

Ausgangssituation

Im Workflow- und Dokumentenmanagementsystem MARiS zur Vorgangsbearbeitung im Asyl- und Dublinverfahren galt es, die Suche für Namen und deren verschiedene Schreibweisen zu verbessern. Für das BAMF realisierte exorbyte in einem Proof of Concept und erweiterten Probebetrieb die fehlertolerante Suche in MARiS-Daten und die Identifikation von Dubletten. Der PoC stand im Zeichen der Qualitäts-offensive des BAMF.



Fehlertolerante Ähnlichkeitssuche in Stammdaten

Mit sicheren Ergebnissen aus der Personensuche in die Fallbearbeitung einsteigen.

Für ein zuverlässiges und schnelles Finden von Datensätzen zu einer Person, die außerhalb biometrischer Verfahren liegen, wurden über die Technologieplattform [matchmaker](#) Stammdaten für die Ähnlichkeitssuche nutzbar gemacht. Dies erleichtert die Fallbearbeitung und vermindert Risiken durch das Nicht-Finden von heterogenen Personendaten.

Ähnliche Personendaten finden

Für den Einstieg in die Fallbearbeitung in MARiS kann der Anwender über [matchmaker](#) Suchbegriffe unstrukturiert in einem einzigen Formularfeld eingeben und die relevanten Personendatensätze finden. Die approximative Suche berücksichtigt fehlerhafte Eingaben und unterschiedliche Qualitäten sowohl in der Eingabe wie auch in der Referenz – unabhängig von Schrift, Sprache, Format oder Transkriptionen.

Neben gängigen phonetischen Transformationen bietet [matchmaker](#) Lösungen zur Transliteration aus Schriftsystemen wie Dewanagari, Hangul, Katakana und Mandarin. Die eigens entwickelte Kanonisierungsfunktion für arabische Namen abstrahiert verschiedenste Transkriptionsvarianten. So werden z. B. die demselben arabischen Zeichen ‚و‘ entsprechenden lateinischen Buchstaben(-sequenzen) ‚o‘, ‚u‘, ‚w‘, ‚oo‘ und ‚ou‘ äquivalent behandelt. Wird ‚دمحم‘ im angloamerikanischen Sprachraum in der Regel als Muhammad transkribiert, wertet [matchmaker](#) auch die Varianten Mohammed, Mohamed, Muhamed, Muhammed, Muhamet oder Mohammad als Treffer.

Die Grundlage zur Dublettenbereinigung

Mit den gleichen Algorithmen, die die Datenverfügbarkeit gewährleisten, können Cluster von potenziellen Personendubletten erzeugt werden. Zusatzinformationen, z. B. zu Familienverbänden, werden berücksichtigt und visualisiert. Automatisiert können nun bei der Datenanlage Dubletten erkannt und präventive Maßnahmen ergriffen werden.

„In einem mehrwöchigen Proof of Concept mit dem Arbeitstitel „Ähnlichkeitssuche“ stellte exorbyte vor Ort die technologischen und fachlichen Expertisen zum Auffinden von Ähnlichkeiten und der unscharfen Suche eindrucksvoll unter Beweis. Die Ergebnisse der Plattform [matchmaker](#) waren so herausragend gut, dass wir zur weiteren Evaluierung im Frühjahr 2018 einen Pilotbetrieb mit erweitertem Testerkreis gestartet haben. Unabhängig zu sein von Schrift, Sprache oder Transkription, bedeutet einen enormen Gewinn an Sicherheit und Verlässlichkeit für unsere Fallentscheidungen und die betroffenen Personen selbst.“

Michaele Golbostan,
Referatsleiterin Datenqualitätsmanagement,
AZR-Kontaktstelle Asyl, IT-Forschung im BAMF



Das Bundesamt ist das Kompetenzzentrum für Asyl, Migration und Integration in Deutschland. Es ist eine Bundesbehörde im Geschäftsbereich des Bundesministeriums des Innern, für Bau und Heimat (BMI). Mit seinen dezentralen Standorten, darunter Außenstellen, Ankunftscentren sowie Entscheidungszentren, steht es in direktem Kontakt mit allen Akteurinnen und Akteuren des Flüchtlingsschutzes und der Integrationsarbeit.