

Lexicography and Semantic Web: a Demo with LexO

Andrea Bellandi, Anas Fahad Khan

Institute for Computational Linguistics, Via G. Moruzzi, Pisa - Italy
E-mail: name.surname@ilc.cnr.it

Keywords: Semantic web; Lexicography; LexO; OntoLex-Lemon; Linked data

1. Introduction

The purpose of this contribution is to present LexO¹, the first version of a collaborative web editor for easily building and managing lexical and terminological resources in the context of the Semantic Web. The adoption of Semantic Web technologies and the Linked Data paradigm has been driven by the need to ensure the construction of resources that are interoperable and can be shared and reused by the scientific community. LexO's primary

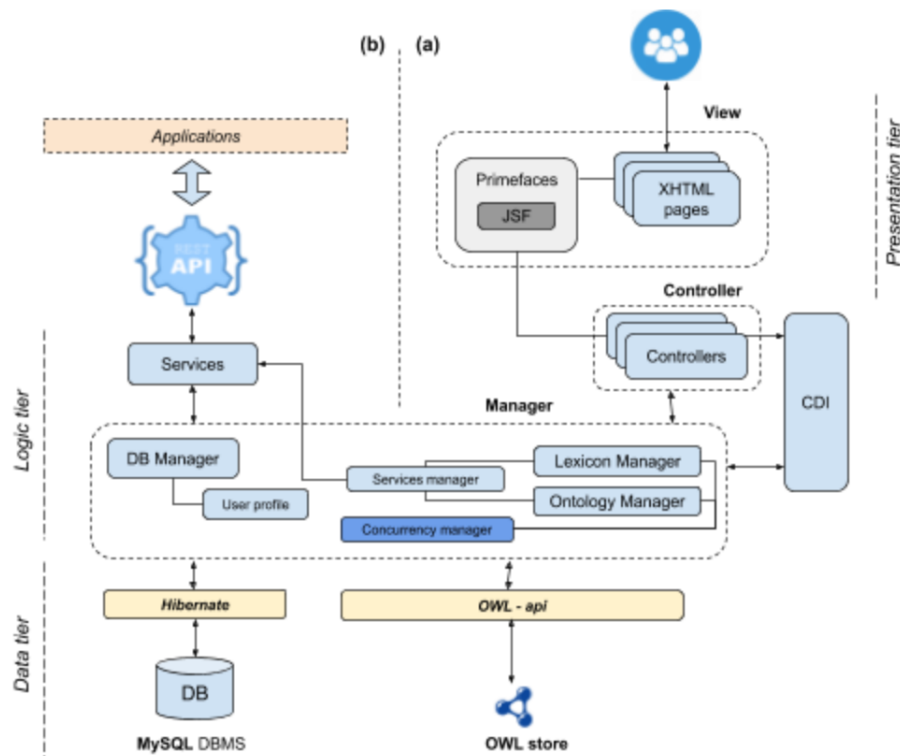


Figure 1: Multi-tier architecture of LexO. It is based on the software design pattern known as “three-tier architecture”, and which exploits Apache Tomcat v8.0 as webservice. The component-based architectural structure was implemented by the object-oriented J2SE framework, enhanced with Contexts and Dependency Injection annotations (CDI). (a) LexO as web application - (b) LexO as rest services.

objective is to enable terminologists and lexicographers to create a resource ex novo this is by means of the adoption of a lexical model that allows the association of detailed and structured lexical information (Bellandi et al., 2018); (Khan et al., 2016) to ontological

¹ The source code is available at <https://github.com/cnr-ilc/LexO-lite>.

concepts. In this respect, the lemon lexical model (McCrae et al., 2012), later renamed Ontolex-lemon (McCrae et al., 2017), is currently regarded as the de facto standard for enriching Semantic Web ontologies with lexical information. LexO can provide a support for creating, managing, publishing lexical and terminological resources as Linked Open Data, that is, typically, a complex task especially for those who have not yet mastered Semantic Web-based standards and technologies, such as RDF and OWL. However, the long-term ambition of LexO would be to make a deeper contribution to e-lexicography.

2. LexO

LexO arises out of several DH research projects² that aimed to construct lexical resources and terminologies, and also takes into account experiences coming from international projects dealing with e-lexicography, such as ELEXIS³. It tries to consider the necessity

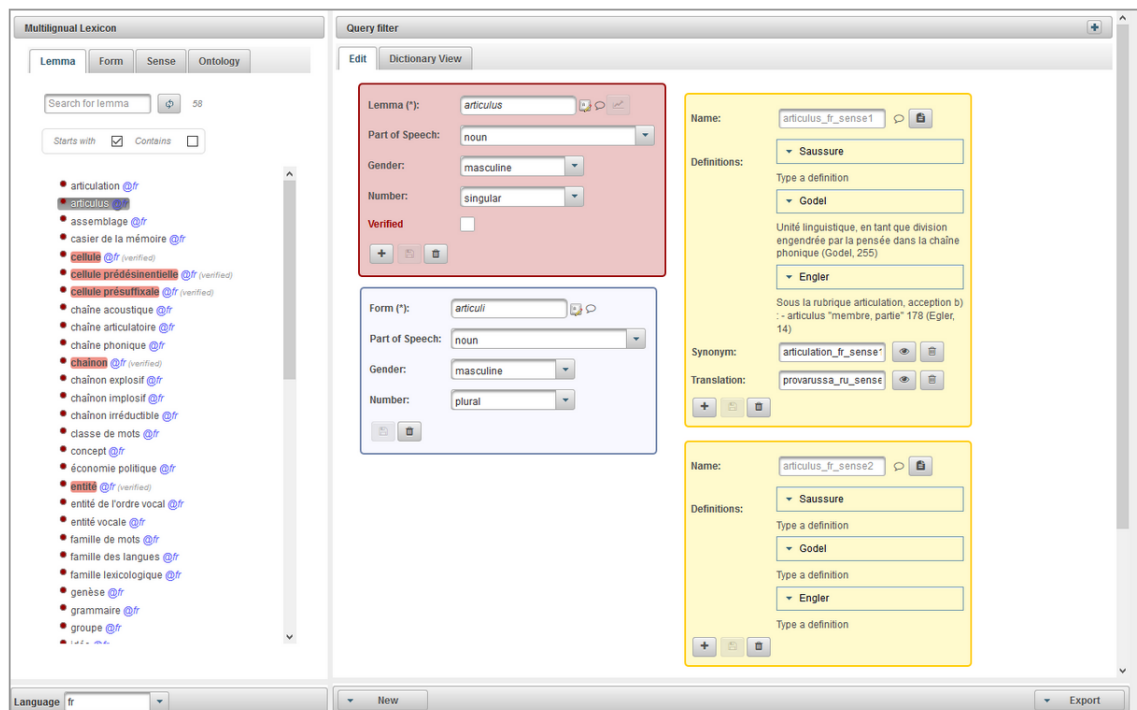


Figure 2: Main interface of LexO. On the left, a column shows the list of lemmas composing the resource, the forms, the lexical senses, and the concepts belonging to the ontology of reference. The information related to the selected entry is shown in the central panel. It is possible to link senses to the related ontological concept.

of making the main formats (TEI, LMF (Francopoulo et al., 2006), Ontolex-lemon) interoperable, and in particular looking at a broader perspective of an ecosystem, where

² LexO has been mainly developed in the context of the DiTMAO project - multilingual resource of medico-botanical terminology focussed on Old Occitan, and it has already been used in several projects, such as, i) Ferdinand De Saussure (<http://www.ilc.cnr.it/en/content/saussure>) - multilingual diachronic resource of Saussure's terminology, ii) Totus Mundus - bilingual chinese-italian resource on Matteo Ricci's Atlas (<http://www.ilc.cnr.it/en/content/ilc-iit-totus-mundus>).

³ ELEXIS is an ambitious project within the domains of NLP and e-lexicography with the aim of creating a European wide lexicographic infrastructure. It is based on a previous Cost Action ENeL - aiming to establish a European network of lexicographers and a common approach to e-lexicography that forms the basis for a new type of lexicography (<http://www.elexicography.eu/>).

different standards can coexist and mutually enrich each other. In the following, we list the main features of LexO, also in the light of the results presented in the “User Needs” task⁴ of the ELEXIS project: i) it hides all the technical complexities related to markup languages, language formalities and other technology issues, facilitating access to the Semantic web technologies to non expert users; ii) it allows for a team of users, each one with his/her own role (lexicographers, domain experts, scholars, etc.) to work on the same resource collaboratively; iii) it adheres to international standards for representing lexica and ontologies in the Semantic Web (such as ontolex-lemon and OWL), so that lexical resources can be shared easily or specific entities can be linked to existing datasets; iv) it provides a set of services implemented by means of RESTful Web Services that allow software agents to access to the resources managed by means of LexO.

The demo session will be organised in two parts. First, we will introduce LexO and how it falls in the framework of e-lexicography, its architecture, the data model, and how the tool can help users on the basis of their needs. Finally, we will start with the demo software by means of different real use cases.

3. Acknowledgements

This work has been conducted in the context of the cooperation agreement between Guido Mensching, director of the DiTMAO project at the Seminar für Romanische Philologie of the Georg-August-Universität Göttingen, and the Istituto di Linguistica Computazionale “A. Zampolli” of the Italian National Research Council (29 August 2016).

The authors are also supported by the EU H2020 programme under grant agreements 731015 (ELEXIS - European Lexical Infrastructure).

4. References

- Bellandi, A., Giovannetti, E. & Weingart, A. (2018). *Multilingual and Multiword Phenomena in a lemon Old Occitan Medico-Botanical Lexicon*. *Information*, 9(3), 52.
- Francopoulo, G., George, M., Calzolari, N., Monachini, M., Bel, N., Pet, M. & Soria, C. (2006). Lexical markup framework (LMF).
- Khan, F., Bellandi, A. & Monachini, M. (2016). Tools and Instruments for Building and Querying Diachronic Computational Lexica. In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*. pp. 164–171.
- McCrae, J., Aguado-de Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., G.J., Hollink, L., Montiel-Ponsoda, E., Spohr, D. & Wunner, T. (2012). Interchanging Lexical Resources on the Semantic web. In *Proceedings of the Language Resources and Evaluation (LREC) 46(4)*. pp. 701–719.
- McCrae, J., P., B.G., J., Gracia, J., Buitelaar, P. & Cimiano, P. (2017). The OntoLex-Lemon Model: Development and Applications. In *Proceedings of eLex 2017 conference*. pp. 19–21.

⁴ See deliverable D1.1 “LEXICOGRAPHIC PRACTICES IN EUROPE: A SURVEY OF USER NEEDS” at <https://elex.is/deliverables/>