Metadata-driven Semantic Coordination

Manuel Fiorelli [0000-0001-7079-8941] [1], Armando Stellato [0000-0001-5374-2807] [1], Tiziano Lorenzetti[1], Peter Schmitz[2], Enrico Francesconi[2], Najeh Hajlaoui[2], Brahim Batouche[2]

[1] University of Rome Tor Vergata, Department of Enterprise Engineering, via del Politecnico 1, 00133 Roma, Italy
fiorelli@info.uniroma2.it, stellato@uniroma2.it, tiziano.lorenzetti@gmail.com

[2] Publications Office of the European Union, Luxembourg
{Peter.SCHMITZ, Enrico.FRANCESCONI}@publications.europa.eu
{Najeh.HAJLAOUI, Brahim.BATOUCHE}@ext.publications.europa.eu

Abstract. Reuse and combination of disparate datasets on the Semantic Web require semantic coordination, i.e. the ability to match heterogeneous semantic models. Systematic evaluations raised the performance of matching systems in terms of compliance and resource consumption. However, it is equally important to be able to identify diverse matching scenarios, covering a range of variations in the datasets such as different modeling languages, heterogeneous lexicalizations, structural differences and to be able to properly handle these scenarios through dedicated techniques and the exploitation of external resources. Furthermore, this should be achieved without requiring manual tinkering of low-level configuration knobs. As of the Semantic Web vision, machines should be able to coordinate and talk to each other to solve problems. To that end, we propose a system that automates most decisions by leveraging explicit metadata regarding the datasets to be matched and potentially useful support datasets. This system uses established metadata vocabularies such as VoID, Dublin Core and the LIME module of OntoLex-Lemon. Consequently, the system can work on real-world cases, leveraging metadata already published alongside self-describing datasets.

Keywords: ontology matching, metadata, OntoLex-Lemon.

1 Introduction

The Semantic Web [1,2] and, followingly, Linked Open Data (LOD) best-practices [3] brought knowledge representation, sharing and reuse to the web scale. At such scale, proliferation of different semantic models for overlapping domains is inevitable and even positive, being connected to autonomy and diversity, and to the complementary needs for specialization and experimentation [4]. Moreover, traditional data integration based on the upfront definition of a mediated schema fails on the web, as the web deals with any domain, while a mediated schema about everything is impossible to construct and in any case very brittle [5]. Integration on the web should be afforded in a pay-as-you-go manner [5], only when tighter integration between some data sources appears necessary. This lazy approach to integration marks a departure from consolidated databases towards the novel concept of dataspace [6]. Indeed, Linked Open Data has been evolving the web into a global dataspace [7].

Initially conceived in the context of distributed systems using a game theoretic perspective, semantic coordination is conceptually close to ontology matching [8]: the problem of finding a set of correspondences (i.e. an alignment) between semantically related concepts in two (or more) ontologies. The innumerous approaches to ontology[1] matching can be classified according to different criteria [9]. An important distinction is then between approaches that rely solely on the content of the input ontologies (internal) and those that benefit from other information sources (external). Indeed, matching with background knowledge was identified as one of the future challenges for ontology matching [10], together with – to mention just another example – matcher selection, combination and tuning.

Annual campaigns for the evaluation of ontology matching systems have greatly sustained the improvement of matching techniques, especially for what concerns compliance to the task (measured in terms of precision and recall) and resource consumption (e.g. limiting the amount of time and memory required to solve large matching problems). However, these campaigns are strongly targeted at evaluating the approaches, often allowing data to be cleaned, uniformed and made generally "more easily processable" [11,12]. It is thus equally important that matching systems are flexible enough to identify diverse matching scenarios, covering a range of variations in the datasets such as different modeling languages, heterogeneous lexicalizations, structural differences and that are able to properly handle these scenarios through dedicated techniques and the exploitation of external resources

While configurability is a necessary condition for flexibility, we contend that a usable matching system should be smart enough to do most configuration decisions on its own. Furthermore, as the assessment of an alignment scenario should be based upon the combined analysis of resources' characteristics, we believe that these characteristics should be made evident a priori, in the form of exploitable metadata.

In this paper, we propose a platform for semantic coordination that aims to achieve this goal, by relying on metadata about the input ontologies and potentially relevant third-party resources. We extended our previous work on MAPLE [11], such semantic coordination system, improving its architecture and providing a use case inside the VocBench 3 [13] collaborative RDF editor.

The paper is structured as follows. Section 2 discusses related work focusing on setup of matching processes. Section 3 presents our framework, while Section 4 provides a use case within a real collaborative editor of ontologies, thesauri and lexicons. Section 5 discusses our work. Finally, Section 6 concludes the paper.

[1] The expression "ontology matching" is often used in a broader sense than the one the first word of the term would suggest. "Ontology" is in this case a synecdoche for ontologies, thesauri, lexicons and any sort of knowledge resources modeled according to core knowledge modeling languages for the Semantic Web. The expression ontology matching thus defines the task of discovering and assessing alignments between ontologies and other data models of the RDF family; alternative expressions are ontology mapping or ontology alignment. In the RDF jargon, and following the terminology adopted in the VoID metadata vocabulary [29], a set of alignments is also called a Linkset.

2 Related Work

Nowadays, most matching systems use an ensemble of matching techniques, often relying on different features of the input ontologies, and combined according to varying topologies. As pointed out in the introduction, there is a need for automating the selection, tuning and combination of these approaches. Hereafter, we report some works to showcase the main approaches: weighting, rule-based systems, and machine learning.

RiMOM [14], a multi-strategy ontology matching system, relies on automatically computed metrics about a given matching task to decide the relative weight of lexical and structural approaches, to tune the construction of some representations (in particular, the inclusion of structural features in the virtual documents associated with the ontology concepts) and to decide which edges are considered for similarity propagation. These decisions are based on two metrics about the matching task that are computed jointly against the input ontologies: lexical similarity and structural similarity.

MOMA [15] uses a rule-based approach to select the appropriate matcher for the given match task from a repository of matchers. A set of rules (implemented in SWRL) captures the correspondence between characteristics of the matchers and characteristics of the input ontologies. Both characteristics are modeled through dedicated ontologies: metadata about the input ontologies are computed automatically, whereas metadata about the matchers were obtained through an online survey.

Cruz *et al* [16] framed the selection of a matcher configuration as a classification problem, evaluating a number of supervised learning algorithms, and eventually concluding that k-NN performs best (given the limited amount of training data). In their formulation, the class to predict is the appropriate configuration (among a few of them), while the features are derived from the profiles of the input ontologies. They extended the OntoQA [17] metrics originally developed for ontology evaluation.

MOMA is surely close to our framework because of its use of explicit metadata about the input ontologies and (differently from us) about the available matchers. Like us, MOMA addresses the diversity of the core models (e.g. OWL ontology vs SKOS thesaurus); however, while we strive to support the adaptation of the same matcher to different models, the primary aim of MOMA is to check the compatibility between matchers and models.

Our work focuses on characterizing a matching task to enable the exploitation of information in the input ontologies and, if available, in external resources. Indeed, our work is propaedeutic to matcher selection, combination and tuning, which were dealt with in detail in the above-mentioned works. Moreover, those works disregard the selection of support resources (e.g. ontologies, thesauri, lexicons) that may provide additional clue for the creation of mappings.

As we already pointed out, RiMOM computes metrics about the matching task, by taking into consideration the pair of ontologies. Our framework, instead, is primarily concerned with metrics computed on the input ontologies, took into isolation. The metadata model used in MOMA has probably the widest coverage of syntactic (i.e. modelling constructs) and semantic features (e.g. subject domain, level of formality, natural language). Conversely, OntoQA provides a very detailed picture of the structure of an ontology, while offering just a metric, called "readability", telling the existence

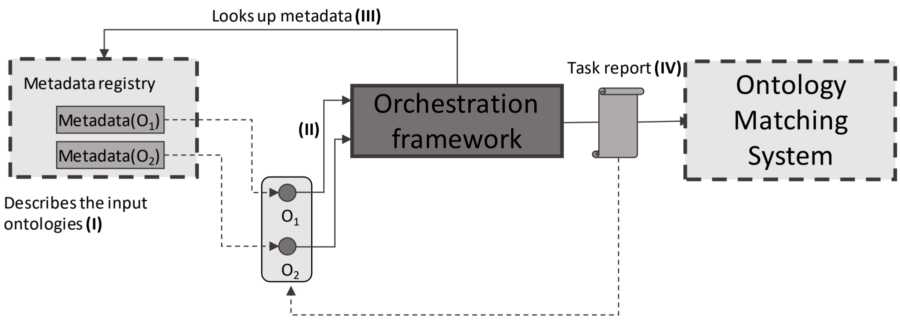

Fig. 1. Overall architecture of the framework. The picture shows dependencies (dashed rectangles) and how components interact (solid arrows).

of `rdfs:labels` and `rdfs:comments`. This metric is bound to one lexicalization model (RDFS) and disregards the language in which information is expressed.

External resources can be useful because they may provide background knowledge that didn't find its way into the ontologies being matched. Following Faria *et al* [18], external resources include ontologies or thesauri, lexical databases, textual corpora and websites. There are a lot of works that focus on some (wide coverage) resource, while – in the best cases – describing it a specific instance of a generic oracle. Mascardi *et al* ran some experiments [19] of indirect matching using SUMO-OWL, OpenCyc and DOLCE, mining some rules for the use of upper ontologies as background knowledge in ontology matching. BLOOMS [20] uses the categories associated with Wikipedia pages to compare the results found with different class names. WikiMatch [21] also uses Wikipedia, but it only compares the search results as sets. Princeton WordNet [22] is a notable lexico-semantic resource used in several systems. Actually, relying on Princeton WordNet alone bounds the system to the English language, ignoring the opportunities offered by similarly modeled resources for other languages [23]. Furthermore, the publication of language resources on the web increasingly uses the linked data paradigm [24], and often relies on OntoLex-Lemon [25]. Following again Faria *et al* [18], we report on a few attempts at discovering (or at least automating the choice of) suitable background ontologies. Sabou *et al* [26] do element-level ontology matching by recursive search of class names using semantic search engines. Other works identify relevant ontologies by looking at the input ontologies as a whole, and trying to optimize different metrics, such as similarity between the input ontologies and the background ontology [27], effectiveness of the background ontology [28] (i.e. the mapping overlap between each input ontology and the background ontology), and mapping gain (i.e. fraction of new mappings generated using the background ontology) [18].

3 Our Framework

Fig. 1 illustrates our orchestration framework and its interactions with collaborators.

A *matching task* (rounded rectangle) is defined by a pair of ontologies (O_1 and O_2) to match. The very first step is to add their metadata into the *metadata registry*. The *orchestration framework* depends on an implementation of this registry, to obtain metadata about the input ontologies and third-party resources that may support the matching task. In Section 4, we discuss a specific implementation of the registry that was developed as a part of a use case. The framework is independent from the implementation of the registry and, moreover, from the strategy for the generation of metadata (e.g. manual insertion, automatic profiling, etc.); nonetheless, it mandates a specific metadata profile, based on popular standards (DCAT [29], VoID [30], OntoLex-Lemon LIME [31], Dublin Core [32]).

Our *orchestration framework* looks up metadata about the input ontologies in the *metadata registry* and by analyzing the discovered metadata becomes aware of the characteristics of the *matching task* (e.g. knowledge/lexicalization models, overlap between supported natural languages, potentially useful external resources, etc.) and produces a *task report* to transfer such awareness to an *ontology matching system*. While the choice of a specific *matching approach* is delegated to the downstream matching system, our framework makes general assumptions, such as the use of lexicalizations to seed the matching process or looking for synonyms or translations within language resources.

Fig. 2 contains a sample *task report* for the alignment of TESEO[2] and EuroVoc[3], respectively, the thesaurus of the Italian Senate of the Republic and the thesaurus of the European Union. A preliminary observation is that this report conforms to the JSON-LD standard, and consequently, can be mapped quite easily to RDF. Specifically, most properties in the JSON object correspond to properties with the same name in the metadata vocabularies just mentioned. A notable exception is the property `languageTag` that should be mapped to `lime:language`.

At the beginning of the report, we can find the description of the datasets to be matched as values of the properties `sourceDataset` and `targetDataset`, respectively. The description contains the identifier of the dataset (`@id`) in the *metadata registry*, which provides a consistent name to reference that dataset unambiguously. The metadata (`dcterms:`)`conformsTo` tells the nature of the dataset, e.g. to distinguish between an ontology, a thesaurus (as in this case), etc. Obviously, this is important in order to interpret the input, to establish the goal of the alignment (in the example, establish correspondences between `skos:Concepts`), and to fine tune the matching strategy (in the example, hierarchy-based techniques should consider the property `skos:broader` rather than `rdfs:subClassOf` as it happens with ontologies). The property (`void:`)`sparqlEndpoint` holds the address of a SPARQL endpoint where the actual content of the dataset can be found, which is, clearly, a must-have for an ontology matching system.

The property `supportDatasets` contains other datasets that are referenced elsewhere in the report as they may be useful to solve the matching task. The description of these datasets is, in general, an extension of the one already discussed for the input datasets, including additional properties depending on the type of the dataset (`@type`). The dataset typology as well as the properties providing various metrics are defined by the LIME module of the OntoLex-Lemon vocabulary.

[2] `http://www.senato.it/3235?testo_generico=745`
[3] `http://eurovoc.europa.eu/`

```
{
"sourceDataset": { "@type": "Dataset",
    "@id": "http://.../void.ttl#TESEO",
    "conformsTo": “http://www.w3.org/2004/02/skos/core”,
    "uriSpace": "http://www.senato.it/teseo/tes/",
    "sparqlEndpoint": http://localhost:7200/repositories/TESEO_core },
"targetDataset": {"@id": "http://.../void.ttl#EuroVoc", … },
"supportDatasets": [{
        "@id": "http://.../omw/MultiWordNet-it-lexicon",
        "@type": "http://www.w3.org/ns/lemon/lime#Lexicon",
        "sparqlEndpoint": "http://localhost:7200/repositories/OMW_core",
        "languageTag": "43011",  "lexicalEntries": 43011
    }, {
        "@id": "http://.../omw/pwn30-conceptset",
        "@type": "http://www.w3.org/ns/lemon/ontolex#ConceptSet",
        "sparqlEndpoint": "http://localhost:7200/repositories/OMW_core",
        "concepts": 117659
    }, {
        "@id": "http://.../void.ttl#OMW_ConceptualizationSet
        "@type": "http://www.w3.org/ns/lemon/lime#ConceptualizationSet",
        "sparqlEndpoint": "http://localhost:7200/repositories/OMW_core",
        "lexiconDataset": "http://.../omw/MultiWordNet-it-lexicon",
        "conceptualDataset": "http://.../omw/pwn30-conceptset",
        "conceptualizations": 63133, "concepts": 35001, "lexicalEntries": 43011,
        "avgSynonymy": 0.537, "avgAmbiguity": 1.468
    }, {
        "@id": "http://.../void.ttl#TESEO_it_lexicalization_set",
        "@type": "http://www.w3.org/ns/lemon/lime#LexicalizationSet"
        "sparqlEndpoint": "http://localhost:7200/repositories/EuroVoc_core",
        "referenceDataset": "http://.../void.ttl#TESEO",
        "lexicalizationModel": "http://www.w3.org/2008/05/skos-xl",
        "lexicalizations": 18545,  "references": 7282, "avgNumOfLexicalizations": 2.546,
        "percentage": 1.0,
        "languageTag": "it",
    },{ "@id": "http://.../void.ttl#EuroVoc_it_lexicalization_set", …} ],
"pairings": [{
        "score": 0.7836831074710862,
        "source": {
            "lexicalizationSet": "http://.../void.ttl#TESEO_it_lexicalization_set",
            "synonymizer": {
                "lexicon": "http://.../omw/MultiWordNet-it-lexicon",
                "conceptualizationSet":   "http://.../void.ttl#MultiWordNet-it-lexicon_pwn30-con-
ceptset_conceptualization_set"
            }
        },
        "target": {
            "lexicalizationSet": "http://.../void.ttl#EuroVoc_it_lexicalization_set",
            "synonymizer": { … }
        }
    } ]
}
```


Fig. 2. Task report for the alignment of TESEO and EuroVoc

The `pairings` property contains an ordered list, the items of which contain a pair of lexicalization sets, respectively, for the source and target datasets. The underlying assumption made by our framework is that an important, if not primary, source of evidence for ontology matching is represented by the labels, which are grouped into lexicalization sets. Referenced through its identifier, the description of a lexicalization set includes useful information. First, the property `languageTag` identifies the natural language in which the labels are provided. This information is important for two reasons: i) to apply language-specific processes (e.g. lemmatization requires the lexicon of a given language), ii) to distinguish between a mono-lingual pairing (as in the example) and a cross-lingual one (currently under development).

In the example, the framework suggests working on labels in Italian, the sole language in common between the two thesauri. Furthermore, the framework instructs the matcher to extract them by applying the pattern for the SKOS-XL lexicalization model. In the case of a monolingual pairing, the framework also suggests a *synonymizer*, which can be understood as a wordnet-like assembly of a *lime:Lexicon* (providing the words) and a *lime:ConceptualizationSet* (linking words to lexical concepts): synonyms are words that share at least one lexical concept. Moreover, we assume that these lexical datasets conform to OntoLex-Lemon, which is gaining momentum for the representation of lexical resources. The score of a paring is computed through an empirical formula combining the metrics of the lexicalization sets and these optional resources.

$$score = \left(\prod_{x \in \{source, target\}} percentage_{lexicalizationSet_x} \right) \left(1 - \alpha e^{-\beta \max(expr_{source}, expr_{target})}\right)$$

where

$$expr_x = \frac{avgNumOfLexicalizations_{lexicalizationSet_x}}{percentage_{lexicalizationSet_x}} (1 + lrContribution)$$

and

$$lrContribution = \frac{conceptualizations_{conceptualizationSet}}{\max\limits_{x \in \{source, target\}} (lexicalizations_{lexicalizationSet_x})} avgAmbiguity_{conceptualizationSet} \cdot$$

$$\cdot\, avgSynonymy_{conceptualizationSet} \frac{lexicalEntries_{concetualizationSet}}{lexicalEntries_{lexicon}}$$

Looking at the *score*, the first factor means that the *score* increases as the *percentage* (expressed as a number between 0 and 1) of each input dataset covered by the paired lexicalization sets increases. The other factor is a number between 0 and 1 that tends to 1 as the expressivity ($expr_x$) of any of the paired lexicalization sets increases. α and β are parameters that were set during development. Here, expressivity is intended as the mean number of labels for entities that have at least one $\left(\frac{avgNumOfLexicalizations_{lexicalizationSet_x}}{percentage_{lexicalizationSet_x}}\right)$, possibly boosted with a factor depending on the synonymizer, $lrContribution$. If no *synonymizer* is suggested, then $lrContribution = 0$. It is important to observe that the addition of a synonymizer can only increase the expressivity of each hand of a pairing. The first factor in the definition of $lrContribution$ correlates to the chance that a given sense of either lexicalization set matches one in the conceptualization set. The subsequent two factors boost language

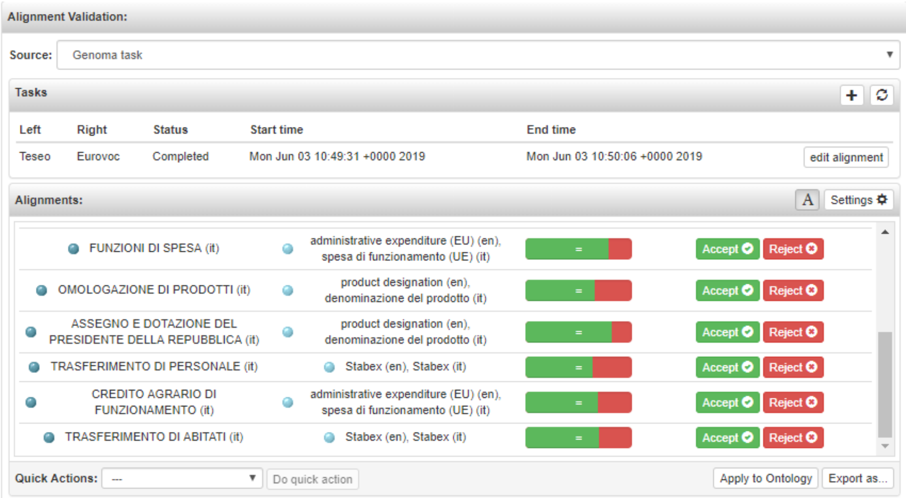

Fig. 3. Alignment validation panel in VocBench 3. It supports the interaction with external ontology matching systems that understand the task report generated by our framework.

resources with high ambiguity and synonymy, since they are more representative of the actual linguistic problems (i.e. it is better to be aware that dog is ambiguous rather than to believe that it has just one sense). The last factor weights how much the conceptualization set covers the underlying lexicon.

4 Use Case: VocBench 3

We showed the usefulness of our framework by integrating it into VocBench 3. With respect to **Fig. 1**, VocBench 3 implements the *metadata registry* and consumes the *task report* or delivers it to a matching system. It can obtain metadata as follows:

1. Manual addition of a (remote) dataset description
2. Discovery of a (remote) dataset description using the VoID backlink mechanism
3. Limited profiling of (remote) dataset SPARQL endpoints
4. Harvesting of (automatically generated) metadata about local projects

The fourth strategy supports the alignment of two local projects. In this case, VocBench 3 delivers the *task report* to an *ontology matching system*, which can obtain the actual data (input ontologies and support datasets) by means of the SPARQL endpoints contained in the report itself. Potentially requiring a non-trivial amount of time (depending on the input size), the generation of the alignment is treated as an asynchronous task, which does not block the user interface, so that the user requesting the alignment can perform other actions in the meanwhile (even logging out from the system and connecting to it later). When the alignment is ready, it is possible to open it inside a validation panel (see **Fig. 3**). This panel enables the interaction with external ontology matching systems, orchestrating the matching process using the task report produced by our framework. We collaborated with the team that worked on GENOMA [33], so that their system could understand our *task report*. Once the process is successfully completed, the panel lists the generated correspondences, showing the confidence on each of them through a progress bar (and, optionally, as a number). The correspondences can be validated individually: in case of acceptance, the mapping relation can be refined (e.g. equivalence between SKOS concepts can be refined as `skos:exactMatch` or `skos:closeMatch`). Bulk validation is also supported: e.g., it is possible to accept all correspondences scored above a user-supplied threshold. The alignment can be exported in the Alignment API format [34] or can be applied to the source dataset (i.e. adding the triples for the accepted correspondences).

5 Discussion

The use case presented in Section 4 validates our framework, showing that it can be instantiated in a real application. Indeed, this application can be regarded as another contribution on its own, since it is freely available and not restricted to an experiment.

The separation between the *metadata registry* and the *orchestration framework* isolates the latter from the generation of metadata, enabling several, interchangeable strategies. Most strategies might rely on facilities provided by the framework or ancillary libraries, for example, to profile datasets automatically. In the proposed use case, however, the harvesting of metadata about the datasets managed by the editing tool integrating the framework surely goes beyond what is offered by the framework itself.

Many works on ontology matching rely on predefined support resources (e.g. language-specific lexicons, domain terminologies, etc.), sometimes replaceable by a power-user. Conversely, our framework suggests the use of a (language) resource, if the *metadata registry* includes one that is compatible with the matching task at hand.

Another important observation is that the *lexicalization sets* in the *pairings* contained in the *task report* (see Section 3) are really *support datasets*, just like the language resources mentioned before. This means that the lexicalizations used by the ontology matcher may not come from either input dataset, but rather obtained from a third-party one. Indeed, this is a distinctive feature among ontology matching systems, which particularly fits the vision behind OntoLex-Lemon, which foresaw the possibility for the lexicalization of some dataset to be published independently from the dataset itself.

Adopting a data warehouse approach, our framework assumes that metadata about every potentially useful dataset has been collected beforehand. However, our methodology is compatible with the dynamic discovery of useful datasets. Indeed, our reliance on explicit metadata makes querying easy and efficient. Furthermore, since we use standard metadata vocabularies, we might use almost unmodified metadata published by existing repositories or by publishers alongside their datasets. Our use case has already achieved a similar goal through the discovery of online VoID descriptors.

In the related work, background ontologies were selected by optimizing some metric (similarity, effectiveness, mapping gain) defined in relation to one or both input datasets. Our methodology for selecting language resources is different and based on a two-step process. The first step is to identify language resources that are compatible with the given matching task, comparing their metadata (mostly the natural language) to the ones of the input datasets. However, the subsequent ranking is determined by intrinsic metrics on language resources, preferring essentially the large ones. This selection criterion clearly presupposes that candidate resources are homogenous, so that the chances that relevant information can be found increase with the size of the resource. For example, this assumption holds between general vocabularies for some natural language. Conversely, our approach can't accurately discriminate domain-specific terminologies. An interesting future development is whether we can address this issue with additional metadata (e.g. telling the domain of a resource) or whether we can incorporate some lexical overlap metrics. Indeed, it is probable that the latter remains confined to purposed vocabularies oriented at supporting the orchestration task, as this sort of metadata is unlikely to be published together with their described datasets, involving two datasets, and being very specific for a given matching task.

6 Conclusions

We discussed an orchestration framework addressing the need for robustness and adaptation in ontology matching. The framework analyzes metadata about the input ontologies and external resources, compiling a report that summarizes the matching scenario at hand. This report can thus guide the selection of the appropriate matching strategy, exploiting information contained in the input ontologies and in external resources. We validated our approach applying this framework to the knowledge editing platform VocBench 3, in order to satisfy its requirements on sematic coordination.

Acknowledgements. This work has been drafted under the 2016.16 action of the ISA2 Programme (`https://ec.europa.eu/isa2`).

References

1. Berners-Lee, T., Hendler, J.A., Lassila, O.: The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American 284(5), 34-43 (2001)
2. Shadbolt, N., Berners-Lee, T., Hall, W.: The Semantic Web Revisited. IEEE Intelligent Systems 21(3), 96-101 (2006)
3. Berners-Lee, T.: Linked Data. In: Design Issues. (Accessed 2006) Available at: https://www.w3.org/DesignIssues/LinkedData.html
4. Wiederhold, G.: Interoperation, Mediation and Ontologies. In : Proceedings International Symposium on Fifth Generation Computer Systems (FGCS94), Workshop on Heterogeneous Cooperative Knowledge Bases, Tokio, Japan, pp.33-48 (1994)
5. Madhavan, J., Jeffery, S.R., Cohen, S., Dong, X.(., Ko, D., Yu, C., Halevy, A.: Web-scale data integration: You can only afford to pay as you go. In : Proceedings of CIDR 2007, pp.342-350 (2007)

6. Halevy, A., Franklin, M., Maier, D.: Principles of Dataspace Systems. In : Proceedings of the twenty-fifth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems , pp.1-9 (2006)
7. Heath, T., Bizer, C.: Linked data: Evolving the web into a global data space. Synthesis Lectures on the Semantic Web: Theory and Technology 1(1), 1-136 (2011)
8. Euzenat, J., Shvaiko, P.: Ontology Matching 2nd edn. Springer-Verlag Berlin Heidelberg (2013)
9. Euzenat, J., Shvaiko, P.: Classifications of ontology matching techniques. In : Ontology Matching. Springer, Berlin, Heidelberg (2007), pp.61-72
10. Shvaiko, P., Euzenat, J.: Ontology Matching: State of the Art and Future Challenges. IEEE Transactions on Knowledge and Data Engineering 25(1), 158-176 (January 2013)
11. Fiorelli, M., Pazienza, M.T., Stellato, A.: A Meta-data Driven Platform for Semi-automatic Configuration of Ontology Mediators. In Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S., eds. : Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland (May 2014)
12. Stellato, Armando: A Language-Aware Web will Give Us a Bigger and Better Semantic Web. In : Proceedings of the 4th Workshop on the Multilingual Semantic Web, co-located with the 12th Extended Semantic Web Conference - ESWC2015, Portoroz, Slovenia (2015)
13. Stellato, A., Turbati, A., Fiorelli, M., Lorenzetti, T., Costetchi, E., Laaboudi, C., Van Gemert, W., Keizer, J.: Towards VocBench 3: Pushing Collaborative Development of Thesauri and Ontologies Further Beyond. In Mayr, P., Tudhope, D., Golub, K., Wartena, C., De Luca, E.W., eds. : 17th European Networked Knowledge Organization Systems (NKOS) Workshop. Thessaloniki, Greece, September 21st, 2017, pp.39-52 (2017)
14. Li, J., Tang, J., Li, Y., Luo, Q.: RiMOM: A Dynamic Multistrategy Ontology Alignment Framework. IEEE Transactions on Knowledge and Data Engineering 21(8), 1218-1232 (2009)
15. Mochol, M., Jentzsch, A.: Towards a Rule-Based Matcher Selection. In Gangemi, A., Euzenat, J., eds. : Knowledge Engineering: Practice and Patterns. EKAW 2008 (Lecture Notes in Computer Science) 5268. Springer, Berlin, Heidelberg (2008), pp.109-119
16. Cruz, I.F., Fabiani, A., Caimi, F., Stroe, C., Palmonari, M.: Automatic Configuration Selection Using Ontology Matching Task Profiling. In Simperl, E., Cimiano, P., Polleres, A., Corcho, O., Presutti, V., eds. : The Semantic Web: Research and Applications. ESWC 2012 (Lecture Notes in Computer Science) 7295. Springer, Berlin, Heidelberg (2012), pp.179-194
17. Tartir, S., Arpinar, I.B.: Ontology Evaluation and Ranking using OntoQA. In : International Conference on Semantic Computing (ICSC 2007). IEEE, pp.185-192
18. Faria, D., Pesquita, C., Santos, E., Cruz, I.F., Couto, F.M.: Automatic Background Knowledge Selection for Matching Biomedical Ontologies. PLOS ONE 9(11), 1-9 (2014)
19. Mascardi, V., Locoro, A., Rosso, P.: Automatic Ontology Matching via Upper Ontologies: A Systematic Evaluation. IEEE Transactions on Knowledge and Data Engineering 22(5), 609-623 (2010)
20. Jain, P., Hitzler, P., Sheth, A.P., Verma, K., Yeh, P.Z.: Ontology Alignment for Linked Open Data. In Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhan, L., Pan, J.Z.,

Horrocks, I., Glimm, B., eds. : The Semantic Web – ISWC 2010. ISWC 2010 (Lecture Notes in Computer Science) 6496. Springer, Berlin, Heidelberg (2010), pp.402-417

21. Hertling, S., Paulheim, H.: WikiMatch: Using Wikipedia for Ontology Matching. In : Proceedings of the 7th International Conference on Ontology Matching - Volume 946, pp.37-48 (2012)
22. Fellbaum, C.: WordNet: An Electronic Lexical Database. WordNet Pointers, MIT Press, Cambridge, MA (1998)
23. Bond, F., Paik, K.: A survey of wordnets and their licenses. In : Proceedings of the 6th Global WordNet Conference (GWC 2012). Matsue, Japan, January, 9-13, 2012 , pp.64-71
24. Chiarcos, C., Nordhoff, S., Hellmann, S., eds.: Linked Data in Linguistics. Springer (2012)
25. McCrae, J.P., Bosque-Gil, J., Gracia, J., Buitelaar, P., Cimiano, P.: The OntoLex-Lemon Model: Development and Applications. In Kosem, I., Tiberius, C., Jakubíček, M., Kallas, J., Krek, S., Baisa, V., eds. : Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference., pp.587-597 (2017)
26. Sabou, M., d'Aquin, M., Motta, E.: Exploring the Semantic Web as Background Knowledge for Ontology Matching. In Spaccapietra, S., Pan, J.Z., Thiran, P., Halpin, T., Staab, S., Svatek, V., Shvaik, P., Roddick, J., eds. : Journal on Data Semantics XI (Lecture Notes in Computer Science) 5383. Springer, Berlin, Heidelberg (2008), pp.156-190
27. Quix, C., Roy, P., Kensche, D.: Automatic Selection of Background Knowledge for Ontology Matching. In : Proceedings of the International Workshop on Semantic Web Information Management. ACM, New York, , USA (2011), pp.5:1--5:7
28. Hartung, M., Groß, A., Kirsten, T., Rahm, E.: Effective mapping composition for biomedical ontologies. In : Workshop on Semantic Interoperability in Medical Informatics (SIMI)
29. World Wide Web Consortium (W3C): Data Catalog Vocabulary (DCAT). In: World Wide Web Consortium (W3C). (Accessed January 16, 2014) Available at: http://www.w3.org/TR/vocab-dcat/
30. Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J.: Describing Linked Datasets with the VoID Vocabulary (W3C Interest Group Note). In: World Wide Web Consortium (W3C). (Accessed March 3, 2011) Available at: http://www.w3.org/TR/void/
31. Fiorelli, M., Stellato, A., Mccrae, J.P., Cimiano, P., Pazienza, M.T.: LIME: the Metadata Module for OntoLex. In Gandon, F., Sabou, M., Sack, H., d'Amato, C., Cudré-Mauroux, P., Zimmermann, A., eds. : The Semantic Web. Latest Advances and New Domains (Lecture Notes in Computer Science). Proceedings of the 12th Extended Semantic Web Conference (ESWC 2015), Portoroz, Slovenia, May 31 - 4 June, 2015. 9088. Springer International Publishing (2015), pp.321-336
32. DCMI Usage Board: DCMI Metadata Terms. In: Dublin Core Metadata Initiative (DCMI). (Accessed June 14, 2012) Available at: http://dublincore.org/documents/dcmi-terms/
33. Enea, R., Pazienza, M.T., Turbati, A.: GENOMA: GENeric Ontology Matching Architecture. In Gavanelli, M., Lamma, E., Riguzzi, F., eds. : Lecture Notes in Computer Science. Proceedings of the IA*IA 2015 Conference. 9336. Springer International Publishing (2015), pp.303-315
34. David, J., Euzenat, J., Scharffe, F., Trojahn dos Santos, C.: The Alignment API 4.0. Semantic Web Journal 2(1), 3-10 (2011)