Assignment: Smart Data Systems for Engineering

Type: Internship / Graduation Project

Discipline(s): ICT / Software Engineering / Data Engineering

Level: Bachelor / Master

Assignment description

How can structured component data help engineers find and compare parts more easily? At AME, electronic product development involves thousands of components—from resistors and ICs to connectors, sensors, screws and bolts. These components are recorded in a central Product Data Management (PDM) system.

Manually written, single-line descriptions vary in structure and detail, making it difficult to group components, compare specifications, identify potential alternatives and automate analysis. You will develop a tool that extracts standardised metadata from datasheets, databases and online catalogues and prepares it for integration into the PDM system.

Together with your supervisor, select representative component categories and define a manageable scope for implementation and testing.

Objective

Develop and demonstrate a metadata enrichment pipeline that converts component information into consistent, machine-readable PDM records. Compare AI/large language model (LLM), rule-based and hybrid extraction approaches, and determine which provides sufficiently accurate, traceable and maintainable results.

Tasks

- Explore: Review existing component descriptions and discuss data needs with engineers. Identify inconsistencies, prioritise component categories and agree the project scope and acceptance criteria.
- Define the metadata: Develop a schema with category-specific fields, such as manufacturer part number, package, dimensions, voltage and current ratings, tolerance and lifecycle status. Define consistent units and rules for missing, conflicting or uncertain information.
- Connect sources: Investigate datasheet libraries, SiliconExpert and other suitable APIs, manufacturer documents and online catalogues. Establish reliable component matching and record the source and retrieval date of extracted information.
- Compare extraction approaches: Evaluate rule-based, LLM-based and hybrid methods against a manually verified reference dataset. Compare accuracy, field coverage, processing time and cost. Ensure unsupported values are flagged rather than presented as facts.
- Build and integrate: Implement a pipeline for new or updated components that retrieves source information, extracts and validates metadata, and prepares updates for the PDM system

through an agreed interface. Include review of uncertain results, logging and protection against unintended overwrites.

- Test and hand over: Demonstrate the pipeline on representative components and test incomplete documents, conflicting sources, duplicate records and failed connections. Document the implementation and recommend a maintenance and update strategy.

Project resources

- Access to the internal PDM system and representative component records.
- Datasheet libraries and SiliconExpert API integration.
- Support from AME's IT and hardware engineering teams.
- Access to suitable LLM tools, if included in the selected approach.
- Integration permissions and a suitable test environment, to be agreed with the supervisor.

Key deliverables

- Requirements and feasibility assessment with a justified technical approach.
- Defined metadata schema, validation rules and PDM field mappings.
- Working prototype of the metadata enrichment pipeline.
- Verified sample dataset and a controlled PDM integration demonstration.
- Comparative evaluation of extraction accuracy, coverage, performance and cost.
- Source code, architecture and integration documentation, maintenance guidance and a final graduation report.