

Towards the Assessment of Gold-Standard Alignments between Legal Thesauri

Armando STELLATO^a, Andrea TURBATI^a,
Manuel FIORELLI^a, Tiziano LORENZETTI^a,
Peter SCHMITZ^b, Enrico FRANCESCONI^{b,c},
Najeh HAJLAOUI^b, Brahim BATOUCHE^b,

^a*University of Rome Tor Vergata, Department of Enterprise Engineering, Rome, Italy*

^b*Publications Office of the European Union, Luxembourg*

^c*Institute of Legal Information Theory and Techniques of CNR (ITTIG-CNR), Italy*

Abstract. In this paper we report on the experience gathered in producing two gold-standard alignment datasets between the European Union thesaurus EuroVoc and two other notable resources adopted in legal environments: the thesaurus of the Italian Senate TESEO and the IATE European terminological resource. The realization of these two resources has been performed in the context of the PMKI project, an European Commission action aiming at creating a Public Multilingual Knowledge management Infrastructure to support e-commerce solutions in a multilingual environment. As of the numerous lexical and terminological resources involved in this project, ontology and thesaurus alignment and, as a consequence, the evaluation of automatically generated alignments, play a pivotal role for the success of the project.

Keywords. legal thesauri, gold-standard alignments, semantic interoperability.

1. Introduction

The Semantic Web [1] has offered a powerful stack of standard languages and protocols for modeling, sharing and reuse of knowledge on the Web. However, the advantages brought by metadata standards and infrastructures for shareability of information cannot avoid (but can support) the reconciliation work needed on the information content. Different, domain-overlapping, redundant to different extents, ontologies, thesauri, vocabularies, datasets etc. are expected to emerge on the Web in order to satisfy specific needs and exigencies and, at different points in time, are expected as well to be – somehow – “reconciled” out of their heterogeneities, for interoperability’s sake.

This “reconciliation” takes the form of alignments, that is, sets of correspondences between the different entities that populate lexical and semantic resources on the Web. The expression “ontology alignment” is often used in a broader sense than the one that the first word of the term would suggest. “Ontology” is in this case a synecdoche for ontologies, thesauri, lexicons and any sorts of knowledge resources modeled according to core knowledge modeling languages for the Semantic Web, which shared and made available on the Web itself. The expression ontology alignment 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 matching* (as the produced alignments are also referred to as *matches*). In the RDF jargon, and following the terminology adopted in the VoID metadata vocabulary [2], a set of alignments is also called a *Linkset*.

The production of alignments is an intensive and error prone task; for this reason, several approaches for automating the task have been devised [3] since the early years of the Semantic Web. An Ontology Alignment Evaluation Initiative¹ [4] is also held every year since 2004 with the intent of evaluating available tools against benchmarks consisting in well-assessed alignments between notable semantic resources (mostly ontologies, and some thesauri). The task are also divided into T-Box/Schema matching, dealing – as the name suggests – with the alignment of ontology vocabularies, i.e. word models, including class and properties, and instance matching, involving the creation of links between domain objects represented in different datasets.

In this work, we report on the experience gathered in producing two gold-standard alignment datasets between EuroVoc² – EU’s multilingual thesaurus covering the activities of the European Union – and two other notable resources adopted in legal environments: the thesaurus of the Italian Senate TESEO³ (TEsauro SE nato per l’Organizzazione dei documenti parlamentari) and IATE⁴ (InterActive Terminology for Europe), the EU’s multilingual terms base. The realization of these resources has been performed in the context of the PMKI⁵ project, a European Commission action launched to promote the Digital Single Market in the European Union (EU) by creating a Public Multilingual Knowledge management Infrastructure. PMKI aims at sharing maintainable and sustainable public multilingual knowledge systems making them interoperable in order to be reused by different services that support the creation of a Multilingual Digital Single Market; proposing a common data model for the representation of multilingual terminologies based on existing recognised standards and developing semantic capabilities to enable the automated alignment of data from different sources in order to further increase their interoperability. Such infrastructure will consist in a set of tools able to create interoperability between multilingual classification systems (like thesauri) and other language resources, so that they can be easily accessible through a Web dissemination platform and reusable by small and medium-sized enterprises (SMEs), as well as by public administrations. In this paper, the standards used to represent aligned resources and a methodology for the production of gold-standard alignments between the language resources mentioned above are presented.

2. Motivation

As far as producing mappings is an intensive and error prone task, it is also a very difficult one which cannot be easily automated. If alignments could be logically inferred, there would be processes generating them automatically with guaranteed precision. In fact, the necessity for alignments comes into play when there is no a-priori semantic agreement between two resources. The identification of proper alignments is thus a discovery process (some approaches [5] treat it as an Information Retrieval problem indeed) based on the analysis of the content. This content is mostly, on a first step, pure textual information, in order to gather the first anchors between the two resources, as common language expressions are indeed the last stand for creating hypothesis about alignments [6, 7], which is later followed by inference mechanisms based on the created hypotheses. This kind of analysis obviously benefits from background knowledge –

¹ <http://oaei.ontologymatching.org/>

² <http://eurovoc.europa.eu/>

³ http://www.senato.it/3235?testo_generico=745

⁴ <http://iate.europa.eu>

⁵ <https://ec.europa.eu/isa2/actions/overcoming-language-barriers>

owned by the mapping agent – that goes beyond the one available in both datasets and from the capability to process natural language content.

As of the state of the art, the importance and main contribution of automatic alignment consists in providing in a short time a seed that would be hard to produce manually, in order to support the subsequent refinement performed by domain experts. Even though meant to be completed by means of human supervision, the quality of the automatically produced alignments is important: besides the trivial aspect that a good automatic process shortens the following validation, it is also true that, especially in alignments of large corpora, an unprecise seed will in any case provide a bias on the validator: a single “almost perfectly matched” concept would drift the attention of the validator away from some best representative that is still laying somewhere in the target dataset. Even more important, an incomplete mapping between two datasets (i.e. missing some important matches) can be validated in the precision of the produced matches, but missing matches would be no less hard to be discovered than in a point-blank situation. For this reason, even though precision is important, recall (as in every Information Retrieval process) is fundamental, as users may easily discard wrong results but their trust in the system depends on the feeling that no important information is left unretrieved.

A proper evaluation of alignment processes demands for gold-standard alignments, guaranteeing maximum precision (quality-produced matches) and recall (completeness of the alignment). Another possible exploitation of such resources is their use as training sets for mapping algorithms based on learning techniques. Also, to avoid the negative effects of in-vitro experimentation, alignments between real datasets should be produced.

With these requirements in mind and with the legal domain as an important topic for the project that has been the context for this work; we chose to create two alignments for subsets of three important resources of the European Union: EuroVoc, TESEO and IATE. In order to create reliable gold-standard alignments, we selected a range of circa 300 concepts per each resource, so that the task of manual alignment could be performed entirely by humans without any bias nor errors induced by machine processing.

3. Standards for the Representation of Semantic Resources and their Alignments

PMKI considers resources that are modeled and published according to Semantic Web standards, thus adopting RDF (and the various vocabularies for creating ontologies, thesauri, etc.) and Linked Open Data best practices for publication. Different actions of the project also aim at providing services (transformation, hosting etc.) for bringing non-RDF resources into the PMKI platform, thus leveling all of them to the desired standards.

Within this context, the kind of resources that have been considered (at least for a first bootstrap of the project) are multilingual language resources, including thesauri and lexicons. W3C offers two core modeling vocabularies for dealing with these kind of resources: for thesauri there is the W3C recommendation SKOS [8] (and its extension SKOS-XL [9], for modeling reified labels) while a community group under the W3C umbrella has recently developed the OntoLex-Lemon (<https://www.w3.org/2016/05/ontolex/>) suite of vocabularies for modeling lexicons.

SKOS is a vocabulary defined in OWL introducing a set of terms characterizing thesauri: skos:Concepts are described as representing ideas or notions, units of thought. However, as the reference manual clarifies, what constitutes a unit of thought is subjective, and this definition is meant to be suggestive, rather than restrictive. In a few words, while OWL classes provided a formal classification mechanism for describing entities, adopting a first-level logic perspective made of objects and predicated over them,

```

<map>
  <Cell>
    <entity1 rdf:resource="http://iasted#Student registration fee"/>
    <entity2 rdf:resource="http://sigkdd#Registration Student"/>
    <measure rdf:datatype="xsd:float">0.7555555555</measure>
    <relation>=</relation>
  </Cell>
</map>

```

Figure 1. an excerpt of the RDF/XML code for an alignment expressed for the Alignment API

SKOS concepts are meant to provide a simplified view, with shallow semantics, where entities can be generally described with values. Concepts are, in this sense, all objects with respect to an OWL perspective (they are indeed instances of the `skos:Concept` class), whether they are merely conceptual entities or concrete objects of a domain. A more detailed description of differences between SKOS and OWL (and what to model as a thesaurus or as an ontology) is provided in [10].

Another important aspect of SKOS lies in its terminological properties. Preferred labels (`skos:prefLabel`) describe the lexical expressions that are most recurring for referring to a certain concept. Alternative labels (`skos:altLabel`) provide alternative expressions, less common than the preferred ones. Hidden labels (`skos:hiddenLabel`) cover common misspells and other expressions that should not be explicitly shown while being at the same adopted for technical aspects such as indexing and retrieval. SKOS also provides several properties for expressing notes about the elements of a thesaurus. These notes can be in-domain or extra-domain, representing respectively the description (e.g. `skos:definition`) of the object of the domain represented by the SKOS concept or notes taken by the editors (e.g. `skos:editorialNote`) that concern the concept as an element of the thesaurus construct and not as its denoted domain object.

A third important aspect of SKOS lies in its mapping vocabulary: a set of properties describing relationships of similarity or, at least, connection, between concepts from two different thesauri. `skos:exactMatch` provides a shallow semantics (i.e. no logical entailment) approach to identity, expressing the fact that two SKOS concepts denote the same object of the world. `skos:closeMatch` implies a strong closeness between two SKOS concepts even though they are not considered to be exactly the same thing. `skos:broad/narrowMatch` express a “more specific”/“more generic” relationship between two concepts from different thesauri while `skos:related` connects two concepts denoting two things that, while being clearly different, are somehow related to each other.

The vocabulary that has been used for the alignments is the one created and adopted for the Alignment API [11], and it is a base for the more elaborated EDOAL mapping language. In the Alignment API ontology, mappings are reified into resources (see figure 1), in order to be decorated with further metadata: for instance, besides the type of relation (expressed by the property `:relation`) the vocabulary foresees a property `:measure`, representing the confidence of the mapping. However, other additional information can then be added as well (e.g. `rdfs:comments`). The vocabulary is also extensible, as new relations can be added in custom extensions of the vocabulary.

For instance, the basic vocabulary only foresees an equivalence relation, probably due to the fact that it is not expected for a machine to be so precise in distinguishing among subtle variations of a match, thus postponing this specification to later refinement by means of human supervision. However, in the context of the maintenance of the EuroVoc thesaurus, the Publications Office of the European Union adopted an extension

of the alignment ontology where a property semantically close to `skos:closeMatch` has been added to the range of available ones, thus making the porting to SKOS mapping relations a simple 1-1 transformation.

4. Alignment Completeness and Minimality

As outlined in section 2, an important aspect when considering an alignment effort is, besides the (trivially clear, yet not easy to achieve) precision of the outcome, its completeness. Most alignment vocabularies consider at least three kind of relations: equivalence and the two inverse more specific/more general relations, plus usually a relation for non-matchable entities. In an alignment effort between two datasets D_A and D_B it would be theoretically possible to assert the cartesian product of $|D_A| \times |D_B|$ possible alignments. However, there are some aspects that, intuitively, should be taken into consideration:

- *Direction of the alignment.* An alignment has a direction, from a “source” dataset to a “target” dataset
- *“Nobleness” of the relationship.* If a concept $c_A \in D_A$ is aligned with a relation of a certain “nobleness” to a concept $c_B \in D_B$, then there is no need to establish other “less noble” relationships between c_A and any other concept in D_B . “nobleness” is a total order over the available relations in a given mapping vocabulary. E.g. in the SKOS mapping properties, the following order would hold: `exactMatch`, `closeMatch`, `narrowMatch`, `broadMatch`, `related`. We consider then two relevant sets of relations: *equivalence* relations and *tolerance* relations. *Equivalence* relations are reflexive, symmetric and transitive. *Tolerance* relations are reflexive, symmetric but not necessarily transitive. `skos:exactMatch` match is an example of a equivalence relation, while `skos:closeMatch` is a tolerance relation. Notably, in SKOS neither `:exactMatch` nor `:closeMatch` have been defined as reflexive. The absence of reflexivity is due to historical reasons (the SKOS model is older than OWL2, where reflexivity has been introduced).
- Note that more relationships of the same type (if allowed by the relation itself) would be admitted. For instance, a concept $c_A \in D_A$ could be declared to be in a `skos:narrowMatch` relation with two concepts $c_{1B}, c_{2B} \in D_B$. The rationale for this double assignment is that being no better (more noble) match for c_A in D_B , then as many concepts from D_B , which (together) provide a best representation for c_A , are linked to it.

We can thus provide the following definitions:

Definition 1 (Linkability and Reachability). *If a concept $c_A \in D_A$ is aligned with a tolerance relation to a concept $c_B \in D_B$, (expressed as $l(c_A, r, c_B)$) then $\forall c_i \in D_A : c_i < c_A$ it holds that $c_i \in \text{reach}(c_B)$.*

In other words, in absence of an explicit relationship for c_i , c_i can still be considered linked to c_B (thus being able to “reach” c_B) through a traversal of the hierarchy in D_A until concept c_A that is perfectly matching c_B is met.

Definition 2 (link to a Dataset). *Given two datasets D_A and D_B linked through linkset L , a set of alignment relations R , a concept $c_A \in D_A$ is said to be linked to D_B iff at least one of the two following statements is true:*

- 1. $\exists l(c_A, r, c_B) : c_A \in D_A, c_B \in D_B, r \in R$

- 2. $\exists c_B \in D_B: c_A \in \text{reach}(c_B)$

In particular, given R^T the set of tolerance relations:

- c_A is said to be perfectly linked to D_B iff c_A is linked to D_B and statement 1 is false if it is not possible to find an explicit link $l(c_A, r, c_B)$ ⁶
- c_A is said to be redundantly linked to D_B iff both statements 1 and 2 are true and the asserted alignment is not a match based on a tolerance relation

Definition 3 (properly unlinked concepts). Given two datasets D_A and D_B linked through linkset L , a concept $c_A \in D_A$ is said to be properly unlinked from D_B if there is no possible assignment of alignments $\in L$ so that c_A can be linked to D_B .

So, the notion of *perfectly linked* concept guarantees that no concept is simply linked as the consequence of a tree-traversal when it could have been better linked through an equivalence link. This is necessary for defining the completeness of an alignment. Conversely, the notion of *redundantly linked* concept supports the definition of minimal and complete alignment.

Definition 4 (Mapping completeness). A linkset L between two datasets D_A and D_B is said to be complete with respect to dataset D_A iff each concept in D_A is perfectly linked to or is properly unlinked from D_B .

Definition 5 (Minimal complete mapping). A linkset L between two datasets D_A and D_B is said to be complete and minimal with respect to dataset D_A iff each concept in D_A is either perfectly and non-redundantly linked to D_B or is properly unlinked from D_B .

5. The Procedure for Creating the Datasets Corpus

As the objective for the golden-standard corpus is to reach a number of 300 concepts per dataset, a procedure for selecting the 300 concepts without de facto performing a mapping (which we wanted to perform on a second phase, after the cut for the corpus had been done) had to be established.

1. Having the two concept trees at hand or simply, by knowing the domains covered by the two thesauri, a domain is chosen so that the selected concepts will be generally expected to be connected with it.
2. In order to generate cuts of the thesauri that can be still considered as real resources, macro portions of the hierarchy should be selected, without performing too much pruning. The user performing the cut should thus select a concept relevant for the domain (and general enough to subsume other relevant concepts) and perform a bird's eye inspection (again, the real alignment has to be performed later) to verify that roughly a non-trivial number of concepts is relevant to the domain (and thus possibly linkable to concepts in the cut of the other thesaurus). If the number of concepts in the retrieved branch is too high, some cuts will be performed on it. Conversely, if it is too low, a new branch can be selected – with the same approach – from the original thesaurus in order to increment the number of chosen units.

⁶ To understand this apparent tautology, we have to distinguish links which are asserted in a concrete linkset, and links that could possibly be found. So, a link is perfect either if it is explicitly asserted (and sound) or if it is expressed by a *reach*, having established that there is no possible explicit alignment. This *having established* is a truth value that is associated to the presence of an *oracle*, representing all that is known about the relations between two given datasets. The objective of realizing such an oracle is thus the scope of this work, and these properties describe how this oracle can be used to describe alignments computed by machines.

3. Once the two cuts have been produced, a further bird's eye over their content should confirm if there is an acceptable overlap between the two selected portions of thesauri.

An important aspect to clarify, when working with alignment corpora based on real thesauri cuts, is that a perfectly sound linkset which is also minimal and complete with respect to the aligned excerpts, cannot be considered as a subset of a linkset (with same properties of soundness and completeness) between the two original thesauri. The act of cutting both source and target thesauri implies that the search for alignments will incur in local minimums, dictated by the restricted search space.

6. The experience in aligning EuroVoc and TESEO

The alignment work has been organized as follows:

- The pivot language for the alignment has been the Italian language, as TESEO is only expressed in this idiom.
- A domain expert has manually developed a first version of the linkset by analyzing all concept combinations from the two thesauri. The search space is not trivial ($300 \times 300 = 90000$ combinations). For this reason, a quick a-priori identification of palely exact-matching concepts has been performed in order to reduce its dimension.
- A Semantic Web expert has revised the work performed by the domain expert, by both validating/rejecting/modifying the alignments produced in the first step and by performing a further search (notably reduced in size thanks to the first step and the following validation), discovering new alignments.

The following table provides some figures for the outcome of the alignment process:

Table 1. Results of the alignment	
alignments produced by the domain expert	74
alignments discarded by the SW expert	18
alignments modified (type of relation) by the SW expert	3
alignments added by the SW expert	45
total amount of alignments at the end of SW expert's review	101

As it is possible to observe, the Semantic Web expert discarded a non-trivial fraction (~25%) of the original alignments, while being able to discover many more that had been overlooked by the domain expert (though he later re-confirmed these additions).

Had we more resources, a further improvement we would have added to the procedure would have consisted in doubling the first step with two domain experts. This would have allowed us to report on standard measures, such as inter-annotator agreement [12], which are a valid instrument for evaluating the easiness of the task (e.g. if humans agree on a given percentage of the results after carefully performing the job, we cannot expect a machine to perform better than them). In retrospect, considering the numerous changes and improvements that the Semantic Web expert brought during the second step, the inter-annotator agreement would have been extremely low.

This probably reflects the inherent difficulty for humans in performing the alignment task (especially if their familiarity with the domain is not matched by an equal proficiency with thesauri and computer systems) more than the specific case study. In

this case, the inter-annotator agreement could probably have not been considered a reliable upper bound for machine performance.

6.1. Report on observed phenomena

Besides the direct outcome of producing the linkset resource, aligning two thesauri is also an occasion to delve into the details of the examined resources and to discover many inconsistencies and issues (because of the deep attention in identifying the nature and identity of concepts) that would have been overlooked otherwise.

The experience in linking the two (cut) thesauri EuroVoc and TESEO has presented a series of results that are indeed not surprising, if some background knowledge about the origins, invested human resources and profile of the two datasets is known a priori. We list here a few considerations emerged during the alignment work:

Rigor. Compared to TESEO, the EuroVoc thesaurus is characterized by a stricter rigor and better precision in the representation and organization of concepts. Vice versa, despite multilinguality is one of its stronger characteristics, EuroVoc loses in precision when we consider labels for languages that are not English or French (respectively, the de facto lingua franca spoken in EU and the first language used in the offices of the European Commission in Luxembourg). This is probably due to the natural interpretation gap generated by different users dealing with the conceptual content and with its translation in different languages (whereas, in a monolingual thesaurus such as TESEO, the two roles, at least for the preferred labels selected at concept creation, converge into the same user). For instance, the domain expert initially expressed a match between “aviazione militare” from EuroVoc and “aeronautica militare” from TESEO. However, despite the two terms show an apparently almost-synonymic expression in Italian, “aviazione militare” in EuroVoc revealed to be a wrong translation for “military aircraft”, whereas the best representative in EuroVoc should have been “forze aeree” (“air force”).

Different practices. While SKOS is already a very permissive standard, leaving much liberty in terms of modeling choices (which is also a limit in terms of universal understandability and thus shareability of content), a further degree of freedom is given by different practices at content level that characterize thesauri since the dawn of their time. For instance, in TESEO it is possible to find many cases in which, rather than modeling a domain concept and its more specific interpretations as different formal concepts in SKOS, they are all converged into a single SKOS concept where the preferred label represents the most general interpretation in the domain and alternative labels provide its specializations. In terms of mere representation this approach might be considered a severe mistake. However, if we consider thesauri merely as tools supporting retrieval of annotated resources (usually documents), this conglomeration seems to be justifiable in all cases in which there is no particular interest in exploring the details of the domain concept (thus considering one element as enough), while it is still considered important to broaden the range of its potential lexical expressions. For instance, in TESEO the concept of “conventional weapon” is represented by several labels including “artillery” that is indeed a lexical expressions denoting only one of the possible conventional weapons. In EuroVoc, there seems to be no use (at least, not as broad as in TESEO) of this conceptual collapse, probably due to the larger resources invested in developing the resource, and in the wider scope of its application. TESEO has indeed been explicitly developed for indexing the documents of the Italian Senate, while EuroVoc represents a conceptual hub for all member states, possibly reused in different contexts other than indexing of the documents published by the EU Publications Office.

Indeed, in cases such as this one, it is logically appropriate to consider the concept's extension as the one invoked by its most general lexical representations (e.g. "conventional weapon") and, consequently, to map it as such as, implicitly, a concept extension contains the extensions of its narrower concepts. It is also true that the broader/narrower relation in SKOS is not formally defined as in OWL and this extension inclusion is not guaranteed. On the other side, the described phenomenon seems to consistently appear in the case of really inclusive representation of concepts.

Differences in the Domain. Aligning two thesauri does not require to deal only with different modeling and lexical choices. The domains of the two resources might not be perfectly overlapping, especially with respect to concrete objects and the classifying concepts that include them. For instance, the domain expert simply mapped two concepts, both labeled as "esercito", as an exact match. However, by looking at the structure of EuroVoc and other translations of the term, it appears that EuroVoc's "esercito" was actually a representation of "armed forces", whilst it is "esercito di terra" that most closely represents the "army", that is what truly the "Esercito" is in Italian. Until now, it could seem another mere translation issue. However, still, if "esercito di terra" has to be interpreted in its broader sense of "ground force" then, strictly speaking, different corps may include ground forces (e.g., US' navy seals belong to the navy while being a ground force, as much as Italian "battaglione San Marco" does). Simply, there is a blurred definition for these elements, as each country has its own military organization and it is not possible to make a 1-1 transposition. For this reason, the new alignment replacing the wrong one has in any case been represented as a close match. Analogous issues happened with all job-related concepts related to workers, rights, separation of legal and physical persons and their abilities. Another troubled area lies in government aspects, as the different perspectives given by the governance of the EU and of a specific government of a member state, brought clearly diverging bias affecting even the most general concepts. However, these phenomena mostly pose a threat to the quality of the alignment, offering several pitfalls to the annotators during the alignment process, and do not represent inherent representation issues.

7. A new challenge: EuroVoc and the IATE terminology

The alignment between EuroVoc and IATE is being produced at the time of writing. We mention here the main differences in terms of knowledge representation models being adopted by this third resource, IATE. IATE is a terminology, giving more emphasis on the lexical aspect rather than the conceptual one. Despite that, lexical entries in IATE are still organized around meanings, which are however, not structured into a hierarchy like a traditional thesaurus. IATE is thus being modeled as a OntoLex-lemma lexicon. The lexical entries are described in their details (e.g. term composition) thanks to the above standard, and the meanings are represented as `ontolex:LexicalConcepts`, a dedicated class of OntoLex (subclassing `skos:Concept`) expressly thought for representing units of thought in lexicons. The presence of (lexical) concepts allowed us to keep a "concept to concept" approach to alignment representation. Conversely, the lack of a conceptual hierarchical structure forced us to adopt a different strategy for the pre-selection of a set of concepts for the initial development of the corpus to be aligned. Luckily, IATE includes an explicit concept of domain called "subject field", thus we have selected the field "Civil Law" which includes 270 concepts (so we could take them all), and we could proceed with the already experimented process of branch-selection for EuroVoc.

8. Conclusions

In this paper, we have described our guidelines for the production of golden-standard alignments with a reasonably-limited amount of resources and reported on our application of the guidelines for the development of such alignments. We hope that this work can provide both a theoretical and practical foundation for similar initiatives.

Acknowledgements

This work has been supported by the PMKI project, under the 2016.16 action of the ISA² Programme (<https://ec.europa.eu/isa2/>). ISA² is a EUR 131 million programme of the European Commission, supporting the modernization of public administrations in Europe through the development of eGovernment solutions.

The produced alignments will be published, when ready, on the PMKI page.

References

- [1] T. Berners-Lee, J. A. Hendler and O. Lassila, "The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities," *Scientific American*, vol. 279, no. 5, pp. 34-43, 2001.
- [2] K. Alexander, R. Cyganiak, M. Hausenblas and J. Zhao, "Describing Linked Datasets with the VoID Vocabulary (W3C Interest Group Note)," World Wide Web Consortium, 3 March 2011. [Online]. Available: <http://www.w3.org/TR/void/>. [Accessed 16 May 2012].
- [3] J. Euzenat and P. Shvaiko, *Ontology Matching*, 2nd ed. ed., Springer-Verlag Berlin Heidelberg, 2013.
- [4] Z. Dragisic, K. Eckert, J. Euzenat, D. Faria, A. Ferrara, R. Granada, V. Ivanova, E. Jimenez-Ruiz, A. O. Kempf, P. Lambrix, S. Montanelli, H. Paulheim, D. Ritze, P. Shvaiko, A. Solimando, C. Trojahn, O. Zamazal e B. Cuenca Grau, «Results of the ontology alignment evaluation initiative 2014,» in *9th International Workshop on Ontology Matching, October 20, 2014, Riva del Garda, Trentino, Italy, 2014*.
- [5] P. Schmitz, E. Francesconi, N. Hajlaoui e B. S. A. Batouche, «Semantic Interoperability of Multilingual Language Resources by Automatic Mapping,» in *Electronic Government and the Information Systems Perspective (Lecture Notes in Computer Science)*, vol. 11032, E. Francesconi, A cura di, Springer, Cham, 2018, pp. 153-163.
- [6] M. T. Pazienza, S. Sguera and A. Stellato, "Let's talk about our "being": A linguistic-based ontology framework for coordinating agents," *Applied Ontology, special issue on Formal Ontologies for Communicating Agents*, vol. 2, no. 3-4, pp. 305-332, 26 December 2007.
- [7] M. T. Pazienza, A. Stellato and A. Turbati, "Linguistic Watermark 3.0: an RDF framework and a software library for bridging language and ontologies in the Semantic Web," in *5th Workshop on Semantic Web Applications and Perspectives (SWAP2008), Rome, Italy, December 15-17, 2008, CEUR Workshop Proceedings*, FAO-UN, Rome, Italy, 2008.
- [8] World Wide Web Consortium (W3C), "SKOS Simple Knowledge Organization System Reference," World Wide Web Consortium, 18 August 2009. [Online]. Available: <http://www.w3.org/TR/skos-reference/>. [Accessed 22 March 2011].
- [9] World Wide Web Consortium (W3C), "SKOS Simple Knowledge Organization System eXtension for Labels (SKOS-XL)," World Wide Web Consortium, 18 August 2009. [Online]. Available: <http://www.w3.org/TR/skos-reference/skos-xl.html>. [Accessed 22 March 2011].
- [10] A. Stellato, "Dictionary, Thesaurus or Ontology? Disentangling Our Choices in the Semantic Web Jungle," *Journal of Integrative Agriculture*, vol. 11, no. 5, pp. 710-719, May 2012.
- [11] J. David, J. Euzenat, F. Scharffe and C. Trojahn dos Santos, "The Alignment API 4.0," *Semantic Web Journal*, vol. 2, no. 1, pp. 3-10, 2011.
- [12] R. Artstein, "Inter-annotator Agreement," in *Handbook of Linguistic Annotation*, J. Pustejovsky, Ed., Springer, Dordrecht, 2017.