Untapping The Hidden Potential of Non-Textual Maritime Data Using AI

Lamin Jatta, Maritime Technology, Novia University of Applied Sciences, first.last@novia.fi.
Supervisors: Andreas Lundell (ÅAU), Johan Westö (Novia UAS), and Mikael Manngård (Novia UAS).

This thesis addresses the challenge of organizing, filtering, searching, and retrieving information from unstructured, non-textual data, including audio, video, images, and 3D models. The research is applied in maritime simulation and virtual commissioning to extract and organize data used for automated testing.

Research Problem and Goals

This research addresses the following research problems:

Problem 1: As industries like maritime and automotive generate massive amounts of audio and video data, only a small fraction of this data proves relevant for training artificial intelligence (AI) models. Current methods for storing and querying unstructured data, such as traditional databases, struggle to efficiently manage the volume of video data produced.

Problem 2: Engineering data is rarely available in a machine-readable structured form. It is often represented in formats like figures, drawings, or diagrams, making data extraction and integration challenging. Thus, automatic extraction of information and AI integration in engineering tools is challenging.

This thesis aims to apply AI-driven techniques to organize, filter, and retrieve relevant information from non-textual data and represent it in a machine-readable format.

Methods

AI methods, including **natural language processing**, **multimodal embeddings**, and **semantic search** methods are applied to address the research problems.

Implementation

The research will be conducted in the Virtual Sea Trial project for the period 1.1.2024--31.12.2026. The planned publications are presented in chronological order.

Paper I: Maritime Automatic Speech Recognition and Key-Phrase Extraction.

Paper II: Semantic Search for Maritime Unstructured Data.

Paper III: Graph Representations of Maritime Systems.

Paper IV: Extracting Information from Engineering Graphs and Diagrams.

Discussion

The maritime industry and technology providers are expected to benefit from the results through increased operational efficiency and cost-effectiveness. Results happen through publication in (preferably) open-access journals and conference series. Dissemination of results happens through the Virtual Sea Trial project.

The Research is Funded by Business Finland | Co-Innovation Project Virtual Sea Trial | virtualseatrial.fi