Okkam Synapsis: a community-driven hub for sharing and reusing mappings across vocabularies

Stefano Bortoli[1], Paolo Bouquet[12], and Barbara Bazzanella[2]

[1] Okkam SRL
Via Segantini 23, I-38121 Trento, Italy
[2] University of Trento - DISI
Via Sommarive, 14 I-38123 Povo di Trento, Italy
bortoli@okkam.it, bouquet@disi.unitn.it, barbara.bazzanella@unitn.it

Abstract. In the past 10-15 years, a large amount of resources have been devoted to develop highly sophisticated and effective tools for automated and semi-automated schema-vocabulary-ontology matching and alignment. However, very little effort has been made to consolidate the outputs, in particular to share the resulting mappings with the community of researchers and practitioners, support a community-driven revision/evaluation of mappings and make them reusable. Yet, mappings are an extremely valuable asset, as they provide an *integration map* for the web of data and the "glue" for the Global Giant Graph envisaged by Tim Berners-Lee. Aiming at kicking-off a positive endeavor, we have developed *Synapsis*, a platform to support a community-driven lifecycle of contextual mappings across ontologies, vocabularies and schemas. Okkam Synapsis offers utilities to load, create, maintain, comment, subscribe, and define levels of agreement over user defined contextual mappings available also through REST services.

Acknowledgement. This work is partially supported by TAG CLOUD (Technologies lead to Adaptability and lifelong enGagement with culture throughout the CLOUD) FP7 EU Funded project, Grant agreement nr: 600924.

1 Introduction and Motivation

In the promising vision of the Semantic Web proposed by Tim Berners-Lee [2], the collaborative and distributed creation of semantically annotated documents would enable software agents to perform time-consuming activities on behalf of human users (see [1]). The community that gathered to corroborate and develop this ambitious vision achieved many relevant results with the definition of important standards such as OWL[19, 18, 14], RDF[15], and the important Linked Data publication principles [3, 4]. The combination of these principles with the more recent open data initiative across many countries is generating a considerable amount of publicly available RDF and OWL data. In recent years, enterprises are attracted by the promise of using such big and rich data to develop new products and services for their customers (e.g. [7]). However, exploiting and mining data rises many challenges including the problems of entity matching, ontology matching, and making the data accessible and usable by non-expert users. In the past ten years many efforts were spent in the definition of sophisticated tools for automated ontology matching. These often provided very effective solutions in narrow domains, but a generic automatic reliable solution to the problem is still an open research problem [20]. Furthermore, in [26] it was recently discussed how often even experts have problems in finding agreement on defined ontology mappings. We argue that this is due to two main problems: 1) the intrinsic complexity and heterogeneity of existing ontologies, and 2) the inconsistency and fuzziness in usage of such ontologies due to contextually interpretable semantics. Namely, concepts and relations expressed in natural language are interpreted outside the original context of definition, and therefore prone to contextual interpretation. In fact, besides the effort of researchers in formal ontology [13, 11, 10] the process of ontology definition is driven by specific domain requirements and often ontology engineering practices are neglected [16].

Under these premises, we decided to take one of the ten challenges of ontology matching described in [23] and confirmed in [20], and propose a novel platform to support a collaborative ontology mappings definition and reuse [27]. The idea to take this challenge is rooted in the pragmatic need of resolving the problem of semantic heterogeneity affecting a knowledge-based solution of the entity matching in the context of the Semantic Web [5]. In particular, in this work we argue that collecting and maintaining ontology mappings as contextual bridge rules [6] to harmonize the semantic of entities' attributes can provide great benefits by enabling the application of knowledge-based solution to an entity matching problem [5]. Therefore, in our attempt to solve the entity matching problem in the linked data, we produced several thousands of mappings from existing ontologies, schemas and vocabularies towards a target ontology we named Identification Ontology[3] ([5] Chap. 5). Often these mappings were produced without considering the original, or intended, semantic of the properties, but rather relying on its actual function looking directly into the data. This approach, besides being practical and concrete, interprets the ontology mappings as contextual analogies as suggested in [21]. Namely, when producing mappings, rather than considering the similarity among original intended functional purpose of the properties (homology), we consider also its real function (analogy) so that the mapping relation holds primarily on the instances level. On the one hand, we are aware that this approach will create mappings that might not be absolutely coherent and correct across several contexts, but as long as they serve the purpose we can live with this limitation. On the other hand, we want to use a first core set of mappings to kick-off a positive endeavor for the definition of a platform to support a community-driven lifecycle of contextual mappings

[3] http://models.okkam.org/identification_ontology.owl

between ontologies, vocabularies and schemas that could serve the definition of new applications exploiting open linked data.

In this work we describe Okkam Synapsis, a web application conceived to support the linked data community in creating, sharing and reusing contextual ontology mappings to support the creation of novel services based on the linked data consumption. Okkam Synapsis offers utilities to load, create, maintain, comment, subscribe, and manage levels of agreement over user defined contextual mappings. Most importantly, endorsing the recommendations described in [26], we support different fine-grained typing models for the definition of the mappings (e.g. OWL and SKOS) and compute level of agreement according to different metrics to support filtering based on them. The mappings produced will be available also through REST services, providing several levels of selection to support diverse and unforeseen application scenarios. The purpose of the application is to enable the users of Okkam Synapsis to collaborate in the definition of mappings, commenting, rating, and subscribing them. Furthermore, we want to allow users to explicitly define the context of use of the defined mappings, so that others can take informed decision about reusing.

The underlying assumptions are:

- real linked data is in general too messy to rely on a unique set of mappings in different contexts of use
- linked data may change in time, therefore contextual mappings must be subject to specific lifecycle
- the number of existing vocabularies is growing, but reuse practices make the manual mapping process feasible (see Linked Open Vocabulary[4])
- perfect agreement about defined mappings is unlikely to happen [26], better let users to select what they need

The reminder of the paper is organized as follows: in Section 2 we overview the related works dealing with crowd-sourcing of ontology mappings, and other community-driven approaches; in Section 3 we describe in detail the platform, discussing functions and services. In Section 4 an overview of the architecture of the application is provided and finally in Section 7 we describe future work and outline some concluding remarks.

2 Related Work

According to the most recent survey we are aware of [20], there are not many tools supporting collaborative creation of ontology mappings. In [27] is described a system for community-driven ontology matching, embedding provenance, freshness and other metadata suitable for the selection of the mappings. Besides the a low-resolution screenshot presented in the paper, the system does not seem to be available anymore. In [17] Noy et al. describe a system for the collection of biomedial ontologies supporting the definition of mappings among them. Having

[4] `http://lov.okfn.org/dataset/lov/`

collected more than 30.000 mappings, the authors propose a systems for filtering and searching mappings. Furthermore, they argue about the concept of *mappings as bridges*, and outline the need of specifying the type of relations (e.g. equivalence). Currently the system is up and running, serving more than 370 biomedial ontologies and several million of concepts. In [22] is described CrowdMap, a solution for ontology matching based on crowd-sourcing. The ontology matching task are decomposed in micro-tasks and submitted to workers of crowd-sourcing platforms such as CrowdFlower and MTurk for manual evaluation. The results obtained were compared with the one of automatic tools showing the feasibility of the process. In [8] the authors discuss about the need of managing and reducing uncertainty related to crowdsourcing of ontology matching tasks, proposing different ways to create micro-tasks suitable to increase possible agreements. In [7] the authors describe Helix as a tool for creating ontology mapping as a pay-as-you-go task while consuming linked data. In [12] the Correndo and Alani describe OntoMediate, a project of the University of Southampton aiming at supporting, among other functions, the creation and sharing of ontology mappings. Unfortunately, the project is over and to the best of our knowledge there is no service available. Another trend in managing collective ontology matching is through gamification. In [16] and [24] are described Guess What?! and SpotTheLink proposing the solution of ontology matching tasks in form of games to give incentives and foster engagement to ease the cognitive effort of users and stimulate the creation mappings and links in the linked data cloud. Noticeably, to the best of our knowledge these systems are not currently available. In this context we do not consider papers presenting automatic solutions to the ontology matching for which we refer to the aforementioned survey [20].

In light of the analysis presented, to the best of our knowledge, the only system available providing the services comparable with the one of Okkam Synapsis is BioPortal [17]. However, given the vertical purpose of BioPortal and the limited collaborative features, we can safely affirm that there is room for a solution such as the one proposed in this paper.

3 User Interface and Features

The current version of Synapsis distinguishes between two kinds of users: administrators and end users. Administrators are users that have unrestricted access to all the user-level functions, including uploading a source ontology, creating new mappings for concepts and properties, deleting existing mappings, setting/changing the status of defined mappings, evaluating existing mappings and reusing/exporting mappings. End users have only access to social functions to express their level of agreement on previously created mappings and reusing them. They can endorse and comment existing mappings, follow mappings they are interested in, rate mappings and export mappings.

Figure 1 shows a snapshot of the User Interface of Okkam Synapsis which presents three main areas: the (target) ontology on the left, the mappings in the central part and the mapping filters on the right.

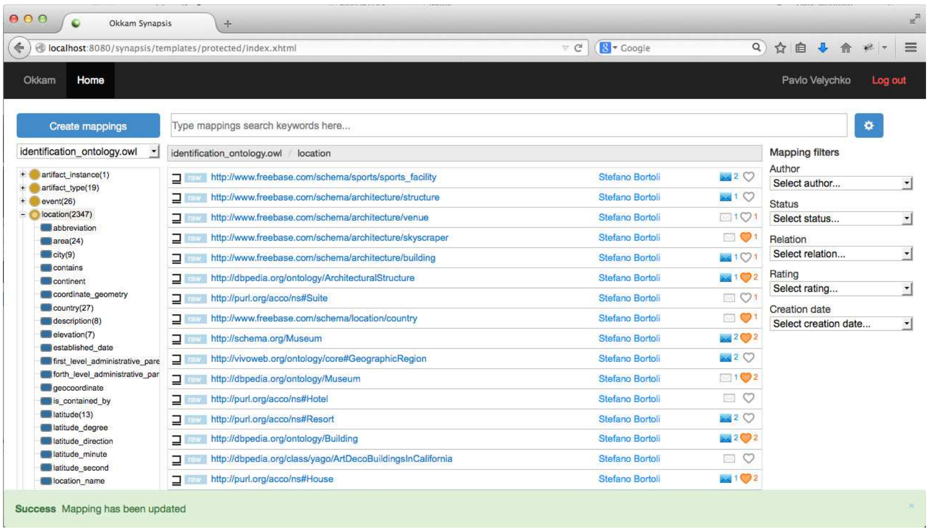

Fig. 1. Synapsis User Interface

After logging in, the user can select one of the ontologies/vocabularies currently present in the platform from the drop-down menu on the top-left corner of the page or import a new ontology selecting the Import function from the Function button. Following [17], we call the selected/uploaded ontology the Target Ontology[5], which is the ontology whose concepts/properties the user wants to map to target concepts/properties. After having selected it, the target ontology is loaded, processed and represented as an indented tree on the left side of the interface. The choice of using an indented tree is based on the study described in [9], where users evaluated this representation model as easier to use and more understandable than alternative models such as the graphs. With the primary objective of enabling users in defining mappings, we decided to flatten the ontology to a list of concepts and present the properties attached to them. In the current version, we rely on a simple RDF processor implemented relying on Apache Jena API[6]. The selection of a node of the source ontology triggers the loading of all the mappings defined for that concept or property in the central part of the window. Each mapping is composed by the following attributes:

- Resource URI: the URI of the resource mapped towards the element of the target ontology.
- Relation Type: the type of relation between the Resource URI and the Target URI. The user can select among an enumeration of relations types including

[5] According to this naming convention, a mapping can be seen as a relationship between two concepts/properties in different ontologies. Each mapping has a source concept/property, a target concept/property, and a mapping relationship.

[6] `https://jena.apache.org/`

OWL meta-relations such as *owl:EquivalentProperty*, *owl:EquivalentClass*, *owl:SubClass*, *owl:SubProperty* and SKOS meta-relations *skos:exact*, *skos:close*, *skos:broader*, *skos:narrower*. In this context, we neglect *skos:related* and *skos:unrelated* because we believe these types of relations are not interesting in our context. In order to help the user in choosing the right type of relation we refer to the guidelines proposed in [26] and still available at [25] as appendix A2.

- Status: a label among *Raw*, *Edited*, *Closed*, *Accepted*, declaring the status of a mapping. These labels are assigned by administrators of Okkam Synapsis keeping into consideration time and opinions expressed by the members of the community.
- Author: the author of the mapping.
- Description: A description of the resource mapped possibly coming from official documentation.
- License: a statement declaring the licensing model under which the mapping is made available to the community.
- Agreement Metrics: every user is enabled in stating whether she/he agrees or not with the proposed mapping. The level of agreement may be estimated using different metrics as suggested in [26].
- Number of Watchers: any mapping can be watched by a member of the community. Watching a mapping allows users to be notified about activities concerning the mapping.
- Number of Likes: any mapping can be *liked* by a member of the community. A like essentially implies an agreement and a subscription to possible events related to the mapping.
- Comments: members of the community are enabled in commenting and discussing about a mapping. We foresee cases where people may ask for clarifications and argue about the validity of the mapping.
- Contextual Tags: any mapping is annotated with a set of tags which identify fuzzy contexts of application of the mappings. These tags can be used to search and filter mappings.

In figure 1, one can see a graphical representation of all the mappings about the Location concept of the Identification Ontology. The first two graphical elements of the interface describe the relation and the status. The, after the URI of the mapping, one can see the author of the mapping, the whether the mapping was watched and by how many users. Finally, we show the number of people care about that specific mapping. Then, on the right side of the interface, a user can filter mappings according to these main dimensions, and typing on the top input field, can filter mappings relying on the namespaces or the local part of the mappings URIs.

Clicking on each mapping, the user can visualize all the details about the mapping in a specific detail page (as shown in Figure 2). This page allows to add comments, rate, subscribe and add possible contextual tags in a collaborative manner. Ratings are made on a 6-item scale including the following options: approved (i.e. the source and target concepts both mean the same thing), broader (the target concept should be a broader term than the source concept), narrower (i.e. the target concept should be a more specific term than the source concept), related (i.e. the two concepts are not an exact match but they are closely related), not sure (i.e. there is a relationship between the two concepts but none of the above relations are appropriate or the term is used in a confusing or contradictory fashion), rejected (i.e. the two concepts are definitely not the same, nor do they have any other direct relationship with each other as listed above). The mapping detail page essentially aims to provide tools for the collaborative interaction for each single defined mapping. If a user subscribes a mapping, any notification will include a link to the specific mapping detail page.

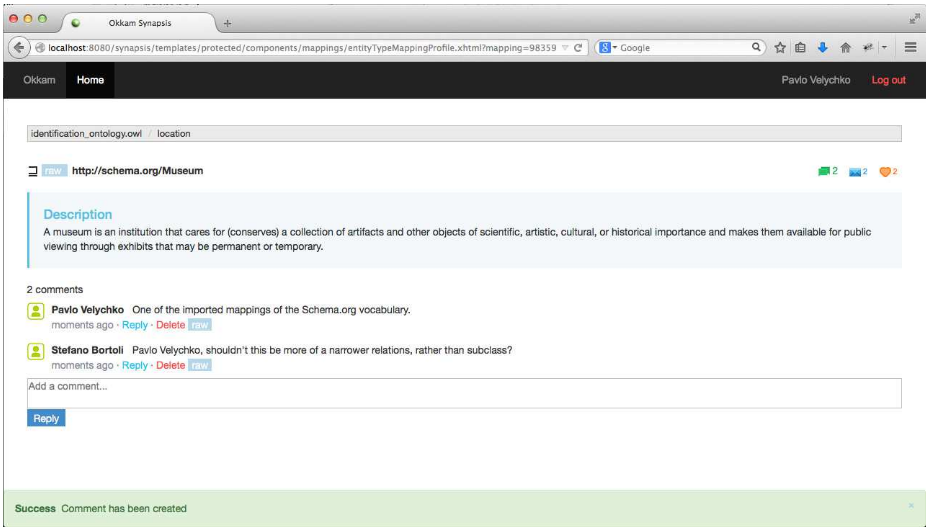

Fig. 2. Mapping page

Once selected the target ontology, the user is enabled in filtering mappings according to different features. On the right part of the page, the filter features are displayed, and the user is enabled in selecting them. Each selection triggers an action on the list of mappings, removing the filtered ones. Mappings can be filtered by author name, status, relation type, rating and creation date. It is also possible to select all the mappings that have comments.

4 Architecture and Data Model

The application is designed according to the traditional MVC design pattern, relying on J2EE JSF framework[7] for the Web interaction part. The mappings

[7] `http://docs.oracle.com/javaee/5/tutorial/doc/bnaph.html`

are also available through rest service, which are implemented relying on Jersey framework[8]. The AJAX based user interface interaction grants quick response and easy interaction for the user. A view of the architecture of the application is presented in figure 3.

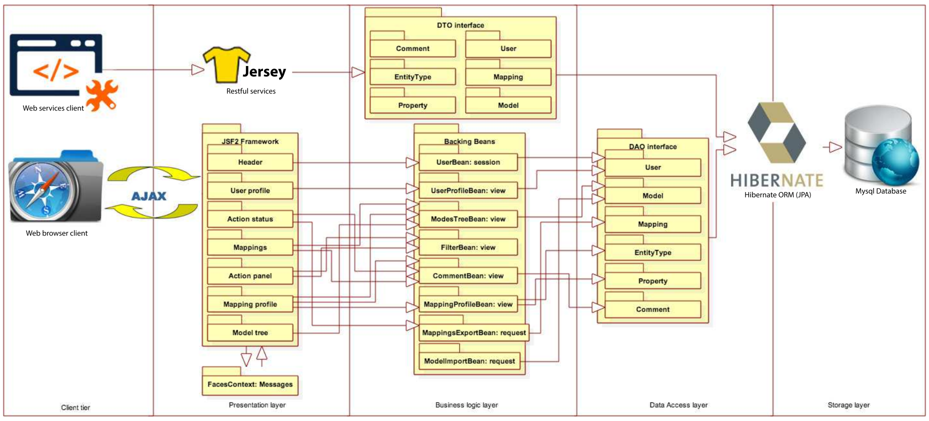

Fig. 3. A graphical view of the Synapsis architecture

Both the J2EE Backing Beans and the REST services access the mappings through standard Data Access Objects (DAO), which rely on the Hibernate ORM JPA provider[9] to interact with a relational database containing all data about mappings, users and supported models. A detailed view of the data model underlying the database is presented in figure 4.

5 Licensing Model

Any mapping created and shared through Okkam Synapsis is realeased under the very popular Common Creative Attribution Share-alike 4.0 International license[10] (CC BY-SA 4.0). The adoption of this copy-left license is to guarantee correct attribution and sharing without affecting the re-usability of the mappings (including commercial purposes). Therefore, contributors and consumers of mappings are granted what we believe is the ideal level of flexibility to accommodate requirements both of researchers and companies. Notice that all the mappings created through the application as subject to this license. Future evolution of the application may allow the selection of other licensing models to be compliant with loading of batches of mappings created else-where and under different licensing model including the share-alike.

[8] `https://jersey.java.net/`
[9] `http://hibernate.org/orm/`
[10] `http://creativecommons.org/licenses/by-sa/4.0/`

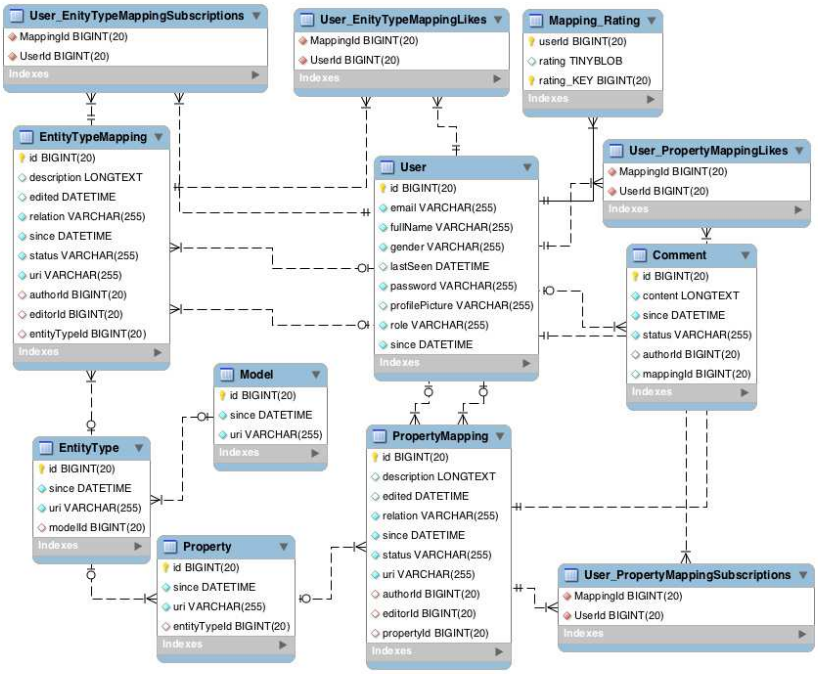

Fig. 4. Synapsis Data Model

6 Kick-off Mappings Dataset

Currently, Synapsis stores 22 mapping for equivalent class, and 205 mapping for subclasses of the entity type Person; 22 mappings for equivalent classes, and 2322 mappings for sub classes of the type Location; and finally we defined 20 mappings for equivalent classes and 2468 mappings for subclasses of the type Organization. These mappings were generated as contextual bridge rules to support semantic harmonization tasks in the knowledge-based solution described in [5]. In particular, the reader can find details about the process leading to the creation of such mappings from existing vocabularies towards the Identification Ontology[11] in Chapter 7 of [5]. We believe that this first core set of mappings can help to kick of a positive endeavor in the adoption of the Okkam Synapsis as a platform to create, share and manage mappings among vocabularies.

7 Conclusion and Future Work

In this paper we have presented a platform called Synapsis which provides a gateway to collaboratively-defined ontology mappings. Looking for a pragmatic

[11] `http://models.okkam.org/identification_ontology.owl` ([5] Chap. 5)

solution to the real world heterogeneity, complexity and inconsistencies, we decided enable users to define mappings as contextual bridge rules, and enable peers to comment and discuss about them. We believe that rating and estimation of level of agreement about mappings would allow to filter commonly shared mappings, and at the same marginalize odd ones. A beta version of the application is available at `http://api.okkam.org/synapsis`, and can be preliminarily tested and evaluated. In the next future, we plan to extend the support for the definition of mappings around applications, to support Linked Data application developer to select the set of mappings of interest and have them available for the application through the defined rest services.

References

1. G Antoniou and F. van Harmelen. *A Semantic Web Primer.* MIT Press, 2004.
2. T. Berners-Lee, J. A. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, May, 2001. http://www.sciam.com/2001/0501issue/0501berners-lee.html.
3. Tim Berners-Lee. Design Issues – Linked Data. Published online, May 2007. `http://www.w3.org/DesignIssues/LinkedData.html`.
4. C. Bizer, R. Cyganiak, and T. Heath. How to publish linked data on the web. online tutorial, July 2007.
5. Stefano Bortoli. *Knowledge Based Open Entity Matching.* PhD thesis, International Doctoral School in ICT of the University of Trento (Italy), 2013.
6. Paolo Bouquet, Fausto Giunchiglia, Frank Harmelen, Luciano Serafini, and Heiner Stuckenschmidt. C-owl: Contextualizing ontologies. In Dieter Fensel, Katia Sycara, and John Mylopoulos, editors, *The Semantic Web - ISWC 2003*, volume 2870 of *Lecture Notes in Computer Science*, pages 164–179. Springer Berlin Heidelberg, 2003.
7. Jason B. Ellis, Oktie Hassanzadeh, Kavitha Srinivas, and Michael J. Ward. Collective ontology alignment. In *OM*, pages 219–220, 2013.
8. Jrme Euzenat. Uncertainty in crowdsourcing ontology matching. In *OM*, pages 221–222, 2013.
9. Bo Fu, Natalya F. Noy, and Margaret-Anne Storey. Indented tree or graph? a usability study of ontology visualization techniques in the context of class mapping evaluation. In *The Semantic Web ISWC 2013*, pages 117–134. Springer Berlin Heidelberg, 2013.
10. A. Gangemi and V. Presutti. *Ontology Design Patterns*, pages 221–243. Springer Berlin Heidelberg, 2009.
11. Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider. Sweetening ontologies with dolce. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*, EKAW '02, pages 166–181, London, UK, UK, 2002. Springer-Verlag.
12. Harith Alani Gianluca Correndo. Collaborative support for community data sharing. In *Proceedings of The 2nd Workshop on Collective Intelligence in Semantic Web and Social Networks*, 2008.
13. Nicola Guarino and Chris Welty. An overview of ontoclean. In Steffen Staab and Rudi Studer, editors, *The Handbook on Ontologies*, pages 151–172. Springer-Verlag, 2004.

14. Pascal Hitzler, Markus Kroetzsch, Bijan Parsia, Peter F. Patel-Schneider, and Sebastian Rudolph. *OWL 2 Web Ontology Language Primer (Second Edition)*. W3C, December 2012.
15. Frank Manola, Eric Miller, and Brian McBride. *RDF 1.1 Primer*. W3C, w3c working group note edition, June 2014.
16. Thomas Markotschi and Johanna Völker. Guess What?! human intelligence for mining linked data. In *Proceedings of the Workshop on Knowledge Injection into and Extraction from Linked Data (KIELD) at the International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, 2010.
17. NatalyaF. Noy, Nicholas Griffith, and MarkA. Musen. Collecting community-based mappings in an ontology repository. In Amit Sheth, Steffen Staab, Mike Dean, Massimo Paolucci, Diana Maynard, Timothy Finin, and Krishnaprasad Thirunarayan, editors, *The Semantic Web - ISWC 2008*, volume 5318 of *Lecture Notes in Computer Science*, pages 371–386. Springer Berlin Heidelberg, 2008.
18. W3C OWL Working Group. OWL 2 Web Ontology Language: Document Overview, 27 October 2009. Available at `http://www.w3.org/TR/owl2-overview/`.
19. P.F. Patel-Schneider, P. Hayes, and I. Horrocks. Web Ontology Language (OWL) Abstract Syntax and Semantics. Technical report, W3C, February 2003. `http://www.w3.org/TR/owl-semantics/`.
20. Shvaiko Pavel and Jerome Euzenat. Ontology matching: State of the art and future challenges. *IEEE Trans. on Knowl. and Data Eng.*, 25(1):158–176, January 2013.
21. Elie Raad and Joerg Evermann. Is ontology alignment like analogy? – knowledge integration with lisa. In *Proceedings of Symposium On Applied Computing (SAC), Korea, Republic Of (2014)*, 2014.
22. Cristina Sarasua, Elena Simperl, and Natalya F. Noy. Crowdmap: Crowdsourcing ontology alignment with microtasks. In *Proceedings of the 11th International Conference on The Semantic Web - Volume Part I*, ISWC'12, pages 525–541, Berlin, Heidelberg, 2012. Springer-Verlag.
23. Pavel Shvaiko and Jérôme Euzenat. Ten challenges for ontology matching. In *Proceedings of the OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008. Part II on On the Move to Meaningful Internet Systems*, OTM '08, pages 1164–1182, Berlin, Heidelberg, 2008. Springer-Verlag.
24. Stefan Thaler, Elena Simperl, and Katharina Siorpaes. Spotthelink: Playful alignment of ontologies. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, SAC '11, pages 1711–1712, New York, NY, USA, 2011. ACM.
25. Anna Tordai. *On Combining Alignment Techniques*. PhD thesis, Vreije Universiteit Amsterdam, 2012-12-03.
26. Anna Tordai, Jacco van Ossenbruggen, Guus Schreiber, and Bob Wielinga. Let's agree to disagree: On the evaluation of vocabulary alignment. In *Proceedings of the Sixth International Conference on Knowledge Capture*, K-CAP '11, pages 65–72, New York, NY, USA, 2011. ACM.
27. Anna V. Zhdanova and Pavel Shvaiko. Community-driven ontology matching. In *Proceedings of the 3rd European Conference on The Semantic Web: Research and Applications*, ESWC'06, pages 34–49, Berlin, Heidelberg, 2006. Springer-Verlag.