

Topic Description Research Internship

Web crawler and chemical industry sites (de/en)

Background

Chemical industry belongs to energy-intensive industry sectors and therefore significantly shapes the overall industrial energy consumption. Its geographical position determines the load distribution in the electricity grid and infrastructural requirements for the feedstock supply. However, the information on locations and capacities of chemical industries is not systematically available but rather split on webpages of many individual companies. A web crawler or web spider is a computer program that's used to search and automatically index website content and other information over the internet. Can we use a web crawler to identify chemical sites and their capacities?

Goal: Answering the questions below

- What are the legal frameworks for using a web crawler? May we use it for this topic?
- What methodology techniques are the web crawlers using to search the web?
- Should we next to English use further languages – Spanish, Chinese etc.?
- Finally: **Where are the chemical sites world-wide?**
- Optionally: which capacities per product do they have?

Requirements

- Programming skills: Python, C, C++, MATLAB, Java... or similar
- Interest in energy modelling and economy

Contact

M.Sc. Andjelka Kerekes

Lichtenbergstr. 4a, 85748 Garching b. München, Raum 2013

E-Mail andelka.kerekes@tum.de