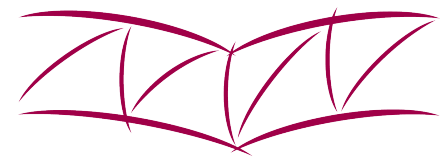




Prof. Dr. Oliver Niggemann
Institut für Automatisierungstechnik

Fakultät für Maschinenbau



HELMUT SCHMIDT
UNIVERSITÄT

Universität der Bundeswehr Hamburg

Studienarbeit

Entwicklung von Web-Scrapern zur Datenakquise

Themengebiet: Datenverarbeitung und Web-Scraping

In der Professur für Informatik im Maschinenbau wird ab sofort eine Studienarbeit im Bereich "Datenverarbeitung und Web-Scraping" angeboten. Die Studienarbeit soll sich mit der Datenakquise für das fakultätsübergreifende Projekt *Open Intelligence Platform* aus öffentlich zugänglichen Quellen beschäftigen.

Das Projekt *Open Intelligence Platform* soll Daten aus öffentlich zugänglichen Quellen zentral einem fakultätsübergreifenden Forscherkreis bereitstellen. Die Daten sollen in einem einheitlichen Format abgelegt und über die gleiche Schnittstelle abrufbar sein. Späteres Ziel der Plattform ist es, die Daten mittels maschinellem Lernen zu analysieren und darauf basierend Risikoeinschätzungen zur Sicherheitslage zu treffen.

Ziel der Studienarbeit soll sein, von einer oder mehreren öffentlichen Datenquellen, wie Zeitungsportalen oder sozialen Netzwerken, Artikel, Beiträge o.ä. zu erfassen. Aus diesen die relevanten Informationen, wie Beitragstext, Autor und Veröffentlichungsdatum, zu extrahieren und unter einheitlichem Schema in einer Datenbank abzulegen. Hierbei spielt vor allem die Abwägung rechtlicher Aspekte, sowie der angemessene Einsatz von *Web-Crawlern* eine Rolle.

Interessenten sollten folgende **Voraussetzungen** erfüllen:

- Grundverständnis über die Architektur von Webseiten
- Erfahrung mit einer Programmiersprache, vorzugsweise Python
- optional, aber sehr hilfreich, Erfahrung mit dem *Web scraping* Framework [scrapy](#)



Bei Interesse wenden Sie sich bitte an:
Joshua Schraven
schravej@hsu-hh.de