A Dialectical Approach to Selectively Reusing Ontological Correspondences

Terry R. Payne[1] and Valentina Tamma[1]

Department of Computer Science, University of Liverpool,
Liverpool L69 3BX, United Kingdom
{T.R.Payne, V.Tamma}@liverpool.ac.uk

Abstract. Effective, contextualised communication between autonomous knowledge systems is dependent on the correct interpretation of exchanged messages, based on the entities (or vocabulary) within the messages, and their ontological definitions. However, as such systems cannot be assumed to share the same ontologies, a mechanism for autonomously determining a mutually acceptable alignment between the ontologies is required. Furthermore, the ontologies themselves may be confidential or commercially sensitive, and thus neither systems may be willing to expose their full ontologies to other parties (this may be pertinent as the transaction may only relate to part, and not all of the ontology). In this paper, we present a novel inquiry dialogue that allows such autonomous systems, or *agents* to selectively assert, counter, accept and reject those correspondences the agents previously acquired from past encounters, or from publicly available alignment systems. Thus, such knowledge is *asymmetric* and *incomplete* (i.e. not all agents may be aware of some correspondences, and their associated utility can vary greatly). By strategically selecting the order in which correspondences are disclosed, the two agents can jointly construct a bespoke alignment whilst minimising the disclosure of private knowledge. We show how partial alignments, garnered from different alignment systems, can be aggregated through our dialectical approach, and illustrate how argumentation can be used to eliminate ambiguities (i.e. when an entity in one ontology is mapped to several other entities in another ontology). We empirically evaluate the performance of the resulting alignment compared to the use of randomly selected alignment systems.

1 Introduction

Within open, distributed computing environments, effective communication is dependent on the ability of agents to reach a mutual understanding of the entities found in the exchanged messages. These entities, which are typically defined within an *ontology* may be private to the owner (an agent, institution, commercial organisation, etc), and thus not fully exposed or shared. This may be due to the knowledge encoded within the ontologies being confidential or commercially sensitive. Furthermore, disclosed ontological axioms could be exploited by other self-interested agents (and thus have intrinsic value to the owner whilst undisclosed), where agents may compete over multiple transactions. Thus, the lack of explicitly shared semantics can impede comprehension of the exchanged messages. Knowledge integration has traditionally depended on the creation of *alignments* between pairs of ontologies (consisting of sets of *mappings* between the corresponding entities). However, most systems that align ontologies rely on the respective ontologies to be fully shared [6], and no single approach can provide a panacea for all ontology pairs. Although such systems can support limited knowledge integration within closed or controlled scenarios, they cannot readily facilitate autonomous integration within open, dynamic and opportunistic environments (such as in commerce, linked open-data systems or mobile systems). However, once constructed, the alignments can be exchanged, shared and reused by other agents (given the right context), and thus provide some support within such open environments.

Two divergent approaches have emerged whereby agents align their respective ontologies. Agents can *exchange* messages that consist of conceptual definitions (including their axioms and related concepts), so that each agent can evolve its ontology to include the exchanged concepts [1, 3]. Alternatively, various negotiation and argumentation techniques have been exploited to discover mutually acceptable alignments [11, 7]; typically using a course-grained decision metric based on the *type* of correspondence, rather than whether each correspondence is *acceptable* to each agent or whether the resulting alignment is unambiguous or coherent. However, these approaches assume that alignments are all publicly available, and shared amongst each of the agents.

We present a novel inquiry dialogue that allows agents to assert, counter, accept and reject correspondences shared by different agents. It assumes that agents have acquired correspondences from past encounters, or from publicly available alignment systems, that they keep private, and that each agent associates some *utility* to each known correspondence. However, this knowledge is typically *asymmetric* and *incomplete* (i.e. not all agents may be aware of some correspondences, and their associated utility can vary greatly). Therefore, agents need to engage in an inquiry dialogue where they select which correspondences to disclose in order to ascertain the joint viability and acceptability of each correspondence. Furthermore, as different correspondences may map a single entity in one ontology to different entities in other ontologies (and vice versa), this ambiguity (in the form of one-to-many mappings) needs to be resolved for the ontologies to be effectively used. We empirically demonstrate that compared to *reference* alignments, our approach increases the precision of the resulting alignment by up to 40% whilst only slightly affecting recall.

The remainder of this paper is organised as follows: the *Correspondence Inclusion Dialogue* is presented in Section 2 where the performatives are illustrated through examples. It is then empirically evaluated with respect to the alignments produced in Section 3. Related work is presented in Section 4, before concluding in Section 5.

2 The Correspondence Inquiry Dialogue

The *Correspondence Inclusion Dialogue* enables two agents to exchange knowledge about ontological correspondences through a dialogical game that satisfies the following: 1) each agent is aware of a set of correspondences[1], each with an associated *utility*; 2) there should be no *ambiguity* with respect to either the source entities in the resulting

[1] We restrict ourselves to correspondences between atomic entities, and do not consider correspondences that map single entities in one ontology to compound formula in another (also known as complex correspondences [6]).

Table 1. The set $\mathcal{M}$ of legal moves permitted by the Correspondence Inquiry Dialogue.

Syntax	**Description**
$\langle x, join, \text{nil}, \text{nil}\rangle$	Agents assert the *join* move to participate within the dialogue.
$\langle x, assert, \phi, \text{nil}\rangle$	The agent x will *assert* the belief ϕ for a correspondence c that is believed to be viable for inclusion into the final alignment AL, and is the undisclosed belief with highest personal utility.
$\langle x, object, \phi, \phi^{att}\rangle$	An agent can *object* to some correspondence c^{att} if it knows of another correspondence c that shares one of the two entities in c^{att}, i.e. ambiguous(ϕ, ϕ^{att}), and $\kappa_c^{est} \geq \kappa_{c^{att}}^{joint}$. The agent utters the *object* move to: 1) inform the recipient of the senders personal utility of the disclosed correspondence c^{att} through the belief ϕ^{att}; and 2) propose an alternative correspondence c by asserting the belief ϕ.
$\langle x, accept, \phi, \text{nil}\rangle$	If the agent received a belief ψ for c in the previous move, and $\kappa_c^{joint} \geq \epsilon$, then the agent can confirm this by accepting the correspondence and sharing its own personal utility in ϕ, where ϕ and ψ represent the different beliefs about the same correspondence.
$\langle x, reject, \psi, \text{nil}\rangle$	If the agent received a belief ψ for c in the previous move, but was not viable (i.e. $\kappa_c^{joint} < \epsilon$), then the agent can reject this simply by returning the original belief ψ.
$\langle x, endassert, \text{nil}, \text{nil}\rangle$	If an agent has no more objections to make about the correspondences negotiated since the previous assert, it can then indicate this by uttering an *endassert* move. Once both agents have uttered this move sequentially, a new *assert* move can be uttered, or the dialogue can close.
$\langle x, close, \text{nil}, \text{nil}\rangle$	If an agent has no more correspondences that could be viable, but that have not been disclosed, then it can utter a *close* move. However, the dialogue does not terminate until both agents utter a sequence of *close* moves (known as a *matched-close*).

alignment, or the target entities; 3) if alternative choices of correspondences exist, the selection is based on the *joint utility* of both agents; 4) no correspondences should be selected where their joint utility is less than some defined *admissibility threshold*; and 5) the alignment should be generated by disclosing as few beliefs as possible.

2.1 The Inquiry Dialogue Moves

The dialogue consists of a sequence of communicative acts, or *moves*[2], whereby agents take turns to assert the viability of some correspondence c for inclusion in a mutually acceptable final alignment, AL, and respond to such assertions by: confirming the acceptability of c; rejecting the acceptability of c; or proposing alternate correspondences in the case of ambiguity. Each agent discloses its private *belief* regarding the viability of c for the alignment AL, and the agents' goal is to rationally identify an unambiguous set of correspondences deemed viable by both agents, given an *admissibility threshold* ϵ. It assumes there are always exactly two agents, $Alice$ and Bob, who participate in the dialogue, and that each agent plays a role in each dialogue move, i.e. an agent is either a *sender* x or *recipient* $\hat{x}$.

[2] The moves of the *Correspondence Inclusion Dialogue* are formally presented in [9].

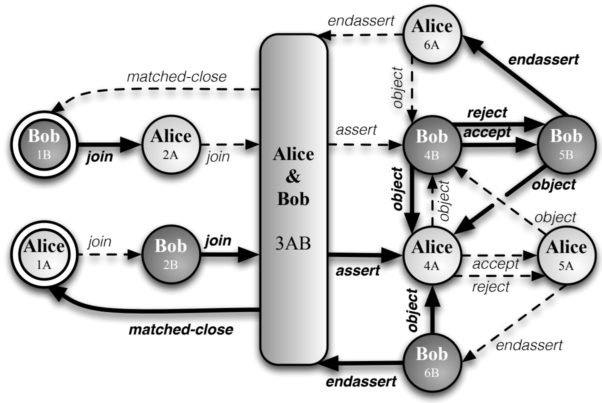


Fig. 1. The dialogue as a state diagram. Nodes indicate the agent whose turn it to utter a move. Moves uttered by Alice are labelled with a light font / dashed edge, whereas those uttered by Bob are labelled with a heavy font / solid edge.

The *moves* of the dialogue are summarised in Table 1. The syntax of each move at time s is of the form $m_s = \langle x, \tau, \phi, \phi^{att} \rangle$, where x represents the identity of the agent making the move; τ represents the move type; ϕ is a tuple that represents the belief that agent x has for a correspondence and the utility it associates to that correspondence; whereas ϕ^{att} represents a belief for some correspondence that the agent is countering. For some moves, it may not be necessary to specify one or either beliefs; in which case they will be empty or unspecified (represented as *nil*).

Agents take turns to utter *assert* moves (i.e. to transition from state 3AB in Fig. 1). A sender x can also make multiple moves in certain circumstances, such as an *accept* or *reject* move (see states labelled 4A for *Alice* and 4B for *Bob* in Fig. 1). This enables an agent to accept or reject a disclosed correspondence before making some other move (such as raising a *object* move), or signalling its intention to end a negotiation round (through an *endassert* move).

2.2 Ontologies, Correspondences and Beliefs

The agents negotiate over the viability of different correspondences that could be used to align the two agents' ontologies. The dialogue therefore assumes that each agent maintains an ontological model $\mathcal{O}$, which represents the agent's knowledge about the environment, and its background knowledge (domain knowledge, beliefs, tasks, etc.). $\mathcal{O}$ is modelled as a set of axioms describing classes and their relations.[3]

During any given encounter, the sender and the recipient use only part of their ontological model (i.e. they use their "working" ontology $\mathcal{W}$) to communicate. This could be based on an ontology module relevant to the task [2, 4], or the agent's ontological knowledge may comprise several different ontologies (as illustrated in Fig. 2). To avoid confusion, the sender's ontology is denoted $\mathcal{W}^x$, whereas the recipient's ontology is $\mathcal{W}^{\hat{x}}$, and $\widetilde{\mathcal{W}}$ denotes the *ontology signature*; i.e. the set of class and property names

[3] Here we restrict the ontology definition to classes and roles.

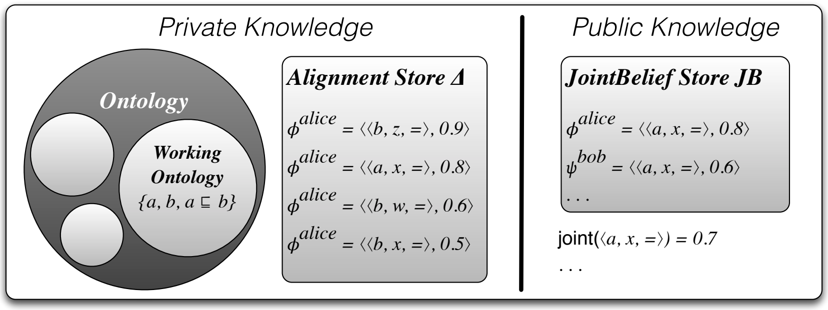


Fig. 2. The knowledge model assumed by each agent. Only alignments grounded in the Working Ontology $\mathcal{W}$ are listed in *Alice's* Alignment Store Δ. As the correspondence $\langle a, x, = \rangle$ has been *asserted* and *accepted* by the agents, they appear in *Alice's* Joint Belief Store JB. As Alice knows both her, and *Bob's* belief for this correspondence, she can calculate the joint belief using the *Avg* aggregation function.

used in $\mathcal{W}$. Both $\mathcal{W}^x$ and $\mathcal{W}^{\hat{x}}$ are fragments[4] of the sender and recipient ontologies (respectively) that denote each agent's private subset of the ontology used to model the corresponding entities used in the transaction. We also assume that agents do not disclose their "working" ontologies, and hence the participants involved in the encounter have no knowledge whether these ontologies overlap completely, partially, or in the worst case not at all (which would imply that no interaction would be possible [4]).

For agents to interoperate in an encounter, they need to determine an *alignment* AL between the two working ontology fragments $\mathcal{W}^x$ and $\mathcal{W}^{\hat{x}}$ for that encounter. An alignment [6] consists of a set of *correspondences* that establish a logical relationship[5] between the entities (classes, properties or roles, and instances) belonging to each of the two ontologies. The universe of all possible correspondences is therefore denoted $\mathcal{C}$.

Definition 1: *A* **correspondence** *is a triple denoted* $c = \langle e, e', r \rangle$ *such that* $e \in \widetilde{\mathcal{W}^x}$*,* $e' \in \widetilde{\mathcal{W}^{\hat{x}}}$*,* $r \in \{=\}$.

Agents associate a private, static *utility* κ_c to a correspondence, which represents the viability of c appearing in AL. The tuple $\langle c, \kappa_c \rangle$, where $0 \leq \kappa_c \leq 1$, is a *belief* an agent holds on a correspondence c. We refer to beliefs sent by x as ϕ, the beliefs sent by $\hat{x}$ (to x) as ψ, and the set of all beliefs is denoted $\mathcal{B}$. The function $\mathsf{corr} : \mathcal{B} \mapsto \mathcal{C}$ returns the correspondence c for some belief. The aim of the dialogue is to select an *unambiguous* (i.e. where no entity appears more than once) set of viable correspondences, $AL \subseteq \mathcal{C}$, which maps between the entities in $\mathcal{W}^x$ and $\mathcal{W}^{\hat{x}}$, and whose joint utility is at least as great as the admissibility threshold ϵ. The function $\mathsf{ent}(c)$ returns the set of entities e and e' for a correspondence c.

---

[4] We do not prescribe the logical properties exhibited by the fragment, but refer to the work on ontology modularisation, e.g. [2].

[5] We only consider *logical equivalence* (as opposed to *subsumption* ($\sqsubseteq$) and *disjointness* ($\perp$)), as it has the property that correspondences are symmetric; i.e. $\langle e, e', = \rangle$ is logically equivalent to $\langle e', e, = \rangle$, and thus can be easily used by either agent. Furthermore, the majority of alignment generation systems only computes equivalence correspondences

Each agent manages a private knowledge base, known as the *Alignment Store* (Δ), which holds the beliefs an agent has over its correspondences, and a public knowledge base, or *Joint Belief Store* (JB), which contains correspondences that have been shared (see Fig. 2). We distinguish between the sender's stores, Δ^x and JB^x, and the recipient's stores, $\Delta^{\hat{x}}$ and $JB^{\hat{x}}$, respectively. The sender's joint belief store JB^x ($JB^{\hat{x}}$ for the receiver) contains beliefs that are exchanged as part of the dialogue and hence contains beliefs sent and received by x (conversely $\hat{x}$). Throughout the dialogue, both agents will be aware of all of the beliefs shared[6]; i.e. $JB^x = JB^{\hat{x}}$.

2.3 Aggregating Beliefs and the Upper Bound

Within the dialogue, the agents try to ascertain the unambiguous, mutually acceptable correspondences to include in the final alignment AL by selectively sharing those correspondences that are believed to have the highest utility. Once each agent knows of the other agent's utility for a given correspondence c, it can then calculate c's *joint utility*, and check if it is greater than or equal to the *admissibility threshold*, ϵ. This threshold is used to filter out correspondences with a low κ_c, whilst minimising the number of beliefs disclosed. The function $\mathsf{joint} : \mathcal{C} \mapsto [0,1]$ returns the *joint utility* for some correspondence $c \in \mathcal{C}$. This results in either: 1) κ_c^{joint} calculated as an aggregation function of the utilities for both agents (if both utilities have been disclosed); or 2) κ_c^{est} for a conservative, upper estimate if only one of the two utilities is known, such that $\kappa_c^{est} \geq \kappa_c^{joint}$. The intuition behind this aggregation function here is that the mean utility is representative of the beliefs of both agents; if one agent believes strongly in a correspondence, but the other agent believes the same correspondence has little utility, then this aggregation function reflects a balanced, joint perspective by both agents on the same correspondence.

Definition 2: *The function* $\mathsf{joint} : \mathcal{C} \mapsto [0,1]$ *returns the mean* joint utility *for some* $c \in \mathcal{C}$*, where* $c = \mathsf{corr}(\phi) = \mathsf{corr}(\psi)$*:*

$$\mathsf{joint}(c) = \begin{cases} avg(\kappa_c^x, \kappa_c^{\hat{x}}) & \psi \in JB^x; \phi \in \Delta^x \\ \frac{1}{2}(\kappa_c^{\hat{x}}) & \psi \in JB^x; \phi \notin \Delta^x \wedge \phi \notin JB^x \\ avg(\kappa_c^x, \kappa_{upper}^x) & \phi \in \Delta^x; \phi, \psi \notin JB^x \end{cases}$$

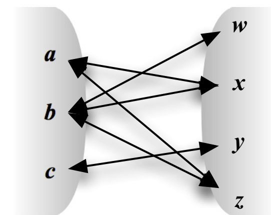
When the sender x receives a belief ψ from $\hat{x}$ ($\psi \in JB^x$) on a correspondence c, it can assess the joint utility for c as the average between its own utility and the one by $\hat{x}$, assuming that x holds a belief on c, i.e. $\phi \in \Delta^x$ (**Case 1**). If, however, x has no prior knowledge of c (i.e. $\phi \notin \Delta^x$), then $\kappa_c^x = 0$, and the joint utility depends only on $\kappa_c^{\hat{x}}$ (**Case 2**). Finally, if x holds a belief on c that has not yet been disclosed to $\hat{x}$ ($\phi \in \Delta^x; \phi \notin JB^x$) and if ψ has not been disclosed by $\hat{x}$ ($\psi \notin JB^x$), then $\kappa_c^{\hat{x}}$ can only be estimated (**Case 3**). The *upper bound*, κ_{upper}^x is explained below.

Example 1: Bob *makes the move* $\langle bob, assert, \langle\langle a, x, =\rangle, 0.6\rangle, nil\rangle$*. Alice adds the belief* $\psi^{\mathsf{bob}} = \langle\langle a, x, =\rangle, 0.6\rangle$ *into her Joint Belief store,* JB *(Fig. 2). She then responds with the move* $\langle alice, accept, \langle\langle a, x, =\rangle, 0.8\rangle, nil\rangle$*, which is also stored in* JB*. Now*

[6] We will not distinguish between the two stores JB^x and $JB^{\hat{x}}$ in the remainder of this paper.

Table 2. The individual and joint utilities for the correspondences in the examples, and how they map between ontological entities.

c	κ_c^{Alice}	κ_c^{Bob}	joint(c)
$\langle a, x, = \rangle$	0.8	0.6	**0.7**
$\langle b, x, = \rangle$	0.5	0.8	**0.65**
$\langle b, w, = \rangle$	0.6	0.4	**0.5**
$\langle b, z, = \rangle$	0.9	—	**0.45**
$\langle c, y, = \rangle$	—	0.2	**0.1**
$\langle a, z, = \rangle$	0.1	—	**0.05**



both beliefs have been disclosed, the joint belief joint$(\langle a, x, = \rangle)$ *can be calculated; i.e.* $avg(0.8, 0.6) = 0.7$.[7]

Each agent takes turns to propose a belief regarding some correspondence, and the other participant confirms if the actual joint utility $\kappa_c^{joint} \geq \epsilon$. Proposals are made by identifying an undisclosed correspondence with the highest degree of belief κ_c^x. As the dialogue proceeds, each subsequent correspondence asserted will have an equivalent or lower degree of belief than that previously asserted by the same agent.

Example 2: *It is* Bob's *turn to assert a new correspondence. If he were starting a new dialogue, he would assert the correspondence* $\langle\langle b, x, = \rangle, 0.8\rangle$*, as this would be the one with the highest utility* κ_c^{Bob}*. But in a separate dialogue, the correspondences* $\langle\langle b, x, = \rangle, 0.8\rangle$ *and* $\langle\langle a, x, = \rangle, 0.6\rangle$ *have already been disclosed. So the next one to assert is* $\langle\langle b, w, = \rangle, 0.4\rangle$*, as this is the undisclosed correspondence with the highest utility. However, he first has to determine if the estimated joint utility is* $\geq \epsilon$*.*

Whenever a correspondence is asserted, or included in an objection, the agent should check that its estimated joint utility is not less than the *admissibility threshold*, ϵ. Because the estimate is an upper estimate, the actual joint utility could be lower, and the correspondence still rejected. Agents determine this upper estimate by exploiting the fact that assertions are always made on the undisclosed correspondence with the highest utility. Thus, if one agent asserts some correspondence, the other agent's utility for that asserted correspondence will never be greater than their own previous assertion. Therefore, each agent maintains an *upper bound*, κ_{upper}^x, corresponding to the other agents assertions (prior to the dialogue, $\kappa_{upper}^x = 1.0$).

Example 3: *In example 2,* Bob *wanted to assert* $\langle\langle b, w, = \rangle, 0.4\rangle$*, but needed to determine if the estimated joint utility* $\geq \epsilon$*.* Alice's *previous assertion was for the correspondence* $\langle\langle a, x, = \rangle, 0.8\rangle$*, and therefore* $\kappa_{upper}^x = 0.8$*. If we assume that* $\epsilon = 0.5$*,* Bob *can determine that the estimated* $joint(\langle b, w, = \rangle) = avg(0.4, 0.8) = 0.6 \geq \epsilon$*, and thus makes the assertion.*

2.4 Ambiguity and Argumentation

Ambiguities occur when more than one correspondence maps several entities in the source ontology to a single entity in the target ontology (or vice versa). This can result

[7] Note that these examples illustrate properties of the dialogue. However, the actual moves in the dialogue would depend on the agent starting the dialogue. See [9] for a fully documented example walkthrough.

in logical incoherence (i.e. generate unsatisfiable concepts). Objections can be made to an ambiguous belief, once it has been asserted. An ambiguity can be determined if there is some entity that exists in the correspondences of two beliefs.

Definition 3: **Ambiguity** *occurs given beliefs* ϕ, ϕ', $\phi \neq \phi'$ *(denoted* $\mathsf{ambiguous}(\phi, \phi')$*) iff* $\mathsf{ent}(\mathsf{corr}(\phi)) \cap \mathsf{ent}(\mathsf{corr}(\phi')) \neq \varnothing$.

A belief ϕ *attacks* another belief ϕ' if they result in an *ambiguity*, and the joint utility for ϕ is greater than or equal to that of ϕ', and above the admissibility threshold, ϵ.

Definition 4: *Given two beliefs* $\phi, \phi', \phi \neq \phi'$, $\mathsf{attacks}(\phi, \phi')$ *is true iff* $\mathsf{ambiguous}(\phi, \phi') \wedge \mathsf{joint}(\mathsf{corr}(\phi)) \geq \mathsf{joint}(\mathsf{corr}(\phi')) \geq \epsilon$.

Attacks are represented in the dialogue as *object* moves. There are three scenarios where an agent responds with *object*:

1. when a new correspondence has appeared in a previous *assert* or *object* move. In this case, the sender needs to respond with its own belief of the correspondence, but may also want to raise its own objection.
2. when there is an undisclosed correspondence that could be used to attack a previously disclosed correspondence. This is where agents can identify other attacks on ambiguous correspondences.
3. when a disclosed correspondence could attack another disclosed correspondence, but the attack does not appear in the attack graph Ag. This ensures that all possible attacks have been identified within Ag.

In each case, an additional *ambiguous* correspondence is added to the attack graph Ag.

Example 4: Alice *asserts the correspondence* $\langle\langle b, w, =\rangle, 0.6\rangle$. Bob *realises that he has an alternate correspondence,* $\langle\langle b, x, =\rangle, 0.8\rangle$*, which shares the entity "b". He estimates the joint utility for this alternate correspondence as* $joint(\langle b, x, =\rangle) = avg(0.8, \kappa^x_{upper}) = 0.7$ *(here,* $\kappa^x_{upper} = 0.6$*), which is* $\geq joint(\langle b, w, =\rangle) = 0.5$*, i.e. the joint utility for* Alice's *asserted correspondence. Therefore* Bob *makes the move*[8] $\langle bob, object, \langle\langle b, x, =\rangle, 0.8\rangle, \langle\langle b, w, =\rangle, 0.6\rangle\rangle$.

Thus, the dialogue generates a set of possibly ambiguous joint-beliefs, $b \in Ag$ *such that* $b = \langle c, \kappa^{joint}_c\rangle$, that should be included in the final alignment AL, where each b represents the belief regarding the inclusion of some correspondence c, and $\mathsf{joint}(c)$ represents the joint degree of belief for that correspondence.

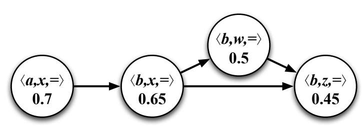


Fig. 3. The Final Attack Graph.

Once the dialogue has closed (i.e. after a *matched-close*), the agents resolve the attack graph to determine which of the correspondences disclosed should be included in the final alignment. A heuristic approach, based on Dung's argumentation framework [5] is used for resolving the graph. Recall that, as $JB^x = JB^{\hat{x}}$, both agents will form identical attack graphs. Each graph (see the example in Figure 3) includes only those correspondences which appeared in an *assert* or *object* move, and then were subsequently accepted, or themselves objected to (i.e. attacked). The objections are represented as edges between the vertices, such that vertices with a higher joint degree of belief will

[8] Note that *Bob* responds by providing his belief over *Alice's* asserted correspondence, and also provides his belief for the attacking correspondence.

attack vertices with lower values. The attack graph is traversed, starting with the highest value vertex; if the highest value vertex *attacks* another vertex, that other vertex is then removed from the graph. This continues until all of the remaining vertices have been traversed. For example, in Figure 3, $\langle a, x, = \rangle$ attacks $\langle b, x, = \rangle$, which can no longer attack either $\langle b, w, = \rangle$ or $\langle b, z, = \rangle$. Furthermore, $\langle b, w, = \rangle$ attacks $\langle b, z, = \rangle$, which is then removed from the graph. The correspondences found in the remaining vertices (once all attacks have been resolved) are included within the final alignment AL.

3 Empirical Evaluation

We empirically evaluate the effectiveness of the proposed dialogue by investigating the alignment solutions found. We also explore how the use of the admissibility threshold ϵ can affect the resulting alignments, by eliminating possibly spurious or erroneous correspondences (i.e. those with little evidence to support their validity). The following hypothesis has been tested using OAEI [9] data sets:

> *Selecting and combining correspondences taken from a number of different alignment methods can yield comparable performance to existing alignment methods, when measured using the* precision*,* recall *and* f-measure *metrics.*

The OEAI 2012 Conference Track comprises various ontologies describing the same conceptual domain (conference organisation) and alignments between pairs of ontologies, generated by 17 different ontology alignment approaches. Seven ontologies were selected as these were accompanied by *reference alignments* (defined by a domain expert), resulting in 17 different alignments for each pair of ontologies, and $\frac{7!}{(7-2)!2!} = 21$ ontology combination pairs.

The empirical evaluations were conducted over each of the 21 ontology pairs (i.e. for each experiment, an agent would be assigned an ontology, but would have no knowledge of the ontology of the other agent), with a random selection of 16 alignments divided between the two agents (such that each agent knows of 8 alignments). The allocation was random, and ensured that no alignment belonged to both agents. We exploited the fact that, as the alignments were generated independently, some correspondences could be found in more than one alignment. Whilst this is no guarantee that such correspondences are necessarily correct, it provides a mechanism whereby agents can assign some κ_c for each correspondence c based on the *probability* of finding that correspondence in the alignments each agent possessed. Experiments were also repeated for different admissibility thresholds; as 16 alignments were divided between the two agents, the threshold was varied in sixteenths; e.g. $\epsilon = \frac{2}{16}$ required there to be at least two instances of a correspondence c to be found (i.e. $\mathsf{joint}(c) \geq \frac{2}{16}$) before c was considered. Each experiment was repeated 500 times. The resulting alignments were evaluated using the *precision*, *recall* and *f-measure* metrics, where: **precision** (p) is the proportion of correspondences found by the dialogue that also appear in the *reference alignment*; **recall** (r) is the proportion of correspondences in the *reference alignment* that were also found by the dialogue; and the **f-measure** (f) represents the harmonic mean of the precision p and recall r.

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

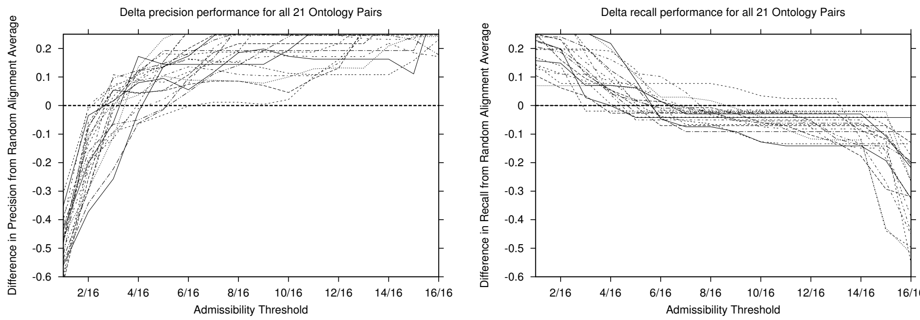


Fig. 4. Delta precision ($\Delta p_{\mathcal{D}}$) and recall ($\Delta r_{\mathcal{D}}$) for all 21 ontology pairs.

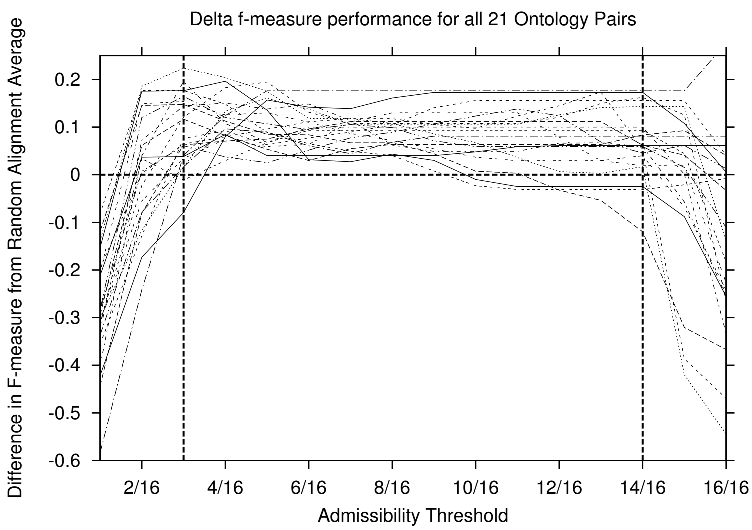


Fig. 5. Delta f-measures ($\Delta f_{\mathcal{D}}$) for all 21 ontology pairs.

A baseline was generated by assuming that a naive approach for finding an alignment would consist of an agent randomly picking and using one of the pre-defined alignments. To evaluate the alignments generated by the dialogue, we first calculate the precision ($\overline{p_{\mathcal{A}}}$), recall ($\overline{r_{\mathcal{A}}}$) and f-measure ($\overline{f_{\mathcal{A}}}$) of this average alignment for each ontology pair. These are then used to generate the difference between these measures and the corresponding ones for the alignments generated by the dialogue: i.e. $\Delta p_{\mathcal{D}} = \overline{p_{\mathcal{D}}} - \overline{p_{\mathcal{A}}}$, $\Delta r_{\mathcal{D}} = \overline{r_{\mathcal{D}}} - \overline{r_{\mathcal{A}}}$ and $\Delta f_{\mathcal{D}} = \overline{f_{\mathcal{D}}} - \overline{f_{\mathcal{A}}}$. Thus, values above zero indicate better results compared to the baseline, whereas those below are worse. These values have been plotted against different admissibility thresholds, ranging from $\epsilon = \frac{1}{16}$ (i.e where no filtering occurs - the existence of one correspondence is sufficient for consideration by the dialogue) to $\epsilon = \frac{16}{16}$ (i.e. only those correspondences that were found in every alignment known by both agents were considered). The graphs for differences in precision and recall are given in Figure 4, whereas Figure 5 plots the differences in F-measures for all 21 ontology pairs. Two vertical limits are also given in Figure 5: $\epsilon_l = \frac{3}{16}$ and $\epsilon_u = \frac{14}{16}$

represent the lower and upper ranges where the average f-measure for the dialogue's alignment is significantly higher (using a one-sided paired t-test where $\alpha = 0.05$) than $\overline{f_{\mathcal{A}}}$. Figure 5 demonstrates how, in most cases, the f-measure performance of the dialogue is significantly higher than that achieved from selecting an alignment at random, when $\epsilon_l \leq \epsilon < \epsilon_u$. Thus, the dialogue produces alignments that are more precise than selecting an original alignment at random in this range, although r is worse when $\epsilon < \frac{5}{16}$. At $\epsilon = \frac{5}{16}$, p ranges from -5.5% to a 40.3% increase for different ontology pairs, whereas r ranges from -12% to 19% increase. The maximum p occurs at $\epsilon = \frac{6}{16}$ for *cmt-ekaw* (47.37%), whereas r falls in general to between 0% and -12.19%.

The dialogue performance degrades for low and high thresholds. When $\epsilon = \frac{1}{16}$ (i.e. no filtering), a large number of correspondences appear in the alignment, yielding a high r but very low p, suggesting that although the correct correspondences were found, a high number of incorrect ones were also included. Whilst this could be a property of the dataset used (several alignments include a number of rare, but erroneous correspondences), it demonstrates the value of eliminating low utility correspondences (e.g. in this case, those that were found in the source alignments with low frequency). When $\epsilon > \frac{14}{16}$, a high number of correspondences are eliminated (an average of 2.8% of each agent's correspondences appear in the final alignment when $\epsilon = \frac{16}{16}$), resulting in a sharp drop in $\Delta f_{\mathcal{D}}$, which supports our hypothesis.

4 Related Work

A number of different approaches have addressed the reconciliation of heterogeneous ontologies by using some form of rational reasoning. In [1] an ontology negotiation protocol was discussed to enable agents to exchange parts of their ontology by a process of successive interpretations, whereas [3] presented an approach whereby agents shared an explicit goal to collaboratively evolve a common ontology. Whilst these approaches resolve semantic interoperability through negation to achieve semantic *homogeneity*, other approaches attempt to align the heterogeneous ontologies through negotiation; typically through the use of *argumentation* [7, 11]. Argumentation has been used as a rational means for agents to select ontology correspondences based on the notion of partial-order preferences over their different properties (e.g. structural vs terminological) [7]. A variant was also proposed [11] which represented ontology mappings as disjunctive queries in Description Logics. Typically, these approaches have used a course-grained decision metric based on the *type* of correspondence, rather than whether or not each correspondence was *acceptable* to each agent (given other mutually accepted correspondences), and do not consider the notion of private, or asymmetric knowledge (the correspondences are assumed to be publicly accessible). A formal dialogue that allows agents to debate in order to reach an agreement over conflicting concept representations and assertions and the dialogues for overcoming semantic heterogeneity is proposed in [8]. This dialogue detects conflicting representations and tries to resolve reasoning with conflicting information in the commitment store, rather than trying to reach an agreement over the correspondences, hence our model handles cases where entities with different representations are explicitly declared equivalent. A Max-Sum algorithm was used in [10] for synthesising ontology alignment methods whilst maximising social welfare in a group of interacting agents. Although similar to the aims of our study, [10] assumes that all agents have knowledge of the ontologies to align, and each agent is associated with an alignment method with its own preferences on the assessed relation, and quantified by a degree of confidence.

5 Conclusions

We present and empirically evaluate an inquiry dialogue that facilitates negotiation over asymmetric and incomplete knowledge of ontological correspondences. Two agents can selectively disclose private beliefs regarding the utility of previously acquired and reused ontological correspondences. Ambiguities are resolved through the use of argumentation, that identifies solutions based on the combined beliefs. The dialogue was implemented and empirically evaluated using correspondences found in alignments sourced from 17 different approaches over 21 ontology pairs, and using a set of reference alignments. The results supported the hypothesis that, by filtering low probability correspondences, alignments generated by our dialogue performed as well as selecting an existing alignment approach at random, when compared to reference alignments.

References

1. Bailin, S., Truszkowski, W.: Ontology negotiation: How agents can really get to know each other. In: Innovative Concepts for Agent-Based Systems, LNCS, vol. 2564, pp. 320–334. Springer (2003)
2. Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Modular reuse of ontologies: Theory and practice. J. of Artificial Intelligence Research (JAIR) 31, 273–318 (2008)
3. van Diggelen, J., Beun, R.J., Dignum, F., van Eijk, R.M., Meyer, J.J.C.: Anemone: an effective minimal ontology negotiation environment. In: Proceedings of AAMAS 2006. pp. 899–906 (2006)
4. Doran, P., Tamma, V., Payne, T.R., Palmisano, I.: Dynamic selection of ontological alignments: a space reduction mechanism. In: Proceedings of IJCAI 2009. pp. 2028–2033 (2009)
5. Dung, P.: On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-person Games. Artificial Intelligence 77(2), 321–358 (1995)
6. Euzenat, J., Shvaiko, P.: Ontology Matching. Springer-Verlag (2007)
7. Laera, L., Blacoe, I., Tamma, V., Payne, T., Euzenat, J., Bench-Capon, T.: Argumentation over ontology correspondences in MAS. In: Proceedings of AAMAS 2007. pp. 1285–1292 (2007)
8. Morge, M., Routier, J.C.: Debating over heterogeneous descriptions. Applied Ontology 2(3), 333–349 (2007)
9. Payne, T.R., Tamma, V.: Negotiating over ontological correspondences with asymmetric and incomplete knowledge. In: Proceedings of the 13th International Conference on Autonomous Agents and Multi-Agent Systems. pp. 517–524 (2014)
10. Spiliopoulos, V., Vouros, G.A.: Synthesizing ontology alignment methods using the max-sum algorithm. IEEE TKDE 24(5), 940–951 (2012)
11. Trojahn dos Santos, C., Quaresma, P., Vieira, R.: Conjunctive queries for ontology based agent communication in MAS. In: Proceedings of AAMAS 2008. pp. 829–836 (2008)