Socface to projekt badawczy uruchomiony we wrześniu 2021 r., którego realizacja zakończy się w marcu 2025 r. Projekt ma na celu zbadanie zmian zachodzących w społeczeństwie francuskim w okresie stulecia 1836 - 1936 w oparciu o zachowane i zdigitalizowane spisy ludności, przechowywane w archiwach miejskich i departamentalnych. W ramach projektu, zdigitalizowane spisy ludności, za pomocą sztucznej inteligencji, będą automatycznie transkrybowane i przenoszone do bazy danych. Baza ma zawierać docelowo 700 milionów rekordów opisujących żyjących we wskazanym czasie ludzi. Ta olbrzymia ilość rekordów znajduje się obecnie na 15 milionach skanów rozproszonych w ponad stu archiwach. Wykorzystując fakt powtarzalności źródła stworzone zostaną modele automatycznego przetwarzania obrazu w taki sposób, aby wyodrębniać w obrazie źródłowym tabele, kolumny i wiersze, a wreszcie rozpoznać tekst w tak aby identyfikować jego znaczenie logiczne (nazwa, wiek, miejscowość, zawód itd.) Tak pozyskane informacje będą przechodziły walidację w oparciu o testy przygotowane przez archiwistów, historyków i demografów. Baza ma pozwolić nie tylko na badania nominatywne, genealogiczne, ale ma umożliwić także prace z zakresu demografii, historii gospodarczej i społecznej (np.: zmiany na rynku pracy, mobilność społeczną, różnice społeczne itp.).

Projekt kierowany jest przez Narodowy Instytut Studiów Demograficznych (INED - l'Institut national d'études démographiques i firmę Teklia; wspiera go Narodowa Agencja Badań (ANR - l'Agence nationale pour la Recherche, we współpracy z Międzyministerialną Służbą Archiwów Francji (SIAF - Service interministériel des Archives de France) oraz Paris School of Economics.

Więcej:

France Archives: Socface, l'intelligence artificielle appliquée aux archives

TEKLIA: Automatic recognition of 100 years of French Census: the SOCFACE project

Adam Baniecki
AP Wrocław O/Bolesławiec