

# GoldenProfile

---

- *GoldenProfile*
  - Motivation
  - Ziel von *GoldenProfile*
  - Mögliche Use-Cases
  - Technischer Ansatz
    - Datensammlung
    - Datenaggregation
    - Bereitstellung
    - Pattern zur Analyse und Aggregation der Daten
    - Datenquellen
  - Datenschutz/Rechtliche Fragestellungen

## Motivation

Heutzutage geben die meisten Menschen sämtliche Informationen über sich im Internet preis, ohne sich Gedanken darüber zu machen, wer alles auf diese Daten zugreifen kann. Inzwischen ist es möglich, anhand von ein paar wenigen Daten im Internet sehr viel über einen Menschen herauszufinden.

## Ziel von *GoldenProfile*

Mit dem Projekt *GoldenProfile* soll ein Tool geschaffen werden, das auf Basis einiger weniger Stammdaten, wie zum Beispiel dem Namen und das Geburtsdatum einer Person, im Internet nach weiteren Daten sucht.

Durchsucht werden hierbei neben den vorhandenen Social Media Netzwerken, auch Zeitungen und diverse weitere Internetplattformen. Ist der Crawling-Vorgang abgeschlossen, werden die Daten aufbereitet, aggregiert und dem Anwender zur Verfügung gestellt.

## Mögliche Use-Cases

(Im Folgenden wird aus Gründen der besseren Lesbarkeit die männliche Form verwendet und ist in keiner Form wertend zu verstehen)

1. Eine Privatperson möchte herausfinden, was über sie im Internet zu finden ist.
2. Ein Bewerber möchte prüfen, was die HR-Abteilung über ihn herausfinden kann.
3. Ein Mitarbeiter in der HR-Abteilung möchte einen Bewerber überprüfen und schauen, was die von sich im Internet veröffentlicht haben.
4. Ein Mitarbeiter in der HR-Abteilung möchte wissen, wie sensibel ein Bewerber mit seinen persönlichen Daten umgeht.
5. Ein Mitarbeiter in der HR-Abteilung möchte die bestehenden Mitarbeiterdaten mit weitergehenden Informationen anreichern.
6. Ein Key-Account-Manager bereitet sich auf einen Kundentermin mit einem neuen Kunden vor und benötigt dazu ein möglichst umfangreiches berufliches als auch persönliches Profil des Kunden.
7. Ein Key-Account-Manager möchte die bestehenden Informationen über einen Kunden um weitere Informationen anreichern.

# Technischer Ansatz

## Datensammlung

In der Regel stellen die verschiedenen Plattformen und Social Media Netzwerke Schnittstellen bereit, über die auf potentielle Online-Profile zugegriffen werden kann. Ein weiteres Vorgehen ist das Scraping von Informationen von den Webseiten selbst.

## Datenaggregation

Die eingegebenen Daten werden in mehreren Durchläufen mit den gewonnenen Informationen angereichert und erneut für eine Datensammlung verwendet. Dieser Vorgang wird so lange wiederholt, bis keine neuen Daten mehr erfasst werden.

## Bereitstellung

Die gewonnenen Informationen werden nach der Datenaggregation gefiltert und bereinigt, um sie im Anschluss dem Anwender bereitzustellen.

## Pattern zur Analyse und Aggregation der Daten

- **Scoring** - Gewichtung von Informationen
- **Bildererkennung** - Extraktion von Informationen aus Bildern
- **Videoanalyse** - Extraktion von Informationen aus Videos
- **Gesichtserkennung** - Abgleich und Erkennung von Personen auf Bildern und in Videos
- **Textanalyse** - Extraktion von Informationen aus Texten
- **Phonetische Ähnlichkeiten** - Erkennung von ähnlichen Schreibweisen (Bsp. Stephan, Steffan, Stefan)

## Datenquellen

Social Media Netzwerke:

- Xing
- LinkedIn
- Facebook
- Twitter
- Instagram
- askfm
- tumblr
- myspace

Weitere Plattformen:

- Online Telefonbücher
- Zeitungen
- Suchmaschinen
- ...

## Datenschutz/Rechtliche Fragestellungen

- Dürfen öffentliche bzw. personenbezogene Daten gescraped werden?
- Dürfen personenbezogene Daten von einer API-Schnittstelle gescraped werden?
- Dürfen die Daten auf unseren Servern gespeichert werden?
- Dürfen die Daten auf unseren Servern verarbeitet werden?
- Dürfen die Daten auf gemieteten Servern gespeichert werden?
- Dürfen die Daten auf gemieteten Servern verarbeitet werden?
- Dürfen die Daten im Rahmen einer öffentlichen Webanwendung, für andere Benutzer dargestellt werden?
- Dürfen die Daten im Rahmen einer geschlossenen Webanwendung, für Unternehmen dargestellt werden?
- Dürfen Unternehmen die Applikation nutzen, um bereits existierende Nutzerdaten anzureichern?
- (Im Bezug auf Use-Case 4) Welche (privaten) Daten eines Mitarbeiters dürfen Unternehmen speichern?