Online Based Enterprise Characteristics Experimental Statistics: Austria

Aims

Statistics Austria (STAT) aims to produce experimantal statistics about enterprises for the following indicators

- Use-case 1 URLs Inventory
    - Rate of enterprises having a website
- Use-case 2 Variables in the ICT usage in enterprise survey
    - Social Media Presence - rate of enterprises that are present on social media
    - E-commerce - Rate of enterprises engaged in web sales on their website

Data Sources

For the production of the statistical information the following dat a sources were used

- Statistical Business Register (SBR)
    - Enterprise ID
    - Enterprise Name
    - Enterprise post address
    - Enterprise VAT and or commercial register number
    - Enterprise grouping ID, to link enterprises belonging to the same enterprise group
- Bing Search API
    - 10 suggestions for enterprise URL
- Enterprises web sites
    - up to 25 pages from the enterprise web site

Methodology

Use-case 1 URLs Inventory

Rate of enterprises having websites

1. Get enterprises from the ICT survey sample with the following data: IDs, Names, post address.
2. Create search string for Bing Search API.
3. Write out input file for Bing Search API and run Bing Search, see https://github.com/ EnterpriseCharacteristicsESSnetBigData/UrlSearcher.
4. Read in URLs from Bing Search and discrad blacklisted URLs like yellow pages, ect. . . .
5. Scrape each URL, specifically the impressum at each URL and save the scraped data, see https: //github.com/EnterpriseCharacteristicsESSnetBigData/webStats.
6. Load scraped data and extract unique identifier, like VAT or commercial register number.
7. Match found VAT or commercial register number with SBR and consider grouping ID.

For a more detailed step by step explanation with exampel code see https://github.com/EnterpriseCharacteristicsESSnetBigData/ webStats.

Use-case 2 Variables in the ICT survey

1. Load scraped data
2. Retrieve social media links found in scraped data, see https://github.com/EnterpriseCharacteristicsESSnetBigData/ webStats.

For a more detailed step by step explanation with exampel code see https://github.com/EnterpriseCharacteristicsESSnetBigData/ webStats.