



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 17170

The contribution was presented at ISWC 2016 :
<http://iswc2016.semanticweb.org/>

To cite this version : Megdiche-Bousarsar, Imen and Teste, Olivier and Trojahn, Cassia *An Extensible Linear Approach For Holistic Ontology Matching*. (2016)
In: 15th International Semantic Web Conference (ISWC 2016), 17 October 2016 - 21 October 2016 (Kobe, Japan).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

An Extensible Linear Approach For Holistic Ontology Matching

Imen Megdiche, Olivier Teste, and Cassia Trojahn

Institut de Recherche en Informatique de Toulouse (UMR 5505),
Toulouse, France
{Imen.Megdiche, Olivier.Teste, Cassia.Trojahn}@irit.fr

Abstract. Resolving the semantic heterogeneity in the semantic web requires finding correspondences between ontologies describing resources. In particular, with the explosive growth of data sets in the Linked Open Data, linking multiple vocabularies and ontologies simultaneously, known as holistic matching problem, becomes necessary. Currently, most state-of-the-art matching approaches are limited to pairwise matching. In this paper, we propose a holistic ontology matching approach that is modeled through a linear program extending the maximum-weighted graph matching problem with linear constraints (cardinality, structural, and coherence constraints). Our approach guarantees the optimal solution with mostly coherent alignments. To evaluate our proposal, we discuss the results of experiments performed on the Conference track of the OAEI 2015, under both holistic and pairwise matching settings.

Keywords: Ontology Matching, Holistic Matching, Linear Approach

1 Introduction

Ontology matching is an essential task in the management of the semantic heterogeneity problem in several scientific disciplines and applied fields, notably to support data exchange, schema/ontology evolution, data integration, and data linkage. The typically high degree of semantic heterogeneity reflected in different ontologies makes this task an inherently complex task [21]. Several approaches for automatic or semi-automatic ontology matching have emerged [6] in the literature, which exploit in many different ways the knowledge encoded within each ontology when identifying correspondences between their features or structures.

Despite the different proposals in the field, most ontology matching approaches have been designed to deal with pairs of ontologies, a task so-called *pairwise matching*. However, with the continuously increasing amount of data sources being produced by the Linked Open Data community, designing solutions to deal with the simultaneously matching of different schemas and ontologies is becoming necessary [19, 27]. This task is called *holistic ontology matching* [21]. The holistic ontology matching problem is one of the key challenges proposed in [19] in its future research agenda. The proposal of the paper falls within the scope of holistic approaches.

Broadly speaking, the matching process takes as input a set of ontologies, denoted by Ω , and determines as output a set of correspondences, called alignment. The *pairwise*

ontology matching process takes as input two ontologies, $\Omega = \{O_1, O_2\}$, and determines as output a set of correspondences denoted as $A = \{c_1, c_2, \dots, c_x\}$. A correspondence c_i can be defined as $\langle e_1, e_2, r, n \rangle$, such that: e_1 and e_2 are ontology entities (e.g. properties, classes, instances) of O_1 and O_2 , respectively; r is a relation holding between e_1 and e_2 (usually, $\equiv, \supseteq, \perp, \sqcap$); and n is a confidence measure in the $[0,1]$ range assigning a degree of trust on the correspondence. The correspondence $\langle e_1, e_2, r, n \rangle$ asserts that the relation r holds between the ontology entities e_1 and e_2 with confidence n . The higher the confidence value, the higher the likelihood that the relation holds. Within an individual mapping entity, one or more O_1 entities can match with one or more O_2 entities. Alignments have different cardinalities; we distinguish 1:1 (one-to-one), 1:m (one-to-many), n:1 (many-to-one) or n:m (many-to-many). An alignment may be a simple alignment 1:1, or a multiple alignment 1:n or n:1, and n:m.

The *holistic ontology matching* process extends the ontology pairwise matching using a set $\Omega = \{O_1, \dots, O_N\}$ of ontologies with $N \geq 2$. For instance, if $\Omega = \{O_1, O_2, O_3\}$, then the alignment is defined as $A = A_{12} \cup A_{13} \cup A_{23}$ where

- $A_{12} = \{ \langle e_1, e_2, r_{12}, c_{12} \rangle \mid e_1 \in O_1 \wedge e_2 \in O_2 \}$,
- $A_{13} = \{ \langle e_1, e_3, r_{13}, c_{13} \rangle \mid e_1 \in O_1 \wedge e_3 \in O_3 \}$,
- $A_{23} = \{ \langle e_2, e_3, r_{23}, c_{23} \rangle \mid e_2 \in O_2 \wedge e_3 \in O_3 \}$.

Triple correspondences between entities of O_1 , O_2 , and O_3 can be deduced from A by detecting *cliques*; e.g., each subset of adjacent correspondences $\langle e_1, e_2, r, c_{12} \rangle$, $\langle e_1, e_3, r, c_{13} \rangle$ and $\langle e_2, e_3, r, c_{23} \rangle$.

The main limitation of the pairwise approaches regard to the holistic approaches is that in the former, A is considered as a local solution depending of the order with which the ontology matching is carried out; e.g. $A_{12} \cup A_{(12)3} \neq A_{13} \cup A_{(13)2} \neq A_{23} \cup A_{(23)1}$. Thus the set of correspondences in A differs according to the order users apply the ontology matching pairwise approach. Our holistic approach resolves the problem globally thus the solution is unique and considered as a global solution.

In this paper, we tackle the challenges of providing an extensible holistic ontology matching solution at schema-level. We provide an holistic approach which is able to link multiple ontologies simultaneously from Ω with $N \geq 2$. The approach guarantees to find always the same A global optimal solution. Our solution is extensible to operate with simple and multiple correspondences. To identify the best correspondences, a normalized degree of similarity between 0 and 1 is calculated using various similarity metrics. We develop a linear program based on an extension of the maximum-weighted graph matching problem [23], which is solved in polynomial time [15]. Our linear program encompasses different constraints related to the ontology matching problem. The constraints are used to guarantee the structural coherence between matched ontologies.

The main contributions of this paper are as follows :

- We provide an efficient approach to determine holistic correspondences between multiple ontologies. We model the approach within a linear program by reducing the ontology matching problem to the maximum-weighted graph matching problem, which is solvable in polynomial time.
- The approach is extensible with different structural similarity strategies and several linear constraints, which ensure mostly coherent alignments. We provide four constraints allowing the matching of classes and properties between ontologies.

- This approach extends a contribution [1] in the field of schema matching, especially designed to hierarchical schema structures like XML. The flexibility of the employed technique has allowed us to adapt the previous model with new constraints in order to take into account the specificities of the ontology matching problem.

The rest of the paper is organised as follows. Section 2 discusses related work. Section 3 presents our extensible linear approach for matching multiple ontologies. Section 4 discusses the experiments conducted on the *Conference* track of the Ontology Alignment Evaluation Initiative Campaign (OAEI) 2015, under both pairwise and holistic settings. Finally, Section 5 concludes the paper and discusses future directions.

2 Related work

This paper concerns the problem of holistic ontology matching, which is modelled through the maximum-weighted graph matching problem with constraints and techniques from the combinatorial optimisation field.

Graph-based approaches. In [28], an association graph is built from two input ontologies, where nodes represent candidate correspondences and edges as affinities between them. The selection of correspondences is formulated as a node ranking in the association graph using a Markov random walk process [3]. An iterative matcher (GMO) using bipartite graphs to represent ontologies is proposed in [11]. It computes structural similarities between entities by recursively propagating their similarities in the graphs. A similar representation is adopted by OLA [7], where the selection of alignments is reduced to a weighted bipartite graph matching problem. This approach models structural similarity computation as a set of equations of the different properties of ontologies.

Combinatorial optimisation strategies. S-Match [8] reduces the semantic matching to the propositional validity problem, which is theoretically a co-NP hard problem. The elements of schemes are translated into logical formulas and the matching consists of resolving propositional formula constructed between entities. Similarity Flooding (SF) [18] reduces the selection of correspondences to the stable marriage problem, which returns a local optimal solution. SF proposes a graph-based structural-matcher which propagates similarities between neighbourhood nodes until a fixed point computation. CODI [12] implements the probabilistic markov logical framework, transforming the matching problem into a maximum-a-posteriori (MAP) optimization problem which is equivalent to Max-Sat problem (NP-hard). Recently, [20] proposes a multi-cultural taxonomies matching that is modelled as a combinatorial optimisation problem using integer linear programming and quadratic programming. Mamba [17] is another system applying a combinatorial optimization approach with constraints and Markov Logic.

Holistic approaches. While state-of-the-art matching proposals mainly focus on pairwise matching, most works on holistic matching give special attention to pairwise-attribute matching. In [9], a probabilistic framework for hidden model discovery is used for determining an underlying unified model capturing the correspondences between attributes in different schemes. Given the input schemas as observations, it reconstructs the hidden generative distribution by selecting consistent models with highest probability. For dealing with complex attribute correspondences, [10] exploit co-occurrence

information across schemes and a correlation mining approach. It is based on the observations that frequent attribute co-presence indicates a synonym relationship and rare ones indicates a grouping relationship. This approach has been extended in [24] improving accuracy and efficiency, by reducing the number of synonymous candidates (assuming that two attributes co-present in the same schema cannot be synonymous candidates). [22] present an approach for incrementally merging 2-way schemes and generating an integrated one by clustering the nodes based on linguistic similarity and a tree mining technique. Under a different perspective, [27] proposes a holistic matching approach for aligning large ontologies from different domains, by grouping concepts in topics that are aligned locally. The topic identification is based on TF-IDF applied on Wikipedia pages related to concepts, resulting in a category trees (forests), while the similarity of topics is based on Jaccard, resulting in a graph containing topically related forest nodes. The correspondences between forests are determined using a tree overlap measure, before applying logical reasoning for removing conflicting correspondences.

Discussion. While the alignment selection strategy in [28] is based on paths in the graph, we reduce the selection to the maximum-weighted bipartite graph matching (MWGM) problem like OLA and we adopt a different structural similarity strategy from [11]. The complexity of MWGM with linear programming is known to be polynomial [23] even with the simplex algorithm [23]. Compared to OLA we do not compute structural similarities but encode structural properties as linear constraints. As CODI, we perform both structural matching (without additional structural similarity computation) and alignment extraction phases. Compared to CODI, we consider disjointness for all types of entities. Unlike CODI whose pairwise approach is reduced to a NP-Hard problem, our solution extends a polynomial problem in both pairwise and holistic versions. While MAMBA can be reduced to an NP-Hard problem, our approach is reduced to a polynomial problem. In a holistic and monolingual setting, we apply a combinatorial optimisation problem using linear programming, as done in [20] in pairwise. The constraints proposed by [20] for multiple correspondences can be simply added to our model to enhance the matching of multiple correspondences in the relaxed version of our model (i.e with relaxed decision variables). While most holistic approaches focus on pairwise-attribute matching [9, 10, 24, 27], our approach is not restricted to attributes. These holistic approaches handle simple attributes compared to the more structured schemes of ontologies. Differently from [27], we do not perform cross-domain holistic matching. Compared to [10], our approach can also return simple and multiple correspondences. Finally, as some pairwise matchers [13, 14], we adopt constraints that reduce the possibility of generating incoherent alignments. In that sense, an interesting direction concerns applying repair techniques in holistic ontology matching [5].

3 Extensible holistic approach

3.1 Global overview

Our approach is based on a well-known combinatorial optimisation problem, the maximum-weighted graph matching (MWGM) problem [23]. The idea consists in generalizing the pairwise matching on a set of N input ontologies through generic decision variables and generic linear constraints modelled in a linear program. The MWGM problem aims at

finding a set of disjoint edges having the maximum weights in a weighted graph G . Here, we reduce the ontology matching to the MWGM problem¹. Indeed, we consider that G expresses the potential candidate correspondences between the input ontologies and has (i) three types of nodes representing classes, object and data properties, and (ii) edges representing virtual connections between the same types of nodes (i.e classes related to classes, object properties to object properties and data properties to data properties). These edges have weights represent similarities between the nodes and can be establish using different strategies. In our approach, the similarities are calculated in a pre-processing step (Section 3.2). Given this reduction, searching simple correspondences (1:1) with a maximum weight on similarities is equivalent to find a set of disjoint edges with a maximum weight in the MWGM problem.

Our approach processes simultaneously $N \geq 2$ input ontologies. It involves a pre-processing step and a processing step. In the pre-processing step, we apply element-level matchers and then aggregate the results in order to produce similarities between the entities of the ontologies. In the processing step, we instantiate the different elements of the linear program (decision variables and linear constraints) and then resolve the model by using the CPLEX solver².

We will use the following notations in the remainder of this paper:

- $N = |\Omega|$ is the number of input ontologies;
- i, j are internal identifiers of the ontologies O_i and O_j ;
- $\{k, l\}, \{m, n\}, \{q, r\}$ refer respectively to the order of classes, object properties, data properties in the ontology (the order refers to an internal identifier of the entity in the set of ordered entities of the same type);
- C_i, OP_i, DP_i refer respectively to the set of classes, object properties and data properties in the ontology O_i ;
- $nbC_i, nbDP_i, nbOP_i$ refer respectively to the cardinality of classes ($|C_i|$), the cardinality of object properties ($|OP_i|$) and the cardinality of data properties ($|DP_i|$) in O_i ;
- c_{i_k} is the class of order k in the ontology O_i ;
- op_{i_m} is the object property of order m in the ontology O_i ;
- dp_{i_q} is the data property of order q in the ontology O_i ;

Running example. In order to illustrate our approach, we have chosen three ontologies from the OAEI Conference track [26]. These ontologies are *Cmt*, *Sigkdd* and *Conf-of*. For the sake of brevity, we present only some fragments of these ontologies as depicted in Figure 1. This example will be used in the remainder of the paper.

The objective of our model is to resolve simultaneously the set of alignments given the ontologies O_1 (*Cmt*), O_2 (*Sigkdd*) and O_3 (*Conf-of*). It will resolve in a unique run the alignments for A_{12} , A_{23} and A_{13} . As depicted in Figure 1, O_1 is composed of $nbC_1 = 8$ classes, $nbOP_1 = 2$ object properties and $nbDP_1 = 2$ data properties. $C_1 = \{c_{1_1}, c_{1_2}, \dots, c_{1_8}\}$, $OP_1 = \{op_{1_1}, op_{1_2}\}$, $DP_1 = \{dp_{1_1}, dp_{1_2}\}$.

¹ Note that we do not transform an OWL ontology into a graph but represent all entities as nodes with connections between them representing candidate correspondences.

² <http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/>

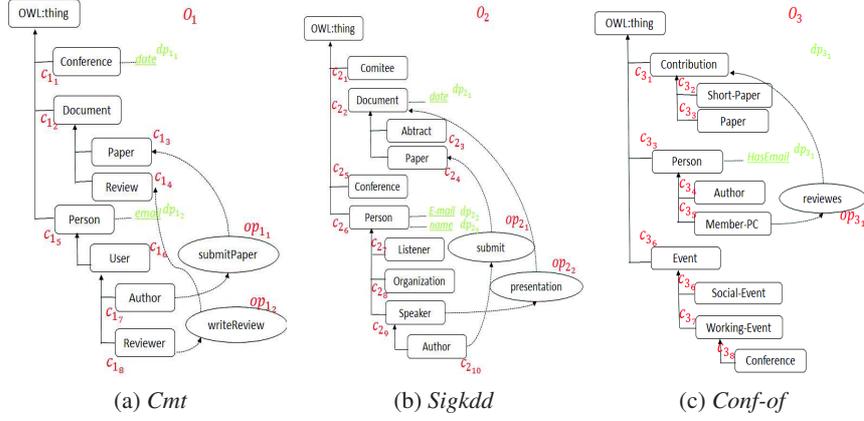


Fig. 1: Example of three ontologies from conference track in OAEI.

3.2 Pre-processing step

Our linear program takes as input a set of $N \geq 2$ ontologies $O_i = C_i \cup OP_i \cup DP_i$, $i \in [1, N]$, and a set of $N(N-1)/2$ similarity matrices representing the average results of different element-level matchers. These matrices are computed between each pair of ontologies for classes, object properties and data properties. For instance, sim_{i_k, j_l} denotes a similarity measure computed between the classes c_{i_k} and c_{j_l} , which belong respectively to ontology O_i and O_j . We have selected a restrictive set of element-level matchers according to their time performance and to their quality in the recent comparative study of [25]. The selected metrics are as follows: (1) from the *character-based category* metrics we have chosen ISUB and 3-gram to compute similarity between tokens and we have applied the generalized Mongue-Elkan method on these metrics to get the similarity between entities, (2) from the *token-based category*, we have applied Jaccard and (3) from the *language-based category* we applied Lin measure. These metrics are aggregated according to the average function in order to keep a balanced result.

3.3 Linear program

In this section, we describe the formalization of our linear program for holistic matching named LPHOM. The formalization is generalizable for $N \geq 2$ graphs.

Decision Variables. Our model is composed of three types of binary decision variables referring respectively to the three simple types of alignments in ontologies:

- The first type refers to the possible correspondences between classes. For each O_i and O_j , $\forall i \in [1, N-1]$, $j \in [i+1, N]$, x_{i_k, j_l} is a binary decision variable equals to 1 if the class c_{i_k} in the ontology O_i aligns with the class c_{j_l} in O_j and 0 otherwise.

- The second type refers to the possible correspondences between object properties. For each O_i and O_j , $\forall i \in [1, N-1]$, $j \in [i+1, N]$, y_{i_m, j_n} is a binary decision variable equals to 1 if the object property op_{i_m} in the ontology O_i aligns with the object property op_{j_n} in the ontology O_j and 0 otherwise.
- The third type refers to the possible correspondences between data properties. For each O_i and O_j , $\forall i \in [1, N-1]$, $j \in [i+1, N]$, z_{i_q, j_r} is a binary decision variable equals to 1 if the data property dp_{i_q} in the ontology O_i aligns with the data property dp_{j_r} in the ontology O_j and 0 otherwise.

Example 1. For the concept c_{1_1} in the ontology O_1 , we have the following decision variables: $x_{1_1, 2_1}, x_{1_1, 2_2}, \dots, x_{1_1, 2_9}, x_{1_1, 3_1}, \dots, x_{1_1, 3_8}$. For the object property op_{1_1} in the ontology O_1 , we have the following decision variables: $y_{1_1, 2_1}, y_{1_1, 2_2}, y_{1_1, 3_1}$. For the data property dp_{1_1} in the ontology O_1 , we have the following decision variables: $z_{1_1, 2_1}, \dots, z_{1_1, 3_1}$.

Linear Constraints. LPHOM involves four types of constraints:

- Constraints of type C1 express the matching cardinality, we apply this type of constraint on classes, object properties and data properties;
- Constraints of type C2 allow reducing the incoherences by limiting the correspondences to non-disjoint entities;
- Constraints of type C3 express restrictions in aligning object properties considering that classes represent ranges and domains of object properties;
- Constraint of type C4 express the relationships between data properties and classes by mean of involving the domain restrictions of data properties. We have not considered ranges because they are less restrictive than domains.

In the following, we detail and illustrate each constraint. For binary decision variables, we propose to use this classical *C1 constraint* in order to resolve 1:1 alignments. This constraint is equivalent to resolve a set of disjoint edges in the MWGM problem.

Definition 1 (C1 constraint). We define a C1 constraint for each type of decision variables. Each class c_{i_k} (respectively object property op_{i_m} , data property dp_{i_q}) in the ontology O_i could match with at most one class c_{j_l} (respectively object property op_{j_n} , data property dp_{j_r}) in the ontology O_j , $\forall i \times j \in [1, N-1] \times [i+1, N]$. These constraints are defined as follows:

- C1 constraints for classes are : $\sum_{l=1}^{nbC_j} x_{i_k, j_l} \leq 1, \quad \forall k \in [1, nbC_i]$
- C1 constraints for object properties are : $\sum_{n=1}^{nbOP_j} y_{i_m, j_n} \leq 1, \quad \forall m \in [1, nbOP_i]$
- C1 constraints for data properties are: $\sum_{r=1}^{nbDP_j} z_{i_q, j_r} \leq 1, \quad \forall q \in [1, nbDP_i]$

Example 2. Applying C1 for object properties in O_1 , O_2 and O_3 generates the following constraints:

$$y_{1_1, 2_1} + y_{1_1, 2_2} \leq 1; y_{1_2, 2_1} + y_{1_2, 2_2} \leq 1; y_{1_1, 2_1} + y_{1_2, 2_1} \leq 1; y_{1_1, 2_2} + y_{1_2, 2_2} \leq 1$$

$$y_{2_1, 3_1} + y_{2_2, 3_1} \leq 1; y_{1_1, 3_1} + y_{1_2, 3_1} \leq 1$$

The *C2 constraint* aims at reducing the possibility of producing incoherent alignments by considering the disjointness between entities. If we suppose that we have two

disjoint classes c_{i_k} and $c_{i_{k'}}$ in the ontology O_i ($c_{i_k} \sqsubseteq \neg c_{i_{k'}}$) so each class c_{j_l} in the ontology O_j should align either with c_{i_k} or $c_{i_{k'}}$. By this mean, we take into consideration the disjointness between classes, object and data properties.

Definition 2 (C2 Constraint). For each pair of ontologies $O_i, O_j \forall i \times j \in [1, N-1] \times [i+1, N]$ such as $i \neq j$, we define C2 constraint for each type of decision variables :

- For disjoint classes, $\forall k, k' \in [1, nbC_i] \forall l \in [1, nbC_j]$, C2 constraint is defined as follows: $x_{i_k, j_l} + x_{i_{k'}, j_l} \leq 1$
- For disjoint object properties, $\forall m, m' \in [1, nbOP_i] \forall n \in [1, nbOP_j]$, C2 constraint is defined as follows: $y_{i_m, j_n} + x_{i_{m'}, j_n} \leq 1$
- For disjoint data properties, $\forall q, q' \in [1, nbDP_i] \forall r \in [1, nbDP_j]$, C2 constraint is as follows: $z_{i_q, j_r} + x_{i_{q'}, j_r} \leq 1$

Example 3. In O_1 , *Person* is disjoint with *Document*, ($c_{1_5} \sqsubseteq \neg c_{1_2}$). A part of instantiated C2 constraints is as following:

$$x_{1_5, 2_1} + x_{1_2, 2_1} \leq 1; x_{1_5, 2_2} + x_{1_2, 2_2} \leq 1; x_{1_5, 3_1} + x_{1_2, 3_1} \leq 1; x_{1_5, 3_2} + x_{1_2, 3_2} \leq 1$$

The C3 constraint takes the advantage of the restrictions of domain and range of each object property in order to make a sense between aligned object properties and aligned classes. We have noticed that when some object properties are aligned, we have either domains aligned or ranges aligned or both of them aligned. The following constraint aims to guide alignments according to this observation. If we suppose that we have some object property op_{i_m} in the ontology O_i and some other object property op_{j_n} in the ontology O_j , such that $T \sqsubseteq \forall op_{i_m}^- . c_{i_k}$ and $T \sqsubseteq \forall op_{i_m} . c_{i_{k'}}$ and $T \sqsubseteq \forall op_{j_n}^- . c_{j_{l'}}$ and $T \sqsubseteq \forall op_{j_n} . c_{j_{l''}}$. Supposing that op_{i_m} aligns with op_{j_n} so either $c_{i_{k'}}$ aligns with $c_{j_{l'}}$ (i.e domain of op_{i_m} aligns with domain of op_{j_n}) or c_{i_k} aligns with $c_{j_{l''}}$ (i.e range of op_{i_m} aligns with range of op_{j_n}) or both of them.

Definition 3 (C3 constraint). For each pair of ontologies $O_i, O_j \forall i \times j \in [1, N-1] \times [i+1, N]$ such as $i \neq j$ and $\forall m \in [1, nbOP_i] \forall k', k'' \in [1, nbC_i]$ and $\forall n \in [1, nbOP_j] \forall l', l'' \in [1, nbC_j]$, we express C3 constraints as follows:

$$y_{i_m, j_n} \leq x_{i_{k'}, j_{l'}} + x_{i_k, j_{l''}}$$

Example 4. In O_1 and O_2 , the properties *submit* and *submitPaper* are similar. By applying the constraint C3 between these object properties we obtain: $y_{1_1, 2_1} \leq x_{1_7, 2_{10}} + x_{1_3, 2_4}$. This constraint leads to aligning both domains and ranges. We can also observe that due to the similarity between *reviews* in O_3 and *writeReview* in O_1 we obtain *Member_PC* aligned to *Reviewer* by the following constraint: $y_{3_1, 1_2} \leq x_{3_5, 1_8} + x_{3_1, 1_4}$

Finally, for the C4 constraint, we investigate the domains of the data properties. The idea consists of making classes be aligned when data properties gets aligned. If we suppose that some data property dp_{i_q} in the ontology O_i get aligned with another data property dp_{j_r} in the ontology O_j , such that $T \sqsubseteq \forall dp_{i_q}^- . c_{i_{k'}}$ and $T \sqsubseteq \forall dp_{j_r}^- . c_{j_{l'}}$ so the class $c_{i_{k'}}$ in ontology O_i will align with the class $c_{j_{l'}}$ in ontology O_j .

Definition 4 (C4 constraint). For each pair of ontologies $O_i, O_j \forall i \times j \in [1, N - 1] \times [i + 1, N]$ such as $i \neq j$ and $\forall q \in [1, nbOP_i] \forall k' \in [1, nbC_i]$ and $\forall r \in [1, nbOP_j] \forall l' \in [1, nbC_j]$, C4 constraints are defined as follows:

$$z_{i_q, j_r} \leq x_{i_{k'}, j_{l'}}$$

Example 5. We can illustrate the constraint C4 through the similar data properties *hasEmail* in O_3 and *email* in ontology O_1 : $z_{3,1,2} \leq x_{3,1,5}$ as *hasEmail* and *email* are similar, their domains, which are also similar will also be aligned.

We summarize our linear program for holistic ontology matching (LPHMO) as depicted in Figure 2. We emphasize that our model focuses on 1:1 alignments by using binary decision variables. We must however also point out that by relaxing the decision variables in the $[0,1]$ interval, this model is able to find n:m alignments. Moreover, we have to emphasize too that by using thresholds for entities similarities, we reduce significantly the size of the generated problem.

4 Experimental evaluation

In the following we present the results of our approach in both pairwise and holistic matching settings. For both settings, our approach has been evaluated for similarities higher than a fixed threshold equals to 0.65 for both classes and properties. Furthermore, all generated correspondences have a confidence degree of 1.0.

4.1 OAEI Conference data set

The evaluation of LPHOM is carried out using the OAEI *Conference track*³. The intent of this track is to provide expressive and real-world matching problems over expressive ontologies covering the same domain [2]. This data set is composed of 16 ontologies covering the domain of conference organization and a subset of 21 reference alignments involving 7 ontologies. The track evaluation is based on crisp reference alignments (RA1) and two other entailed alignments (RA2 and RAR2) deduced from RA1. Our evaluation is restricted to the RA1 alignments as they are the only publicly available set. RA1 is divided into three sub-evaluations, as follows:

- In RA1-M1 only alignments between classes are evaluated;
- In RA1-M2 only alignments between properties (object and data) are evaluated;
- In RA1-M3 both alignments between classes and properties are evaluated.

4.2 Pairwise matching evaluation

Here, we compare the results of our approach with the results of the 14 matchers participating in the 2015 OAEI campaign. These results have been obtained from the Web page describing the results of the campaign⁴. With exception of MAMBA, that applies

³ <http://oaei.ontologymatching.org/2015/conference/>

⁴ <http://oaei.ontologymatching.org/2015/conference/eval.html>

$$\begin{aligned}
& \max \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sum_{k=1}^{nbC_i} \sum_{l=1}^{nbC_j} sim_{i_k,j_l} x_{i_k,j_l} + \sum_{m=1}^{nbOP_i} \sum_{n=1}^{nbOP_j} sim_{i_m,j_n} y_{i_m,j_n} + \sum_{q=1}^{nbDP_i} \sum_{r=1}^{nbDP_j} sim_{i_q,j_r} z_{i_q,j_r} \\
& \text{s.t.} \quad \sum_{l=1}^{nbC_j} x_{i_k,j_l} \leq 1, \forall k \in [1, nbC_i] \quad (C1 \text{ Classes}) \\
& \quad \quad \quad \forall i \in [1, N-1], j \in [i+1, N] \\
& \quad \quad \quad \sum_{n=1}^{nbOP_j} y_{i_m,j_n} \leq 1, \forall m \in [1, nbOP_i] \quad (C1 \text{ Object Properties}) \\
& \quad \quad \quad \forall i \in [1, N-1], j \in [i+1, N] \\
& \quad \quad \quad \sum_{r=1}^{nbDP_j} z_{i_q,j_r} \leq 1, \forall q \in [1, nbDP_i] \quad (C1 \text{ Data Properties}) \\
& \quad \quad \quad \forall i \in [1, N-1], j \in [i+1, N] \\
& \quad \quad \quad x_{i_k,j_l} + x_{i_{k'},j_l} \leq 1 \quad (C2 \text{ Classes}) \\
& \quad \quad \quad \forall i \in [1, N-1], j \in [i+1, N] \\
& \quad \quad \quad \forall k, k' \in [1, nbC_i], \forall l \in [1, nbC_j] \\
& \quad \quad \quad y_{i_m,j_n} + x_{i_{m'},j_n} \leq 1 \quad (C2 \text{ Object Properties}) \\
& \quad \quad \quad \forall i \in [1, N-1], j \in [i+1, N] \\
& \quad \quad \quad \forall m, m' \in [1, nbOP_i], \forall n \in [1, nbOP_j] \\
& \quad \quad \quad z_{i_q,j_r} + x_{i_{q'},j_r} \leq 1 \quad (C2 \text{ Data Properties}) \\
& \quad \quad \quad \forall i \in [1, N-1], j \in [i+1, N] \\
& \quad \quad \quad \forall q, q' \in [1, nbDP_i], \forall r \in [1, nbDP_j] \\
& \quad \quad \quad y_{i_m,j_n} \leq x_{i_{k'},j_{n'}} + x_{i_{k''},j_{n''}} \quad (C3) \\
& \quad \quad \quad \forall i \in [1, N-1], j \in [i+1, N] \\
& \quad \quad \quad \forall m \in [1, nbOP_i], \forall n \in [1, nbOP_j] \\
& \quad \quad \quad \forall k', k'' \in [1, nbC_i], \forall l', l'' \in [1, nbC_j] \\
& \quad \quad \quad z_{i_q,j_r} \leq x_{i_{q'},j_{r'}} \quad (C4) \\
& \quad \quad \quad \forall i \in [1, N-1], j \in [i+1, N] \\
& \quad \quad \quad \forall q \in [1, nbDP_i], \forall r \in [1, nbDP_j] \\
& \quad \quad \quad \forall k' \in [1, nbC_i], \forall l' \in [1, nbC_j] \\
& \quad \quad \quad x_{i_k,j_l} \in \{0, 1\} \quad \forall i \in [1, N-1], j \in [i+1, N] \\
& \quad \quad \quad \forall k \in [1, nbC_i], \forall l \in [1, nbC_j] \\
& \quad \quad \quad y_{i_m,j_n} \in \{0, 1\} \quad \forall i \in [1, N-1], j \in [i+1, N] \\
& \quad \quad \quad \forall m \in [1, nbOP_i], \forall n \in [1, nbOP_j] \\
& \quad \quad \quad z_{i_q,j_r} \in \{0, 1\} \quad \forall i \in [1, N-1], j \in [i+1, N] \\
& \quad \quad \quad \forall q \in [1, nbDP_i], \forall r \in [1, nbDP_j]
\end{aligned}$$

Fig. 2: LPHOM Linear model

an optimization method with constraints and Markov Logic, these matchers apply different matching strategies than us. For example, AML is based on lexical similarities, external resources and alignment coherence; XMAP applies both lexical and structural contexts and exploits external resources; LogMap applies consistency and locality principles while their variants LogMap-C further implements the conservativity principle and LogMapLite essentially applies string matching techniques; GMAP uses a sum-product network encoding the similarities on individuals and disjointness axioms and a noisy-or model encoding probabilistic matching rules; RSDLWB exploits lexical and structural heuristics and machine learning on statistical patterns and DKP-AOM is based on linguistic, synonym and axiomatic based alignment. Although MAMBA is very performing in this track, it can not deal with more than 1000 classes [17], contrarily to LPHOM. (e.g., for the anatomy track, 2744 and 3305 classes, LPHOM spent about 36sec with a F-measure of 0,76). However, it is out of the scope of this paper to provide a deep analysis of the results obtained by each tool.

The evaluation is based on Precision (P), Recall (R), F1-measure (F1), F2-measure (F2) and F0.5-measure (F0.5) computed for the threshold that provides the highest average F1-measure computed for each matcher. F1-measure is the harmonic mean of precision and recall. F2-measure weights recall higher than precision and F0.5-measure weights precision higher than recall.

RAI-M1. For this evaluation, we have evaluated LPHOM with the constraints exclusively dedicated to classes (C1 and C2). The average results for the 21 pairs of alignments are summarized in the table of Figure 3a. We observe that our results are situated in the middle, we are better than 8 participants but lower than the other 6 participants. The best approaches benefit from more elaborate strategies and external resources to compute similarities. Even if our approach uses simple average similarities between known measures in the literature, we can observe that the strategy to find the best set of alignments checking coherence seems returning very good results on recall. These results are slightly closer, see even better than the recall of XMap and LogMap-C participants.

RAI-M2. Here, only one type of constraints exclusively dedicated to properties is applied (C1 on data properties and on object properties). As shown in the table of Figure 3b, we observe that except Mamba and AML perform well in this task, all the other approaches have difficulties in aligning properties. The results of our approach are once again in the middle. We have noticed that the chosen threshold (65%) applied on properties is not a very good compromise for this task. We have observed several properties having similarities equal to 0 (according to our measures), that we have not been able to capture. The results of baseline approaches Edna and StringEquiv confirm our observations, since these approaches use very high similarity threshold.

RAI-M3. Finally, Table 1 summarizes our results compared to the results of the other participants for the evaluation on both classes and properties. We have evaluated our model with all the constraints (C1, C2, C3, C4). Our approach keeps a stable rank compared to other approaches. We notice that GMAP or RSDLWB have non stable positions through the evaluations. Using all constraints seems advantageous for recall more than for precision because of the noise caused by the false positive aligned properties.

	P	F0.5	F1	F2	R
AML	0.83	0.8	0.76	0.72	0.7
Mamba	0.84	0.8	0.74	0.69	0.66
XMAP	0.86	0.8	0.73	0.67	0.63
GMAP	0.76	0.75	0.73	0.72	0.71
LogMap	0.82	0.78	0.73	0.68	0.65
LogMap-C	0.84	0.78	0.71	0.65	0.62
our approach	0.76	0.73	0.69	0.66	0.64
DKP-AOM	0.84	0.77	0.69	0.63	0.59
Edna	0.88	0.78	0.67	0.59	0.54
COMMAND	0.84	0.76	0.66	0.58	0.54
RSDLWB	0.88	0.78	0.66	0.58	0.53
LogMapLite	0.84	0.76	0.66	0.58	0.54
ServOMBI	0.64	0.64	0.64	0.65	0.65
StringEquiv	0.88	0.76	0.64	0.55	0.5
Lily	0.59	0.6	0.61	0.62	0.63
CroMacher	0.72	0.67	0.6	0.54	0.51
JarvisOM	0.88	0.73	0.59	0.49	0.44

(a) RA1-M1

	P	F0.5	F1	F2	R
Mamba	0.89	0.79	0.67	0.59	0.54
AML	0.89	0.78	0.58	0.46	0.41
LogMap-C	1	0.51	0.39	0.32	0.28
LogMap	0.65	0.5	0.39	0.31	0.28
CroMatcher	0.62	0.31	0.34	0.37	0.39
JarvisOM	0.3	0.31	0.34	0.37	0.39
GMAP	0.3	0.46	0.31	0.23	0.2
our approach	0.23	0.24	0.25	0.26	0.26
LogMapLite	0.29	0.27	0.25	0.23	0.22
ServOMBI	0.29	0.27	0.24	0.21	0.2
XMAP	0.67	0.37	0.22	0.15	0.13
Edna	0.24	0.19	0.15	0.12	0.11
COMMAND	0.18	0.11	0.07	0.05	0.04
RSDLWB	0.03	0.04	0.05	0.1	0.24
StringEquiv	0.08	0.05	0.03	0.02	0.02

(b) RA1-M2

Fig. 3: Average results for RA1-M1 and RA1-M2 in the Conference Track. Results are ranked according to the F1-Measure.

Semantic evaluation. As we have observed that our generated alignments seem semantically close to the crisp reference, we have evaluated our results and those of the other approaches, using the semantic measures [4] (Table 1). Indeed, our results are semantically very interesting. In particular, we observe an improvement in the recall, which is equivalent to the the recall of AML and MAMBA. We note also that the semantic evaluation reveals a slight change in the ranking of systems.

To sum up, our approach reaches promising results for its first comparison with regard to the pairwise ontology matching problem. Our model is more efficient when we use all the proposed constraints (RA1-M3). The interaction between constraints leads to semantically significant results closer to gold references which are illustrated by a good recall on semantic distances. The constraints proposed for reducing incoherence are experimentally efficient. We applied the ALCOMO [16] to evaluate if there is incoherence in our results and we get the following average results (for the 21 combinations we removed between 3 and 0 correspondences per alignment): (1) for RA1-M1 we have 0,95 removed correspondences, (2) for RA1-M2 we have 0 removed correspondences and (3) for RA1-M3 we have 0,85 removed correspondences.

Finally, the average runtime of LPHOM (pre-processing, linear program generation and resolution), over 21 pairs of the conference track was 2,84 sec using the different types of measures and 0,24 sec using only the token-based measure Jaccard.

	Precision	F0.5-Measure	F1-Measure	F2-Measure	Recall
AML	0.84	0.8	0.74	0.69	0.66
AML (semantic)	0.84	0.79	0.79	0.69	0.67
Mamba	0.83	0.78	0.72	0.67	0.64
Mamba(semantic)	0.84	0.79	0.79	0.68	0.66
XMAP	0.85	0.77	0.68	0.6	0.56
XMAP (semantic)	0.87	0.79	0.79	0.62	0.58
LogMap	0.8	0.75	0.68	0.62	0.59
LogMap (semantic)	0.82	0.77	0.77	0.65	0.62
LogMap-C	0.82	0.75	0.67	0.61	0.57
LogMap-C (semantic)	0.83	0.77	0.77	0.63	0.6
GMAP	0.66	0.66	0.65	0.65	0.65
GMAP (semantic)	0.68	0.68	0.68	0.69	0.7
DKP-AOM	0.84	0.74	0.63	0.54	0.5
DKP-AOM (semantic)	0.86	0.76	0.76	0.56	0.52
Our approach	0.65	0.63	0.61	0.59	0.58
Our approach (semantic)	0.65	0.65	0.66	0.66	0.67
ServOMBI	0.61	0.6	0.59	0.59	0.58
ServOMBI (semantic)	0.58	0.6	0.6	0.69	0.73
COMMAND	0.78	0.69	0.59	0.51	0.47
COMMAND (semantic)	0.6	0.6	0.6	0.63	0.65
LogMapLite	0.73	0.67	0.59	0.53	0.5
LogMapLite (semantic)	0.75	0.7	0.7	0.58	0.56
Edna	0.79	0.7	0.59	0.51	0.47
Lily	0.59	0.58	0.56	0.54	0.53
Lily (semantic)	0.58	0.58	0.58	0.61	0.62
StringEquiv	0.8	0.68	0.56	0.47	0.43
CroMatcher	0.59	0.57	0.54	0.52	0.5
CroMatcher (semantic)	0.61	0.59	0.59	0.54	0.53
JarvisOM	0.84	0.67	0.51	0.42	0.37
JarvisOM (semantic)	0.84	0.69	0.69	0.45	0.41
RSDLWB	0.25	0.28	0.33	0.41	0.49
RSDLWB (semantic)	0.32	0.36	0.36	0.59	0.76

Table 1: Average results for RA1-M3 in the Conference Track

4.3 Holistic matching evaluation

The ontology matching field lacks in benchmarks dedicated to the evaluation of holistic ontology matching. In order to be able to evaluate our holistic approach, we analyse:

- the differences between cliques manually deduced from reference alignments and the cliques generated by our holistic approach (remember that cliques define correspondences between N ontologies, which have to be matched);
- the differences between the results of pairwise and holistic matching settings.

In the following, we denote a clique as $Cl_i = \langle e_1, \dots, e_N \rangle$, such as each e_j belongs to ontology O_j . In the first part of this evaluation, we compare the cliques generated by

LPHOM with the cliques that we have manually identified from the reference alignments involving 3 ontologies. For the 7 available ontologies in the Conference Track, which are classified into types (Tool, Insider and Web), we selected 3 ontologies from the ‘Tool’ type (Cmt, Conf-Of, Edas). In order to maximize the chance to have cliques in the reference alignments, we have tried to find $N \geq 2$ ontologies of the same type. The only combination of ontologies verifying that was Cmt, Conf-Of and Edas, for which the reference alignments are available. Given O_1 (Cmt), O_2 (Conf-Of) and O_3 (Edas), we have manually identified the following cliques from the reference alignment:

- $Cl_1(\text{reference}) = \langle \text{author}_1, \text{author}_2, \text{author}_3 \rangle$
- $Cl_2(\text{reference}) = \langle \text{hasBeenAssigned}_1, \text{revives}_2, \text{isReviewing}_3 \rangle$
- $Cl_3(\text{reference}) = \langle \text{person}_1, \text{person}_2, \text{person}_3 \rangle$
- $Cl_4(\text{reference}) = \langle \text{hasAuthor}_1, \text{writtenBy}_2, \text{isWrittenBy}_3 \rangle$

Applying our approach, we have found the following cliques:

- $Cl_1 = \langle \text{author}_1, \text{author}_2, \text{author}_3 \rangle$
- $Cl_2 = \langle \text{paper}_1, \text{paper}_2, \text{paper}_3 \rangle$
- $Cl_3 = \langle \text{person}_1, \text{person}_2, \text{person}_3 \rangle$
- $Cl_4 = \langle \text{hasAuthor}_1, \text{writtenBy}_2, \text{isWrittenBy}_3 \rangle$
- $Cl_5 = \langle \text{writePaper}_1, \text{writes}_2, \text{hasRelatedPaper}_3 \rangle$
- $Cl_6 = \langle \text{email}_1, \text{hasEmail}_2, \text{hasEmail}_3 \rangle$

We first notice that cliques Cl_1 , Cl_3 and Cl_4 are the same as the cliques identified in the reference alignments whereas the clique $Cl_2(\text{reference})$ has not been found by our approach. However, our model has found three other significant cliques Cl_2 , Cl_5 and Cl_6 . Cl_2 is composed of the same concept *Paper* occurring in all ontologies. In the reference alignments, the correspondences in which *Paper* occur does not form a clique. We emphasize here the benefit of holistic matching which inspects simultaneously all ontologies. The Cl_5 clique is particularly interesting since that the properties of Cl_5 are the inverse of the properties of Cl_4 . Finally, Cl_6 is composed of similar data properties which is also relevant and strangely not provided in the reference alignments.

We also analyse the differences between the results of pairwise and holistic matching settings, applied on the example of the Figure 1 (O_1 is Cmt, O_2 is Sigkdd and O_3 is Conf-Of). The holistic approach discovers simultaneously alignments for N ontologies, from all combinations of pairs of input ontologies. The resulting alignments are collected from a simultaneous resolution of A_{12} , A_{13} and A_{23} . Here we focus on main differences occurring in the alignments:

- If we match O_1 and O_2 by producing A_{12} , then we match with O_3 by producing $A_{(12)3}$, we get the following alignments:
 $A_{12} = \{ \langle \text{Conference}, \text{ConferenceHall}, \equiv, 0.63 \rangle, \langle \text{ConferenceMember}, \text{Conference}, \equiv, 0.66 \rangle, \langle \text{Paper}, \text{Paper}, \equiv, 1 \rangle \}$
and $A_{(12)3} = \{ \langle \text{Conference}, \text{Conference}, \equiv, 1 \rangle, \langle \text{Paper}, \text{Paper}, \equiv, 1 \rangle \}$
- If we produce A_{13} , then we produce $A_{(13)2}$, we get the following alignments:
 $A_{13} = \{ \langle \text{Paper}, \text{ShortPaper}, \equiv, 0.63 \rangle, \langle \text{PaperFullVersion}, \text{Paper}, \equiv, 0.66 \rangle, \langle \text{Conference}, \text{Conference}, \equiv, 1 \rangle \}$
and $A_{(13)2} = \{ \langle \text{Conference}, \text{Conference}, \equiv, 1 \rangle, \langle \text{ShortPaper}, \text{AuthorOfPaper}, \equiv, 0.5 \rangle, \langle \text{Paper}, \text{Paper}, \equiv, 1 \rangle \}$

Applying the holistic matching for O_1 , O_2 and O_3 , we get the following alignments:

- $A_{12} = \{ \langle \textit{Conference}, \textit{Conference}, \equiv, 1 \rangle, \langle \textit{Paper}, \textit{Paper}, \equiv, 1 \rangle \}$
- $A_{23} = \{ \langle \textit{Conference}, \textit{Conference}, \equiv, 1 \rangle, \langle \textit{Paper}, \textit{Paper}, \equiv, 1 \rangle \}$
- $A_{13} = \{ \langle \textit{Conference}, \textit{Conference}, \equiv, 1 \rangle, \langle \textit{Paper}, \textit{Paper}, \equiv, 1 \rangle \}$

From these alignments, two cliques are deduced:

- $Cl_1 = \langle \textit{Paper}_1, \textit{Paper}_2, \textit{Paper}_3 \rangle$
- $Cl_2 = \langle \textit{Conference}_1, \textit{Conference}_2, \textit{Conference}_3 \rangle$

To sum up, the results presented in this section show the subtleties between a local and global investigations on $N \geq 2$ ontologies, which confirm the usefulness of holistic approaches for ontology matching.

5 Conclusion and future work

In this paper, we have presented an extensible linear model named LPHOM performing holistic ontology matching. The main contribution of this approach consists in allowing simultaneous matching of multiple ontologies. We model the approach within a linear program by reducing the ontology matching problem to the maximum-weighted graph matching problem, which is solvable in polynomial time. Our approach is extensible with different linear constraints handling classes and properties of ontologies. These constraints are used to reduce the logical incoherence in generated alignments, what is not done systematically by all matching systems. We experimented LPHOM on the OAEI Conference set on both pairwise and holistic settings. For future work, we intend to deeply study the similarity computation of entities with more accurate external resources. With respect to the constraints, we plan to add the constraint that classes can also match with properties and other hypothesis concerning incoherence. We also intend to extend our evaluation on the whole set of Conference and other data sets. Finally, we plan to extend the approach to deal with holistic instance matching and larger data sets.

References

1. A. Berro, I. Megdiche, and O. Teste. A linear program for holistic matching: Assessment on schema matching benchmark. In *Proceedings of the International Conference on Database and Expert Systems Applications*, pages 383–398, 2015.
2. M. Cheatham and P. Hitzler. Conference v2.0: An uncertain version of the oaei conference benchmark. In *Proceedings of the 13th International Semantic Web Conference*, pages 33–48, 2014.
3. M. Cho, J. Lee, and K. M. Lee. Reweighted random walks for graph matching. In *Proceedings of the 11th European Conference on Computer Vision*, pages 492–505, 2010.
4. J. Euzenat. Semantic precision and recall for ontology alignment evaluation. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 348–353, 2007.
5. J. Euzenat. Revision in networks of ontologies. *Artificial Intelligence*, 228:195–216, 2015.
6. J. Euzenat and P. Shvaiko. *Ontology matching*. Springer-Verlag, 2nd edition, 2013.
7. J. Euzenat and P. Valtchev. Similarity-based ontology alignment in OWL-Lite. In *Proceedings of the 16th European Conference on Artificial Intelligence*, pages 333–337, 2004.

8. F. Giunchiglia, M. Yatskevich, and P. Shvaiko. Semantic matching: Algorithms and implementation. *Journal on Data Semantics*, 1:2007, 2007.
9. B. He and K. C.-C. Chang. Statistical schema matching across web query interfaces. In *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, SIGMOD '03, pages 217–228. ACM, 2003.
10. B. He, K. C.-C. Chang, and J. Han. Discovering complex matchings across web query interfaces: A correlation mining approach. In *Proceedings of the 20th International Conference on Knowledge Discovery and Data Mining*, pages 148–157, 2004.
11. W. Hu, N. Jian, Y. Qu, and Y. Wang. Gmo: A graph matching for ontologies. In *K-Cap 2005 Workshop on Integrating Ontologies 2005*, pages 43–50, 2005.
12. J. Huber, T. Sztyley, J. Nöbner, and C. Meilicke. CODI: combinatorial optimization for data integration: results for OAEI 2011. In *Proceedings of the 6th International Workshop on Ontology Matching*, 2011.
13. Y. Jean-Mary, E. Shironoshita, and M. Kabuka. Ontology matching with semantic verification. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3), 2009.
14. E. Jiménez-Ruiz and B. C. Grau. Logmap: Logic-based and scalable ontology matching. In *Proceedings of the 10th International Semantic Web Conference*, pages 273–288, 2011.
15. N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.
16. C. Meilicke. *Alignment incoherence in ontology matching*. PhD thesis, University of Mannheim, 2011.
17. C. Meilicke. MAMBA - results for the OAEI 2015. In *Proceedings of the 10th International Workshop on Ontology Matching*, pages 181–184, 2015.
18. S. Melnik, H. Garcia-Molina, and E. Rahm. Similarity flooding: A versatile graph matching algorithm and its application to schema matching. In *Proceedings of the 18th International Conference on Data Engineering*. IEEE Computer Society, 2002.
19. L. Otero-Cerdeira, F. J. Rodriguez-Martinez, and A. Gomez-Rodriguez. Ontology matching: A literature review. *Expert Systems with Applications*, 42(2):949 – 971, 2015.
20. N. Prytkova, G. Weikum, and M. Spaniol. Aligning multi-cultural knowledge taxonomies by combinatorial optimization. In *Proceedings of the 24th International Conference on World Wide Web*, pages 93–94. ACM, 2015.
21. E. Rahm. *Schema Matching and Mapping*, chapter Towards Large-Scale Schema and Ontology Matching, pages 3–27. Springer Berlin Heidelberg, 2011.
22. K. Saleem, Z. Bellahsene, and E. Hunt. PORSCHE: Performance ORiented SCHEma mediation. *Information Systems*, 33(7-8):637–657, 2008.
23. A. Schrijver. *Combinatorial Optimization - Polyhedra and Efficiency*. Springer, 2003.
24. W. Su, J. Wang, and F. Lochovsky. Holistic schema matching for web query interfaces. In *Proceedings of the 10th International Conference on Advances in Database Technology*, pages 77–94. Springer-Verlag, 2006.
25. Y. Sun, L. Ma, and W. Shuang. A comparative evaluation of string similarity metrics for ontology alignment. *Journal of Information & Computational Science*, 12(3):957 – 964, 2015.
26. V. Svatek and P. Berka. Ontofarm: Towards an experimental collection of parallel ontologies. In *Poster Session at International Semantic Web Conference*, 2005.
27. F. N. Toni Gruetze, Christoph Böhm. Holistic and scalable ontology alignment for linked open data. In *Proceedings of the 5th Linked Data on the Web Workshop at the 21th International World Wide Web Conference*, 4 2012.
28. C. Xiang, B. Chang, and Z. Sui. An ontology matching approach based on affinity-preserving random walks. In *Proceedings of the 24th International Conference on Artificial Intelligence*, pages 1471–1477, 2015.