Consolidating Wind Farm Data in Data Spaces for Training Predictive Maintenance ML

ROBERT DAVID, The Semantic Web Company, Austria
AXEL WEISSENFELD, Austrian Institute of Technology, Austria
PETAR IVANOV, NIKOLA TULECHKI, and VLADIMIR ALEXIEV, Ontotext, Bulgaria

The UNDERPIN project addresses the challenge of training machine learning (ML) models for predictive maintenance of wind farms by leveraging consolidated training data provided by multiple partners. This collaborative effort is facilitated through a European-wide Data Space for Manufacturing for dynamic asset management and predictive and prescriptive maintenance established as part of the project. Data spaces allow trusted participants to share data while maintaining data sovereignty. However, consolidating diverse data sets requires a clear semantics and mapping framework. To solve this, the project employs Semantic Web standards and technologies to model and process the data. Using OWL (Web Ontology Language), vocabularies and ontologies are developed to represent the data with clear, standardized semantics. Two key software components, GraphDB and PoolParty, are utilized to implement these standards and provide a semantic layer to efficiently process the consolidated data. The main benefit is to support ML training by presenting a consolidated and uniform training data set. Our approach aims to enhance the quality and reliability of predictive maintenance models for wind farms. A demonstration from the UNDERPIN project showcases the process of modeling and processing the data, illustrating the practical application of these semantic technologies. This example highlights the project's approach and the potential improvements in predictive maintenance through enhanced data integration and processing.

Additional Key Words and Phrases: data spaces, oil and gas, renewable energy, refineries, windfarms, time series, ontologies, semantic technologies

ACM Reference Format:
Robert David, Axel Weißenfeld, Petar Ivanov, Nikola Tulechki, and Vladimir Alexiev. . Consolidating Wind Farm Data in Data Spaces for Training Predictive Maintenance ML. In *Proceedings of* . ACM, New York, NY, USA, 2 pages.

1 INTRODUCTION

Consolidating wind farm data in data spaces for training predictive maintenance ML - Use Case Description

2 INITIAL SITUATION

In the UNDERPIN project, we face the challenge of training ML for predictive maintenance of wind farms. This scenario benefits from large amounts of training data. Such data can be provided by different partners who collaborate on this ML approach. We support this scenario by implementing a data space for wind farm data. Data spaces enable data sharing between trusted participants while keeping data sovereignty. However, for consolidating different data sets from participants, we need to establish a clear semantics and mapping for the data.

Authors' Contact Information: Robert David, The Semantic Web Company, Austria; Axel Weißenfeld, Austrian Institute of Technology, Austria; Petar Ivanov; Nikola Tulechki; Vladimir Alexiev, Ontotext, Bulgaria.

3 APPROACH AND IT-SOLUTION

To tackle this problem, we introduce Semantic Web standards and technologies to model and process this data. Vocabularies and ontologies, specifically expressed using OWL, can represent such data with a clear semantics based on standards. Furthermore, we introduce GraphDB and PoolParty as two software components which implement these standards to process the data.

The Wind Farm Ontology [**?**] ...

4 BUSINESS VALUE AND BENEFITS OF THE SEMANTIC SOLUTION

When consolidating the data, we need to provide a method to automatically map and process it so this can be done efficiently on large data sets and/or regular updates in data spaces. The benefit provided should support the ML training in such a way that the consolidated data set is exposed as uniform training data set.

5 PROSPECTS AND RECOMMENDATIONS

We provide a first prototype to implement our approach. Next steps are bringing it into the actual data space in practice and representing and integrating further data sets from different data space participants. Also, the ML should actually benefit regarding quality in practice, which still needs to be verified.

6 CONCLUSION

REFERENCES