GoldenProfile

Motivation

Nowadays, most people disclose all information about themselves on the Internet without worrying about who has access to this data. It is possible to find out a lot about a person with just a few data on the Internet.

Goal of *GoldenProfile*

With the project *GoldenProfile* a tool is to be created, which searches the Internet for further data on the basis of a few master data, such as the name and date of birth of a person. In addition to the existing social media networks, newspapers and various other Internet platforms are also searched. Once the crawling process is complete, the data is prepared, aggregated and made available to the user.

Possible Use Cases

- A private individual wants to find out what can be found about him or her on the Internet.
- An applicant wants to check what the HR department can find out about him or her.
- An employee in the HR department wants to check an applicant and see what he or she has published on the Internet.
- An employee in the HR department wants to know how sensitive an applicant is with his or her personal data.
- An employee in the HR department wants to enrich the existing employee data with more detailed information.
- A key account manager prepares for a customer appointment with a new customer and requires a customer's professional and personal profile that is as comprehensive as possible.
- A key account manager wants to enrich the existing information about a customer with further information.

Technical approach

Data collection

Usually, the various platforms and social media networks provide interfaces through which potential online profiles can be accessed. A further procedure is the scraping of information from the websites themselves.

Data aggregation

The entered data is enriched with the obtained information in several runs and reused for a data collection. This process is repeated until no new data is collected.

Data presentation

The information obtained is filtered and cleaned after data aggregation in order to make it accessible to the user.

Pattern for data analysis and aggregation

- **Scoring** - weighting of information
- **Image recognition** - extraction of information from images
- **Video analysis** - extraction of information from Videos
- **Face recognition** - matching and recognizing people in images and videos
- **Text analysis** - extraction of information from texts
- **Phonetic similarities** - recognition of similar spellings (e.g. Stephan, Steffan, Stefan)

Data sources

Social media networks:

- Xing
- LinkedIn
- Facebook
- Twitter
- Instagram
- ASKfm
- Tumblr
- Myspace

Other platforms:

- Online telephone directories
- Newspapers
- Search engines
- ...

Data Protection/Legal Issues

- Is it allowed to scrape public or personal data?
- Is it allowed to scrape personal data from an API interface?
- Is it allowed to store the data on our servers?
- Is it allowed to process the data on our servers?
- Is it allowed to store the data on rented servers?
- Is it allowed to process the data on rented servers?
- Is it allowed to display the data for other users within the scope of a public web application?
- Is it allowed to display the data in a closed web application for companies?
- Is it allowed for companies to use the application to enrich existing user data?
- (With regard to Use-Case 4) Which (private) data of an employee may companies store?