15-532-302

Date: 09.04.2022

Name: Pascal Emmenegger

Keywords: Docker Setup Service Predict, Refactoring Service Predict, Code Fetcher, Service Train, Docker Compose Configuration, MongoDB

Overall Progress: 3

Activities completed:

- US3a: Train Service implemented which is gonna be executed regularly (every 5 minutes) as cron job and does the training if a training batch size is reached for every language. Research has shown that a batch size of 100 sample for a training session is appropriate. We always store a better model on our MongoDB.
- US3b: Splitting training data in train (80%) and validation (20%). Preparing the dataset for the train service (shuffling data, setting epochs, etc...)
- Predict service containerized with Docker and entry in docker compose added
- Code Fetcher implemented which downloads code files from GitHub repositories with the most stars. We use it for testing the training service (US3a & US3b & US12)
- Documentation improved

Current activities:

- US4: Research about how to best switch models that we satisfy US4 and US5 (no downtime)
- Finding out a smart way how to further improve our Docker setup with hot-reloading or automated pushing and pulling images from docker hub. Currently Docker is only running locally on every computer.
- Transferring know-how to Michael Blum on how to train the model manually. He is responsible for the upcoming demo which should be based on an already trained syntax highlighting model (US11 & US12)
- Updating wiki and documentation

- Research about best practices of flask setup with unit and integration tests

Goals or Activities for next deliverable:

- US4: making sure that the user is always interacting with the most accurate model (smallest loss) so that the syntax highlighting will be as accurate as possible. I am doing this together with Nicola.
- US12: possibility to train the model manually on a definable set of code/projects in the three supported languages so that I can verify that the training process works and the model’s accuracy is improved. I am doing this together with Nicola.

Immediate attention: -