SHORT TERM SCIENTIFIC MISSION (STSM) SCIENTIFIC REPORT

This report is submitted for approval by the STSM applicant to the STSM coordinator

Action number: 18209
STSM title: Generation, representation and exploitation of terminologies in the Semantic Web
STSM start and end date: 2021-07-01 to 2021-09-30
Grantee name: Patricia Martín Chozas

PURPOSE OF THE STSM:

Objective

This STSM was motivated by the prior cooperation of the applicant group (Ontology Engineering Group) with the Speech and Language Technology lab of DFKI GmbH (Berlin) in the context of the Lynx project and the current cooperation with the Multilinguality and Language Technology lab of DFKI GmbH (Saarbrücken, which is the hosting institution for the STSM) in the context of the Prêt-à-LLOD project and the NexusLInguarum COST Action.

Such collaboration has demonstrated the affinity of the work performed in both institutions, which, in this case, lies in the generation, representation and exploitation of domain-specific linguistic resources (terminologies). The work plan focussed on any of these three tasks, depending on the progress of the research at the time of the stay.

Specific activities
1. Term Extraction
2. Terminology Exploitation
3. Terminology Representation

Expected outcome
- Promotion of open linked data-based models for linguistic data such as OntoLex-Lemon for the representation of terminological data
- Development of a terminology extension for OntoLex-Lemon
- Discussion, validation and consolidation of existent best practices on Linguistic Linked Open Data (LLOD) for terminologies
- Development of novel methodologies for the creation, linkage and improvement of related language resources (LRs)

DESCRIPTION OF WORK CARRIED OUT DURING THE STSM

The STSM took place in the expected time frame, and while the COVID-19 pandemic imposed restrictions, the involved researchers from both institutions (Patricia Martín Chozas, visitor from UPM and Thierry Declerck, host from DFKI) could have on a regular basis meetings, respecting physical distance, whereas the host institution DFKI offered Corona testing facilities before each meeting. The following activities were carried out :
- Joint discussions on the requirements of both sides:
    - DFKI on the representation of terminological data of the Deutsche Bahn (German Railways)
    - UPM on the representation of enriched terminologies generated by the tool TermitUp
    - Both on the representation of traditional terminological resources such as IATE
- Joint writing of technical proposal to cover the above-mentioned needs
- Joint preparation of the presentation of the terminology module to the "Ontology Lexica" W3C community group.
- Contribution by UPM to a paper submitted by DFKI to the TOTh 2021 conference.

DESCRIPTION OF THE MAIN RESULTS OBTAINED:

The mission consisted mainly in discussions to agree on and implement a first version of a standard model to represent terminologies following Semantic Web formats, specifically the OntoLex-Lemon model. The objective was twofold:
➔ Represent every element in "traditional" terminological resources (i.e. every data appearing in IATE entries)
➔ Represent modern terminological resources (those semi-automatically generated from heterogeneous resources, with the help of the TermitUp tools)

The main result is a data model, represented in OWL, generated and validated with the ontology editor Protégé. More specifically, the STSM resulted in:
- A draft specification of the possible terminology extension module for OntoLex-Lemon to be considered by the W3C "Ontology Lexica" community group, which is available here: https://www.w3.org/community/ontolex/wiki/Terminology. This document contains background and motivation to develop this model (Termlex). Its core contributions, classes and properties are described in detail, and a set of real examples applying this proposal to multilingual IATE data.
- Collaboration on a paper to be published in the proceedings of the TOTh Conference
- A presentation given by the visitor (Patricia) and the host (Thierry) during the Ontolex W3C day within the LDK 21 conference in Zaragoza.
- Consolidation of the cooperation between the two institutes
- A presentation of the results of the STSM given by the visitor to the DFKI
- Plans for future cooperation lines

FUTURE COLLABORATIONS

- Continue collaboration between the two institutes
- Started work on a publication to be submitted to a conference (e.g., LREC conference) and/or to a journal (e.g. Terminology)
- Set a series of teleconferences with the W3C "Ontology Lexica" community group for the official approval of the Termlex module.