"Can we build dynamic curricula for software engineering that meet the fast-paced industry demand for professionals?" "What if we could integrate all e-health records into one single place for better decision-making?" "Is there a way to make machine learning trustworthy and utilize it in safety-critical domains?" These are only some of the research questions that academia is actively working to find answers to. In this edition of GraphDB In Action, we present to you the work of three bright researchers who have set out to find solutions that allow meaningful analysis and interpretation of data, supported by Ontotext GraphDB.

We are proud to see the use of GraphDB has led to breakthroughs in solving real-world challenges with managing and interpreting complex e-health data, entity linking, and data interoperability for education and ontologies for machine learning.

Designing Responsive and Relevant Software Engineering Curricula Using Semantic Interoperability

The first thesis (MS) we want to highlight is Extracting a Body of Knowledge as a First Step Towards Defining a United Software Engineering Curriculum Guideline by Anton Kiselev. It was submitted to the Faculty of Embry-Riddle Aeronautical University in March 2023.

The thesis focuses on developing an automatic approach to creating a software engineering curriculum graph using natural language processing (NLP) and graph databases. It aims to bridge the gap between software engineering education and industry needs. The thesis also discusses the challenges and limitations of existing methods, which often fail to quickly adapt to industry demands for skilled graduates. This leads to a disconnect between educational outcomes and market requirements.

The research highlights the importance of industrial and academic collaboration and the need for a systematic approach to creating and assessing software engineering curricula. The proposed research methodology involves creating a knowledge graph and using NLP algorithms to automate extracting information from existing curriculum guidelines and creating a unified ontology to inform curriculum design.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

As the project involved automatic NLP entity linking, it required a suitable graph storage. After initial processing with Neo4j, the researcher discovered that the effort of learning its query language, Cypher, was not a valid time investment. The labeled property graphs were converted into an RDF-compliant format. It was used with SPARQL in GraphDB to maintain semantic interoperability with external RDF namespaces. Ontotext's leading RDF graph database enhanced the ability to manage and query complex, linked data while ensuring compatibility with broader semantic ecosystems.

The significance of this research lies in its potential to create a responsive and relevant software engineering curriculum that effectively adapts to changing industry needs. The novel approach of using automated NLP techniques to streamline curriculum development reduces the manual effort required in curriculum design and makes it more efficient.

> Overall, even if the suggested technique still necessitates some manual work, its potential benefits for enhancing software engineering education are considerable, and it marks a substantial advancement toward the creation of a more integrated and thorough software engineering curriculum.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

Download Ontotext's GraphDB!

Building Better E-Health Systems By Fusing Semantic Data

The second thesis (MS) we want to discuss is Semantic Data Integration from Heterogeneous Sources in the e-Health Domain by Kontogiannis Stergios. The thesis was submitted for the Master of Science in Data Science degree to the School of Science and Technology in January 2023.

The study discusses the key concepts and technologies related to semantic data integration in the field of brain diseases. It outlines the entire process with examples of how it has been applied in practice to support the analysis and management of brain diseases within the context of the EU-funded research project ALAMEDA. ALAMEDA's objective is to provide personalized rehabilitation treatment assessments.

The methodology applied in the project includes the use of standardized terminologies and

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontologies to enable the mapping of data from different sources to a common representation. They also include the application of semantic reasoning to infer new knowledge from the knowledge graph using domain-specific rulesets. ALAMEDA uses the CASPAR framework and GraphDB to convert data from various sources into RDF triples and populate the knowledge graph with instance data from sensors.

The results successfully demonstrate the benefits of using semantic technologies to support data integration and knowledge sharing in the e-health domain. They show the potential of visualization and analysis tools such as GraphDB as well as the implementation of rules to trigger alerts for doctors.

> We believe that the techniques developed in the ALAMEDA project, as presented in this dissertation, hold great promise for supporting data integration and knowledge sharing in the e-health field and beyond.

Download Ontotext's GraphDB!

Improving Trustworthiness In Machine Learning Models With Ontologies

The last thesis (MS) in this selection is Defining Safe Training Datasets for Machine Learning Models Using Ontologies by Lynn Vonder Haar. It was submitted to the Department of Electrical Engineering and Computer Science Embry-Riddle Aeronautical University in the spring of 2023.

This research addresses the lack of trust in machine learning models in safety-critical applications where life and livelihood are at risk. The thesis outlines the limitations of current performance metrics and explainability methods, which don't provide enough confidence that the model makes fair and balanced decisions. It argues that, if a training dataset can be shown to be safe, this could increase users' trust in a model trained on that data. The solution discussed in the thesis involves validating the domain completeness and image quality robustness, using a domain ontology and an image quality characteristic ontology.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

The thesis presents an experiment in the domain of recognizing emergency road vehicles during autonomous driving as a proof of concept for the proposed method. The first part of the experiment compares a model trained on a full dataset with one trained on a dataset missing all instances of tow trucks. The second part of the experiment compares the full model with one missing all instances of haze, blur, or fog. The RDF triples are imported into GraphDB and then SPARQL queries are performed on the three knowledge graphs to gather information on the representation and completeness of the datasets**.**

The results show that the absence of tow trucks in the training data reduces the model's performance on tow trucks, and increasing the number of instances of haze blur in the training dataset could improve the model's performance on hazy images. The research concludes that ensuring the safety and completeness of the training dataset is essential to increasing trust in machine learning models used in safety-critical domains. It demonstrates that the proposed method is promising for identifying gaps in the training dataset and beginning a transition to human-out-of-the-loop.

The results of the experiment are promising and indicate that this direction of

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

research should continue to be followed. Some possible areas of future work are to confirm the results from this research, expand the domain, and automate the process to make the wide-spread use of such a process easier and less time-consuming for developers.

Download Ontotext's GraphDB!

Semantics Is Inevitable

We've explored how GraphDB technology has helped young researchers address pressing challenges across diverse domains. From enhancing software engineering curricula to integrating complex e-health data, semantic technology consistently proves its capacity to create meaningful connections and tame the complexity of today's world.

Whenever semantics comes into play to solve multilayered problems, an envelope is pushed.

By making data interoperable, easy to manage, and interconnected, semantic technology helps overcome the challenges of handling heterogeneous data. It enables seamless information flow across different systems, breaking down barriers and improving data integration.

And, we are always happy to see Ontotext GraphDB serving as the backbone of solving data and knowledge management challenges.

To learn more about the diversity of use cases and research domains where GraphDB is used, see the the Zotero library Ontotext GraphDB and this publication: Diverse Uses of a Semantic Graph Database for Knowledge Organization and Research, European Data Conference on Reference Data and Semantics (ENDORSE 2021).

Do you think GraphDB can help you in your use case?

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

Ontotext's GraphDB

Give it a try today!

Download Now

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com