

Anforderungsprofil für den Bewerber-Informationscrawler:

Zielsetzung:

Entwickeln eines automatisierten Crawlers, der Informationen über vielversprechende Bewerberinnen und Bewerber sammelt. Der Crawler soll die zeitintensive manuelle Suche ersetzen, indem er relevante Informationen von verschiedenen Plattformen und Websites aggregiert. Der Crawler soll den gesamten Prozess der Informationsbeschaffung automatisieren und ohne menschliches Eingreifen funktionieren.

Dateninput:

Der Crawler soll in der Lage sein, Informationen anhand der folgenden Eingabedaten zu ermitteln:

- E-Mail-Adresse (optional)
- Vor und Zuname
- Wohngegend / Umgebung

Datensammlung:

Der Crawler soll auf verschiedene öffentlich zugängliche Quellen zugreifen und Informationen zu den Bewerberinnen und Bewerbern sammeln. Dazu gehören, sind aber nicht beschränkt auf:

Social-Media-Profile (z. B. LinkedIn, Twitter, Facebook, Instagram)
Berufliche Netzwerke (z. B. Xing, LinkedIn)
Blogs und persönliche Websites
Online-Artikel und Publikationen

Output und Speicherung der Daten:

Datenstrukturierung:

Die gesammelten Daten sollten strukturiert in einer Datenbank abgelegt werden. Dies beinhaltet:

- Kategorisierung der Informationen nach Quellen (z. B. LinkedIn-Daten, Twitter-Posts).
- Speicherung von Meta-Informationen, wie Datum des Crawlings oder der URL der Informationsquelle.

REST API:

Starten der Suche: Durch eine Anfrage über eine REST API kann die Suche gestartet werden. Als Antwort erhält der Nutzer eine spezifische ID für diesen Auftrag, die bei nachfolgenden Anfragen als Referenz dient.

Abfrage von Status und Daten: Mit der erhaltenen ID kann der Nutzer über einen Endpunkt der API den aktuellen Stand der Suche prüfen. Wenn Daten vorhanden sind, werden Sie hier auch mitgeliefert.

Datenschutz:

Der Crawler muss datenschutzkonform arbeiten und darf nur öffentlich zugängliche Informationen sammeln. Es dürfen keine personenbezogenen Daten erfasst werden, die nicht frei verfügbar sind oder gegen geltende Datenschutzrichtlinien verstoßen.