generation*. First, the matchers work in an independent manner, applying the specific matching approach and generating the alignment set. A mapping $m$ is described as a 5-tuple $m$ = *(e,e',h,R,s)*, where $e$ corresponds to an entity in the ontology 1, *e'* corresponds to an entity in the ontology 2, $h$ is one of {-,+} depending on whether the mapping does or does not hold, $R$ is the matching relation resulting from the matching between these two terms, and $s$ is the *strength* [3] associated to the mapping. Each mapping $m$ is encapsulated into an argument. An *argument* $\in AR$ is a 2-tuple $x$ = *(m,a)*, where $m$ is a mapping; $a \in V$ is the value of the argument, depending of the matcher generating that argument (i.e, matcher 1, 2 or 3).

After generating their set of arguments, the matchers exchange with each other their arguments – the dialogue between them consists of the exchange of individual arguments. When all matchers have received the set of arguments of the each other, they generate their *attacks* set. An *attack* (or counter-argument)

[2] As a matter of fact, in [9] we have carried out experiments with a voting framework that considered these strengths – and was performing some normalization of these. But this did not bring conclusive results.

[3] The *strength* of an argument is defined by the matcher when applying the specific matching approach

will arise when we have arguments for mapping the same entities but with conflicting values of $h$. For instance, an argument $x$ = ($m_1$,M1), where $m_1$ = (e,e',+,*equivalence*,1.0), has as a counter-argument an argument $y$ = ($m_2$,M2), where $m_2$ = (e,e',-,*equivalence*,1.0). $m_1$ and $m_2$ refer to the same entities $e$ and $e'$ in the ontologies. The argument $y$ also represents an *attack* to the argument $x$.

When the set of arguments and attacks have been produced, the matchers need to define which of them must be accepted, with respect to each audience. To do this, the matchers compute their preferred extension, according to the audiences and strength of the arguments.

Note that for S-VAF and V-VAF, we choose to have the values $v \in$ V represent different matching approaches used by the agents (i.e., different matching systems). For instance, when three matchers are used, matcher 1 (M1), matcher 2 (M2), and matcher 3 (M3), then $V = \{M1,M2,M3\}$.

Each audience has an ordering preference between the *values*. For instance, the matcher 1 represents an audience where the value *M1* is preferred to the values *M2* and *M3*. The idea is not to have an individual audience with preference between the agents (i.e., matcher 1 is preferred to all other matchers), but to try accommodate different audiences and their preferences. The idea is to obtain a consensus when using different matching techniques, which are represented by different preference between values.

## 4 Experiments

The evaluation of the argumentation frameworks is carried out focusing on real ontologies portion of the Ontology Alignment Evaluation Initiative (OAEI)[4] evaluation data set. Next, the data set is described, the configuration of the matchers is presented, and the results are discussed. The argumentation models are compared with the best matchers for each test case and with the baseline, based on the union of the all individual matcher results.

### 4.1 Dataset Description

The Ontology Alignment Evaluation Initiative is a coordinated international initiative to establish a consensus for evaluation of ontology matching methods. It organizes evaluation campaigns on the basis of controlled experiments for comparing competitive techniques performances.

A systematic benchmark[5] is provided by the OAEI community. The goal of this benchmark is to identify the areas in which each algorithm is strong and weak. A first series of testes is based on one particular (reference) ontology dedicated to the domain of bibliography [6].

[4] http://oaei.ontologymatching.org/

[5] http://oaei.ontologymatching.org/2007/benchmarks/

[6] This ontology contains 33 named classes, 24 object properties, 40 data properties, 56 named individuals and 20 anonymous individuals

We however chose to focus our evaluation on a second series of tests, which is formed by a group of *real ontologies* (tests 301, 302, 303, and 304). We consider that this would represent a more realistic evaluation scenario, regarding the presentation of several competing approaches. In these tests the reference ontology is compared with four real ontologies: BibTex MIT[7] (test 301), BibTex UMBC[8] (test 302), BibTex Karlsruhe[9] (test 303), and INRIA[10] ontology (test 304).

### 4.2 Matchers Configuration

The experiments are carried out using the group of OAEI matchers, which had participated of the OAEI Benchmark Track 2007[11]: *ASMOV* ([11]), *DSSim* ([17]), *Falcon* ([8]), *Lily* ([27]), *Ola* ([7]), *OntoDNA* ([12]), *PriorPlus* ([16]), *RiMOM* ([22]), *Sambo* ([21]), *SEMA* ([19]), *TaxoMap* ([28]), and *XSOM* ([3]).

*DSSim, OntoDNA, PriorPlus, TaxoMap*, and *XSOM* are based on the use of ontology-level information, such as labels of classes and properties, and ontology hierarchy, while *ASMOV, Falcon, Lily, Ola, RiMOM, Sambo*, and *SEMA* use both ontology-level and data-level (instances) information.

When considering the techniques used in the matching process, *DSSim, PriorPlus* and *XSOM* are based on edit-distance similarity, where *DSSim* and *X-SOM* combine the string-based approaches with the synonymous relations provided by WordNet[12]. Regarding the structural approaches, several heuristics are used, such as number of common descendants and the number of similar nodes in the path between the root and the element (*PriorPlus*). A variety of strategies to combine individual matching techniques is used by the systems. The techniques can be executed in parallel (*DSSim, Falcon, Lily, Ola, PriorPlus, RiMOM, Sambo, XSOM*), or sequentially (*TaxoMap*). To combine the results of these executions, several ways are proposed: weighted formula (*PriorPlus and Sambo*), Dempster's rule of combination (*DSSim*), combination using a feed-forward neural network (*XSOM*), systems of equations (*OLA*), linear interpolation (*RiMOM*), and experimental weighted (*Lily*). *Falcon* executes sequentially a TFIDF linguistic matcher that combines concepts and instances, together a graph-based matcher. *ASMOV* iteratively combines several matchers using a single weighted sum to combine the individual results. Instance-based matchers are commonly based on Naive-Bayes classifiers (*RiMOM*), statistics (*Falcon* and *SEMA*), or probabilistic methods (*Sambo*).

### 4.3 Evaluation Measures

To evaluate mapping quality, we measure precision, recall and f–measure with respect to (manually built) reference alignments provided in the OAEI bench-

[7] http://visus.mit.edu/bibtex/0.1/
[8] http://ebiquity.umbc.edu
[9] http://www.aifb.uni-karlsruhe.de/ontology
[10] fr.inrialpes.exmo.rdf.bib.owl
[11] http://oaei.ontologymatching.org/2007/results/benchmarks/
[12] http://wordnet.princeton.edu/

marks. Such measures are derived from a contingency table (Table 1).

**Table 1.** Contingency table for binary classification.

| | manual h = + | manual h = - |
|---|---|---|
| output h = + | $m_{++}$ | $m_{+-}$ |
| output h = - | $m_{-+}$ | $m_{--}$ |

*Precision* (P) is defined by the number of correct automated mappings ($m_{++}$) divided by the number of mappings that the system had returned ($m_{++} + m_{+-}$). It measures the system's correctness or accuracy. *Recall* (R) indicates the number of correct mappings returned by the system divided by the number of manual mappings ($m_{++} + m_{-+}$). It measures how complete or comprehensive the system is in its extraction of relevant mappings. *F–measure* (F) is a weighted harmonic mean of precision and recall.

$$P = \frac{m_{++}}{(m_{++} + m_{+-})}, \qquad R = \frac{m_{++}}{(m_{++} + m_{-+})}, \qquad F = \frac{(2 * P * R)}{(P + R)}$$

For all comparative results, a significance test is applied, considering a confidence degree of 95%. The best values are indicated in bold face in the tables below. When there is reference for results *slightly* better, we mean that some true positive mappings are retrieved while some false positive mappings are discarded, however without having so significantly differences in the results.

### 4.4 Individual Matchers Results

Table 2 shows the results for each OAEI matcher[13], considering values of Precision (P), Recall (R), and F–measure (F).

Looking for each individual test in terms of F–measure, different groups of best matchers can be ranked:

- Test 301: *ASMOV, PriorPlus, Falcon, Lily*, and *Sambo*;
- Test 302: *PriorPlus, Lily, XSOM, Falcon, DSSim*, and *RiMOM*;
- Test 303: *OntoDNA, DSSim, Sambo, XSOM, ASMOV, PriorPlus* and *Falcon*;
- Test 304: *ASMOV, Falcon, DSSim, Lily, PriorPlus, RiMOM, Ola and Sambo, XSOM*, and *OntoDNA*.

Only PriorPlus and Falcon are in all rankings, but in different positions. In average, PriorPlus, Falcon, ASMOV, Lily, OntoDNA, and XSOM are in the group of the best matchers.

[13] http://oaei.ontologymatching.org/2007/results/benchmarks/ HTML/results.html

**Table 2.** Individual matcher results.

| | ASMOV | | | DSSim | | | Falcon | | | Lily | | | Ola | | | OntoDNA | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Test | P | R | F | P | R | F | P | R | F | P | R | F | P | R | F | P | R | F |
| 301 | 0.93 | **0.82** | **0.87** | 0.82 | 0.30 | 0.44 | 0.91 | **0.82** | **0.86** | 0.89 | **0.80** | **0.84** | 0.70 | 0.66 | 0.68 | 0.88 | 0.69 | 0.77 |
| 302 | 0.68 | **0.58** | 0.63 | 0.85 | **0.60** | **0.70** | 0.90 | **0.58** | **0.71** | 0.82 | **0.65** | **0.73** | 0.51 | 0.50 | 0.50 | 0.90 | 0.40 | 0.55 |
| 303 | 0.75 | **0.86** | **0.80** | **0.85** | **0.80** | **0.82** | 0.77 | 0.76 | **0.76** | 0.58 | 0.69 | 0.63 | 0.41 | **0.82** | 0.54 | **0.90** | 0.78 | **0.84** |
| 304 | **0.95** | **0.96** | **0.95** | **0.96** | **0.92** | **0.94** | **0.96** | **0.93** | **0.95** | **0.91** | **0.97** | **0.94** | 0.89 | **0.97** | **0.93** | **0.92** | 0.88 | **0.90** |
| Average | 0.83 | **0.80** | **0.81** | 0.87 | 0.65 | 0.73 | **0.89** | **0.77** | **0.82** | 0.80 | **0.78** | **0.79** | 0.63 | **0.74** | 0.66 | **0.90** | 0.69 | **0.77** |
| | PriorPlus | | | RiMOM | | | Sambo | | | SEMA | | | TaxoMap | | | XSOM | | |
| Test | P | R | F | P | R | F | P | R | F | P | R | F | P | R | F | P | R | F |
| 301 | 0.93 | **0.82** | **0.87** | 0.75 | 0.67 | 0.71 | 0.95 | 0.69 | **0.80** | 0.70 | **0.75** | 0.72 | **1.0** | 0.21 | 0.35 | 0.91 | 0.49 | 0.64 |
| 302 | 0.82 | **0.67** | **0.74** | 0.72 | **0.65** | **0.68** | 0.90 | 0.19 | 0.32 | 0.62 | **0.60** | 0.61 | **1.0** | 0.21 | 0.35 | 1.0 | **0.58** | **0.73** |
| 303 | 0.81 | **0.80** | **0.80** | 0.45 | **0.86** | 0.59 | **0.90** | 0.76 | **0.82** | 0.55 | 0.80 | 0.65 | 0.80 | 0.24 | 0.38 | **0.90** | 0.73 | **0.81** |
| 304 | 0.90 | **0.97** | **0.94** | 0.90 | **0.97** | **0.94** | **0.96** | 0.89 | **0.93** | 0.77 | 0.93 | 0.85 | **0.93** | 0.34 | 0.50 | **0.96** | 0.87 | **0.91** |
| Average | 0.87 | **0.81** | **0.84** | 0.71 | **0.79** | 0.73 | **0.93** | 0.63 | 0.72 | 0.66 | **0.77** | 0.71 | **0.93** | 0.25 | 0.39 | **0.94** | 0.67 | **0.77** |

### 4.5 Baseline and Argumentation Results

The use of argumentation models aims to obtain a consensus between the matchers, improving or balancing the individual results. This section presents the results using VAF, S-VAF, and V-VAF, considering as input the results from the previously described matchers. We compare the results of these frameworks with the baseline – which is composed by the union of all individual mappings – and with the results of the best matchers.

The argumentation results contain only the arguments objectively acceptable. It means that only the mappings strictly acceptable for all matchers are evaluated. The audiences represent the following complete preference order (pattern), which has been defined according to the individual performance of the matchers (i.e., the best matcher has higher preference, and so on): *ASMOV* audience – ASMOV > Lily > RiMOM > Falcon > Ola > PriorPlus > SEMA > DSSim > XSOM > Sambo > OntoDNA; *Lily* audience – Lily > ASMOV > RiMOM > Falcon > Ola > PriorPlus > SEMA > DSSim > XSOM > Sambo > OntoDNA; and so on.

Specially for S-VAF, two arbitrary values are used to represent the strength of the counter-argument of a positive mapping, 0.5 and 1.0. The generation of counter-arguments is a step that we implement on top of the positive mappings generated by the matchers. Different values of strength for such arguments can be specified by the user.

Basically, the OAEI matchers produce arguments for positive mappings with strength between 0.80 and 1.0. Using 0.5 as strength for a negative argument will not lead to (many) successful attacks for the positive mappings. Therefore, this results in better values of recall (the majority of the true positive mappings are selected). However, some positive arguments corresponding to wrong mappings are selected as acceptable because generated negative arguments do not lead to successful attacks, resulting in lower precision.

When using a value of 1.0, the positive arguments corresponding to false mappings from the matchers with lower strength are attacked and not selected as objectively acceptable (the false positive mapping is not acceptable for the audience of the true negative mapping). In this way, the precision is better. On the other hand, the resulting recall represents the lowest recall of the matchers. Moreover, a notable problem when using the value 1.0 when one matcher outputs no mapping is that if all others have true positive mappings with strength below 1.0, such true positive mappings are successfully attacked by the negative mappings. So, in this set of experiments, the value of 0.5 is considered to the strengths of (negative) counter-arguments in the S-VAF.

*However, the strength of arguments is an important issue that must be explored in more detail, as well as the preference order, which can have great impact in the results.*

Table 3 shows the results of baseline and argumentation, considering the three frameworks.

**Table 3.** Baseline and argumentation results.

| | | | | Argumentation | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Baseline | | | VAF | | | S-VAF | | | V-VAF | | |
| Test | P | R | F | P | R | F | P | R | F | P | R | F |
| 301 | 0.46 | **0.85** | 0.60 | **1.0** | 0.13 | 0.24 | 0.56 | **0.83** | 0.67 | 0.94 | **0.78** | **0.85** |
| 302 | 0.32 | **0.72** | 0.44 | - | 0.0 | - | 0.43 | **0.70** | 0.53 | **0.97** | 0.60 | **0.74** |
| 303 | 0.22 | **0.86** | 0.35 | **1.0** | 0.2 | 0.34 | 0.42 | **0.86** | 0.56 | 0.93 | **0.80** | **0.86** |
| 304 | 0.63 | **0.97** | 0.76 | **1.0** | 0.3 | 0.46 | 0.72 | **0.97** | 0.83 | **0.97** | **0.95** | **0.96** |
| Average | 0.41 | **0.85** | 0.54 | 0.75 | 0.16 | 0.26 | 0.53 | **0.84** | 0.65 | **0.95** | **0.78** | **0.85** |

Tables 4 and 5 show a comparison among the best argumentation model and the best matchers, taking into account the values of F–measure for each case. Best matchers vary for different ontologies – e.g., OntoDNA is the first better matcher for test 303, while it is the last one for test 304.

**Table 4.** Best argumentation vs. best matcher results.

| | Argumentation | Best matcher |
|---|---|---|
| Test | F | F |
| 301 | 0.85 (V-VAF) | 0.87 (ASMOV, PriorPlus) |
| 302 | 0.74 (V-VAF) | 0.74 (PriorPlus) |
| 303 | 0.86 (V-VAF) | 0.84 (OntoDNA) |
| 304 | 0.96 (V-VAF) | 0.95 (ASMOV, Falcon) |

**Table 5.** Best argumentation vs. best matcher results (average for each best matcher).

| | Argumentation | ASMOV | PriorPlus | OntoDNA | Falcon |
|---|---|---|---|---|---|
| Test | F | F | F | F | F |
| Average | 0.85 | 0.81 | 0.84 | 0.77 | 0.82 |

### 4.6 Discussion

As expected, baseline produces good values of recall – all true (positive) mappings are retrieved – while precision is lower – all false (positive) mappings are retrieved. By argumentation, false positive mappings can be are filtered out, improving the precision, while true positive mappings are also discarded, reducing the recall.

In average, the V-VAF performs better than VAF and S-VAF. In terms of averaged F-measure, V-VAF slightly outperforms the best matcher (Table 5), while having comparable level of quality in respect to the best matcher, for each test (Table 4). The VAF, since the preferences in the audiences are specified by the individual performance of the matchers, produces high values of precision.

The irregular performance of S-VAF confirms that one cannot fully rely on strengths output by matchers. As we had explained in [9] for motivating the introduction of consensus-based argumentation frameworks, these confidence levels are usually derived from similarity assessments made during the alignment process, and are therefore often based on objective grounds. However, there is no theory or even guidelines for determining such confidence levels. Using them to compare results from different mappers is therefore questionable, especially because of potential scale mismatches. For example, a same strength of 0.8 may not correspond to the same level of confidence for two different matcher. The approach we have taken in V-VAF, which does not rely on strengths, has been confirmed to perform better in our tests.

Analyzing the results of the individual matchers, the consensus achieved by the cooperative models is a balancing between the individual results. Consensus does not improve over every individual result, but instead delivers an intermediary performance, which is close to the one of the best matcher but represents a considerable improvement over the worst matchers.

**When comparing our results with the closer state-of-the-art argumentation proposal, namely from [13] (with a report of the results for the four cases used in our paper), better results are are obtained by the V-Voting framework.**

**Using notions of acceptability of arguments based on voting is a more promising option than using quantitative aspects as strengths, especially when a "good" number of matchers is available. Our experiments indeed confirm previous observations in the ontology matching field, according to which mappings that are found by several matchers**

**on a same case are generally more precise.[14] In fact in this paper we have put such observation in practice, by devising a matcher combination framework that can be deployed on top of existing matchers. It is important to notice that even though the implementation we have tested is dependent on a priori knowledge of matcher performance, we claim this dependence to be minimal. First, the "performance knowledge" required just consists of a simple order relation. Second, this preference order is used just when votes do not lead to a choice between contradictory arguments, which limits its application.**

**Regarding the field of argumentation, in general, in another cases where argumentation is applied, such as law reasoning, using confidence is a reasonable issue to be considered, as well the mechanism of voting, already quoted in the law field, but not at the level of argumentation.**

## 5 Related Work

In the field of ontology argumentation few approaches are being proposed. Basically, the closer proposal is from [14][13], where an argument framework is used to deal with arguments that support or oppose candidate correspondences between ontologies. The candidate mappings are obtained from an Ontology Mapping Repository (OMR) – the focus is not how the mappings are computed – and argumentation is used to accommodate different agent's preferences. Differently from Laera and colleagues, that use the VAF, our approach considers different quantitative issues on ontology matching, such as confidence level and voting on the arguments.

We find similar proposals in the field of ontology negotiation. [20] presents an ontology to serve as the basis for agent negotiation, the ontology itself is not the object being negotiated. A similar approach is proposed by [26], where agents agree on a common ontology in a decentralized way. Rather than being the goal of each agent, the ontology mapping is a common goal for every agent in the system. [1] presents an ontology negotiation model which aims to arrive at a common ontology which the agents can use in their particular interaction. We, on the other hand, are concerned with delivering mapping pairs found by a group of agents using argumentation. [18] describes an approach for ontology mapping negotiation, where the mapping is composed by a set of semantic bridges and their inter-relations, as proposed in [15]. The agents are able to achieve a consensus about the mapping through the evaluation of a confidence value that is obtained by utility functions. According to the confidence value the mapping rule is accepted, rejected or negotiated. Differently from [18], we do not use utility functions. Our model is based on cooperation and argumentation, where the agents change their arguments and by argumentation they select the preferred mapping.

[14] It is worth noting that we obtain in our experiments results that are way more conclusive than those we previously obtained with much less matchers [9].

## 6 Concluding Remarks and Future Work

This paper has presented the evaluation of three argumentation frameworks for ontology matching. Using argumentation, it is possible to use the *values* to represent preferences between the matchers. Each approach represents a *value* and each agent represents an audience, with preferences between the *values*. The *values* are used to determine the preference between the different matchers. Based on these notions, extended frameworks using confidence levels and number of supports were also considered. These extended frameworks, respectively, take into account the importance of using arguments with strength, reflecting the confidence the matcher has in the similarity between the two entities (the matching tools actually output mappings with a confidence measure), and the notion of that more often a mapping is agreed on, the more chances for it to be valid.

It is hard to improve the best matcher, especially when there is a large intersection between the individual results. In the experiments carried out, the best individual matcher varies depending on the specific characteristics of each set, while considering voting on arguments the results are similar to the best matchers for all sets.

The results obtained in this paper are more conclusive than the results of our previous paper[9]. First, much more mappers are involved, and their quality is better. We indeed had hinted for these previous experiments the results were inconclusive because there were not enough matchers performing well enough. When one takes more mappers, and the case becomes easier (the library one, as a Dutch one was hard), the proportion of mappers that really fail is lower. This results in less consensus for accepting wrong mappings. In this case voting has really helped, and performs better than baseline and allow for performance close to the best matcher.

An important issue is that we have results similar to the best matcher, but we are aware of the best matcher when obtaining them: in fact this knowledge has been used for the preference order. One possible experiment in the future is thus to check (i) whether a random order achieves good result, or (ii) if an order that is obtained for one test case can achieve good results in another case; and (iii) **explore the strength of arguments is a more detail**.

## References


1. S. C. Bailin and W. Truszkowski. Ontology negotiation between intelligent information agents. *Knowledge Engineering Review*, 17(1):7–19, 2002.
2. T. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.
3. C. Curino, G. Orsi, and L. Tanca. X-som results for oaei 2007. In *Proceedings of the 2nd Ontology Matching Workshop*, pages 276–285, Busan, Korea, 2007.
4. P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n–person games. *Artificial Intelligence*, 77(2):321–357, 1995.

5. J. Euzenat, M. Mochol, P. Shvaiko, H. Stuckenschmidt, O. Sváb, V. Svátek, W. R. van Hage, and M. Yatskevich. Results of the ontology alignment evaluation initiative 2006. In P. Shvaiko, J. Euzenat, N. F. Noy, H. Stuckenschmidt, V. R. Benjamins, and M. Uschold, editors, *Proceedings of the 1st International Workshop on Ontology Matching*, volume 225 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2006.
6. J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer-Verlag, Heidelberg (DE), 2007.
7. J. Euzenat and P. Valtchev. Similarity-based ontology alignment in OWL-Lite. In *Proceedings of the European Conference on Artificial Intelligence*, volume 16, pages 333–337, 2004.
8. W. Hu and Y. Qu. Falcon–AO: A practical ontology matching system. *Journal of Web Semantics*, 6(3):237–239, 2008.
9. A. Isaac, C. Trojahn, S. Wang, and P. Quaresma. Using quantitative aspects of alignment generation for argumentation on mappings. In *Workshop on Ontology Matching (OM'08) at 7th International Semantic Web Conference (ISWC'08) (to appear)*, 2008.
10. A. Isaac, L. van der Meij, S. Schlobach, and S. Wang. An empirical study of instance-based ontology matching. In *The Semantic Web Conference 2007*, volume 4825/2008 of *Lecture Notes in Computer Science*, Busan, Korea, 2008. Springer Berlin / Heidelberg.
11. Y. Jean-Mary and M. Kabuka. Asmov: Ontology alignment with semantic validation. In *Joint SWDB-ODBIS Workshop on Semantics, Ontologies, Databases*, pages 15–20, Vienna, Austria, 2007.
12. C.-C. Kiu and C.-S. Lee. Ontodna: Ontology alignment results for oaei 2007. In *Proceedings of the 2nd Ontology Matching Workshop*, pages 196–204, Busan, Korea, 2007.
13. L. Laera, I. Blacoe, V. Tamma, T. Payne, J. Euzenat, and T. Bench-Capon. Argumentation over ontology correspondences in mas. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1–8, New York, NY, USA, 2007. ACM.
14. L. Laera, V. Tamma, J. Euzenat, T. Bench-Capon, and T. Payne. Reaching agreement over ontology alignments. In *The Semantic Web - ISWC 2006*, volume 4273/2006 of *Lecture Notes in Computer Science*, pages 371–384. Springer Berlin / Heidelberg, 2006.
15. A. Maedche, B. Motik, N. Silva, and R. Volz. Mafra - a mapping framework for distributed ontologies. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*, pages 235–250, London, UK, 2002. Springer-Verlag.
16. M. Mao and Y. Peng. The prior+: Results for oaei campaign 2007. In *Proceedings of the 2nd Ontology Matching Workshop*, pages 219–226, Busan, Korea, 2007.
17. M. Nagy, M. Vargas-Vera, and E. Motta. DSSim – managing uncertainty on the semantic web. In *2nd International Workshop on Ontology Matching*, Busan, Korea, 2007.
18. N. Silva, P. Maio, and J. Rocha. An approach to ontology mapping negotiation. In *Proceedings of the Third International Conference on Knowledge Capture Workshop on Integrating Ontologies*, Banff, Canada, 2005.
19. V. Spiliopoulos, A. Valarakos, G. Vouros, and V. Karkaletsis. Sema: Results for the ontology alignment contest oaei 2007. In *Proceedings of the 2nd Ontology Matching Workshop*, pages 244–254, Busan, Korea, 2007.

20. V. Tamma, M. Wooldridge, I. Blacoe, and I. Dickinson. An ontology based approach to automated negotiation. In *Revised Papers from the Workshop on Agent Mediated Electronic Commerce on Agent-Mediated Electronic Commerce IV, Designing Mechanisms and Systems*, pages 219–237, London, UK, 2002. Springer-Verlag.
21. H. Tan and P. Lambrix. Sambo results for the ontology alignment evaluation initiative 2007. In *Proceedings of the Second Ontology Matching Workshop*, pages 236–243, Busan, Korea, 2007.
22. J. Tang, B. Liang, J.-Z. Li, and K. Wang. Risk minimization based ontology mapping. In *Advanced Workshop on Content Computing*, volume 3309 of *Lecture Notes in Computer Science*, pages 469–480. Springer Verlag, 2004.
23. C. Trojahn, P. Quaresma, and R. Vieira. An argumentation framework based on strength for ontology mapping. In *Argumentation in Multi-Agent Systems: 4th International Workshop on Argumentation in Multi-Agent Systems (ArgMAs2007)*, volume 4. Springer-Verlag, Berlin, Germany, 2008.
24. C. Trojahn, P. Quaresma, and R. Vieira. An extended value-based argumentation framework for ontology mapping with confidence degrees. In *Argumentation in Multi-Agent Systems: 4th International Workshop on Argumentation in Multi-Agent Systems (ArgMAs2007)*, volume 4946, pages 132–144. Springer-Verlag, Berlin, Germany, 2008.
25. C. Trojahn, P. Quaresma, R. Vieira, and M. Moraes. A cooperative approach for composite ontology mapping. *LNCS Journal on Data Semantic X (JoDS)*, 4900(0):237–263, 2008.
26. J. van Diggelen, R.-J. Beun, F. Dignum, R. M. van Eijk, and J.-J. Meyer. Anemone: An effective minimal ontology negotiation environment. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 899–906, New York, NY, USA, 2006. ACM.
27. P. Wang and B. Xu. Lily: The results for the ontology alignment contest oaei 2007. In *Proceedings of the 2nd Ontology Matching Workshop*, pages 179–185, Busan, Korea, 2007.
28. H. Zargayouna, B. Safar, and C. Reynaud. Taxomap in the oaei 2007 alignment contest. In *Proceedings of the 2nd Ontology Matching Workshop*, pages 268–275, Busan, Korea, 2007.

This article was processed using the LaTeX macro package with LLNCS style