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

Action number:

Grantee name:

Details of the STSM

Title: Neural Premise Selection for Agda

Start and end date: 05/09/2023 to 12/09/2023

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.

We manually reviewed the prototype model's prediction in juxtaposition with the gold-standard programs, and identified fundamental issues related to the problem statement. Concretely, the model works as well as it possibly ever could for ad-hoc programs, which are under-specified with respect to their contents. In light of this, it might make sense to adjust the problem statement and confine our proof search to more constrained propositions, i.e. proofs of properties of programs. Further, we performed a manual inspection of the data pipeline prior to the model, and found a collection of minor bugs that might have cascading performance effects (including inconsistent de Bruijn indexing, underspecified Sort and Level types, operator overloading, etc.). Finally, we discussed alternative ways to align the model with the data structure, alternative (1) being to use de Bruijn indices as real indexing operators in the NN rather than ad-hoc learned embeddings and alternative (2) being to drop type structure altogether and revert to a simplified system that only handles co-referencing, the idea being to use the latter as a naive baseline.

**Funded by
the European Union**

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)

The STSM was successful in multiple ways. First, it allowed us to assess the behavior of our prototype and to confirm its acquisition of Agda's syntax, regardess of the raw perofrmance metrics which can prove misguiding. It allowed us to identify potential issues in what happens outside the model, and the ways to alleviate them.Even though these are not the targeted outcome of a fully-functional interface and a finalized paper, they certainly bring us closer to them.

In light of the above, our plans are to (1) apply another iteration of error-correction, data sanitation and network optimization to evaluate our progress (2) open-source and release a version of the current prototype in the short-term future so as to extend an invitation to other interested parties, (3) write and publish a technical report detailing our efforts and (4) continue our active collaboration in the coming months with the end-goal of a high-quality publication and a fully-fledged model.