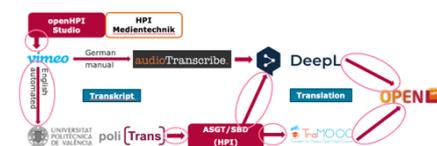


# Developing a Pipeline for the Automated Transcription and Translation of Videos

Enabling Multi-Lingual Courses on openHPI, openSAP, and OpenWHO

## MOTIVATION

On our digital learning platforms—openHPI, openSAP, and OpenWHO—all courses are offered in a certain language, e.g. German, French, or English. Particularly, in the case of OpenWHO, courses often need to be provided in more than one language—English and French and additionally in several native African languages. Here, the automatic generation of subtitles in different languages would reduce the course production effort a lot. The rise of machine learning technologies during the recent decade has provided the world with many opportunities to automate these translation tasks. The challenge, however, is that these systems are not really connected yet. From the moment when a video is uploaded to the Internet to the presentation of the video in a course—including transcript and translations—many manual tasks have to be done by the course administrators. The tools for each of these



steps are offered by different providers. Videos, e.g., can be hosted on Vimeo, Azure, or Youtube. We use several tools to generate the video transcripts. The transcripts are post-processed by a tool that has been developed at the HPI and is also hosted here. Translations are offered by tools, such as DeepL, TraMOOC, or Google.

## PROJECT GOAL

Develop a tool that offers an easy and intuitive user interface and allows to configure each step of this pipeline. The providers for each step need to be easily configurable. The pipeline provides an end-to-end solution and relieves the course administrators from the in-betweens.

## APPROACH

The tool to be developed will be a standalone tool that connects existing services. Although the context is in the realm of machine learning, the actual tasks are not. To connect the existing services, the tool to be developed will have to provide interfaces that are able to communicate with the interfaces of the services to be connected. An agile development process that includes the end user (platform/course admins) and service providers is key.

## OUR EXPECTATIONS

You should be able to work in a team and be open for agile development methods, co-innovation with customers and other stakeholders. Previous knowledge of web protocols, such as REST or SOAP and a web programming language, such as Ruby or Python would be helpful.

## CONTACT

**Thomas Staubitz**  
H-1.37  
thomas.staubitz@hpi.de  
0331-5509-528

**Sebastian Serth**  
H-1.17  
sebastian.serth@hpi.de

**Joseph Bethge**  
H-1.21  
joseph.bethge@hpi.de

**Prof. Dr. Christoph Meinel**  
H-1.5  
christoph.meinel@hpi.de