

Automatic Extraction of Information out of technical datasheet using LLMs and ontologies

Semesterthesis / Masterthesis / Forschungspraxis

Robot Systems Group

Laboratory for Product Development and Lightweight Design

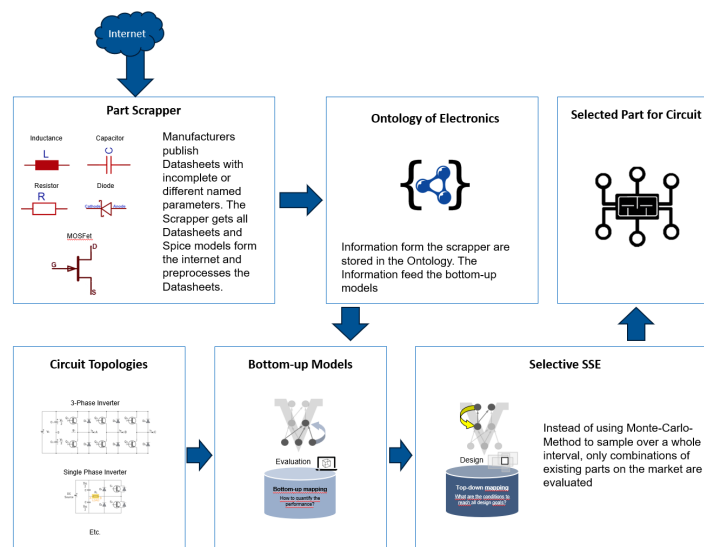


Figure 1 Datasheet Extraction Procedure Model

Project Description: This thesis will explore the innovative application of large language models (LLMs) and ontologies for automatic information extraction from technical datasheets. Technical datasheets are crucial sources of information for engineers, researchers, and product developers, but their complex structure and domain-specific language often make manual data extraction time-consuming and error-prone. This research aims to develop a system that leverages the advanced natural language understanding capabilities of LLMs in conjunction with structured knowledge representation through ontologies to automate this process. The student will investigate various LLM architectures suitable for information extraction tasks, explore methods for mapping extracted data to relevant ontology concepts, and evaluate the performance of the developed system on benchmark datasets or real-world technical datasheets. This thesis offers a unique opportunity to contribute to the advancement of AI-powered information retrieval techniques in a domain with significant practical implications.

Desired Skills: We seek highly motivated students with a passion for natural language processing, knowledge representation, and data analysis to contribute to this exciting research project. Ideal candidates will possess strong programming skills in Python and familiarity with machine learning libraries like TensorFlow or PyTorch. A solid understanding of semantic web technologies, ontologies, and knowledge graphs is also desirable. Excellent analytical and problem-solving skills are essential, as is the ability to work both independently and collaboratively within a research team. Prior experience with large language models (LLMs) would be advantageous, but not mandatory. Most importantly, we are looking for students who are eager to learn, contribute innovative ideas, and push the boundaries of automated information extraction.

Application Process: Interested candidates should submit their application via email, including a detailed CV and a current transcript of records. Please send your applications to the Email mentioned below.

We look forward to reviewing your background and discussing how you can contribute to this innovative project.

Contact

Maximilian Amm, M.sc. mult.

Room: MW2631

E-mail: Maximilian.Amm@tum.de