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

Action number: CA18209

Grantee name: Giedre Valunaite Oleskeviciene

Details of the STSM

Title: Researching integrated explicit and implicit offensive language taxonomy

Start and end date: 05/07/2022 to 10/07/2022

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.

(max. 500 words)

Grantee enters max 500 word summary here.

The aim of the STSM was extending the available resources and providing the taxonomy for linguistic processing of offensive language identification working on refining the annotation model of explicit and implicit offensive language tagging by cooperating with the team of researchers from Centre of Linguistics of the University of Porto who are engaged in the scientific research on lexical resource construction, lexicons, text and speech, translation, computational linguistics, etc. The Centre of Linguistics of the University of Porto joined the initiative of creating an integrated model of explicit and implicit offensive language taxonomy and carried out annotation experiments on offensive discourses in selected English social media materials which the research team is working on in Use Case 4.1.1. The aim of the STSM was closely related to the main aim of NexusLinguarum action pointed out in the Memorandum of Understanding, which is construction of multilingual and semantically interoperable linguistic data extending the set of available resources. It was also related to the objectives of work group 4 working on the use cases and focusing especially on offensive language in media texts and work group 2 working towards low-resourced languages.

The aim of the STSM was achieved by analysing the offensive discourses on the publicly available 25 web-based hate speech datasets and performing the semantic annotation experiment in the INCEpTION tool. The steps of the research performed included the following. First, the integrated model of explicit and implicit offensive language taxonomy was applied to perform the semantic annotation on web-based hate speech datasets in the INCEpTION tool. Then, the experiment data

---

**Funded by
the European Union**

was processed refining our annotation categories. The results led to the more refined offensive explicitness and implicitness categorization criteria.

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.

(max. 500 words)

Grantee enters max 500 word summary here.

The main result of the STSM is the refined integrated model of explicit and implicit offensive language semantically interoperable and applicable in multilingual contexts. We are intensively cooperating with other researchers from our COST action who expressed their interest in working on the use case of offensive language identification. The resulting explicit and implicit offensive language taxonomy is one of the inputs working towards the objectives of NexusLinguarum action pointed out in the Memorandum of Understanding, which is construction of multilingual and semantically interoperable linguistic data extending the set of available resources. It is also an input towards the objectives of work group 4 working on the use cases and focusing especially on offensive language in media texts Also, the results of the experiment of refining offensive explicitness and implicitness categorization were presented at the Braga conference on hate speech research (https://sites.google.com/view/hate-speech-conference/call-for-papers) and also, they are going to be presented at our COST organized conference in Vilnius September 2022 and other conferences.

Joining the team of researches of Centre of Linguistics of the University of Porto and working on offensive language taxonomy building and analysis research was a great opportunity for me to grow as a researcher and expand the set of available resources in Lithuanian language looking for the ways making the resources co-searchable.

The collaboration with the host institution will continue. We are expecting to continue analysing offensive language explicit and implicit contexts and refining the offensive language taxonomy. We are aiming at a paper presenting the results of research within this STSM and it will be sent either to the prominent journals or conferences/workshops in the field. We are also planning to present our work at the following COST events.