Call for Master Thesis

Stamp Detection and Segmentation in Historical Documents

With the increasing amount of digitised documents, automatic document analysis is gaining more and more attention. Despite all the progress in the field, historical documents still pose challenges to document analysis systems. They differ from modern documents in many aspects, such as digitisation format, paper quality, degradation, font type, and amount of noise.

An information retrieval pipeline can benefit greatly from the detection of stamps, especially in the context of digital cultural heritage, as stamps usually contain information relating to creation, distribution, and storage of the documents [1]. The goal of this thesis is to detect stamps on historical documents using a combination of computer vision and machine learning techniques and algorithms.

In this thesis:

1. The existing approaches of stamp detection will be reviewed and investigated.
2. A synthesised dataset with pixel-level labels will be created for historical documents.
3. A deep learning-based pipeline for stamp detection and segmentation will be adapted and applied.
4. The results will be evaluated based on ground truth data.

This thesis will be supervised by **Prof. Dr. Harald Sack, Mahsa Vafaie and Russa Biswas, Information Service Engineering at Institute AIFB, KIT, in collaboration with FIZ Karlsruhe.**

Keywords:
Machine Learning, Computer Vision, Pattern Recognition

Pre-requisites:
- Good programming skills in Python
- Basic knowledge of Machine Learning
- Familiarity with Computer Vision

Contact persons:
Mahsa Vafaie
mahsa.vafaie@kit.edu
Russa Biswas
russa.biswas@kit.edu

[1] Younas, Junaid, et al. "D-star: A generic method for stamp segmentation from document images." *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. Vol. 1. IEEE, 2017.