A Language-Aware Web will Give Us a Bigger and Better Semantic Web

Armando Stellato

ART Group, Dept. of Computer Science, Systems and Production
University of Rome, Tor Vergata
Via del Politecnico 1, 00133 Rome, Italy
stellato@info.uniroma2.it

Abstract. The role of natural language is becoming in these years a more and more acknowledged aspect of the Semantic Web. Not limited to mere modeling and representation proposals, but backed by concrete use-cases and scenarios, the use of natural language is emerging through a plethora of approaches and solutions. Now that we have languages and protocols for modeling and publishing content, for querying, and for efficiently describing datasets and repositories (metadata), it is time for natural language to regain its due space and become a first-class-citizen in the Web of Data. Lexical resources need to comply with standard, unifying vocabularies upon which they can be discovered, chosen, evaluated and ultimately queried upon need. At the same time, NLP systems and components should pull their head out of their esoteric corner and become classifiable, discoverable and interactive elements in the Semantic Web, so that many language related tasks can be carried on more easily thanks to coordinating modules/agents sensible to this information.

In this position paper, I will provide by first a quick outlook into the last years of language and ontologies and describe what the community has achieved by the state-of-the-art. I will then discuss open points, and try to draw conclusions, based on my perspective and contributions to this research field, towards the future of a more Language-Aware Semantic Web.

1 Introduction

More than a decade has passed since the first OntoLex workshops [1,2] and the earliest attempts at modeling lexical resources and language aspects of ontologies in the Semantic Web, and a whole community has grown around these intents.

Even though no specific criticisms exists in literature, the objectives of this community are sometimes being claimed by critics and detractors to be unreasonably complex, looping in a never-ending lack of consensus and providing no benefit to the Web of Data. However, the role of natural language has become in recent years a more and more acknowledged aspect of the Semantic Web. Not limited to mere modeling and representation proposals, but backed by concrete use-cases and scenarios (e.g. question answering [3,4], ontology-based information extraction [5], ontology learning

[6], semi-automatic lexical enrichment [7,8,9]), "ontolexical ideas" are now emerging as a plethora of approaches and solutions

Still the question arises: what can we do with a more linguistically-aware Web? Can it contribute to the Web of Data, which is meant to be food for machines, or will it be doomed to consume itself, in a niche of language-oriented tasks?

As we have learned from the slow evolution of the Semantic Web itself, its vision has not lost any of its original power, yet some of the related innovations and problems took time to be acknowledged, or required foundations to be laid before they even started to make sense. For many years, "language" has always been the seasoning added by experts for improving performances and quality in data-intensive tasks such as ontology alignment, question answering over ontologies and knowledge acquisition, decorating them with evocative terms such as "language-based" or "language-driven", and still being out of any specification on the Web architecture. Now that we have languages and protocols for modeling and publishing content, for querying, and for efficiently describing datasets and repositories (metadata), it is time for natural language to regain its due space and become a first-class citizen in the Web of Data. lexical resources need not only to be represented as Linked Data, but to comply with standard, unifying vocabularies upon which they can be discovered, chosen, evaluated and finally queried upon need. At the same time, NLP systems and components should pull their head out of their esoteric corner, exploit the stream of innovation brought by software provisioning mechanisms such as Maven Repositories and OBR, breed it with Linked Open Data (LOD) principles, and become classifiable, discoverable and interactive elements in the Semantic Web. With such an elaborated web-of-language architecture, those same tasks mentioned above, which today require lot of pre-processing and fine-tuning, should be carried out more easily (if not on-the-fly) thanks to coordinating modules/agents sensitive to the lexical information present in ontologies, datasets, lexical resources and in NLP software and services.

In this position paper, I will provide by first a quick outlook into the last years of language and ontologies and see what the community has achieved by now. I will then discuss open points, and try to draw conclusions, based on my perspective and contributions to this research field, towards the future of a more Language-Aware Semantic Web.

2 A bit of history…

The first events in the area of ontologies and lexical resources were the OntoLex series of workshops [1,2]. OntoLex covered, *tout court*, all research dealing with the representation of lexical resources (or linguistic content to a larger extent) in the Semantic Web world, with the relationships between formal ontologies and lexical semantics in general, and more specifically in the construction of lexical knowledge bases. The workshop also dealt with various applications of lexical semantics to information retrieval, information extraction and related fields.

With the end of the OntoLex experience (Ontolex 2010, [10]), other events inherited its mission. Among the various initiatives, we mention the series of workshops "Linked

Data in Linguistics (LDL)" [11], which is focused on discussing principles, case studies, and best practices for representing, publishing and linking linguistic data collections, and infrastructure such as corpora, dictionaries, lexical nets, translation memories, thesauri. The workshops on natural language Processing and Linked Open Data (NLP&LOD, [12]) focus on all the problematics related to the adoption of NLP techniques in (and for) the LOD world. These include precision and quality of information, the definition of standard vocabularies for both the syntactic and semantic targets of extraction (such as Part of Speech vocabularies) and for describing NLP chains as well. Finally, this same series of workshops, the Multilingual Semantic Web [13] focus on the multilingual aspects of the Web of Data, and on the social, political and economic implications that multilingual scenarios imply.

In parallel with events, many proposals have come out in these years, in terms of models and vocabularies for representing lexical resources, for a better lexical representation of ontologies, and for trying to standardize the plethora of efforts that have been conducted in the NLP area, in the context of linked data.

A number of models have been proposed to enrich ontologies with information about how vocabulary elements are expressed in different natural languages, including the Linguistic Watermark framework [14,15], LexOnto [16], LingInfo [17], LIR [18], LexInfo [19] and more recently *lemon* [20] and Lime [21].

Due to these efforts, concrete proposals emerged and were refined into new ones through experience and discussion, and community group emerged to concretize forthcoming standards for language-related aspects of the Web of Data.

The OntoLex W3C Community Group[1] has the goal of providing an agreed-upon standard by building on the aforementioned models, the designers of which are all involved in the community group. Additionally, linguists have acknowledged [22] the benefits that the adoption of the Semantic Web technologies could bring to the publication and integration of language resources. As such, the Open Linguistics Working Group[2] of the Open Knowledge Foundation is contributing to the development of a LOD (Linked Open Data) (sub)cloud of linguistic resources[3].

These complementary efforts by Semantic Web practitioners and linguists are in fact converging: the ontology lexicon model being defined in OntoLex provides a principled way [23] to encode even notable resources such as the Princeton WordNet [24,25] and other similar ones for other languages. At the same time, the Lexical Linked Data Cloud [26] is providing a parallel Linked Open Data cloud aimed at linguistics and natural language processing, by the same principles of the traditional LOD. To complete the whole picture, and with an eye on computation and not only on resources, the NLP Interchange Format (NIF, [27]) provides an RDF/OWL-based format that allows to combine and chain several NLP tools in a flexible, light-weight way.

[1] http://www.w3.org/community/ontolex/
[2] http://linguistics.okfn.org/
[3] http://nlp2rdf.lod2.eu/OWLG/llod/llod.svg

3 Is language so necessary in what is called "Web of Data" ?

By seeing all of these efforts from the very neutral perspective of someone approaching the world of the Web of Data, one question would naturally arise: if the Semantic Web is all about machine-understandable data, why should we care at all about language? Can a more linguistically-aware Web contribute to the Web of Data, which is meant to be food for machines, or will it be doomed to consume itself, in a niche of language-oriented tasks? What can we do with it?

We could admit that large part of the research done in the area is more biased by the (linguistic) background of the researchers working on it, than driven by the general interest in the problems it tries to tackle. What puts in the same category all of the works and efforts described until now, is the scientific and cultural background of the communities working on them. As "Semantic Web" and "Linked Data" are not properly "foundations of science" but applied sectors of fields of computer science such as knowledge representation, knowledge and data management, and many others.., researchers who contributed to any aspect of natural language in the Semantic Web have their foundational background as well. Specifically, they flowed into the semantic river as Computational Linguists (or pure Linguists), or, in any case, language experts who wanted to give their contribution to the Semantic Web cause.

The risk of proposing solutions to issues which do not characterize the addressed scenario is high, and this has probably occurred more than occasionally. For sure, a lot of effort has been spent in addressing specific issues seen with the glasses of traditional computational linguists. Without mentioning the specific works, any short look on the literature of the Semantic Web area will return a plethora of articles on converting the n^{th} lexical resource to the k^{th} OWL porting, or on leveraging existing "less noble" resources to ontologies. Such is the case of the frenzied plethora of articles on lifting the so-called "folksonomies" to ontologies ("folksonomies" which were strangely more recent than the concept of ontology itself). All efforts that, by applying well known techniques for knowledge synthesis, but lacking an overall perspective on the matter, left neither new theoretical foundations nor concrete results (e.g. in term of resources) for the future to come.

On the other side however, as it often happens in research, problems take their time to be acknowledged, or may require foundations to be laid before they even start to make sense. The efforts described in the previous section all aimed at providing general frameworks for describing lexical resources, common loci for finding their descriptions on the Web and metadata for properly selecting them. As well as much desired results for trying to convey order in the plethora of models and processes which years of NLP research have produced at the expenses of any engineering attempt in the field. Together with the above, a few application scenarios have proven the usefulness of a more aware representation of natural language content in the Web of Data. In the next section I will describe the importance of such scenarios in the context of the whole Web of Data, and how they can be improved by a more systematic inclusion of natural language in the Semantic Web architecture.

4 Scenarios

I describe here three scenarios that prove the advantages brought by a more language-aware approach to the Web of Data. These are, respectively:

- Linguistic Enrichment of Ontologies
- Ontology Alignment
- Knowledge Acquisition

4.1 Linguistic Enrichment of Ontologies

It may seem redundant and self-referential to prove the importance of a more linguistically-aware Web, by showing a scenario for linguistic enrichment.

Fact is that the provision of natural language descriptors (for how simple and de-structured these can be) for ontological entities is a primary necessity which comes far before the more elaborated features that we are trying to highlight. From the dawn of computer programming, providing "meaningful and evocative names" for functions (and later for classes/objects and methods) is a primary necessity for, to say the least, improving the readability of source code. In this sense, ontologies and their RDF code are no different from computer programs, and they need human understandable descriptors in order to be maintained and evolved. Furthermore, in the case of ontologies, the connection between *concepts* and their *referents* [28] even more so needs proper *symbols* for being recognized and understood.

A complete ontology development process should thus consider the formal aspects of conceptual knowledge representation, as well as guarantee that the same knowledge be recognizable amongst its multiple expressions that are available on real data. Thus, the importance of properly enriching ontologies with lexical content does not need further support, while what we need to prove is that more attention to the representation of language can improve the quality (and time to realization) of this necessary practice.

Ontology Development tools should support this task, providing users with dedicated interfaces for efficient carrying it out. It appears obvious that the reuse of linguistic resources creates added value, as they can contribute additional descriptors, as well as suggesting additional knowledge which was not considered in the conceptual development phase. Views over lexical resources must be integrated with classic views over knowledge data such as class trees, property and instance lists, offering a set of functionalities for linguistically enriching concepts and, possibly, for building new ontological knowledge starting from linguistic one.

By considering past experiences [4,29,30] with knowledge based applications dealing with concepts and their lexicalizations, a few basic functionalities for browsing linguistic resources have emerged to be mandatory:

- *Searching for term definitions (glosses)*
- *Asking for synonyms*
- *Separating different senses of the same term*
- *Exploring genus and differentia*
- *Exploring resource-specific semantic relations*

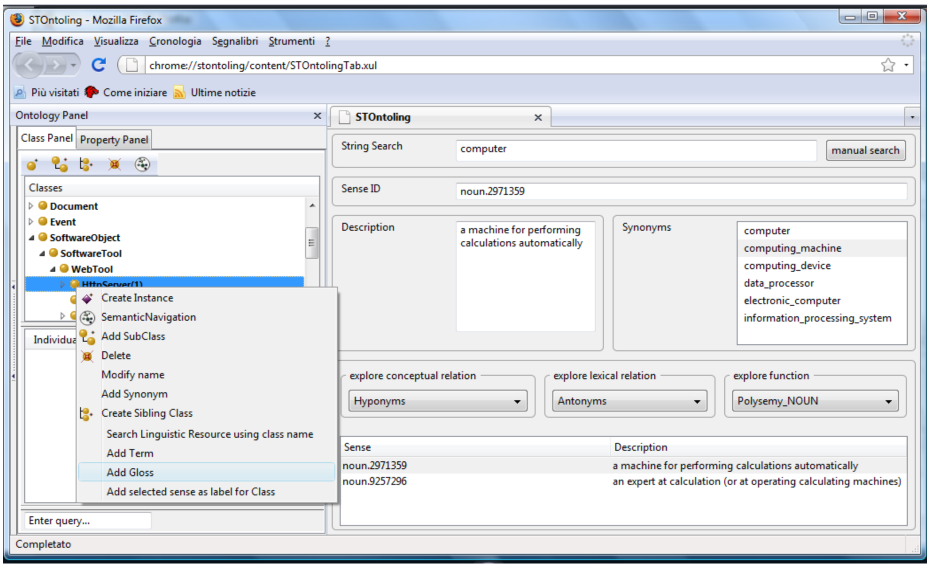

Fig. 1 Exploiting lexical resources in Ontoling in order to linguistically enrich ontologies

as well as some others for ontology editing:

– *Adding synonyms (or translations, for bilingual resources) as additional labels for identifying concepts*
– *Adding glosses to concepts description (documentation)*
– *Using notions from linguistic resources to create new concepts*

However, one question emerges: how are these tools supposed to interact with lexical resources? In years of literature in NLP, many systems have adopted resource-based approaches in order to perform given tasks, however the "knowledge" about these resources, i.e., their semantics, were always "hardwired", "embedded" within that same systems that were exploiting their content. Besides doing the same work again and again, these systems were lacking at the same time the capability to scale up to a wider scenario including similar resources, or being adaptively capable of reusing different ones…in different ways. In a word, they lacked a common lexical model, upon which to see each (lexical) resource not as a world-per-se, but as the emanation of a more general theory.

In [31] I have shown a general model for describing lexical resources that can be used as a driver in different tasks, and for the linguistic enrichment of ontologies in particular, by binding the enrichment functions to those same abstract descriptors, and by adaptively shaping enrichment processes according to the (description of the) lexical resource being exploited. This greatly enhances the possible application of the system, as much as facilitates its adaptivity to different scenarios. In that work, the usefulness of model for lexical resources is extended to adaptive algorithms for semi-automatic enrichment as well.

4.2 (Semi-)Automatic Ontology Alignment

Ontology Matching [32] is the task of finding (semantic) correspondences between a pair of ontologies. When any form of semantic agreement is missing – and this is the scenario of ontology alignment, as reconciling semantic heterogeneity is its purpose – the first common resource is generally natural language. The reconciliation of different ontologies is thus driven by their linguistic grounding, then more advanced methods have been devised, combining structural, extensional and semantic features. Some approaches exploit background knowledge, such as the Web, Wikipedia, domain corpora, lexical resources and upper-ontologies.

Since 2004, the OAEI[4] organizes annual campaigns for the evaluation of automatic ontology matchers. The general scenario for the contests is simple: two datasets, and the objective to align their elements as much as possible. The contest then comprises different challenges, differentiating by the type of dataset (an ontology, a thesaurus instance data), the natural language(s) involved (same language, two different languages, sets of languages available for each dataset), the elements to match etc.

While the participants to the contest are focused on developing more and more complex techniques for alignment (and the results can obviously be considered scientific as they are run on repeatable experimental settings), the question whether it is important to improve performances on worst case scenarios – while the conditions of these scenarios could be improved at the Web level – arises. Currently, there is no widely adopted metadata vocabulary providing resuming information about the available natural languages, their coverage of the analyzed resources and the modeling vocabulary adopted to represent this information (are terms expressed as `rdfs:labels`, through SKOS, SKOS-XL, or other advanced models such as those presented in section 2). While metadata can obviously be recomputed as a first processing step in a deep alignment procedure, chances that these alignments could be applied "on the fly" are severely reduced (in quality, if not on feasibility at all). If we want the Web of Data to succeed, it is obvious that proper publication of data should be encouraged by providing quality indicators, such as the Five Stars Open Data reference[5]. Metadata should be a fundamental part of the Web, and vocabularies such as VoID provide useful information that all publishers should provide. While it is important to foster research in robust procedures considering worst cases, it should be also important to understand what would be the benefit of a better acknowledgement of the available information in a repository, and how this can be exploited.

By analyzing the conclusions from each of the workshops and related literature, we can gather some interesting facts:

- We are probably close to the achievable limits (*discussion in OAEI 2013*)
- There is however still an incremental growth by tuning existing techniques [33]
- State of the art systems can *fail*, if they do not understand their input (*as seen in all recent OAEI with SKOS challenges*)

---

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

[5] http://5stardata.info/

Fig. 2 UML Use Case for Linguistic Agents supporting Semantic Coordination

- Performance in evaluations does not correlate straightforwardly with success in real-world scenarios [34]
- In many real-world scenarios, even simple string matching techniques applied to lexically rich datasets may support effective semi-automatic processes [35]. As often happens when human resources are available: overall recall is more important than a high precision, as humans can always filter out spurious results while would feel uncertain if they had doubts about the system filtering out potentially good results.

In particular, in the library track of OAEI, dealing with the alignment of the thesauri TheSoz[6] and STW[7]:

[6] http://www.gesis.org/en/services/tools-standards/social-science-thesaurus/
[7] http://zbw.eu/stw/

- Most of the participants in OAEI do not understand SKOS. Thus, thesauri need to be sweeten into ontologies by means of a lossy and incorrect mapping [36]
- As of 2014 results [37], "the baselines with preferred labels are still very good and can only be beaten by one system". Furthermore, the use of SKOS annotations increases the performance of 7%

It thus seems that the quality of algorithms adopted is an order of magnitude less important than properly filtering, cleaning, managing and, most importantly, being able to exploit the linguistic information available in ontologies and datasets. Simply put, the competing systems are currently not squeezing all the juice out of this very important information, losing much of what would be easily available, while pushing hard on disentangling un-understandable URIs and trying to grasp some knowledge from their structural organization.

Unfortunately, as I remarked, this information is not explicitly available through any dedicated vocabulary and most of these tools require fine-tuning by users, where they should be able to self-configure. In a few words: machines need humans to tell them how to read machine-understandable data!

As I explained in [38], linked open data, in order to become really operable by machines with increasing levels of automation, needs to "talk about itself", and this obviously has to cover the language expressiveness of its content, in order to facilitate alignment process which, as discussed above, would gain a lot from this asset. In that work, I suggest bootstrapping techniques for allowing agents on the Web to start linguistic coordination activities (see fig. 2) as a primary step of semantic coordination. The word "agents" is intended with a wide interpretation, encompassing active agents, such as software agents browsing the linked data and taking decisions on the basis of the gathered results, passive agents such as SPARQL endpoints, and intermediate ones such as semantic web services, which react to requests, but can have complex business logics to satisfy their tasks.

In the context of the OntoLex community group, the scenario envisioned in [38] has been made largely supported, thanks to the effort spent on Lime [21] for providing a definitive metadata vocabulary, as part of the whole OntoLex suite of vocabularies, for representing both lexical resources, and the lexical asset of ontologies, thesauri and datasets in general. I hope that, with a larger adoption of such metadata, the vision of agents able to cope autonomously with their heterogeneities, will be more detailed.

A recent attempt at exploiting information from Lime and from other available metadata (such as the Vocabulary of Interlinked Datasets VoID [39]) is also reported in [40].

4.3 Content Acquisition and Ontology Evolution

The acquisition of knowledge from unstructured content has maturated dramatically – especially in the latest years – on the engineering aspects other than on the techniques for content analysis. The TIPSTER [41] model has provided a reference architecture for text processing, which has been implemented by systems such as GATE [42] and has later influenced the OASIS[8] standard UIMA [43].

Following the advent of RDF, a few attempts at empowering information extraction architectures with data projection mechanisms have been presented. Some examples are the RDF UIMA CAS (Common Analysis Structure) Consumer [44], Clerezza-UIMA integration [45], and Apache Stanbol [46] and CODA [47].

Another important achievement has been the proposal for a NLP Interchange Format (NIF, [48]), an RDF/OWL-based format that aims to achieve interoperability between natural language Processing (NLP) tools, language resources and annotations. Such approaches have been sometimes seen as too constraining by traditional NLP scientist: naturally, studies in NLP have produced a plethora of different models and approaches, which are difficult to put under a common umbrella, that is why all previous attempts have focused on more model-agnostic solutions. However, again, the linked data approach has proven to be successful: whereas UIMA and GATE provided neutral feature-structure based solution, but have never seen the proliferation of vocabularies for storing data according to different models (e.g. different set of part-of-speech tags), the introduction of the linked data approach allows for different models to coexist and to be reused.

However, again, in the same way as a linked data approach can benefit NLP, the benefit can be reciprocal. Until now, NLP has been a very closed world: even when getting out of the scientific field and moving to industry, NLP has been seen as a sort of highly specialized sector where experts work out of ordinary software performance/quality parameters and interoperability needs. This has explained the limited success that it had until now, especially in smaller organization which are have not the power to deal with the complexity that NLP processes require.

In my vision, those same NLP components the behavior of which is advertised as linked data (in environments such as NIF), should be made available as open, downloadable and pluggable components that can be dynamically agglomerated, in order to assemble complete NLP chains. Much in the spirit of provisioning systems such as Maven[9], that are however focused (or, at least, commonly used for the large part) on the provisioning of software libraries and on dependency resolution for Integrated Development Environments, NLP components should be provisioned as interconnectable pieces of a larger puzzle. Their metadata, and metadata coming from linguistic resources and from ontologies as well, should all create a global ecosystem in which it will be possible to target knowledge to be acquired/improved, sources of information, potential supporting resources and the software needed to provide them.

5 Conclusions

In this article, which followed my keynote speech at the 4th Multilingual Semantic Web Workshop (collocated with the 12th edition of the Extended Semantic Web Conference, in Portoroz, Slovenia), I showed a few practical applications of a more linguistically-

[8] Organization for the Advancement of Structured Information Standards

[9] https://maven.apache.org/

aware Semantic Web, in terms of enriched possibilities for well known tasks, such as lexical documentation, ontology alignment and content acquisition, which ultimately can contribute to the growth of the Web itself.

The objective is to remark, once more, that despite (to reuse the words of Tim Berners-Lee about his own creation) the “concept of machine-understandable documents does not imply some magical artificial intelligence which allows machines to comprehend human mumblings”, still this machine-understandability has its limits, dictated by open and environments and heterogeneous needs. Allowing machines to trade their knowledge when they have no better agreement on it, or to support humans in extracting information from textual content, brings about a marriage with the sole form of knowledge exchange that has endured hundreds of years of evolution: natural language.

References

1. K., S., Kiryakov, A., eds.: Proceedings of OntoLex’2000: Ontologies and Lexical Knowledge Bases., Sozopol, Bulgaria (Sept. 8-10, 2000)
2. Simov, K., Guarino, N., Peters, W., eds.: Proceedings of The Ontologies and Lexical Knowledge Bases Workshop (OntoLex02)., Las Palmas, Canary Islands - Spain (27th May 2002)
3. Unger, C., Freitas, A., Cimiano, P.: An Introduction to Question Answering over Linked Data. In Koubarakis, M., Stamou, G., Stoilos, G., Horrocks, I., Kolaitis, P., Lausen, G., Weikum, G., eds. : Reasoning Web. Reasoning on the Web in the Big Data Era 8714. Springer International Publishing (2014), pp.100-140
4. Atzeni, P., Basili, R., Hansen, D.H., Missier, P., Paggio, P., Pazienza, M.T., Zanzotto, F.M.: Ontology-based question answering in a federation of university sites: the MOSES case study. In Meziane, F., Métais, E., eds. : Natural Language Processing and Information Systems (Lecture Notes in Computer Science). 9th International Conference on Applications of Natural Languages to Information Systems, NLDB 2004, Salford, UK, June 23-25, 2004, Proceedings. 3136. Springer-Verlag Berlin Heidelberg, Manchester (United Kingdom) (2004), pp.413-420
5. Basili, R., Vindigni, M., Zanzotto, F.M.: Integrating Ontological and Linguistic Knowledge for Conceptual Information Extraction. In : Web Intelligence, 2003. WI 2003. proceedings of the IEEE / WIC International Conference on Web Intelligence, 13-17 October, 2003, Halifax, Canada. IEEE, Washington, DC, USA (2003), pp.175 - 181
6. Cimiano, P.: Ontology Learning and Population from Text Algorithms, Evaluation and Applications XXVIII. Springer (2006)
7. Bouayad-Agha, N., Casamayor, G., Wanner, L.: Natural Language Generation in the context of the Semantic Web. Semantic Web 5(6), 493-513 (January 2014)
8. Bontcheva, K., Wilks, Y.: Automatic Report Generation from Ontologies: The MIAKT Approach. In Meziane, F., Métais, E., eds. : Natural Language Processing and Information Systems 3136. Springer Berlin Heidelberg (2004), pp.324-335

9. Galanis, D., Androutsopoulos, I.: Generating Multilingual Descriptions from Linguistically Annotated OWL Ontologies: The NaturalOWL System. In : Proceedings of the Eleventh European Workshop on Natural Language Generation, Stroudsburg, PA, USA, pp.143-146 (2007)
10. Oltramari, A., Vossen, P., Qin, L., eds.: Proceedings of the 6th Workshop on "Ontologies and Lexical Resources (OntoLex 2010)., Beijing, China (August, 21, 2010)
11. Chiarcos, C., Cimiano, P., Declerck, T., McCrae, J.P., eds.: Proceedings of the 2nd Workshop on Linked Data in Linguistics (LDL-2013): Representing and linking lexicons, terminologies and other language data., Pisa, Italy (2013)
12. Vossen, P., Rigau, G., Osenova, P., Simov, K., eds.: Proceedings of the Second Workshop on Natural Language Processing and Linked Open Data (NLP&LOD2)., Hissar, Bulgaria (11 September 2015)
13. Proceedings of The Multilingual Semantic Web workshop., Dagstuhl, Germany (September 2 – 7 , 2012)
14. Pazienza, M.T., Stellato, A., Turbati, A.: 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, vol. 426, p.11 (2008)
15. Oltramari, A., Stellato, A.: Enriching Ontologies with Linguistic Content: an Evaluation Framework. In : The role of ontolex resources in building the infrastructure of Web 3.0: vision and practice (OntoLex 2008), Marrakech, Morocco, pp.1-8 (2008) May, 31.
16. Cimiano, P., Haase, P., Herold, M., Mantel, M., Buitelaar, P.: LexOnto: A Model for Ontology Lexicons for Ontology-based NLP. In : In Proceedings of the OntoLex07 Workshop (held in conjunction with ISWC'07) (2007)
17. Buitelaar, P., Declerck, T., Frank, A., Racioppa, S., Kiesel, M., Sintek, M., Engel, R., Romanelli, M., Sonntag, D., Loos, B., Micelli, V., Porzel, R., Cimiano, P.: LingInfo: Design and Applications of a Model for the Integration of Linguistic Information in Ontologies. In : OntoLex06, Genoa, Italy, pp.28-34 (2006)
18. Montiel-Ponsoda, E., Aguado de Cea, G., Gómez-Pérez, A., Peters, W.: Enriching ontologies with multilingual information. Natural Language Engineering 17, 283-309 (July 2011)
19. Cimiano, P., Buitelaar, P., McCrae, J., Sintek, M.: LexInfo: A declarative model for the lexicon-ontology interface. Web Semantics: Science, Services and Agents on the World Wide Web 9(1), 29-51 (2011)
20. McCrae, J., Aguado-de-Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., Wunner, T.: Interchanging lexical resources on the Semantic Web. Language Resources and Evaluation 46(4), 701-719 (2012)
21. Fiorelli, M., Stellato, A., Mccrae, J.P., Cimiano, P., Pazienza, M.T.: LIME: the Metadata Module for OntoLex. In : The Semantic Web. Latest Advances and New Domains. Proceedings of the 12th Extended Semantic Web Conference (ESWC 2015), Portoroz, Slovenia, May 31 - 4 June, 2015. (2015), pp.321-336
22. Chiarcos, C., McCrae, J., Cimiano, P., Fellbaum, C.: Towards Open Data for Linguistics: Linguistic Linked Data. In Oltramari, A., Vossen, P., Qin, L., Hovy, E., eds. : New Trends

of Research in Ontologies and Lexical Resources. Springer Berlin / Heidelberg (2013), pp.7-25 10.1007/978-3-642-31782-8_2.

23. McCrae, J.P., Fellbaum, C., Cimiano, P.: Publishing and Linking WordNet using lemon and RDF. In : Proceedings of the 3rd Workshop on Linked Data in Linguistics, Reykjavik, Iceland (2014)
24. Miller, G.A.: WordNet: A Lexical Database for English. Communications of the ACM 38(11), 39-41 (November 1995)
25. Fellbaum, C.: WordNet: An Electronic Lexical Database. WordNet Pointers, MIT Press, Cambridge, MA (1998)
26. Chiarcos, C., Hellmann, S., Nordhoff, S.: Linking linguistic resources: Examples from the Open Linguistics Working Group. In Chiarcos, C., Nordhoff, S., Hellmann, S., eds. : Linked Data in Linguistics. Representing and Connecting Language Data and Language Metadata. Springer Berlin Heidelberg. (2012), pp.201-216
27. Rizzo, G., Troncy, R., Hellmann, S., Brümmer, M.: Proceedings of LDOW, 5th Workshop on Linked Data on the Web., Lyon, France (April 16, 2012)
28. Ogden, C.K., Richards, I.A.: The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism. Magdalene College, University of Cambridge (1923)
29. Pazienza, M.T., Stellato, A., Vindigni, M., Valarakos, A., Karkaletsis, V.: Ontology integration in a multilingual e-retail system. In Stephanidis, C., ed. : Universal Access in HCI: Inclusive Design in the Information Society. proceedings of HCI International 2003. 4. Lawrence Erlbaum Associates, Crete, Greece (2003), pp.785-789
30. Pazienza, M.T., Stellato, A., Enriksen, L., Paggio, P., Zanzotto, F.M.: Ontology Mapping to support ontology-based question answering. In : Second MEANING workshop, Trento, Italy (2005)
31. Pazienza, M.T., Stellato, A.: An Environment for Semi-automatic Annotation of Ontological Knowledge with Linguistic Content. In Sure, Y., Domingue, J., eds. : The Semantic Web: Research and Applications (Lecture Notes in Computer Science). 3rd European Semantic Web Conference, ESWC 2006, Budva, Montenegro, June 11-14, 2006, Proceedings. 4011. Springer (2006), pp.442-456
32. Euzenat, J., Shvaiko, P.: Ontology Matching. Springer-Verlag New York, Inc., Secaucus, NJ, USA (2007)
33. 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)
34. van Ossenbruggen, J., Hildebrand, M., de Boer, V.: Interactive Vocabulary Alignment. In Gradmann, S., Borri, F., Meghini, C., Schuldt, H., eds. : Research and Advanced Technology for Digital Libraries (Lecture Notes in Computer Science) 6966. Springer-Verlag, Berlin, Germany (2011), pp.296-307
35. Caracciolo, C., Morshed, A., Stellato, A., Johannsen, G., Jaques, Y., Keizer, J.: Thesaurus maintenance, alignment and publication as linked data: the AGROVOC use case. In : Metadata and Semantic Research. Springer Berlin Heidelberg (2011), pp.489-499
36. Ritze, D., Eckert, K.: Thesaurus mapping: a challenge for ontology alignment? In : In Proceedings of the 7th International Workshop on Ontology Matching, November 11, 2012, Boston, MA, USA, vol. 946, pp.248-249 (2012)

37. Dragisic, Z., Eckert, K., Euzenat, J., Faria, D., Ferrara, A., Granada, R., Ivanova, V., Jimenez-Ruiz, E., Kempf, A.O., Lambrix, P., Montanelli, S., Paulheim, H., Ritze, D., Shvaiko, P., Solimando, A., Trojahn, C., Zamazal, O., Cuenca Grau, B.: Results of the ontology alignment evaluation initiative 2014. In : 9th International Workshop on Ontology Matching, October 20, 2014, Riva del Garda, Trentino, Italy, vol. 1317, pp.61-104 (2014)

38. Pazienza, M.T., Sguera, S., Stellato, A.: Let's talk about our "being": A linguistic-based ontology framework for coordinating agents. Applied Ontology, special issue on Formal Ontologies for Communicating Agents 2(3-4), 305-332 (December 2007)

39. 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). (March 3, 2011) Available at: http://www.w3.org/TR/void/

40. Fiorelli, M., Pazienza, M.T., Stellato, A.: A Meta-data Driven Platform for Semi-automatic Configuration of Ontology Mediators. In Chair), N.C.(., 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), May 2014. European Language Resources Association (ELRA), Reykjavik, Iceland (2014), pp.4178-4183

41. Harman, D.: The DARPA TIPSTER project. SIGIR Forum 26(2), 26-28 (1992)

42. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In : In Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia (2002)

43. Unstructured Information Management Architecture (UIMA) Version 1.0. In: OASIS Standard. (Accessed March 2, 2009) Available at: http://docs.oasis-open.org/uima/v1.0/uima-v1.0.html

44. Apache UIMA RDF CAS Consumer documentation. Available at: http://uima.apache.org/downloads/sandbox/RDF_CC/RDFCASConsumerUserGuide.html

45. Clerezza integration with Apache UIMA. Available at: http://incubator.apache.org/clerezza/clerezza-uima/

46. Apache Stanbol official website. Available at: http://stanbol.apache.org/

47. Fiorelli, M., Pazienza, M.T., Stellato, A., Turbati, A.: CODA: Computer-aided ontology development architecture. IBM Journal of Research and Development 58(2/3), 14:1 - 14:12 (March-May 2014)

48. Hellmann, S., Lehmann, J., Auer, S., Brümmer, M.: Integrating NLP using Linked Data. In : The Semantic Web – ISWC 2013 (Lecture Notes in Computer Science). 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part II. 8219. Springer Berlin Heidelberg, Sydney, Australia (21-25 October 2013), pp.98-113