

BOT-11: Universeller Scraper & Download

Der Beuthbot soll einen "universellen" Web-Scraper beinhalten, der als Grundlage für künftige Features dienen soll, die für konkrete Scraping-Funktionalitäten vorgesehen sind. Aufgrund der hohen Diversität an Datenstrukturen unterschiedlicher Webseiten, soll dieser möglichst abstrakte Funktionalitäten zur Extrahierung von Datensätzen bieten.

Initiale Schätzung 1

Technologien Javascript

Abhängigkeiten keine

Anforderungen *Import von HTML- und XML-Dateien

*Daten lassen sich im JSON-Format exportieren

*Datensätze sind per HTML-Tags und CSS-Selektoren

extrahierbar

*Dateien einer Webseite lassen sich downloaden

Tasks * BOT-26 Recherche nach geeignetster Methode

(HTML-JSON)

Nutzungshinweis: Auf dieses vorliegende Schulungs- oder Beratungsdokument (ggf.) erlangt der Mandant vertragsgemäß ein nicht ausschließliches, dauerhaftes, unbeschränktes, unwiderrufliches und nicht übertragbares Nutzungsrecht. Eine hierüber hinausgehende, nicht zuvor durch datenschutz-maximum bewilligte Nutzung ist verboten und wird urheberrechtlich verfolgt.