Ontology Alignment Using Web Linked Ontologies as Background Knowledge

Thomas Hecht, Patrice Buche, Juliette Dibie-Barthelemy, Liliana Ibanescu, Cassia Trojahn dos Santos

Ontology Alignment Using Web Linked Ontologies as Background Knowledge

Thomas Hecht and Patrice Buche and Juliette Dibie and Liliana Ibanescu and Cassia Trojahn dos Santos

Abstract This paper proposes an ontology matching method for aligning a source ontology with target ontologies already published and linked on the Linked Open Data (LOD) cloud. This method relies on the refinement of a set of input alignments generated by existing ontology matching methods. Since the ontologies to be aligned can be expressed in several representation languages with different levels of expressiveness and the existing ontology matching methods can only be applied to some representation languages, the first step of our method consists in applying existing matching methods to as many ontology variants as possible. We then propose to apply two main strategies to refine the initial alignment set: the removal of different kinds of ambiguities between correspondences and the use of the links published on the LOD. We illustrate our proposal in the field of life sciences and environment.

Thomas Hecht
INRA - AgroParisTech, UMR518 MIA-Paris, F-75231 Paris Cedex 05, France, e-mail: thomashecht95@gmail.com

Patrice Buche
INRA & LIRMM, 2 place Pierre Viala, F-34060 Montpellier Cedex 2, France, e-mail: Patrice.Buche@supagro.inra.fr

Juliette Dibie
INRA - AgroParisTech, UMR518 MIA-Paris, F-75231 Paris Cedex 05, France, e-mail: Juliette.Dibie@agroparistech.fr

Liliana Ibanescu
INRA - AgroParisTech, UMR518 MIA-Paris, F-75231 Paris Cedex 05, France, e-mail: Liliana.Ibanescu@agroparistech.fr

Cassia Trojahn dos Santos
IRIT & UTM2, 5 allées Antonio Machado, F-31058 Toulouse Cedex 9, France, e-mail: Cassia.Trojahn@irit.fr

1 Introduction

Ontologies are nowadays used as a common and standardized vocabulary for representing concepts and relations from a particular domain (e.g. life-science, geography). The Linked Open Data (LOD) cloud[1] contains more and more data sources published and linked together on the Web. Publishing and linking scientific data on the Web using ontologies for describing them should facilitate scientific data sharing, such as giving access to data from specific disciplines or data produced within specific geographic regions [Bizer, 2013].

When a new ontology, the source ontology, is published on the LOD, first, the 'target' ontologies, i.e. ontologies from similar domains with similar concepts, has to be identified among the already published ontologies in order to access new entities (concepts, properties or instances) and data sources. The source ontology can then be linked with each target ontology by finding an alignment (i.e. a set of correspondences) between entities. Different approaches have been proposed for the Ontology Matching task [Shvaiko and Euzenat, 2013, Bernstein et al., 2011, Rahm, 2011, Euzenat and Shvaiko, 2007] and a systematic evaluation on data sets from different domains has been carried out over the last ten years by the Ontology Alignment Evaluation Initiative (OAEI)[2].

In this paper, we propose an ontology matching method for aligning a source ontology with different target ontologies already linked and published on the LOD. An ontology can be either a thesaurus, an ontology or an ontological and terminological resource, expressed in different representation languages. Our method is based on the principle of *alignment refinement*: starting from a set of input alignments generated by several existing ontology matching methods, we propose to apply different strategies in order to refine this initial alignment set. One of our strategies is to exploit the links between the target ontologies, already published on the LOD.

We illustrate our method in the field of life sciences and environment. In this field, several thesauri have been created and published on the LOD. The two largest ones are AGROVOC[3] and NALT[4]. AGROVOC was created in the 1980s by FAO (Food and Agriculture Organization of the United Nations) as a structured multilingual thesaurus for agriculture, forestry, fishery, food and related fields (such as environment). It is available in 19 languages, with about 40 000 terms in each language [Caracciolo et al., 2012]. NALT is a bilingual thesaurus comparable with AGROVOC in terms of covered domain and maintained by USDA (United States Department of Agriculture). It is currently composed of approximately 91 000 terms in English and Spanish. For instance, the vocabulary of AGROVOC is currently linked to 15 international resources like GeoNames[5], DB-

[1] http://linkeddata.org

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

[3] http://aims.fao.org/standards/agrovoc/about

[4] http://agclass.nal.usda.gov/agt.shtml

[5] http://www.geonames.org

pedia[6] and GEMET [7]. In addition, 13 390 terms of AGROVOC are currently aligned with NALT [Caracciolo et al., 2012]. In this paper, we focus on the alignment of an ontological and terminological resource NARYQ (n-ary Relations between Quantitative experimental data) [Buche et al., 2013] with AGROVOC and NALT, in order to publish it on the LOD. NARYQ contains about 1 100 concepts structured into several sub-domains, such as food products, microorganisms and packaging.

This paper is organised as follows. Section 2 describes our method for aligning an ontology with linked ontologies on the LOD. Section 3 discusses the results of our experiments in the field of life sciences and environment. Section 4 presents related work and, finally, Section 5 concludes the paper and presents our perspectives.

2 Ontology matching method with linked ontologies

In this section, we present our matching method for aligning a source ontology with two target and linked ontologies. Our method is designed to align ontologies, thesauri or ontological and terminological resources, possibly described in different representation languages with different levels of expressiveness (e.g. OWL DL[8], SKOS[9]). An Ontological and Terminological Resource (OTR) [Reymonet et al., 2007, Roche et al., 2009, McCrae et al., 2011] is a hybrid model that combines a conceptual component and a terminological component: a concept is associated with a set of terms, each term denoting the concept with different lexical functions (e.g. synonyms, abbreviations, etc.). In the following, for the sake of simplicity and the paper's readibility, we abusively use ontology for either ontology, thesaurus or OTR.

The proposed method relies on the refinement of a set of input alignments generated by existing ontology matching methods. Our aim is therefore to be able to apply as much existing matching methods as possible in order to generate as much candidate correspondences as possible. Since the existing ontology matching methods can only be applied to some particular representation languages and the ontologies to be aligned can be expressed in several representation languages with different levels of expressiveness, we propose to apply matching methods on different variants of the ontologies to be aligned. A variant of an ontology corresponds to its expression in a given representation language. The first step of our matching method consists in aligning variants of the source ontology O_s with variants of the two target ontologies O_t^1 and O_t^2 using existing ontology matching methods. It allows the production of an initial set of alignments. In the second step, different refinement strategies are applied to this initial set of alignments, including the exploitation of the links defined on the LOD between the target ontologies. Figure 1 gives the overview of our matching method, which is detailed in the next two subsections.

[6] http://dbpedia.org

[7] http://www.eionet.europa.eu/gemet

[8] http://www.w3.org/TR/owl-guide

[9] http://www.w3.org/TR/2009/REC-skos-reference-20090818

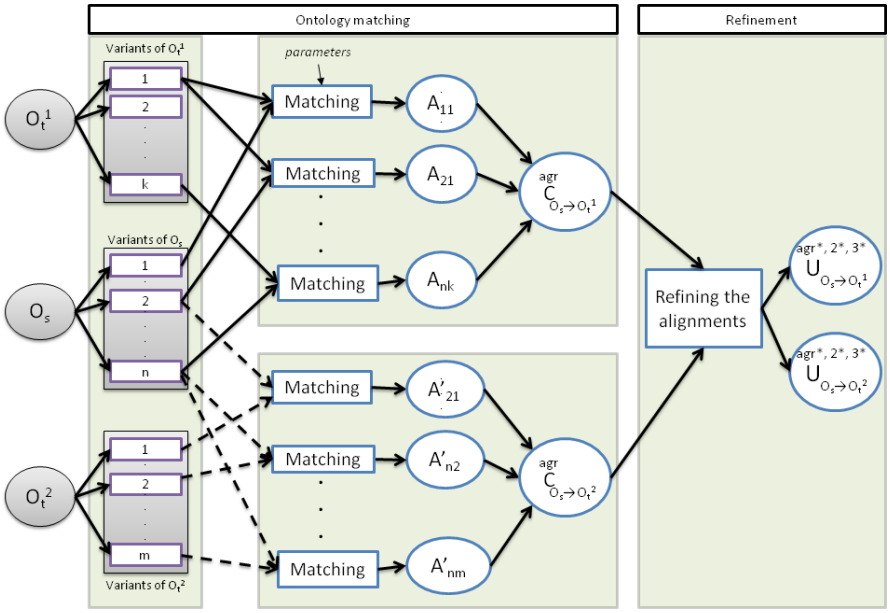

Fig. 1 Overview of our matching method.

2.1 First step : ontology matching

The first step of our method consists in aligning the source ontology O_s with each one of the two target ontologies O_t^1 and O_t^2.

2.1.1 Ontology variants

The source and target ontologies to be aligned can be thesauri, ontologies or OTR and may be expressed in different representation languages. Since the existing matching methods are usually designed for one particular representation language, we propose to associate to each ontology a set of *variants*, defined in the following:

Definition 1 (Set of variants of an ontology). The *set V_O of variants of an ontology O* is composed of its transformations in different representation languages $L_1, L_2, \ldots$. It contains the original version O^{orig} of the ontology.
$V_O = \{O^{orig}, O^{L_1}, O_1^{L_2}, O_2^{L_2}, \ldots\}$, where $O_j^{L_i}$ is the j^{th} transformation of the ontology O using the representation language L_i.

The aim of using variants is twofold. First, matching tools are designed to deal with specific input models (OWL ontologies in most cases). Diversifying the kinds of input, we are able to produce more candidate correspondences by using more tools. Second, representing resources using different constructors (OWL and SKOS)

allows the encoded knowledge to be exploited in different ways. On the one hand, tools can take advantage of OWL models for better exploiting automated reasoning. On the other hand, the lexicalisation of concepts is better expressed in SKOS models than in OWL models. For instance, in classical SKOS to OWL transformations, both skos:prefLabel and skos:altLabel are mapped to rdfs:label, where skos:altLabel are often used to represent synonyms, but also to refer to related terms. As a consequence, without introducing variants to catch this semantic richness, we could lose this information, which can instead be useful for tools able to deal with the specificities of SKOS.

2.1.2 Matching the ontology variants

The ontology matching process takes as input two ontologies and produces as output a set of correspondences between the entities of these two ontologies. According to [Euzenat and Shvaiko, 2007], this process can be defined as follows:

Definition 2 (Matching process [Euzenat and Shvaiko, 2007]). The *matching process* is a function f that, applied to two ontologies O_s and O_t and an (optional) initial alignment A^{orig}, produces a directed alignment $A^f_{O_s,O_t}$ between the two ontologies ($O_s \rightarrow O_t$). This process can use matching parameters p (e.g. weights, thresholds) and external resources r (e.g. common knowledge and domain specific thesauri):

$$A^f_{O_s,O_t} = f(O_s, O_t, A^{orig}, p, r)$$

Definition 3 (Correspondence [Euzenat and Shvaiko, 2007]). Let us consider two ontologies O_s and O_t, a *correspondence* c^f resulting from a matching process f is a relation r between the two entities e_s and e_t, denoted $c^f = \langle id, e_s, e_t, r, n \rangle$, such that : $c^f \in A^f_{O_s,O_t}$; $e_s \in O_s$ and $e_t \in O_t$; $r \in \{\equiv, \sqsubseteq, \sqsupseteq\}$; n is the confidence level (in general, $n \in [0,1]$) indicating the degree of confidence that the relation r holds between e_s and e_t.

Since the structural and the lexical information of the ontologies are exploited in different ways by the matching processes, the use of the most expressive variant of an ontology does not guarantee the best results. Therefore, the first step of our matching method consists in launching several matching processes on several variants of the ontologies to be aligned. Let us consider the source ontology O_s, one of the target ontologies O_t and the set of matching processes $F = \{f_1, f_2, \ldots\}$, which are launched to align the ontologies O_s and O_t, each matching process f_i is launched on a pair of variants (O^j_s, O^k_t) where $O^j_s \in V_{O_s}$ and $O^k_t \in V_{O_t}$ and generates the following alignment (i.e. set of correspondences):

$$A^{f_i}_{O^j_s,O^k_t} = f_i(O^j_s, O^k_t, \emptyset, p, r) \quad (1)$$

The result of our matching method between the source ontology O_s and one of the two target ontologies is a set of sets of correspondences, denoted $\overset{agr}{C}_{O_s \rightarrow O_t}$ for

aggregated set, generated by each matching process of F on each pair of ontology variants. This set, which comes from the concatenation of results from several matching processes on several ontology variants, is denoted:

$$\overset{agr}{C}_{O_s \to O_t} = \bigoplus_{i,j,k} A^{f_i}_{O_s^j, O_t^k} \tag{2}$$

The total number of matching processes launched in order to obtain an alignment between the source ontology O_s and each one of the two target ontologies O_t^1 and O_t^2 is:

$$|V_{O_s}| \times |V_{O_t^1}| \times |F| + |V_{O_s}| \times |V_{O_t^2}| \times |F| \tag{3}$$

2.2 Second step: refining the alignments

The second step of our matching method consists in refining the sets of sets of correspondences $\overset{agr}{C}_{O_s \to O_t^1}$ and $\overset{agr}{C}_{O_s \to O_t^2}$. These two sets of sets contain many correspondences (suggesting a good coverage) but also a lot of noise (i.e. incorrect correspondences) that has to be reduced.

In order to improve the quality of the correspondences found in the first step, we propose two refinement methods: the first one allows the identification of the potentially correct correspondences (see Subsection 2.2.1), the second refinement method allows the deletion of the correspondences considered as ambiguous and therefore potentially incorrect (see Subsection 2.2.2). Finally, we present in Subsection 2.2.3 our refinement process.

2.2.1 Identification of potentially correct correspondences

We distinguish two ways to identify potentially correct correspondences. When redundancies occur between correspondences that have been generated from at least two distinct matching methods, we assume that these correspondences can be considered as having more chances to be correct. We will retain them in a separate set, denoted $\overset{recT}{C}_{O_s \to O_t}$ for *recovering set*. These correspondences will be presented to the user as potentially correct correspondences.

Definition 4 (Recovering set). Let us consider two matching processes f_1 and f_2 applying two distinct matching methods for aligning two ontologies O_s and O_t, the *recovering set* $\overset{recT}{C}_{O_s \to O_t}$ is defined as follows:

If $c^{f_1} = \langle id_1, e_s^1, e_t^1, r_1, n_1 \rangle \wedge c^{f_2} = \langle id_2, e_s^2, e_t^2, r_2, n_2 \rangle \wedge e_s^1 = e_s^2 \wedge e_t^1 = e_t^2 \wedge r_1 = r_2$

then $c^{f_k} \in \overset{recT}{C}_{O_s \to O_t}$, where $c^{f_k} = \begin{cases} c^{f_1} \text{ if } n_1 \geq n_2 \\ c^{f_2} \text{ otherwise} \end{cases}$

Example 1. Let us also consider the correspondence c_1 generated by a matching process f_1 on the variants $\text{NARYQ}^{OWL-SKOS}$ of the source ontology NARYQ, and the variant AGROVOC^{SKOS} of the target ontology AGROVOC (see Subsection 3.1). Let us also consider the correspondence c_2 generated by a matching process f_2 which applies another matching method as the one used in the matching process f_1 on the variants $\text{NARYQ}^{OWL-SKOS}$ and AGROVOC_2^{OWL}. We have:

$c_1 = \langle id_1, \text{sheep}, c\_8854, \equiv, 0.95\rangle$, $c_1 \in A^{f_1}_{\text{NARYQ}^{OWL-SKOS},\text{AGROVOC}^{SKOS}}$

$c_2 = \langle id_2, \text{sheep}, c\_8854, \equiv, 0.75\rangle$, $c_2 \in A^{f_2}_{\text{NARYQ}^{OWL-SKOS},\text{AGROVOC}_2^{OWL}}$

The correspondences c_1 and c_2 generated by two distinct matching methods can be considered as redundant. The *recovering set* $\overset{recT}{C}_{\text{NARYQ}\rightarrow\text{AGROVOC}}$ contains the correspondence c_1 with the highest confidence level.

The second way of identifying potentially correct correspondences relies on the same assumption as above, i.e. 'comparable' correspondences can be considered as having more chances to be correct. Let us consider that there exists an alignment $A^{LOD}_{O_t^1\rightarrow O_t^2}$ defined on the LOD between the target ontologies O_t^1 and O_t^2, correspondences are said 'comparable' if an entity of the source ontology O_s is aligned, by an equivalence relation, with two distinct but linked on the LOD entities of the target ontologies O_t^1 and O_t^2. These correspondences will be kept in two separate sets, denoted $\overset{LOD}{C}_{O_s\rightarrow O_t^1}$ and $\overset{LOD}{C}_{O_s\rightarrow O_t^2}$ as *LOD recovering sets*. These sets will be presented to the user as sets of potentially correct correspondences.

Definition 5 (LOD recovering set). Let us consider $A^{LOD}_{O_t^1\rightarrow O_t^2}$ the result of a matching process between two ontologies O_t^1 and O_t^2 from the LOD, a set of matching processes $F^1 = \{f_1^1, f_2^1, \ldots\}$ applied to two ontologies O_s and O_t^1, and a set of matching processes $F^2 = \{f_1^2, f_2^2, \ldots\}$ applied to O_s and O_t^2, the *LOD recovering sets* $\overset{LOD}{C}_{O_s\rightarrow O_t^i}$, $i \in [1,2]$, are defined as follows:

If $\exists c \in A^{LOD}_{O_t^1\rightarrow O_t^2} \wedge c = \langle id, e_t^1, e_t^2, \equiv, n\rangle \wedge c^{f_i^1} \in A^{f_i^1}_{O_s,O_t^1} \wedge c^{f_j^2} \in A^{f_j^2}_{O_s,O_t^2} \wedge$
$c^{f_i^1} = \langle id_1, e_s^1, e_t^1, \equiv, n_1\rangle \wedge c^{f_j^2} = \langle id_2, e_s^2, e_t^2, \equiv, n_2\rangle \wedge e_s^1 = e_s^2 \wedge e_t^1 \neq e_t^2$,
then $c^{f_i^1} \in \overset{LOD}{C}_{O_s\rightarrow O_t^1}$ and $c^{f_j^2} \in \overset{LOD}{C}_{O_s\rightarrow O_t^2}$.

Hence, a correspondence $c^{f_i^1}$ belongs to the LOD recovering set $\overset{LOD}{C}_{O_s\rightarrow O_t^1}$, if i) the entity source e_s ($e_s = e_s^1 = e_s^2$) is aligned with an entity target e_t^1, ii) there exists a correspondence $c^{f_i^2}$ such that the entity source e_s is aligned with an entity target e_t^2, iii) there exists on the LOD a correspondence c linking e_t^1 and e_t^2.

Example 2. Let us consider the correspondence c_3 generated by the matching process f_1 of Example 1 and the correspondence c_4 generated by the matching process f_3 on the variants $\text{NARYQ}^{OWL-SKOS}$ and NALT^{OWL}. We have:

$c_3 = \langle id_3, \text{surimi}, c\_33271, \equiv, 0.87\rangle$, $c_3 \in A^{f_1}_{\text{NARYQ}^{OWL-SKOS},\text{AGROVOC}^{SKOS}}$

where c_33271 is a concept of AGROVOC;
$c_4 = \langle id_4, \text{surimi}, c\_40365, \equiv, 0.92\rangle$, $c_4 \in A^{f_3}_{\text{NARYQ}^{OWL-SKOS},\text{NALT}^{OWL}}$
where c_40365 is a concept of NALT.
Let us also consider that: $\exists c \in A^{f}_{\text{AGROVOC,NALT}}$, $c = \langle id_c, c\_33271, c\_40365, \equiv, 0.96\rangle$.
Then, we have: $c_3 \in \overset{LOD}{C}_{\text{NARYQ}\rightarrow\text{AGROVOC}}$ and $c_4 \in \overset{LOD}{C}_{\text{NARYQ}\rightarrow\text{NALT}}$.

2.2.2 Deletion of ambiguous correspondences

We distinguish three types of ambiguity between correspondences. The first type covers the correspondences obtained from the same matching method launched on different variants of the source and target ontologies. The correspondences of this type have the same source entity, the same target entity and the same relation. We propose to remove ambiguities of type 1 by keeping the correspondence with the highest confidence level.

Definition 6 (Ambiguous correspondences of type 1). Let us consider two matching processes f_1 and f_2 applying the same matching method to align two ontologies O_s and O_t (with O_s^j and O_t^k its respective variants), two correspondences c^{f_1} and c^{f_2}, from the sets $A^{f_1}_{O_s^{j_1},O_t^{k_1}}$ and $A^{f_2}_{O_s^{j_2},O_t^{k_2}}$, are *ambiguous according to type 1* if:

$$c^{f_1} = \langle id_1, e_s^1, e_t^1, r_1, n_1\rangle \wedge c^{f_2} = \langle id_2, e_s^2, e_t^2, r_2, n_2\rangle \wedge e_s^1 = e_s^2 \wedge e_t^1 = e_t^2 \wedge r_1 = r_2.$$

The set of sets of non ambiguous correspondences according to type 1 is:

$$\overset{agr*}{C}_{O_s\rightarrow O_t} = \bigoplus_{i,j,k}(A^{f_i}_{O_s^j,O_t^k}\backslash\{c^{f_k}\}) \text{ where } c^{f_k} = \begin{cases} c^{f_1} \text{ if } n_1 \geq n_2 \\ c^{f_2} \text{ otherwise} \end{cases}$$

Remark 1. Let us remember that when redundancies occur between correspondences generated by two distinct matching methods, these correspondences are considered as potentially correct (see Definition 4).

Example 3. Let us consider the correspondence c_1 generated by the matching process f_1 of Example 1. Let us also consider the correspondence c_5 generated by a matching process f_4 using the same matching method as the one used in the matching process f_1 but on the variant NARYQ^{SKOS} and the variant AGROVOC^{SKOS}. We have:
$c_1 = \langle id_1, \text{sheep}, c\_8854, \equiv, 0.95\rangle$, $c_1 \in A^{f_1}_{\text{NARYQ}^{OWL-SKOS},\text{AGROVOC}^{SKOS}}$
where c_8854 corresponds to the concept 'caprins' in AGROVOC.
$c_5 = \langle id_5, \text{sheep}, c\_8854, \equiv, 0.88\rangle$, $c_5 \in A^{f_4}_{\text{NARYQ}^{SKOS},\text{AGROVOC}^{SKOS}}$
The set of sets $\overset{agr*}{C}_{\text{NARYQ}\rightarrow\text{AGROVOC}}$ of non ambiguous correspondences according to type 1 only contains the correspondence c_1 with the highest confidence level.

The second type of ambiguity covers the correspondences in which an entity of the source ontology O_s is aligned, by an equivalence relation, with two distinct

entities of the target ontology O_t. We propose, in this case, to keep only the most relevant correspondence, i.e. the one that has *a priori* the highest confidence level. However, considering the fact that these correspondences were not generated by the same matching method, their confidence degrees are not comparable. Therefore, we propose to compute a similarity measure *sim* on the two correspondences to be compared, which is independent on the matching methods used to generate them. This similarity measure can rely, for instance, on syntactic similarity measures implemented in the Alignment API [David et al., 2011]. Here, we use the following syntactic measures: Hamming distance, Levenshstein distance, n-grams and Jaro and Jaro-Winkler to compute the similarity between all the labels, in a given language, of the two entities. The *sim* measure is the average of the computed similarities.

Definition 7 (Ambiguous correspondences of type 2). Let us consider a set of matching processes $F = \{f_1, f_2, \ldots\}$ applied to two ontologies O_s and O_t, two correspondences c^{f_i} and c^{f_j} are *ambiguous according to the type 2* if:

$$c^{f_i} = \langle id_1, e_s^1, e_t^1, \equiv, n_1 \rangle \wedge c^{f_j} = \langle id_2, e_s^2, e_t^2, \equiv, n_2 \rangle \wedge e_s^1 = e_s^2 \wedge e_t^1 \neq e_t^2.$$

The set of sets of non ambiguous correspondences of type 2 is:

$$\overset{agr2*}{C}_{O_s \to O_t} = \bigoplus_{i,j,k} (A^{f_i}_{O_s^j, O_t^k} \backslash \{c^{f_k}\}) \text{ where } c^{f_k} = \begin{cases} c^{f_i} \text{ if } sim(e_s^1, e_t^1) \leq sim(e_s^2, e_t^2) \\ c^{f_j} \text{ otherwise} \end{cases}$$

Remark 2. We only consider the equivalence relation here, because with other relations, the correspondences are not necessarily ambiguous, i.e. both of the correspondences can, in some cases, be considered as correct.

Example 4. Let us consider the correspondence c_1 generated by the matching process f_1 of Example 1 and the correspondence c_6 generated by the matching process f_2 of Example 1. We have:
$c_1 = \langle id_1, \text{sheep}, c\_8854, \equiv, 0.95 \rangle$, $c_1 \in A^{f_1}_{\text{NARYQ}^{OWL-SKOS}, \text{AGROVOC}^{SKOS}}$
$sim(\text{sheep}, c\_8854) = 0.815$, where c_8854 corresponds to the concept 'caprins' in AGROVOC;
$c_6 = \langle id_6, \text{sheep}, c\_9214, \equiv, 0.65 \rangle$, $c_6 \in A^{f_2}_{\text{NARYQ}^{OWL-SKOS}, \text{AGROVOC}_2^{OWL}}$
$sim(\text{sheep}, c\_9214) = 0.621$, where c_9214 corresponds to the concept 'goat' in AGROVOC.
The set of sets $\overset{agr2*}{C}_{\text{NARYQ} \to \text{AGROVOC}}$ of non ambiguous correspondences according to type 2 only contains the correspondence c_1 with the highest similarity measure.

Finally, the third type of ambiguity covers the correspondences where two distinct entities from the source ontology O_s are aligned, by an equivalence relation, with the same entity of the target ontology O_t. We propose, in this case, to keep the most relevant correspondence, i.e. the one with the highest similarity measure *sim*.

Definition 8 (Ambiguous correspondences of type 3). Let us consider a set of matching processes $F = \{f_1, f_2, \ldots\}$ applied to two ontologies O_s and O_t, two correspondences c^{f_i} and c^{f_j} are *ambiguous according to type 3* if:

$$c^{f_i} = \langle id_1, e_s^1, e_t^1, r_1, n_1 \rangle \wedge c^{f_j} = \langle id_2, e_s^2, e_t^2, r_2, n_2 \rangle \wedge e_s^1 \neq e_s^2 \wedge e_t^1 = e_t^2 \wedge r_1 = r_2.$$

The set of sets of non ambiguous correspondences of type 3 is defined as:

$$\overset{agr3*}{C}_{O_s \to O_t} = \bigoplus_{i,j,k} (A^{f_i}_{O_s^j, O_t^k} \backslash \{c^{f_k}\}) \text{ where } c^{f_k} = \begin{cases} c^{f_i} \text{ if } sim(e_s^1, e_t^1) \leq sim(e_s^2, e_t^2) \\ c^{f_j} \text{ otherwise} \end{cases}$$

Example 5. Let us consider the correspondence c_1 generated by the matching process f_1 of Example 1 and the correspondence c_6 generated by the matching process f_1. We have:
$c_1 = \langle id_1, \text{sheep}, c\_8854, \equiv, 0.95 \rangle$, $c_1 \in A^{f_1}_{\text{NARYQ}^{OWL-SKOS}, \text{AGROVOC}^{SKOS}}$
$sim(\text{sheep}, c\_8854) = 0.815$, where c_8854 corresponds to the concept 'caprins' in AGROVOC;
$c_7 = \langle id_7, \text{ewe}, c\_8854, \equiv, 0.55 \rangle$, $c_7 \in A^{f_1}_{\text{NARYQ}^{OWL-SKOS}, \text{AGROVOC}^{SKOS}}$
$sim(ewe, c\_8854) = 0.722$.
The set of sets $\overset{agr3*}{C}_{\text{NARYQ} \to \text{AGROVOC}}$ of non ambiguous correspondences according to type 3 only contains the correspondence c_1 with the highest similarity measure.

2.2.3 The refinement process

Figure 2 gives the overview of our refinement process, detailed in Subsections 2.2.1 and 2.2.2.

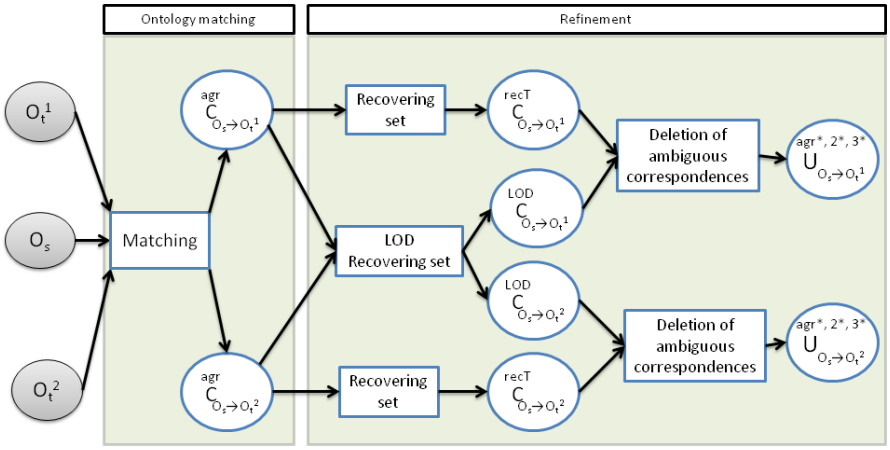

Fig. 2 Overview of our refinement process.

The set obtained by the union of the two recovering sets defined in Subsections 2.2.1 is denoted by:

$$U_{O_s \to O_t} = \overset{recT}{C}_{O_s \to O_t} \cup \overset{LOD}{C}_{O_s \to O_t} \quad (4)$$

We define the set of *potentially correct and non ambiguous correspondences* between a source ontology O_s and a target ontology O_t as follows.

Definition 9. The set of *potentially correct and non ambiguous correspondences* between a source ontology O_s and a target ontology O_t, denoted $\overset{agr*,2*,3*}{U}_{O_s \to O_t}$, is the set obtained after the removing of ambiguities of types 1, 2 and 3 as defined in Definitions 6, 7 and 8 from the set $U_{O_s \to O_t}$ given in Equation 4.

3 Experiments

We illustrate in this section our matching method described above for aligning a source ontology NARYQ (presented in Section 3.1) with each of the two target ontologies AGROVOC and NALT. In the following, the alignment of NARYQ with AGROVOC will be denoted NARYQ $\to$ AGROVOC and the alignment of NARYQ with NALT: NARYQ $\to$ NALT.

3.1 The source ontology NARYQ

The ontology NARYQ (n-ary Relations between Quantitative experimental data) has been created for representing n-ary relations between quantitative experimental data (see [Buche et al., 2013]). The characteristics of this ontology are the following: (i) it is an OTR; (ii) the labels are available in French and in English; (iii) it is represented in OWL DL and SKOS; and (iv) the conceptual component contains about 1 100 concepts structured into several sub-domains, the most important one in number being food products ($\approx$ 460 concepts), microorganisms ($\approx$ 180 concepts) and packaging ($\approx$ 150 concepts).

3.2 Reference alignments

In order to evaluate the quality of the generated alignments and to compare the results of the matching processes, we consider the measures of precision and recall adapted to the ontology matching task [Euzenat and Shvaiko, 2007]. These measures are based on a comparison between an automatically generated alignment A and a reference alignment R. The automatically generated alignment A is in this paper either the individual alignments provided by the matching tools or the alignment resulting from our approach. The construction of a complete reference

alignment R was not possible because it is a time-consuming task and it is difficult to find and to involve experts from the domain. Hence, we have built two partial manually validated reference alignments, denoted $\overline{R}^{+}_{\text{AGROVOC}}$ for the alignment NARYQ $\rightarrow$ AGROVOC, and $\overline{R}^{+}_{\text{NALT}}$ for the alignment NARYQ $\rightarrow$ NALT.

For each ontology and for every concept, we extracted their annotations (e.g. skos:prefLabel, skos:altLabel, rdfs:label, rdfs:comment) in English and in French as well as their structural elements (e.g. skos:broader, rdfs:subClassOf). A first alignment was created using SMOA [Stoilos et al., 2005] (*String Metric for Ontology Alignment*), a syntactic similarity metric for ontology matching. Using this metric and the equivalence relation $\equiv$, an alignment with 1 453 correspondences was generated, which was then manually validated by two experts in a double-blind process, and finally re-conciliated, i.e. the experts reached an *a posteriori* consensus. This first expert validation was performed in four hours using a visualisation tool developed for this specific task. It produced 318 validated correspondences in $\overline{R}_{\text{AGROVOC}}$ and 394 validated correspondences in $\overline{R}_{\text{NALT}}$, among which 233 concepts from NARYQ were aligned with both concepts from AGROVOC and concepts from NALT. In order to enrich these first generated reference alignments, an additional set of potentially correct correspondences was generated using our matching method and was validated by two experts. We therefore obtained two new and enriched reference alignments, denoted $\overline{R}^{+}_{\text{AGROVOC}}$ and $\overline{R}^{+}_{\text{NALT}}$.

- $\overline{R}^{+}_{\text{AGROVOC}}$ has 368 validated correspondences, with 361 concepts of NARYQ aligned with concepts of AGROVOC.
- $\overline{R}^{+}_{\text{NALT}}$ has 428 validated correspondences, with 424 concepts of NARYQ aligned with concepts of NALT.
- 303 concepts of NARYQ are aligned with both concepts of AGROVOC and concepts of NALT.

The alignments $\overline{R}^{+}_{\text{AGROVOC}}$ and $\overline{R}^{+}_{\text{NALT}}$, though partial, are used in the following as reference alignments.

3.3 Experimental protocol

Several matching processes were launched on several variants of the ontologies to be aligned in order to generate as much candidate correspondences as possible. We first present the ontology variants and then the selected matching processes.

3.3.1 The ontology variants

The variants of NARYQ are :

$$V_{\text{NARYQ}} = \{\text{NARYQ}^{OWL-SKOS}, \text{NARYQ}^{OWL}, \text{NARYQ}^{SKOS}\}$$

where the original version of NARYQ, denoted $\text{NARYQ}^{OWL-SKOS}$, is defined using OWL2 DL and SKOS. We also use two variants of NARYQ. The variant NARYQ^{OWL} was generated from its conceptual component by transforming both skos:prefLabel and skos:altLabel into rdfs:label, while the variant NARYQ^{SKOS} was generated using the labels of its terminological component and transforming the conceptual hierarchy into a SKOS hierarchy. The variants of AGROVOC are:

$$V_{\text{AGROVOC}} = \{\text{AGROVOC}^{SKOS}, \text{AGROVOC}_1^{SKOS}, \text{AGROVOC}_2^{OWL}, \text{AGROVOC}_3^{OWL}\}$$

AGROVOC^{SKOS} includes AGROVOC in all languages and we used its version downloaded in April 2013 from the official Web site[10]. AGROVOC_1^{SKOS} is a much smaller version, in English Only, available on the same Web site. The variant AGROVOC_2^{OWL} was used into the 2007 OAEI campaign. The variant AGROVOC_3^{OWL} was generated from AGROVOC^{SKOS} using a SKOSParser[11]. The variants of NALT are:

$$V_{\text{NALT}} = \{\text{NALT}^{SKOS}, \text{NALT}^{OWL}\}$$

where the original version NALT^{SKOS} was downloaded in April 2013 from the official Web site[12] and the variant NALT^{OWL} was generated from NALT^{SKOS} using the same SKOSParser.

3.3.2 The matching processes

Two available ontology matching tools, implementing different matching approaches with good results in the 2011 and 2012 OAEI campaigns [Aguirre et al., 2012], were selected: Aroma[13] [David, 2007] and LogMap[14] [Jiménez-Ruiz and Grau, 2011].

Aroma makes use of the association rule paradigm and a statistical measure assessing the implication intensity of the rules. The matching approach is divided into three steps: (1) pre-processing: each ontology entity, i.e. classes and properties, is represented by a set of terms – bag of words; (2) discovery of association rules between entities, and (3) post-processing: cleaning and enhancing the resulting alignment (i.e. deduction of equivalence relations, suppression of cycles in the alignment graph, suppression of redundant correspondences, and enhancement of the alignment by using equality and string similarity-based methods). Aroma is able to deal with both SKOS and OWL variants. This is not the case for LogMap, which encounters problems to parse SKOS variants.

LogMap adopts an approach based on logical reasoning and inconsistency repair techniques. The matching method follows five main steps: (1) lexical indexation of labels of entities and their lexical variations; (2) structural indexation based on in-

[10] http://aims.fao.org/access-agrovoc

[11] http://oaei.ontologymatching.org/2007/SKOSParser.pdf

[12] http://agclass.nal.usda.gov

[13] http://aroma.gforge.inria.fr/

[14] http://www.cs.ox.ac.uk/isg/projects/LogMap/

terval labeling schema for representing extended class hierarchies; (3) computation of initial anchor correspondences by intersecting the lexical indexes of entities; (4) iterative mapping repair and discovery, by filtering out logical inconsistencies in the mappings computed so far and by computing new mappings using string-based similarity method; and (5) ontology overlapping estimation, where ontology fragments overlap in both input ontologies.

Among 24 matching processes launched for aligning NARYQ and AGROVOC (see equation 3), only 9 produced non-empty alignments. From over 12 matching processes launched for aligning NARYQ and NALT, only 4 produced non-empty alignments. In the following, we assume that a correspondence is 'acceptable' if it has a confidence level greater than or equal to 0.5, a threshold empirically defined. The initial sets of alignments generated using the two matching tools (see equation 2) are: $\overset{agr}{C}_{\text{NARYQ}\rightarrow\text{AGROVOC}}$, denoted $\overset{agr}{C}_{\text{AGROVOC}}$, with 3 196 correspondences, and $\overset{agr}{C}_{\text{NARYQ}\rightarrow\text{NALT}}$, denoted $\overset{agr}{C}_{\text{NALT}}$, with 1 676 correspondences.

3.4 Experimental results

3.4.1 Individual results

Table 1 presents the results obtained from each matching tool, with respect to our two partial reference alignments $\overline{R}^{+}_{\text{AGROVOC}}$ and $\overline{R}^{+}_{\text{NALT}}$ (see Subsection 3.2). We can observe that thanks to variants we are able to overcome the limitations of tools in deadling with specific input models and hence we are able to produce more candidate correspondences. While LogMap is not able to generate alignments for the pairs involving SKOS variants (e.g. NARYQ^{SKOS}, AGROVOC^{SKOS} and NALT^{SKOS}), Aroma produces a set of correspondences with intermediary scores.

Table 2 presents the best scores obtained by the two matching tools and extracted from Table 1. These results are, in fact, an approximation, as they were computed using the partial reference alignments, which may affect the accuracy of the results. The values of each line of Table 2 represent the best score obtained for each indicator (number of correct correspondences, precision, recall or F-measure) by the matching tools. $\#^*$ corresponds to the *highest number of good correspondences*; P^* corresponds to the *best precision*; R^* corresponds to the *best recall*; and $F\text{-}m^*$ corresponds to the *best F-measure*.

3.4.2 Results of our approach

Table 3 presents the evaluation of NARYQ→ AGROVOC generated by different refinement methods, with respect to the partial reference alignment $\overline{R}^{+}_{\text{AGROVOC}}$. On the last row of Table 3, the symbol $\star$ indicates, for the indicator of the column, a better

Table 1 Individual results for the matching tools.

		Matching tools									
		LogMap					**Aroma**				
Alignment		#tot	#	P	R	F-m	#tot	#	P	R	F-m
NARYQ$^{OWL-SKOS}$	AGROVOC^{SKOS}	-	-	-	-	-	459	288	0.627	0.783	0.696
	AGROVOC$_1^{SKOS}$	-	-	-	-	-	386	288	0.746	0.783	0.764
	AGROVOC$_2^{OWL}$	203	180	0.887	0.489	0.630	-	-	-	-	-
	AGROVOC$_3^{OWL}$	185	167	**0.903**	0.454	0.604	-	-	-	-	-
	NALT^{SKOS}	-	-	-	-	-	476	**359**	0.754	0.839	0.794
	NALT^{OWL}	417	334	**0.801**	0.780	0.791	-	-	-	-	-
NARYQ^{OWL}	AGROVOC^{SKOS}	-	-	-	-	-	-	-	-	-	-
	AGROVOC$_1^{SKOS}$	-	-	-	-	-	-	-	-	-	-
	AGROVOC$_2^{OWL}$	312	269	0.862	0.731	0.791	1311	228	0.174	0.620	0.272
	AGROVOC$_3^{OWL}$	341	**300**	0.880	**0.815**	**0.846**	+	+	+	+	+
	NALT^{SKOS}	-	-	-	-	-	-	-	-	-	-
	NALT^{OWL}	456	356	0.781	**0.832**	**0.805**	-	-	-	-	-
NARYQ^{SKOS}	AGROVOC^{SKOS}	-	-	-	-	-	915	268	0.293	0.728	0.418
	AGROVOC$_1^{SKOS}$	-	-	-	-	-	1131	212	0.187	0.576	0.283
	AGROVOC$_2^{OWL}$	-	-	-	-	-	-	-	-	-	-
	AGROVOC$_3^{OWL}$	-	-	-	-	-	-	-	-	-	-
	NALT^{SKOS}	-	-	-	-	-	1011	327	0.323	0.764	0.454
	NALT^{OWL}	-	-	-	-	-	-	-	-	-	-

#tot indicates the total number of correspondences,
# indicates the number of correct correspondences,
+ indicates that the tool generated empty alignments,
− indicates that the tool was not able to deal with the input.

Table 2 Best scores of alignments obtained by the two matching tools.

Alignment	#*	P^*	R^*	$F\text{-}m^*$
NARYQ→ AGROVOC	300	0.90	0.82	0.85
NARYQ→ NALT	359	0.80	0.83	0.81

result than the best result of the matching tools for the same indicator presented in Table 2.

Table 3 Evaluation of NARYQ→ AGROVOC with respect to $\overline{R}^{+}_{\text{AGROVOC}}$.

Set	# total	# good	P	R	F-m
$C^{agr*}_{\text{AGROVOC}}$	1583	366	0.23	0.99	0.37
$C^{recT}_{\text{AGROVOC}}$	582	354	0.61	0.96	0.74
C^{LOD}_{AGROVOC}	336	254	0.76	0.69	0.72
U_{AGROVOC}	620	363	0.58	0.99	0.73
$U^{agr*,2*,3*}_{\text{AGROVOC}}$	447	344*	0.77	0.93*	0.84$^{\simeq}$

Table 4 presents the evaluation of NARYQ→ NALT generated by different refinement methods, with respect to the partial reference alignment $\overline{R}^{+}_{\text{NALT}}$. On the last row of Table 4, the symbol $\star$ indicates, for the indicator of the column, a better result than the best result of the matching tools for the same indicator presented in Table 2.

Table 4 Evaluation of NARYQ→ NALT with respect to $\overline{R}^{+}_{\text{NALT}}$.

Set	# total	# good	P	R	F-m
$\overset{agr*}{C}_{\text{NALT}}$	850	415	0.49	0.97	0.65
$\overset{recT}{C}_{\text{NALT}}$	480	368	0.77	0.86	0.81
$\overset{LOD}{C}_{\text{NALT}}$	337	255	0.76	0.59	0.67
U_{NALT}	551	404	0.73	0.94	0.82
$\overset{agr*,2*,3*}{U}_{\text{NALT}}$	400	348	0.87*	0.81	0.84*

3.4.3 Discussion

As we can notice in Tables 3 and 4 and as we might expect, (1) increasing the set of alignments allows the recall to be improved for most of the produced alignment sets[15], and (2) combining the different methods of refinement gives the best results in terms of precision (set $\overset{agr*,2*,3*}{U}$). Comparing these results with the best scores obtained by the two matching tools (Table 2), we obtained very promising results. Our approach obtains similar results in terms of F-measure for NARYQ → AGROVOC, while it increases F-measure for NARYQ → NALT. Our approach outperforms the best result in terms of recall for NARYQ → AGROVOC and in terms of precision for NARYQ → NALT. This performance produces very encouraging results.

Most matching tools apply strategies for combining different basic methods (i.e. lexical, structural, etc.) within a matching process and for filtering their results (threshold, weighted aggregation, etc.) (see [Euzenat and Shvaiko, 2007]). Our encouraging results can be explained by the fact that we propose in this paper to refine the sets of alignments produced by different matching methods in two different ways. First, we have identified three types of ambiguity to be resolved in order to refine the set of correspondences by deleting some of them. Second, we propose two new methods for discriminating and improving the sets of correspondences. In the first method, the redundant correspondences generated by at least two matching processes applying distinct matching methods are considered as potentially correct. The second method exploits the alignments defined on the LOD in order to reinforce the validity of some correspondences (i.e. correspondences allowing a same entity

[15] Since ambiguous correspondences according to type 1 produce only noises, the evaluation of the best recall is done considering $\overset{agr*}{C}$ and not $\overset{agr}{C}$.

to be aligned with two distinct but linked entities on the LOD are considered as potentially correct). Another original aspect of our approach consists in exploiting ontology variants, taking advantage of the characteristics of the ontologies, which can be ontologies, thesauri, OTR and can be expressed in different representation languages with different levels of expressiveness. This gives us the ability to cover a wide and diverse range of resources.

4 Related work

A key aspect of our proposal is the use of published links on the LOD as background knowledge for refining the results of a matching process. Similar works in this direction have been proposed in recent years in the literature, encouraged by the increasing number of available data sets on the LOD cloud. In [Nikolov et al., 2009], a schema matching approach which uses existing instance-level coreference links, defined in third-party repositories, as background knowledge is proposed. It aims at generating schema-level correspondences to assist the instance coreference resolution process. Rather than producing strict equivalence or subsumption relations, the algorithm produces fuzzy correspondences representing degrees of overlap between different ontologies. In [Pernelle and Sais, 2011], an approach that addresses both link discovery and ontology alignment is proposed, where the results of the link discovery step are exploited to improve the results of the ontology alignment step and *vice versa*. In [Parundekar et al., 2012], the proposal consists of (a) generating more expressive concepts from those already present in the ontologies (i.e. exploiting the space of concepts defined by value restrictions), and (b) aligning these extended concepts by exploiting the links between instances on the LOD. Contrary to our approach, these proposals consider a single representation of ontologies (i.e. OWL) and focus on links between instances.

For dealing with the specifities of community-created LOD data sets, a system for finding schema-level links is proposed in [Jain et al., 2010]. It computes alignments (not limited to equivalence relations) with the help of noisy community-generated data available on the Web, i.e. Wikipedia and Wikipedia category hierarchy. The idea of using Wikipedia category hierarchy, together with a rule-based verification approach, has also been exploited in [Grütze et al., 2012], where a holistic matching approach aims at aligning simultaneously multiple schemes on the LOD. In [Cruz et al., 2011], an extended version of the AgreementMaker system is proposed, aiming at handling subsumption relations and improving its performance when dealing with LOD ontologies. For each source and target concepts, the algorithm searches across several LOD ontologies for all concepts that are defined as subclasses, before applying matching strategies. Contrary to our approach, these works exploit other relations than equivalence and focus on the schema-level of LOD ontologies instead of exploiting the links between them. Contrary to the proposals described above, these latter works do not exploit the instance level within the schema-matching process.

[Mochol and Jentzsch, 2008, Steyskal and Polleres, 2013] propose, like us, to reuse existing tools and to combine their results to align two ontologies. In [Mochol and Jentzsch, 2008] a set of rules to select appropriate methods for a given pair of ontologies to be aligned is proposed. This selecting process is based on the background information describing the available approaches and the input properties of the ontologies. In [Steyskal and Polleres, 2013], an iterative method based on voting is proposed, where at every round, the correspondences accepted by the majority of tools are considered as valid. However, these works do not exploit the alignments on the LOD.

With regards to combining multiple alignments, different approaches have been proposed. In [Ghoula et al., 2014], an approach for normalising, combining and integrating alignments from multiple sources is proposed, where a correspondence can be associated to a set of relations and confidence levels. The algebra defined in [Euzenat, 2008] was applied in order to implement operators like union, composition and intersection. The approach can be applied regardless of the formalism used to represent the ontologies to be aligned. As we do, the normalisation step allows the removing of 'concurrent' (i.e. ambiguous) correspondences. However, we do not apply any combining operator, while they do not exploit LOD alignments. In [Lee et al., 2007], a library of matching components is made available and the user can select which components are to be used within a matching process, how they have to be combined together (average, minimum, maximum, weighed sum, decision trees, etc.), and how the correspondences are finally extracted (from a selection based on thresholding to formulate the selection as an optimisation problem over a weighted bipartite graph). The approach involves well automatic tuning of matching systems in order to find a tuning that optimises the performance of them. Here, we propose a different way for combining multiple alignments. In [Eckert et al., 2009], alignments generated from different matching systems are used as training data for a classifier that learns which combination of results provides the best indication of a correct correspondence. The multiple matchers are treated as a black-box. The assumption on which the approach relies is similar to ours : by using multiple matchers one can benefit from the high degree of precision of some matchers and at the same time the broader coverage of other matchers. In [Spiliopoulos and Vouros, 2012], combining multiple matchers is seen as a problem of maximising the social welfare within a group of interacting agents. Different agents computing alignments using specific methods and considering a specific kind of ontology entity, interact with each other and share constraints on the validity of the correspondences in order to reach an agreement. Although we do not aim at reaching a consensus between matchers, a correspondence is more likely to be correct if it is accepted by more than one matcher, which are not dedicated to find correspondences between specific ontology entities.

Finally, with respect to matching of terminologies in several languages, [Mougin and Grabar, 2013] adopts a notion of refining that is close to ours. The authors present a cross-language approach for matching two biomedical terminologies (MedDRA and SNMI). From a set of correspondences computed using lexical methods, the incorrect correspondences are filtered out using the notion of seman-

tic groups, which correspond to the partition of UMLS concepts. If the semantic groups which belong to UMLS concepts of MedDRA and SNMI terms are not the same, the correspondence is considered as incorrect. Then, they compute the number of correspondences which are common to different languages (correspondences which are more likely to be correct) and suppress the ambiguities by eliminating correspondences found in only one language.

5 Conclusion and perspectives

In this paper, we have proposed a new ontology matching method which can raise one of the challenges of ontology matching stated in [Shvaiko and Euzenat, 2013]: matching with background knowledge. In a first step, our matching method allows to generate many correspondences using and combining existing methods for aligning ontologies, thesauri and OTR expressed in different representation languages. Then, it allows a discrimination of the correspondences by removing some ambiguities and by exploiting the redundancy and existing alignments on the LOD, in order to identify a subset of potentially correct correspondences which will be submitted to the user for validation.

This proposal is a preliminary work for publishing ontologies on the LOD. Ontology matching allows a source ontology not only to be enriched with new concepts and/or terms, but also to be linked with existing and in use ontologies on the LOD in order to contribute to the data sharing in the target domain.

In order to improve our process of refinement, we plan in the short term, (i) to evaluate our approach using another data set, such as the OAEI Library task, which offers variants of their thesauri and for which we can find LOD alignments between them, (ii) to take into account the expressiveness of ontology variants to suppress the ambiguities of type 1; (iii) to exploit other relations than the equivalence in the treatment of ambiguities of type 2 – instead of removing ambiguous correspondence of type 2, we plan to propose a methodology based on reasoning for choosing the best correspondence between the ambiguous ones: by removing, for instance, correspondences which introduce a logical inconsistency; (iv) to remove the ambiguities between correspondences by defining new relations – we can, for instance, use an algebra to define a new correspondence with a new relation which combines the relations involved in the ambiguous correspondences; (v) to study how to use the subsumption relation in order to facilitate the identification of potentially correct correspondences; (vi) to study how to use relations between concepts (e.g. their domains and ranges) and their matching results in order to suppress the ambiguities and/or to identify the potentially correct correspondences. In the long term, we plan to exploit indirect alignments between different sources on the LOD to improve the discrimination on the set of correspondences. We also plan to extend our approach to take into account more complex entities such as units of measurement and n-ary relations.

References

[Aguirre et al., 2012] Aguirre, J., Eckert, K., Euzenat, J., Ferrara, A., van Hage, W. R., Hollink, L., Meilicke, C., Nikolov, A., Ritze, D., Scharffe, F., Pavel Shvaiko, O. S.-Z., Trojahn, C., Jiménez-Ruiz, E., Grau, B. C., and Zapilko, B. (2012). Results of the ontology alignment evaluation initiative 2012. In *Proc. 7th ISWC workshop on ontology matching (OM)*, page 73115.

[Bernstein et al., 2011] Bernstein, P. A., Madhavan, J., and Rahm, E. (2011). Generic schema matching, ten years later. *PVLDB*, 4(11):695–701.

[Bizer, 2013] Bizer, C. (2013). Interlinking scientific data on a global scale. *Data Science Journal*, 12:GRDI6–GRDI12.

[Buche et al., 2013] Buche, P., Dervaux, S., Dibie-Barthelemy, J., Ibanescu, L., Soler, L., and Touhami, R. (2013). Intégration de données hétérogènes et imprecise guide par une resource termino-ontologique. application au domaine des sciences du vivant. *RSTI série Revue dIntelligence Artificielle*, 27(4-5):539–568.

[Caracciolo et al., 2012] Caracciolo, C., Stellato, A., Rajbhandari, S., Morshed, A., Johannsen, G., Keizer, J., and Jaques, Y. (2012). Thesaurus maintenance, alignment and publication as linked data: the AGROVOC use case. *IJMSO*, 7(1):65–75.

[Cruz et al., 2011] Cruz, I. F., Palmonari, M., Caimi, F., and Stroe, C. (2011). Towards "on the go" matching of linked open data ontologies. In *Workshop on Discovering Meaning On the Go in Large Heterogeneous Data 2011 (LHD-11), Barcelona, Spain, July 16, 2011*.

[David, 2007] David, J. (2007). *AROMA: une méthode pour la découverte d'alignements orientés entre ontologies partir de règles d'association*. PhD thesis, Université de Nantes.

[David et al., 2011] David, J., Euzenat, J., Scharffe, F., and Trojahn dos Santos, C. (2011). The alignment api 4.0. *Semantic web*, 2(1):310.

[Eckert et al., 2009] Eckert, K., Meilicke, C., and Stuckenschmidt, H. (2009). Improving ontology matching using meta-level learning. In *The Semantic Web: Research and Applications*, volume 5554 of *Lecture Notes in Computer Science*, pages 158–172. Springer Berlin Heidelberg.

[Euzenat, 2008] Euzenat, J. (2008). Algebras of ontology alignment relations. In *International Semantic Web Conference*, volume 5318 of *Lecture Notes in Computer Science*. Springer.

[Euzenat and Shvaiko, 2007] Euzenat, J. and Shvaiko, P. (2007). *Ontology matching*, volume 18. Springer Heidelberg.

[Ghoula et al., 2014] Ghoula, N., Nindanga, H., and Falquet, G. (2014). Opérateurs de gestion des alignements de ressources de connaissances hétérogènes. In *TO BE COMPLETED*.

[Grütze et al., 2012] Grütze, T., Böhm, C., and Naumann, F. (2012). Holistic and scalable ontology alignment for linked open data. In Bizer, C., Heath, T., Berners-Lee, T., and Hausenblas, M., editors, *WWW2012 Workshop on Linked Data on the Web, Lyon, France, 16 April, 2012*, volume 937 of *CEUR Workshop Proceedings*. CEUR-WS.org.

[Jain et al., 2010] Jain, P., Hitzler, P., Sheth, A. P., Verma, K., and Yeh, P. Z. (2010). Ontology alignment for linked open data. In *Proceedings of the 9th International Semantic Web Conference on The Semantic Web - Volume Part I*, pages 402–417, Berlin, Heidelberg. Springer-Verlag.

[Jiménez-Ruiz and Grau, 2011] Jiménez-Ruiz, E. and Grau, B. C. (2011). Logmap: Logic-based and scalable ontology matching. In *The Semantic WebISWC 2011*, page 273288. Springer.

[Lee et al., 2007] Lee, Y., Sayyadian, M., Doan, A., and Rosenthal, A. S. (2007). etuner: Tuning schema matching software using synthetic scenarios. *The VLDB Journal*, 16(1):97–122.

[McCrae et al., 2011] McCrae, J., Spohr, D., and Cimiano, P. (2011). Linking Lexical Resources and Ontologies on the Semantic Web with Lemon. In Antoniou, G., Grobelnik, M., Simperl, E. P. B., Parsia, B., Plexousakis, D., Leenheer, P. D., and Pan, J. Z., editors, *ESWC (1)*, volume 6643 of *Lecture Notes in Computer Science*, pages 245–259. Springer.

[Mochol and Jentzsch, 2008] Mochol, M. and Jentzsch, A. (2008). Towards a rule-based matcher selection. In Gangemi, A. and Euzenat, J., editors, *Knowledge Engineering: Practice and Patterns*, volume 5268 of *Lecture Notes in Computer Science*, pages 109–119. Springer Berlin Heidelberg.

[Mougin and Grabar, 2013] Mougin, F. and Grabar, N. (2013). Using a cross-language approach to acquire new mappings between two biomedical terminologies. In *Artificial Intelligence in*

Medicine, volume 7885 of *Lecture Notes in Computer Science*, pages 221–226. Springer Berlin Heidelberg.

[Nikolov et al., 2009] Nikolov, A., Uren, V., Motta, E., and Roeck, A. (2009). Overcoming schema heterogeneity between linked semantic repositories to improve coreference resolution. In *Proceedings of the 4th Asian Conference on The Semantic Web*, ASWC '09, pages 332–346, Berlin, Heidelberg. Springer-Verlag.

[Parundekar et al., 2012] Parundekar, R., Knoblock, C. A., and Ambite, J. L. (2012). Discovering concept coverings in ontologies of linked data sources. In Cudré-Mauroux, P., Heflin, J., Sirin, E., Tudorache, T., Euzenat, J., Hauswirth, M., Parreira, J. X., Hendler, J., Schreiber, G., Bernstein, A., and Blomqvist, E., editors, *International Semantic Web Conference (1)*, volume 7649 of *Lecture Notes in Computer Science*, pages 427–443. Springer.

[Pernelle and Sais, 2011] Pernelle, N. and Sais, F. (2011). LDM: Link Discovery Method for new Resource Integration. In Zoé Lacroix, Edna Ruckhaus, M.-E. V., editor, *Fourth International Workshop on Resource Discovery*, volume 737, pages 94–108, Heraklion, Grèce.

[Rahm, 2011] Rahm, E. (2011). Towards large-scale schema and ontology matching. In Bellahsene, Z., Bonifati, A., and Rahm, E., editors, *Schema Matching and Mapping*, pages 3–27. Springer.

[Reymonet et al., 2007] Reymonet, A., Thomas, J., and Aussenac-Gilles, N. (2007). Modelling ontological and terminological resources in OWL DL. In *OntoLex 2007 - Workshop at ISWC07*, Busan, South-Korea.

[Roche et al., 2009] Roche, C., Calberg-Challot, M., Damas, L., and Rouard, P. (2009). Ontoterminology - a new paradigm for terminology. In Dietz, J. L. G., editor, *KEOD*, pages 321–326. INSTICC Press.

[Shvaiko and Euzenat, 2013] Shvaiko, P. and Euzenat, J. (2013). Ontology matching: state of the art and future challenges. *IEEE Trans. Knowl. Data Eng.*, 25(1):158–176.

[Spiliopoulos and Vouros, 2012] Spiliopoulos, V. and Vouros, G. (2012). Synthesizing ontology alignment methods using the max-sum algorithm. *Knowledge and Data Engineering, IEEE Transactions on*, 24(5):940–951.

[Steyskal and Polleres, 2013] Steyskal, S. and Polleres, A. (2013). Mix'n'match: An alternative approach for combining ontology matchers. In Meersman, R., Panetto, H., Dillon, T. S., Eder, J., Bellahsene, Z., Ritter, N., Leenheer, P. D., and Dou, D., editors, *On the Move to Meaningful Internet Systems: OTM 2013 Conferences - Confederated International Conferences: CoopIS, DOA-Trusted Cloud, and ODBASE 2013, Graz, Austria, September 9-13, 2013. Proceedings*, volume 8185 of *Lecture Notes in Computer Science*, pages 555–563. Springer.

[Stoilos et al., 2005] Stoilos, G., Stamou, G., and Kollias, S. (2005). A string metric for ontology alignment. In *The Semantic WebISWC 2005*, page 624637. Springer.