

Wie bekommt man die KORHB APP?

Die KORHB App steht ausschließlich den Kommunen der KO.R-Gemeinschaft zur Verfügung. Die Mitgliedschaft ist kosten-

los und kann relativ kurzfristig beantragt werden. Der Beitritt zu KO.R erfolgt durch eine schriftliche Erklärung gegenüber

der betreuenden Stelle, die unter <https://www.staedtestatistik.de/arbeitsgemeinschaften/kosis/kor> zu finden ist.

Daniel Schürmann

Notiz aus der der KOSIS-Arbeitsgemeinschaft KO.R

Ableitung des Migrationshintergrunds – R-App zur Plausibilisierung des Merkmals Geburtsland

Die Ermittlung des Geburtslandes (Merkmal P10) wird üblicherweise in MigraPro über eine manuell erstellte Abgleichliste durchgeführt. P10 wird anschließend gemeinsam mit anderen Merkmalen verwendet, um den Migrationshintergrund für eine Person abzuleiten. Bei spezielleren Auswertungen, wie z. B. Personen mit spanischer Staatsangehörigkeit und marokkanischer Herkunft, ist ein gut gepflegtes Merkmal Geburtsland essenziell.

Beispielsweise waren die Dortmunder Daten dahingehend weiter zu qualifizieren. Um den manuellen Plausibilisierungsaufwand gleichzeitig zu verringern, wurde ein Verfahren zur teilautomatisierten Bestimmung des Geburtslandes mit R entwickelt.

Grundlage hierfür ist die frei verfügbare Geonames-Datenbank, die um weitere

recherchierte Angaben (z. B. aus CompGen und Wikipedia) zu Namen von ehemaligen deutschen Gebieten sowie Angaben zu Gemeindegemeinschaften (über das Bundesamt für Kartographie und Geodäsie) erweitert wurde. Zur Ableitung des Geburtslandes werden die Merkmale erste und zweite Staatsangehörigkeit (P05 und P07), Zuzugsgebiet (W40) sowie Geburtsort (P12) herangezogen.

Von der Textangabe zum Geburtsort zu einem plausibilisierten Geburtsland: Die Geburtsorte sind als Freitexte im Melderegister hinterlegt. Um trotz Falschschreibweisen, etc. eine Zuordnung zu einem Geburtsland zu ermöglichen, wird für jede Kombination aus Geburtsort, P05, P07 und W40 einerseits sowie dem um Einwohnerzahl und Entfernung zur Basisgemeinde erweiterten Geonames-Katalog andererseits ein ge-

wichteter Ähnlichkeitsscore errechnet. Überschreitet dieser Score einen Schwellenwert, wird der Eintrag mit dem höchsten Score als Treffer verwendet.

Um die Berechnungszeit zu reduzieren, kommt ein mehrstufiges Verfahren zur Anwendung. So werden die Einträge zunächst auf direkte Übereinstimmung bei der Schreibweise, dann auf partielle Übereinstimmung und schließlich auf ähnliche Schreibweisen (fuzzy) überprüft.

Für die Geburtsorte, für die der Algorithmus keinen Treffer mit einem hinreichend großen Score erzielt, wurde eine R-Shiny-App entwickelt, die eine effiziente manuelle Zuordnung anhand von generierten Vorschlägen ermöglicht.