



VE/Amt: Nationalbibliothek (NB)



Kontaktperson: Oliver Sievi



Geschäftsprozess & -kontext:



Beschreibung der Challenge:



Die NB betreibt und leitet den Kollektivkatalog des Vereins Schweizer Plakatsammlungen SPS. In diesem Kontext erschliesst die NB Plakate sowohl inhaltlich (was ist abgebildet) wie auch formal (Text, Grösse, Material, etc.).

Basierend auf den existierenden, katalogisierten Plakaten soll ein ML System entwickelt werden, welches in der Lage ist, Schlagworte für ungesehene (nicht in den Trainingsdaten vorhandene) Plakate zu generieren.

In einem Prototypen wurden kommerzielle Methoden (Amazon, Google) für die Bild- und Texterkennung getestet. Dabei zeigte sich, dass diese Methoden grundsätzlich einsatzfähig wären, sich jedoch aufgrund des Urheberrechts Probleme für die Integration/Verarbeitung von Plakate von nach 1913 ergeben. Diese dürfen nicht auf externe digitale Infrastruktur übermittelt und verarbeitet werden.

Dabei kann auf die Gemeinsame Normdatei (GND) zurückgegriffen werden, das System soll jedoch auch in der Lage sein „neue“ Stichworte zu generieren (d.h. Worte welche nicht im Schlagwortkatalog enthalten sind). Eine zweite Challenge wäre auch die Erkennung von abgebildetem Text auf den Postern (OCR). Hier ist eine der Schwierigkeiten die Diversität der Schriften sowohl in Typographie wie auch Anordnung des Textes.

Wegen Vorgaben des Urheberrechts muss die Lösung on Premise laufen. Die Zeit für die Generierung von Stichwörtern ist dabei nicht von höchster Relevanz.

Zusätzlich soll ein User Interface entwickelt werden, welches Anwender*Innen durch den Prozess leitet. Anwender*Innen sollen Stichworte welche in den Katalog übernommen werden manuell auswählen können.

Entwicklungsstand der Challenge (nicht zutreffenden Text streichen):



- Challenge mit der Führungslinie abgestimmt (keine Ressourcen für Folgeauftrag verfügbar)
- Challenge prototypisch umgesetzt (wegen rechtlichen Einwänden nicht in der Cloud umsetzbar)