UDC: 004.774.2:[004.85:111.1
Professional paper

ONTOLOGY MATCHING AND MANAGEMENT USING MACHINE LEARNING ASPECTS

Gökhan CANTAŞ[1]

[1]*International Balkan University, North Macedonia*
gokhan.cantas@ibu.edu.mk

Abstract:

Industry 4. zero is a brand-new generation of data era that ambitions to expand expertise bases for tracking developments in enterprise 4. zero. This paper proposes a framework to cope with the improvement of Knowledge Bases for Monitoring Trends in Industry4.zero.In this framework, we suggest an ontological model (COInd4 ontology) for the manufacturing area that describes the assets and strategies within side the manufacturing unit, and sensor observations are analysed via way of means of remark affects the use of contextual data past classical reasoning mechanism for stopping the forecasted undesired sensor effects that effect on from concept. The framework is primarily based totally on an information-pushed technique for Knowledge Graphs (CSV, JSON, and diverse styles of information) technique for growing Machine studying interoperability combining principles from diverse current ontologies for discovered predictive fashions and execution of the fashions. Moreover, LOTHBROK is designed for estimating cardinalities and paying attention to information locality. The assessment confirmed that TAO can gain appreciably quicker question processing overall performance as compared to the nation of the artwork while processing difficult queries in addition to while besides, TAO gives better transparency, flexibility, and cognitive ergonomics than its options Hontology and Accommodation Ontology. It affords a custom-designed approach to abide via way of means of the necessities of the Greek Programme Diavgeia and proposes the identical time approach to encode authorities and administrative decisions/acts that might be universally followed to combine public files produced via way of means of different EU Member States, with positive changes content-wise.

Keywords: Semantic Web; Web Protégé ; ontology matching; knowledge base

1. Introduction

I have chosen all my literature from Semantic Web – Interoperability, Usability, Applicability an IOS Press Journal. This journal aims to collect research from different fields for sharing their perspective in sharing information for the future of the internet and Web4.0 etc versions. The journal's editorial board is global especially European countries and the USA. I have chosen to catch new aspects of all of my literature from under review session. This means that all research is newly reported and offers us a current vision which is many important in the Semantic Web area.

On the website of the journal Semantic Web main parts are separated as reviewers, authors, under-reviewed and reviewed, issues, blog, calls, and scient metrics. This website offers special sub-concepts Ontology Matching and Machine Learning, Knowledge Graph Generation from text, and Knowledge Graph construction the deadline for submission varies from February 2024 and July 2024. I have chosen **the "Ontology Matching and Machine Learning"** special topic as a cited paper.1

The journal Semantic Web – Interoperability, Usability, Applicability (published and printed by IOS Press, ISSN: 1570-0844) is peer-reviewed, open access, and published bimonthly. In short Semantic Web journal, brings together researchers from various fields who share the vision and need for more effective and meaningful ways to share information across agents and services on the future internet and elsewhere.

https://doi.org/10.62792/ut.jnsm.v9.i17-18.p2817

This special issue aims to discuss the latest research proposals and the use of machine learning for ontology matching, data interlinking, and data integration in general. Ontology matching enables Semantic web classic data integration tasks to solve semantic heterogeneity problems. It accepts ontologies as input and returns an alignment—that is, a collection of correspondences between the ontology's semantically linked entities—as output. These correspondences can be applied to several activities, including knowledge graph navigation, query response, data interlinking, and ontology merging. As a result, compatible ontologies allow the knowledge and data provided by the compatible ontologies to work together.

As known Semantic Web is a perspective to extend WWW by contributions using software programs especially based on ontologies, to give meaning to published information and forecastable data. By semantic web computer-based systems can interpret meaningful information like humans' information process way. Here semantic means machine-processable data and web is the navigation. In conclusion, the semantic web is an approach to meaning data by machines in the manner of human vision.

In this part ontology concept as a pre-knowledge knowledge to be summarised, and the reasons for creating an ontology, ontology creation will be discussed. Popular ontology tool Protege is used generally on ontology visualization. What is an ontology? In artificial intelligence literature term ontology has so many definitions one popular definition of ontology is a formal explicit description of concepts in a domain of discourse (classes (sometimes called concepts). properties of each concept describing various features and attributes of the concept (slots (sometimes called roles or properties)), and restrictions on slots (facets (sometimes called role restrictions)). As mentioned above ontology is an abstract definition. Instances of a class defined in ontology are applications of ontology definition. An ontology and instances of classes define that ontology constructs a knowledge base. An ontology together with a set of individual instances of classes constitutes a knowledge base. It is hard to define where ontology ends and the knowledge base starts. Classes are the focus of the ontologies. Because instances are just samples of classes and slots have a meaning if a class exists. Facets depend on slots so all the bases in the ontology are classes. Practically we can say that to define classes we first list possible class candidates. Then decide taxonomic (subclass–superclass) relationships between classes. Then to get good instances we define slots and if necessary slot taxonomic (subclass–superclass) relationships. The last step is filling the instances to the ontology. An ontology defines a common vocabulary to share information in a domain.

2. The List of Literature

Heading	Submission Date	Submission Type
d2kg: An Integrated Ontology for Knowledge Graph-based Representation of Government Decisions and Acts	05/07/2023	Ontology Description
InterpretME: A Tool for Interpretations of Machine Learning Models Over Knowledge Graphs	07/25/2023	Tool/System Report
A Semantic Framework for Condition Monitoring in Industry 4.0 based on Evolving Knowledge Bases	06/08/2023	Full Paper
Optimizing SPARQL Queries over Decentralized Knowledge Graphs	11/29/2022	Full Paper
The Role of Ontologies and Knowledge in Explainable AI	08/23/2023	Editorial
Survey on complex ontology matching	04/09/2018	Reviewed
Data-driven Methodology for Knowledge Graph Generation within the Tourism Domain	11/30/2022	Reviewed

d2kg: An Integrated Ontology for Knowledge Graph-based Representation of Government Decisions and Acts

Using technologies explosion of meaningful data from open governance resources by using aspects like Natural Language Processing, Information Extraction, Data mining, and Semantic Web based on knowledge representation models such as ontologies. Here knowledge graph approach is used for a newly offered ontology d2kg aiming to integrate Greek Government decision mechanism documents in global European ontology standards by separating and embedding the parts of valuable data into a meaningful independent of language aspect for shareability of government decisions all over Europe.

InterpretME: A Tool for Interpretations of Machine Learning Models Over Knowledge Graphs

In this paper, an analytic tool InterpretME framework is offered for predictive models for Knowledge Graphs(CSV, JSON, and various types of data) approach for increasing Machine learning interoperability combining concepts from various existing ontologies for learned predictive models and execution of the models. To achieve this goal InterpretME is designed for working on one side with datasets and datasets via web interfaces SPARQL endpoints for querying ontologies. They published their work on Git Hub and Zenodo also including a Python software solution.

A Semantic Framework for Condition Monitoring in Industry 4.0 based on Evolving Knowledge Bases

Using an ontological model for the representation of the factory as all components in the data version and sensor observations are analysed by observation results using contextual information beyond classical reasoning mechanism for preventing the forecasted undesired sensor results that affect from idea to scale up Industry 4.0 approach. To succeed in a highly dynamic model and produce a real-time true data stream to knowledge graph-based ontological mind complex situations involving specific pattern-based problems tried to be solved by this model. To achieve this interoperability standards are ontologized, proposed framework components are figured, a context ontology model for Industry4.0 is created, and object and data properties are detailed in all aspects to achieve a good monitoring for knowledge base condition monitoring.

Optimizing SPARQL Queries over Decentralized Knowledge Graphs

Query optimization, execution, and results contribute to RDF triples network management definition in a new vision as a set of rules this approach is called LOTHBROK and introduces three contributions, that altogether decrease the communication overhead and in doing so increase query processing performance. Ontology-based Semantic Web repositories provide data quarriable by SPARQL endpoints but a lot of experience showed that the nodes beyond the query process generally fail to prevent this problem article focuses on optimization of queries over decentralized knowledge graphs in aspects of cardinality estimation, locality awareness, and data fragmentation.

The Role of Ontologies and Knowledge in Explainable AI

This is a special paper as a conclusion document for a group of papers that can be divided into two groups, the first group aims to propose ontology specifications and extensions to conceptualize systems over various domains from finance to cyberbullying, and the second group approaches logic-based methodology for fostering explainable usage of unseen data that inherits data from the seen side by interoperability and the main idea is summarised as ontology and knowledge in Explainable AI.

Survey on complex ontology matching

The step-by-step workflow in this paper is in summary First of all background definitions are shared, then complex matching approaches are classified. Followed by state-of-the-art approaches and their evaluation is discussed. In ontology alignment side literature, a single entity of a source ontology is often linked to a target ontology. In this work, many methods for the difficult matching problem are discussed. This work evaluates a classification of the complicated matching approaches according to their specificities (i.e., type of correspondences, guiding structure). There is also a discussion of these approaches' shortcomings and evaluation issues. Future directions for this field's work are outlined. Complex alignments complement simple alignments to overcome conceptual heterogeneity between ontologies. This includes the approaches used to address lethality Models with different expressive power from XML to databases and ontologies). The classification of these approaches is based on suggestions made for specific features. The special features of the complex matching approach are based on: Issuance (type of communication) and its process (Lead structure). Evaluation of these approaches Various datasets were also discussed. finally, several perspectives on this area were discussed.

Data-driven Methodology for Knowledge Graph Generation within the Tourism Domain

Tourism and hospitality sectors have especially in the last few years working on providing new innovative services. As the big data is the new oil of this century and knowledge Graphs are emerging as the most natural way to collect, refine, and structure this heterogeneous information. This article includes the integration of data from ooking.com, AirBnB, DBpedia, and GeoNames.In the document, a method is offered for the semi-automatic generation of a Tourism Knowledge Graph (TKG), which can be used for supporting a variety of intelligent services in this space, and a new ontology for modeling this domain, the Tourism Analytics Ontology (TAO). Because of the methodology's modular situation, it is simple to extend and include new data sources for enrichment and refinement functions. As a result, a comprehensive evaluation of the functional, logical, and structural dimensions of TKG and TAO is mentioned in the article. Framework for TKG is presented and a new ontology is defined for modeling this domain named the Tourism Analytics Ontology (TAO).TKG and TAO are evaluated in functional, logical, and structural dimensions. The evaluation showed that TAO offers higher transparency, flexibility, and cognitive ergonomics than its alternatives Hontology and Accommodation Ontology.

3. Methodology

In the World Wide Web, the vast majority of the web content is human-readable formatted. Softbots can't understand this process and information and much of the potential of the Web is so far untapped. Here Semantic Web takes care of this problem by creating and managing in all

aspects based on ontological approaches. Data has structure and ontologies describe the semantics of data. As we marked data by ontologies softbots can more easily understand the semantics of data and location and integration of data for a wide variety of tasks can be more comfortably done. "The Semantic Web thus offers a compelling vision, but it also raises many difficult challenges. Researchers have been actively working on these challenges, focusing on fleshing out the basic architecture, developing expressive and efficient ontology languages, building techniques for efficient marking up of data, and learning ontologies”.
Here Ontology Matching and Machine Learning will be analysed over selected papers. As the concept of the Semantic Web is a very large area I limit my work to the Protégé a Semantic Web Ontology tool all over the articles. To make it more clear Protégé related works are done and reviewed by me. For each of the articles, I will mention the important aspects here.

d2kg: An Integrated Ontology for Knowledge Graph-based Representation of Government Decisions and Acts

The d2kg ontology is a unique integration of existing ontologies combined with core and controlled vocabularies developed based on EU standards. It provides a customized solution to abide by the requirements of the Greek Programme Diavgeia, extending significantly the respective Diavgeia ontology and proposing at the same time a solution to encode government and administrative decisions/acts that could be universally adopted to integrate public documents produced by other EU Member States, with certain adjustments content-wise. A graphical UML-like representation of the main entities and their relationships is provided below.
For querying this ontology certain indicative Competency Questions (CQs) per Use Case (UC) are expressed as SPARQL queries, where the RDF prefix indicates the namespace of the core W3C RDF vocabulary, while the d2kg prefix indicates the namespace of the new ontology. For example, "For a given organization, which are the persons (and their staff category, rank, post) appointed/recruited (within a given period))?" this query in the SPARQL query is listed below.
SELECT distinct ?birthName ?Staff_Category ?Staff_Rank ?Post where {
?doc rdf:type dvg:Appointment; eli:date_publication ?pub_date; d2kg:staff ?Staff;
?Staff d2kg:staffCategory ?Staff_Category; d2kg:appointedIn ?Post; person:birthName ?birthName; epo:appointedBy dvgo:99221922;
FILTER (?pub_date ⩾ "2015−01−01" AAxsd:date)}

InterpretME: A Tool for Interpretations of Machine Learning Models Over Knowledge Graphs

InterpretME integrates knowledge graphs (KG) with machine learning methods to generate interesting meaningful insights. It helps to generate human- and machine-readable decisions to assist users and enhance efficiency. InterpretME is a tool for fine-grained representation of key features of trained machine learning models in KG. It takes as input some KG feature definitions, classes, and constraints. InterpretME receives her JSON input from the user as shown below. Feature definitions are divided into independent and dependent variables that can be used in prediction. The feature definition has the following format *"x": "?x a* http://dbpedia.org/ontology/Person*. \n ", "gender": "Optional { ?x* http://dbpedia.org/ontology/gender *?gender.}* where the first part states the attribute from the KG and the later part describes the definition of that attribute in the KG using SPARQL. By this property, InterpretME to trace the origin of that feature within the KG. SELECT SPARQL query is created to retrieve application domain data. InterpretME also accepts constraints as input from the user and checks whether the entity validates or disables the constraints.

InterpretME is divided into two main quadrants training an interpretable predictive model and documenting an interpretable predictive model.

A Semantic Framework for Condition Monitoring in Industry 4.0 Based on Evolving Knowledge Bases

Background knowledge from one manufacturer may be shared with another and this helps in monitoring and decision-making more efficiently for the same type of process and resources. To achieve this goal an ontological model is proposed in the article. The proposed framework aims to enrich data collected from sensors and contextual information using an ontological model approach. By the way after creating the ontology classical reasoning for possible no-detection problems prevention. To support decision making proposed framework provides a road map from each aspect. Manufacturers are extremely dynamic. Processing heterogeneous data streams in real time is critical. To test the scalability and complexity issues of real-time situation detection, larger case studies involving more complex situations, including observations of more temporally and spatially related properties, should be considered. Different implementations of stream inference engines need to be evaluated in terms of efficiency and scalability, while also considering the size of the knowledge base.

Optimizing SPARQL Queries over Decentralized Knowledge Graphs

To provide a good way for SPARQL queries over decentralized knowledge graphs this paper offers a new aspect of LOTHBROK. The focusing aspects of LOTHBROK are cardinality estimation, locality awareness, and data fragmentation. LOTHBROK can achieve significantly faster query processing performance compared to the state of the art when processing challenging queries as well as when the network is under high load. This is the Protégé used created ontologies that can be queried by SPARQL. To achieve this goal Query optimizer, query execution, and Node relations are schematized. Since RDF is the commonly used format for storing semantic data as in background logical rules are mentioned. LOTHBROK offers three contributions for query processing performance. First fragments are created to answer an entire star pattern just by a single fragment. Secondly a novel indexing schema Semantically Portioned Blooms Filter SPBF Indexes for positioning subject, predicate, and object locations. Third a query optimization technique is designed for estimating cardinalities and taking consideration of data locality.

Survey on complex ontology matching

Protégé uses Manchester syntax for visualization of complex alignment in OWL language. To obtain the target ontology, for example, adding a new domain for a property from the source ontology we need transformations on relations. Here Protégé can be used as a tool to evolve the ontology or identify posteriori through ontology diff tools. Existing ontology matching approaches are generally reused in this task for finding the initial overlap between the two ontology versions. OWL can represent complex alignments as axioms involving logic constructors and entities from the source and target ontologies. These axioms form a merging ontology. The meaning of correspondence in OWL (considering the meaning of aligned ontologies) is limited to SROIQ logic for reasons of decidability.

Data-driven Methodology for Knowledge Graph Generation within the Tourism Domain

To create a Tourism Knowledge Graph, the steps include defining the use cases, finding and studying information sources, defining the ontology, transforming the data, creating triples for RDF, and publishing the data after loading the RDF triples into a Triple Store that can be queried using the SPARQL language. The TAO ontology schema is shown below, where each arrow represents a semantic relationship, starting from its domain and ending in its range. Automatically generating a Tourism Knowledge Graph presents a significant challenge. The primary issue is the lack of a single ontology that offers fine-grained descriptions of touristic lodging (e.g., hotels), accommodations (e.g., family rooms), amenities (e.g., swimming pools), locations (e.g., amusement parks), and destinations (e.g., London). To address this gap, the Tourism Analytics Ontology (TAO) is introduced in the article. After creating the TAO ontology, the owlready2 library in Python is used to programmatically enrich it, avoiding the need for manual editing. The TAO ontology must be capable of modeling information obtained from typical data sources in the tourism sector. These sources, such as Booking.com and AirBnB, provide (semi-)structured data as key/value properties as well as unstructured data. A tree representation of the four hierarchies included in the TAO ontology, expanded to the third level, is shown below.

4. Results

d2kg is a huge monolithic ontology. So, it is even hard to understand it besides its reuse is also very difficult. Online documentation of the ontology can be improved. In d2kg the work of generating automatically does not work in such large ontologies. For instance, when a domain or range restriction is so long that it takes several pages, it becomes impossible to even scroll through the lists of classes and properties reasonably. My suggestion would be to modify the script generating the documentation, to maybe avoid including everything (e.g. using pop-ups of expandable panels on the documentation page), or highlighting some parts while leaving other things to be found in the ontology file itself.

InterpretME's strong sides are it is well motivated, it is an available tool that can be used in many domains for interpreting the result of various ML models, and also to facilitate their predictive improvement. Besides these, all resources are available publicly for reproducibility. A new tool named InterpretME includes a pipeline to enable the interpretability of predictive machine learning models along with a knowledge graph and an ontology to support the process of describing the characteristics of trained ML models.

In Industry 4.0, materials and machines in factories are equipped with sensors that collect data to monitor conditions. This is a challenging task as it requires the integration and processing of disparate data from different sources with different physical resolutions and underlying mechanisms. Ontologies have emerged as a suitable way to handle data integration and represent artificial intelligence in a machine translation process by creating semantic models. In addition, the monitoring of industrial processes depends on the dynamic context of the operation. In these situations, audiological models should provide a way to represent this evolution to show which contextual resources are used during decision-support actions. This paper proposes a framework to address the development of knowledge bases for monitoring trends in Industry 4.0. For this, we propose an ontological model (COInd4 ontology) for the production domain that describes the resources and processes in the factory, with special attention to the context of these resources and processes.

Aspects related to the combination of sensor observations and domain knowledge are also included in the model. This approach uses a proposed language model to enrich data collected from sensors with contextual information. The flow intervention manages the integration of

data from multiple data sources with different time resolutions, as well as the real-time processing of this data. This allows you to generate high-level context from low-level contexts and musical information. In this way, it contributes to supporting decisions by enabling the identification of actions that can correct irregularities and avoid the interruption of the production process.
The Web of Data, as a whole, provides access to a vast array of interconnected data, but the architecture of today's Semantic Web largely relies on data providers to access their data via SPARQL nodes. However, several studies have shown that these endpoints frequently experience downtime, making the data they manage unavailable. Decentralized systems based on P2P (peer-to-peer) technology have been shown to improve access to knowledge graphs, although a large fraction of nodes fail. In such a setup, processing queries can be expensive, as the data required to solve a single query may be distributed across multiple nodes. That's why we've proposed an approach to SPARQL query optimization for decentralized knowledge graphs: **Lothroko**. While there are many factors to consider in optimizing these queries, we've focused on three: Colony estimation, Locality awareness, and Data fragmentation.
In tourism domain paper compared to other methodologies is a good approach but it seems to be trivial when compared to other ontologies. Generating a knowledge graph approach is not clear when discussing criteria about other ontologies. Tourism Analytics Ontology says that existing ontologies do not cover different concepts and properties. The need for creating a new ontology is not clear in all manners. Functional dimensions are based on competency questions that assess the ontology.
However, this only proves that the ontology fulfills its purpose but not its originality compared to the state of the art. Logical dimensions based on inference and error provocation that assess the knowledge graph. Structural dimensions to assess the ontology and the knowledge graph. While these metrics were used in literature, it is not clear what is proven here. The proposed TAO ontology is only compared with Acco (an ontology that is reused by TAO) and only 1 other ontology, the Hontology but not with all other ontologies that were used in the tourism section and for which a comparison is needed to show how TAO positions concerning the state of the art.

5. Conclusions

In conclusion, along with this work, it can be easily observed that nearly all of the papers create their way to provide a new solution for ontology maintenance but here another problem arises. The blackness of previous works is criticized but the new methodology is in baby steps. As Tim Berners Lee's vision, the Semantic Web is a great improving version of the web's future but nowadays there are great problematic approaches. Ontologies especially used Protégé Tool and Pellet Reasoner provide massive visualization into ontology domains. The works selected are generally huge ontology domains so just sustainability also another issue to be taken care of. To achieve Artificial Intelligence on ontology matching and knowledge graphs time is our ally as a conclusion. With my best regards.

References

[1] https://www.semantic-web-journal.net/
[2] https://www.semantic-web-journal.net/content/d2kg-integrated-ontology-knowledge-graph- based-representation-government-decisions-and-1
[3] https://www.semantic-web-journal.net/content/interpretme-tool-interpretations-machine- learning-models-over-knowledge-graphs-1
[4] https://www.semantic-web-journal.net/system/files/swj3481.pdf
[5] https://www.semantic-web-journal.net/system/files/swj3312.pdf
[6] https://www.semantic-web-journal.net/content/role-ontologies-and-knowledge-explainable- ai

[7] https://www.semantic-web-journal.net/content/survey-complex-ontology-matching
[8] https://www.semantic-web-journal.net/content/data-driven-methodology-knowledge-graph-generation-within-tourism-domain