SHORT TERM SCIENTIFIC MISSION (STSM) – SCIENTIFIC REPORT

The STSM applicant submits this report for approval to the STSM coordinator

Action number: CA16105
STSM title: (Some) Ethics by design solutions for language learning platforms
STSM start and end date: 19/11/2017 to 01/12/2017
Grantee name: Karën Fort

PURPOSE OF THE STSM/

The STSM was to address an analysis of the possible ethical issues regarding different scenarios of usage of the enetCollect "platform" and provide ethics by design solutions to at least some of them.

In order to build a really actionable ethical solution for the type of language learning platform enetCollect intends to specify, we need more than mere guidelines, we need ethics by design solutions, i.e. solutions that will be put in place as soon as the specifications are written. The solutions are therefore developed within the platform itself. An example of such a solution to some of the anonymity issues we will face may consist in encouraging users, on registration, not to use their real name, but a pseudo (and suggest practical ways to find a nice one) and to check that their pseudo does not correspond to their Email.

The STSM purpose was to investigate different (not all, as they are not documented yet) platform usage scenari, to identify ethically sensitive areas, and to propose ethics by design solutions to the issues that may arise. Its focus is first on the general platform architecture (User(s) registration and profile(s), data creation, interactions between users), then on some language learning exercises.

For this, we needed to read a lot of research papers, to observe a number of language learning platforms and to interact with people from language learning working on such platforms.

DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

During these 12 days, of which 9 were working days, I (Karën Fort, WG5 leader) worked in collaboration with Branislav Bedi (WG3 leader) in Reykjavik, Iceland. We met almost every day. We also met with other people from the University of Iceland to discuss some issues and benefit from their experience in online language learning and language resources creation and to investigate further collaborations.

Our work was divided into two major parts.

First, we started gathering material about ethical issues in language learning and natural language processing. This includes research papers, books, but also online videos. This review is far from being finished, as a lot has been written about this subject in the last years. Besides, it occurred to us that we should also read about ethics in crowdsourcing platforms, in particular Wikipedia, and this represents a large volume of papers to be

examined.

Second, we organize several sessions of brainstorming about the platform: what it could look like, what types of users might interact with it and with themselves, what types of exercises could be used, etc. This allowed us to identify some ethical "hotspots" (risky features or actions) and potential issues for which we tried to imagine ethics by design solutions.

Branislav organized a meeting for us on November 28th with the person responsible for Icelandic Online platform (https://icelandiconline.com/), Kolbrún Friðriksdóttir, with whom we discussed her experience with a huge number of participants (more than 170 000) and the ethical issues that might have arisen. We also discussed potential collaborations and encouraged her to join the next action meetings, as she is a MC for Iceland.

It was also decided before the STSM started that I would do a presentation to Branislav's research team (http://rim.hi.is/). I decided to present them some of my work on crowdsourcing (*The scope of language intuition: some insights from crowdsourcing,* see http://rim.hi.is/bitabox/) on my last day, November 30th.

Finally, I organized a meeting with researchers from the tagged Icelandic corpus team (http://www.malfong.is/index.php?lang=en&pg=mim), in particular Sigrún Helgadóttir and Kristín Bjarnadóttir, on November 24th, to present them ZombiLingo (dependency syntax), Bisame (POS tagging) and Rigor Mortis (multi-word expressions). They showed some interest and could be willing to collaborate on a Game with a purpose for Icelandic multi-word expressions or to adapt ZombiLingo.

DESCRIPTION OF THE MAIN RESULTS OBTAINED

We gathered a list of scientific references on ethics and natural language processing and ethics in language learning applications and platforms. We identified in particular a special issue of the journal of learning analytics on the subject (http://epress.lib.uts.edu.au/journals/index.php/JLA/issue/view/373) and the ethics in NLP 2017 workshop (http://ethicsinnlp.org/ethnlp-2017).

These references are for now stored in a Framapad, and will soon be compiled in a scientific report. This report will also contain definitions and references concerning ethics in general (consequentialism vs deontology) and ethics by design in particular (at least as an adaptation of Privacy by design), in order to give some more theoretical background on the subject.

We will add to the report the results of our brainstorming sessions (also stored in the Framapad), with the first bits and pieces of specifications of the platform, accompanied with the solutions we propose. Needless to say that this part of the report will have to be updated regularly during the action, if new features, new users, new usages are imagined. One of the main outcome of our sessions is the need for an ethical review board (ERB).

From the final version of the report, we should be able to extract the guidelines for ethics (and maybe part of the ones for legal issues) which are a deliverable for WG5. The report itself will be presented in its first version during a meeting of the action (probably in September 2018).

The fact that we come from very complementary backgrounds for enetCollect (language learning for Branislav and crowdsourcing and NLP for me) helped us to progress rapidly on the issues, with very different views on the platform. This has been a very rich and productive STSM.

FUTURE COLLABORATIONS (if applicable)

Apart from the scientific report we are putting together, Branislav and I consider writing a journal article on how gamification and crowdsourcing can help building an ethical language-learning platform, a subject that has never been addressed before.

Also, the meeting with Kolbrún Friðriksdóttir allowed us to discover we have a number of common research subjects, in particular around retention of the participants and we will probably gather our data to write something together on the subject.

Finally, as already pointed out, the Icelandic corpus team showed some interests in some games with a purpose we developed in my team.