

TEMPLATE

Output factsheet: Trainings

Version 1

Project index number and acronym	CE 1569 - ProsperAMnet
Lead partner	University of Applied Sciences Upper Austria
Output number and title	O.T2.2 - Training of Annotation for the Artificial Intelligent Programs
Responsible partner (PP name and number)	University of Szeged - PP 9
Project website	http://prosperamnet.eu/
Delivery date	03.2021

Summary description of the implemented training measure(s), explaining the specific goal(s) and target groups

The goal of the Export Service Radar is to automatically analyze company websites to make automatic predictions for the kinds of services offered by the companies. The users can provide the list of URLs of the companies they are interested in regarding their service portfolio. As users might not be familiar with the corporate landscape of a new region they are about to expand, there is also a possibility for the users to enter prototypical companies (probably from their home market that they are more aware of) and ask for a summary of the services offered by similar companies at some target market (selected via geolocational constraints).

The aim of the training is to get the annotators familiar with the annotation process and the domain knowledge of the different service categories. The goal for the annotators is to collect sufficient good training and validation data (the annotations of company webpages) to train the AI system through NLP and Machine Learning to collect the correct data on its own in the future. The first training in a workshop format was planned to be held in person at the University of Szeged, but due to the Corona pandemic and travel restrictions, the training mode has been changed. There was a series of online meetings held with each annotator individually. In every session a brief project overview was given, the goal for the annotations was explained, and the template of the annotation scheme including an example of an annotation was discussed. The template also included step-by-step instructions on the annotation process and a description of the service categories. The trainer also annotated examples live together with the annotators. In order to ensure a high quality of the annotations, every annotator had to finish the same set of five test annotations. In the feedback meetings the trainer clarified misconceptions regarding the service categories, answered questions which came up during the annotation process, and shared several good practice examples.

A second phase of annotations was introduced in order to target service categories where more information was still needed. Therefore, a new guideline document for collecting the annotations was created alongside a document with instructions regarding the second phase of annotations. The process of individual training and feedback sessions between the trainer and annotators was repeated in the second phase.

NUTS region(s) where training(s) have been conducted (relevant NUTS level)

The trainings have been conducted in an online mode (skype or MS Teams). Primarily the partners FHOÖ (PP1/LP) and University of Passau (PP2) held the training sessions with annotators of all project partners beside PP8 TGZ Bautzen. Therefore, each involved CE-country resp. NUTS region was included:

AT314 Steyr-Kirchdorf; AT312 Linz-Wels
DE222 Passau, Kreisfreie Stadt; DED2C Bautzen
HU211 Fejér; HU333 Csongrád
SI021 Osrednjeslovenska
CZ031 Jihočeský kraj
ITH42 Udine
SK010 Bratislavský kraj

Expected impact and benefits of the trainings for the concerned territories and target groups

The Service Export Radar can provide an approximate summary of the service landscape of some geographic region based on the automated analysis of company websites at scale. This information can help companies in planning their strategies on new target markets to export their services for.

As the project-BSOs in the project countries ultimately are trained to become Expert Hubs for the usage of the Service Export Radar (O.T2.1), the active participation in the development of the radar is important. In that regard, being part of the collection of the annotations is essential in terms of understanding how the tool works and how an Artificial Intelligent based tool can be developed in general. This empowers the project-BSOs to promote the radar and the use of Artificial Intelligence appropriately well in their country and to support companies in using the radar. Additionally, in the annotation process there was a transfer of knowledge regarding the domains of the service categories. For example, what different types of services are commonly developed in a B2B context, in the transformation of manufacturing companies to become more servitized.

Sustainability of the training(s) and developed training material(s) and their transferability to other territories and stakeholders

The database of companies we created can be updated regularly so that it can reflect up-to-date information, and the machine learning solutions we created were obtained in a way that ensures their easy expansibility. The computing resources necessary for training and making predictions are readily available at our disposal, hence, there is no dependence on our side on any third party services, which helps the sustainability of the proposed framework.

The training and its material support the effort of the sustainable knowledge transfer from the academic institutions towards the BSOs regarding the domain knowledge of the different service categories as well as the general project learnings regarding the development of an AI solution. Additionally, the market inputs from the BSOs in close cooperation with the local companies, serve as a viable sustainable source of information on the current trending topics in the concerned industries and the usability of the AI tool.

Lessons learned from the development and implementation of training measures and added value of transnational cooperation

We have identified a series of domain specific peculiarities with respect to dealing with corporate websites. Our main findings were related to the multilingual processing of webpages and the importance of relying on hybrid forms of representations, i.e., symbolic and vectorial representations of natural language utterances. It is important to provide regular and detailed feedback sessions to the annotators to share good practice examples from other annotators and to constantly clarify misconceptions regarding the domain knowledge. The quality of annotations is essential to the development of the tool, as the Artificial Intelligent solution can only be as good as the provided training and validation data. In that regard, it is important that the annotators not only know the task and how to progress through the steps, but also understand the greater context of what they are doing and what is the purpose. Especially in the online format the interaction between trainers and annotators are limited compared to a meeting in person. That is why showcasing live examples and review questions, to confirm the understanding of the task, becomes even more important.

References to relevant deliverables and web-links If applicable, pictures or images to be provided as annex

The fine-grained details of the development of the annotation process can be reviewed in the deliverable D.T2.3.2. Furthermore, this output is directly linked to O.T2.1, as the annotations are the basis for the AI development. This output is also linked to D.T2.3.1 AI Programming as the results of the annotations are considered when programming the AI solution. The project website (prosperamnet.eu), the contents at NextCloud (folder T2) and a standalone prototype of the automatic service discovery tool (http://staging.prosperamnet.eu/service_classification/cv_2_2/bin_3_5gram_per/).