Python Implementation for S1000D Applicability Depended Processing Model - SALERNO

Authors : Theresia El Khoury, Georges Badr, Amir Hajjam El Hassani, Stéphane N'Guyen Van Ky

Abstract : The widespread adoption of machine learning and artificial intelligence across different domains can be attributed to the digitization of data over several decades, resulting in vast amounts of data, types, and structures. Thus, data processing and preparation turn out to be a crucial stage. However, applying these techniques to S1000D standard-based data poses a challenge due to its complexity and the need to preserve logical information. This paper describes SALERNO, an S1000d AppLicability dEpended pRocessiNg mOdel. This python-based model analyzes and converts the XML S1000D-based files into an easier data format that can be used in machine learning techniques while preserving the different logic and relationships in files. The model parses the files in the given folder, filters them, and extracts the required information to be saved in appropriate data frames and Excel sheets. Its main idea is to group the extracted information by applicability. In addition, it extracts the full text by replacing internal and external references while maintaining the relationships between files, as well as the necessary requirements. The resulting files can then be saved in databases and used in different models. Documents in both English and French languages were tested, and special characters were decoded. Updates on the technical manuals were taken into consideration as well. The model was tested on different versions of the S1000D, and the results demonstrated its ability to effectively handle the applicability, requirements, references, and relationships across all files and on different levels.

Keywords : aeronautics, big data, data processing, machine learning, S1000D

Conference Title : ICACIAS 2023 : International Conference on Artificial and Computational Intelligence in Aerospace Sciences

Conference Location : Rome, Italy

Conference Dates : May 04-05, 2023