

Thema für eine Abschlussarbeit (Bachelor- oder Masterarbeit) in den Bereichen Informatik, Data Science, Geoinformatik, Medieninformatik oder verwandten Studiengängen mit Schwerpunkt auf Datenanalyse, Künstlicher Intelligenz, Natural Language Processing oder Geodatenverarbeitung.

Automatische Extraktion geolokalisierter Bauinformationen aus webgescrapten Kulturdenkmal-Daten

Ziel der Arbeit ist die Entwicklung und Evaluation einer Pipeline, die aus webgescrapten Denkmalbeschreibungen strukturierte Informationen zu Bauzeit, Materialität, Baukonstruktion und architektonischem Stil extrahiert und mit räumlichen Daten verknüpft. Die Ergebnisse sollen in GIS-geeigneten Formaten (z. B. GeoJSON/CSV) bereitgestellt werden, um sie in der Plattform „Colouring Dresden“ weiterzuverwenden. Im Fokus steht der Vergleich eines regelbasierten Ansatzes mit einem transformerbasierten NLP-Verfahren; optional kann ein LLM-basierter Ansatz (Zero/Few-Shot) untersucht werden.

Forschungsfragen:

(1) Wie zuverlässig lassen sich Bauzeit, Materialangaben, Konstruktionstypen und Stilmerkmale automatisiert aus heterogenen Denkmaltextrahieren? (2) Wie unterscheiden sich regelbasierte, transformerbasierte und LLM-basierte Verfahren hinsichtlich Genauigkeit, Robustheit, Coverage und Erklärbarkeit? (3) Welche typischen Fehlerquellen treten auf (uneinheitliche Schreibweisen, Mehrdeutigkeiten, fehlende Angaben) und wie lassen sie sich reduzieren?

Arbeitsschritte:

- Literatur- und Technologieübersicht
- Datenerfassung und Vorverarbeitung inkl. Textbereinigung, Tokenisierung und Aufbau einfacher Gazetteers für Materialien, Stilbegriffe und Datumsangaben
- Erstellung eines annotierten Referenzdatensatzes mit klaren Labelklassen und Dokumentation der Annotationsregeln
- Methodenentwicklung (regelbasierte Extraktion, transformerbasierter Ansatz und optionaler LLM-basierter Ansatz)
- Georeferenzierung und Datenintegration
- Evaluation

Technische Anforderungen und Werkzeuge:

Python, pandas, GeoPandas, Jupyter Notebook, OSM Nominatim (zur Georeferenzierung), QGIS (Visualisierung)

Kontakt

Bei Interesse oder Rückfragen melden bei:

Dr. Robert Hecht, Leibniz-Institut für ökologische Raumentwicklung (IÖR), Dresden, Deutschland; Tel. +49 (0)351 4679 248; Email: r.hecht@ioer.de

Sophie Teichmann, TUD Dresden University of Technology. Center for Interdisciplinary Digital Sciences (CIDS), ScaDS.AI Dresden/Leipzig; Tel. +49 (0)351 4679 248; Email: sophie.teichmann@tu-dresden.de

Datum: 11.5.2026