Report on the outcomes of a Short-Term Scientific Mission[1]

Action number: 18209

Grantee name: Elena Montiel-Ponsoda

Details of the STSM

Title: Analysis of terminological resources in the LLOD cloud and validation of the TermLex model

Start and end date: 01/07/2022 to 01/08/2022

Description of the work carried out during the STSM

Description of the activities carried out during the STSM. Any deviations from the initial working plan shall also be described in this section.

The STSM was carried out during the expected time frame. The visitor (Elena Montiel-Ponsoda) was assigned an office at the "Insight SFI Research Centre for Data Analytics" in Galway, where she regularly met the host (John McCrae). Additionally, the visitor attended the weekly Institute research meetings, the monthly seminar organised by the Cardamom project (Comparative deep models for minority and historical languages, see http://www.cardamom-project.org/) -led by the host- and also met with other researchers who are part of the host's research group (the pre-doctoral researcher Shubhanker Banerjee and the post-doctoral researcher Theodorus Fransen).

The visitor and the host reviewed the plan proposed by the visitor and agreed to proceed with Phase I and Phase III.

In Phase I, the visitor proposed to analyse those resources in the LLOD (https://linguistic-lod.org/) classified as "Terminologies, Thesauri and Knowledge Bases". She also performed a more thorough review of other projects whose objective has been the "aggregation" of terminological resources such as EuroTermBank (https://www.eurotermbank.com) or SKOSMOS (http://skosmos.dev.finto.fi/en/).

At that time, Dr. Francesca Frontini and Dr. Fahad Khan visited Insight and gave a talk on research infrastructures for language resources and language technologies. The visitor also had the opportunity to discuss the objectives of her STSM with them, and this encouraged her to also review research infrastructures such as TermCoord (https://termcoord.eu/), CLARIN Virtual Language Observatory (https://vlo.clarin.eu/), ELRC-SHARE repository (https://elrc-share.eu), or SSHOC (Social Sciences & Humanities Open Cloud, at https://marketplace.sshopencloud.eu/) in search for terminological resources.

The host also shared with the visitor some work his team has performed on terminologies for under-resourced languages in the context of a project with "Translators without borders"[2]. The result of this work, in which terminological resources are created by taking advantage of linguistic linked data shared and exposed on the Web, and other initiatives in which the visitor is involved, reveal the need for a review of the Ontolex-lemon model to accommodate new modeling needs. So, after this meeting, the visitor focused on Phase III, in which the idea was to validate with the host the TermLex model, an extension of the Ontolex-lemon model for terminologies, currently proposed to the W3C Ontology-Lexica Community Group for discussion (see https://www.w3.org/community/ontolex/wiki/Terminology).

The visitor is currently working on a paper about TermLex, led by Patricia Martín-Chozas, together with other NexusLinguarum participants (Thierry Declerck, and Víctor Rodríguez-Doncel). Thus, several meetings with the host involved the discussion of this topic and the agreement to validate it with the data resulting from the project discussed in the above-mentioned paper "Enriching a terminology for under-resourced languages using knowledge graphs" by the host and co-authors, and part of the eLex 2021 proceedings.

As for Phase II (devoted to the review of open source frameworks for the creation and management of terminologies that are designed to meet the needs of semantic web and linked data environments), as defined in the STSM Application Form, host and visitor could only briefly discuss some of the tools that exist to the best of their knowledge (VocBench[3], TermitUp[4], Saffron[5]), but no wider comparative analysis was performed.

(max. 500 words)

Grantee enters max 500 word summary here.

Description of the STSM main achievements and planned follow-up activities

Description and assessment of whether the STSM achieved its planned goals and expected outcomes, including specific contribution to Action objective and deliverables, or publications resulting from the STSM. Agreed plans for future follow-up collaborations shall also be described in this section.

Visitor and host agree on the mutual benefit resulting from this STSM, and the achievement of the main objectives, which can be summarised as follows:

1. A report on the current status of the LLOD as for those resources under the category of "Terminologies, Thesauri and Knowledge Bases", and on current initiatives for the identification, aggregation, and publication of terminologies (TermCoord, EuroTermBank, ELRC-SHARE repository, SSHOC, Skosmos, TerminotecaSKOS)
2. The validation of the TermLex modeling proposal with examples of terminological entries from a subset of representative terminological resources that present a varying degree of terminological needs.
3. The consolidation of joint work between the host and the visitor on models and tools for terminologies, so that they can further contribute to the creation, exposure, and consumption of terminological data in the Web of Data. Both are involved in projects that require terminological work and agree on the benefit of further collaboration.

As for agreed plans for future follow-up collaborations, the visitor and the host intent to work on a joint publication of the initiated report on the current status of Terminologies in the Web of Data, which would

---

[2] Enriching a terminology for under-resourced languages using knowledge graphs. McCrae, J. P., Ojha, A. K., Chakravarthi, B. R., Kelly, I., Buffini, P., Tang, G., Paquin, E., & Locria, M. In Proceedings of The Seventh Biennial Conference on Electronic Lexicography, eLex 2021, pages 560–571, July, 2021.
[3] http://vocbench.uniroma2.it
[4] https://termitup.oeg.fi.upm.es/
[5] http://saffron.insight-centre.org/publications/

contribute to the work being conducted in the framework of WP2. This could be presented at a conference in the area (LREC, LDK) or as a journal paper.

The TermLex model is also expected to be discussed within the "Ontology-Lexica W3C Community Group" in the upcoming months.

Finally, they have also identified other projects or PhD works they are currently supervising and which could benefit from further collaboration.

(max. 500 words)

Grantee enters max 500 word summary here.