

Textify

Project goals

A web-based tool specifically designed to facilitate the entire process of digitizing and translating printed content. Using optical character recognition (OCR), it can accurately extract text from scanned images, for example from blackboards, books or handwritten notes. The recognized texts are then translated into the desired language, which significantly increases accessibility for users of different language groups.

In addition, the tool offers a variety of features that increase the utility and flexibility of the application: a text-to-speech function reads the digitized and translated texts aloud, supporting people who prefer listening to reading. The speech-to-text option makes it possible to convert spoken content into text, which is helpful for the documentation and further processing of spoken information.

In addition, the tool has various export options that allow users to save the results in different formats. It is therefore ideal for students, teachers, researchers and all those who regularly work with printed or handwritten texts and want to reuse them in digital form.

Potential Risks

General risks relating to the models

Difficulties in selecting the right models. There is a lot of choice but most of the open source models are of modest quality.

Handling of the models.

Some models may require unexpectedly complicated handling. E.g. through a lot of customization.

Specific risks

Accuracy and reliability of OCR technology

Low-Quality, Distorted Images or complex images, handwritten text, complex and unusual fonts, especially blackboard images and sketches with free-standing text or non-standard and complex layouts can lead to incorrect text recognition.

Quality of the machine translation

Machine translations are often prone to errors (Accuracy of Translation, Fluency and Naturalness, Handling of Idiomatic Expressions, Domain-Specific Accuracy, Preservation of Meaning and Nuance). Especially with less widely used languages.

Quality of the text to speech models

Here too, the quality in terms of natural voices and intonation is bound to be modest, based on previous experience.

Processing speed and computing power

Ki models require a lot of processing power, depending on their size. We may have to make a trade-off between quality and processing time. Especially if the input data is very large.

Selection of suitable web technologies

Server software (web server, REST-API-Framework) for the backend, a UI framework for the frontend and other tools such as a database and different library for different use cases, must be selected. Their handling and implementation is a general risk.

Project scope

- Requirements analysis and making a plan.
- Preparation of the development environment for the use of the selected technologies (libraries, frameworks) and basic tools (git)
- Tests, validation and selection of models for the different modules:
 - OCR, Picture in picture recognition, translations, Speech synthesis (TTS and Speech2Text)
- Tests, validation and selection of server technologies, frameworks and libraries to use
- Implementation
- Tests of the results, Documentation

Team makeup

Anton Nikolai, Christian Jaide, Joel Koch, Aaron Sonntag

responsibility for each member (Scrum workload)

Anton Nikolai: Backend (L), Speech to Text (S), Text to Speech (M)

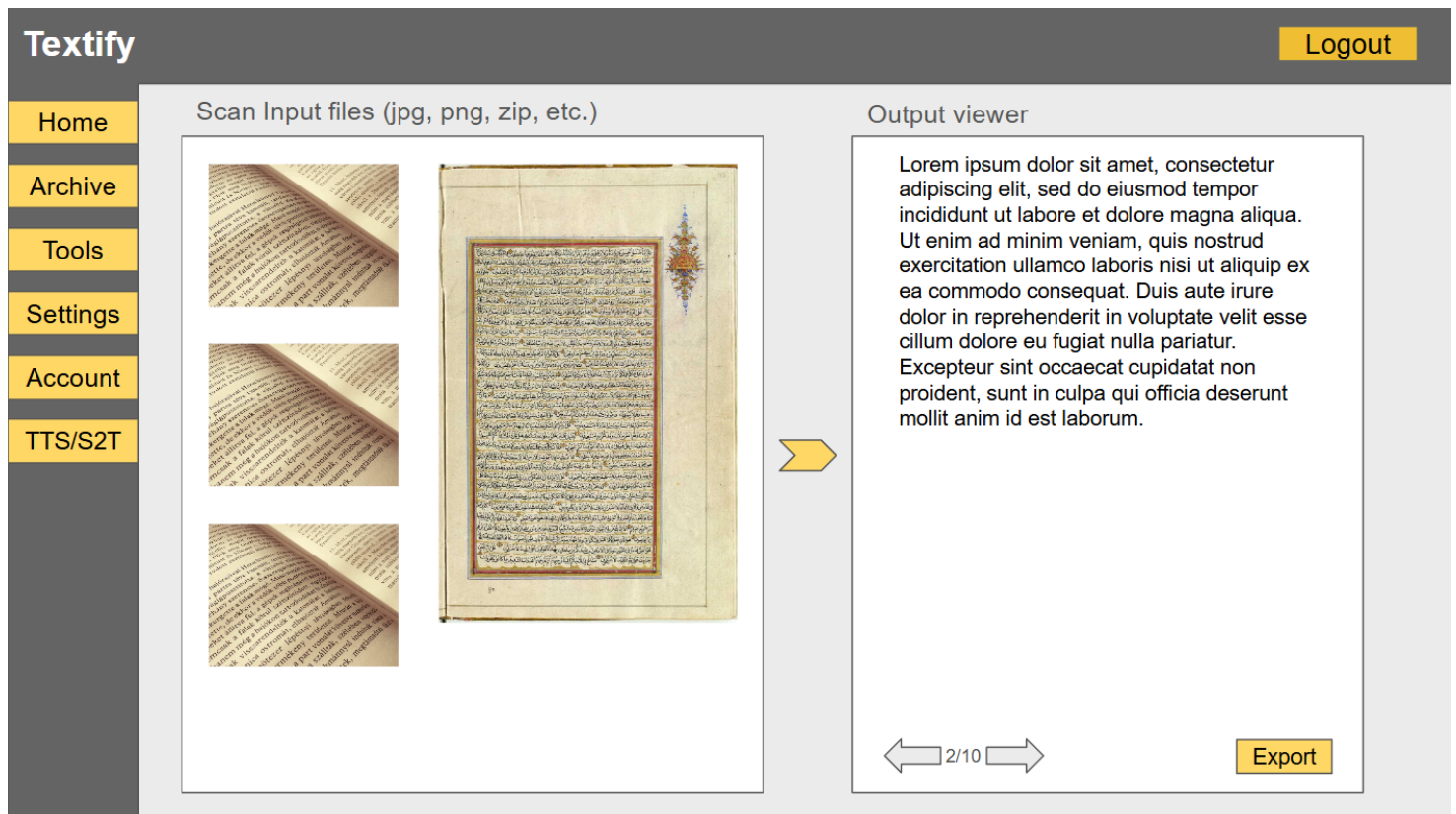
Christian Jaide: Frontend (M), Rest-API (S-M), OCR (M), Exports (S)

Aaron Sonntag: Backend(M-L), OCR(M), Vision(S-M)

Joel Koch: translations (S), picture in picture (M-L), Exports(S)

The choices of the tasks are based on the weaknesses of each team member. So everybody has a task where each member has experience and everybody has a task which he has no experience with.

How it could look like:



Timeline:

