



DigiEduHack Solution

Berlin - #SemesterHack 2.0 (in German & English)

Challenge: Berlin - #SemesterHack 2.0 (in German & English) Challenge 2020

Automatic Diploma Digitalisation

Light at the End of the Paper Tunnel

The aim is to make the digitization of certificates much more efficient. To this end, an application is to be developed that will make it possible to read out image or PDF files of certificates and to transfer the most important data in an XML format that conforms to the XHochschule standards.

Team: Automatic Diploma Digitalisation

Team members

Dobin Dietrich, Daniel Schmedes, Felix Bitterer

Members roles and background

Robin Dietrich, Consultant at [init] in the project XHochschule

- Original Idea and Python Development

Daniel Schmedes, Consultant at [init] in the project XHochschule

- Frontend Prototype Development

Felix Bitterer, Team member of the project „Digital Mobil @ FH Bielefeld“ at University of Applied Science Bielefeld

- Documentation and Quality Controll

Contact details

robin.dietrich@init.de, danielsamuel.schmedes@init.de, felix.bitterer@uni-bielefeld.de

Solution Details

Solution description

Within the framework of the Hackathon "#Semesterhack 2.0" of the Hochschulforum Digitalisierung on November 12 and 13, 2020, the team has already developed two prototypes:

1. a demo for the graphical user interface, how the applicants* could use the application later on.
 2. a first program that is able to read information from images of sample certificates.
2. A video entitled "Automatic certificate recognition - prototype demo" was created for better understanding. We would like to refer to it here in order to better understand the functionality of the programs.

Progress:

After the use of the program is demonstrated on a so-called "click dummy", a demonstration of text recognition and classification is given using a working prototype.

The prototype was written using the open source software Tesseract. This text recognition program was developed by Google, but is available for free use under an Apache 2.0 license.

With the version shown in the video, it is possible to extract data from two sample testimonials and return them as a "dictionary" object. A further processing of the data into the XHochschule XML standard should take place in a later process step.

Findings:

The prototype shows that the program is technically implementable and will produce the desired results with appropriate preparation. The technology used allows potentially large time savings for students and university administration.

However, the script shown will only work for certificates that are structured according to a certain scheme. In order to apply the program to other certificates, the reading function would have to be extended to include additional search patterns and, ideally, semantic plausibility checks of the data.

Solution context

The goal is to standardise student data exchange in the national higher education system so that the necessary interoperability between higher education institution systems (called Campus Management Systems) can be established for the media-consistent processing of future digital administrative services.

Solution target group

Students

- Alumni
- National students (MVP)
- International students (Complete product)

Universities (MVP)

- examination offices
- Student secretariats and services
- International Offices

Solution impact

Students can change between universities more easily, university administrations have more capacity to take care of important matters.

Solution tweet text

A light at the end of the Paper Tunnel: The #XHochschule Project is coming along
<https://youtu.be/RGzclgtpnWA> #studentmobility #ozg #XBildung #tesseract

Solution innovativeness

Updating the university application process using Machine-Learning based Text Recognition

Solution transferability

The solution is based on freely available software and is easily replicable for anyone with a basic grasp of Python

Solution sustainability

Automatic document processing can reduce the need for printing material on paper

Solution team work

We worked in a SCRUM-like Fashion