TECHNICAL UNIVERSITY OF CLUJ-NAPOCA

ACTA TECHNICA NAPOCENSIS

Series: Applied Mathematics, Mechanics, and Engineering
Vol. XX, Issue xx, Month, 20xx

EXPLORATION OF DISTRIBUTED TECHNOLOGIES FOR POSSIBLE INTEGRATION OF VOCABULARIES AND AGENTS IN THE INDUSTRY

Nicolaie CIUBOTARU, Angela REPANOVICI

Abstract: *The growing number of industrial and research actors looking for a common semantic middle ground happens in the advent of themselves becoming actors of the Industry 4.0 paradigm. This involves cloud computing, robotics, the Internet of Things, and Additive manufacturing among other conceptual boundaries. One thing is hampering an organic extension of cooperation between the industry and the active parts of the renewed European Research Area: the closed systems of centralised systems that stand in the way of realising the Semantic Web, a true Linked Open Data vision. This paper looks into the existing communication models and product-related vocabularies keeping in focus the future chains of data and truth sources that distributed ledgers and blockchains are ofering. Possible paths of exploitation and existing practices are taken into analysis to look beyond today's technologies in information, data, and knowledge exchange.*
Key words: *distributes technologies, vocabularies, information technology, industry 4.0*

1. INTRODUCTION

One of the policies of the European Commission (EC) is Digitising industry in the context of a larger Priority called Digital Decade. From this policy description (https://digital-strategy.ec.europa.eu/en/policies/digitising-industry) we retain for future reference through the "digital innovation hubs" and the need for a "regulatory framework". In 2020, EC adopted the industrial strategy[1] where data and metadata analytics are seen as key enablers for transition in the context of the European strategy for data. Although Industry 4.0[2] is the policy catalyst for most European industry related policies and initiatives, a recent brief paper (Industry 5.0: A Transformative Vision for Europe[3]) pushes beyond Industry 4.0 technological paradigm aiming to build cyber-physical objects in a broader context of efficiency, digital connectivity and artificial intelligence. A new dimension concerning social effects that technology monopoly and wealth unbalanced distribution should be taken into consideration if European policies are to succeed. Industry 4.0 is part of the Industry 5.0, and also as the broader context of this paper.

Industry 4.0 is multi tier transformation of the ways manufacturing is done making use of big data, IoT (Internet of Things), IoS (Internet of Services), and cloud computing coupled with adaptive systems.

German Electrical and Electronic Manufacturers' Association (ZVEI) developed a Reference Architecture Model Industrie 4.0 (Industry 4.0) known as RAMI4.0[4] on the bases of DIN SPEC 91345:2016-04. This is a multi-layer tridimensional representation of the Industry 4.0 parts/concepts acting as a reference architecture model. Two of the middle layers are the data/information and communications in the larger context of ever-growing numbers of smart devices. The Life Cycle Value Stream is actually the IEC 62890, and the Hierarchy Levels are IEC 62264 (enterprise-control system integration) / IEC 61512 (batch production records). This effort concerns machines becoming intelligent (Cyber-Physical Systems) as a direct result of adding sensors and wireless connectivity. The coupling with a system able to harmonise the industrial processes is the centre target of Industry 4.0 (I40).

The forth stage in Industry 4.0 development is realising what happens within the various processes. This requires a necessary level of transparency and most of all designing knowledge systems capable to understand what is happening. This level is where knowledge representation systems need to be developed along with machine learning employed in data mining.

Internet of Things (IoT) in this context involves building machinery systems that has part of the management systems augmented with sensory arrays, and software coordination components capable of syncing to a large orchestrated assembly (smart manufacturing). Problems appear at the moment when information systems and different data models with little to none transparent semantics and put to work together in an interoperable manner.

Recently there are efforts put in harmonising different parts of the RAMI4.0 model into a Industry 4.0 Standards Knowledge Graph[5].

2. TRANSFORMING PROCESSES INTO INTELLIGENT ONES

Gaining good insight from data often times implies the use of taxonomies, ontologies, and models for establish certain needed semantics. The objective is modelling a knowledge domain linking data, encoding it into a knowledge graph and make use of it through machine learning algorithms with the intent of obtaining a new insight. Taxonomies are controlled vocabularies set as hierarchical schemas of terms that act as facets and establish relations based on "appropriateness". Taxonomies containing information on how the terms should be uses becomes a thesaurus. Taxonomies have their wide range uses, but what data needs in order to be capable of linking to other data, is a system that allows one to better control the concepts of a subject area. This elevated state of control is given by ontologies. Any knowledge domain has a corpus of terms to operate with the different realities at hand. These are structured in the form of vocabularies. One of the best is AGROVOC - https://agrovoc.fao.org/browse/agrovoc/en/, a multilingual vocabulary for agriculture with a clear path to every concept. It is considered a Linked Open Data set. AGROVOC makes use of SKOS (Simple Knowledge Organisation System).

Ontologies are known to be a bridge between human communication and machine readability. What are best for is formalisation of the taxonomies, classifications and the relation established between the parts that are describing a domain from real life, i.e. scientific research. An abstract view would be seeing an ontology as the relations established between classes (things from real world are considered classes), sub-classes and properties. Instances of ontologies adhere to the principles of modelling knowledge graphs, and this intelligent state in which heterogeneous data is transformed into is the material of Machine Learning exploitation exposing the relations between the nodes.

The moving parts of ontologies are not easy to grasp, and are met with the need for a keen interest in establishing a domain ontology able to describe the entities and the relations they establish among each other and also with other entities belonging to other domain ontologies. The much sought after benefit is the capacity to do automatic reasoning through inferences, and this particular aspect if well balanced leads to making data smart, consistent, and connectable - linked data.

All software applications persist data by the means of a database often Entity Relation Databases. Thus, data is kept into forms that need aggregation and further representation formats in order to being communicated to other applications. The main issue with this model of so called data silos, is the different meaning and diversity of the entities at play. A solution to the issue is the use of ontologies although some efforts were put into creating exploratory software designed to process databases as virtual knowledge graph systems (ontop - https://ontop-vkg.org/). At the current state, in order to gain insight into complex processes that need to be replicated, one needs to make data smart. This means you need to describe and analyse data in the context of certain semantics.

Ontologies are the glue keeping together separate components of a automation system acting as connection bridges between organizations when production and management chains seek tighter integration. According to Szilárd Jaskó et al. study[6], Manufacturing execution systems (MES) ontologies support collaboration and most of all interoperability due to intrinsic modularity. The industrial field is making use of ontologies for certain parts of the production management and because of this approach, the ontologies are incompatible. There is a response to the issue embodied in Industrial Ontologies Foundry accessible at https://www.industrialontologies.org/. The mission states the place as one where ontologies in the domain of digital manufacturing are found.

Building ontologies takes time and is the product of a multidisciplinary team in which subject matter experts and mathematicians should be the nucleus. An ontology engineer is most desirable. The reason is getting a product able to describe as accurate as possible the entities and the processes existing at the level of a domain. The bulk of the iterative work revolves around a translation of definitions that the subject matter expert sets for each individual class or property into formal definitions known in the field of ontologies as First Order Logic definitions. A First Order Logic is the apparatus needed to build a knowledge representation language. In most cases the family of knowledge representation languages OWL (Web Ontology Language) it is used to write the ontologies. The final product of this effort is usually a machine-readable ontology ready for integration within software systems that use knowledge bases (knowledge graph) for process modelling or management. remains in question the issue of multiple ontologies integration, a step necessary to realize the Semantic Web.

One particular important example is The Open Biological and Biomedical Ontology (OBO) Foundry, a place for the filed of life sciences where among many activities, harmonization and sharing ontologies are some important aspects needed to be approached in every field. Ar the core of OBO there is a repository that keeps the metadata files of more than 150 ontologies. This architectural model is backed by OBO set of principles which is the main evaluation instrument, and helps with automatic evaluation of each ontology metadata done through operational tests for conformance. An interesting method to update the files when updates are coming, Github it is used and modification are done as calls (pull requests) for each individual file whenever it is necessary.

There is one particular initiative started in 2002 aim to building high level ontology capable of describing scientific information from all fields. It is known as Basic Formal Ontology (BFO) which is a top level domain neutral ontology. BFO is also a standard as ISO/IEC 21838-2:2021 Information technology - Top-level ontologies (TLO) - Part 2: Basic Formal Ontology (BFO) available at the following link https://www.iso.org/standard/74572.html.

In the manufacturing field Eeva Järvenpää et al. [7] mentioned the efforts put in research of systems capable of harmonizing production processes based on shared semantics, i.e. ontologies. These efforts lead to a new paradigm called Manufacturing-as-a-Service. Eeva Järvenpää et al. (2018) have developed a model of ontology called MaRCO able to describe what are the capabilities of manufacturing resources in contexts where there the communication is done machine-to-machine. Its purpose is to "support automatic match-making between product requirements and resource capabilities".

One potential valuable track is the European OntoCommons project (https://ontocommons.eu/) is an H2020 CSA (Coordination and Support Actions) takes aim at the data documentation in materials and manufacturing domain. This should be achieved through the Ontology Commons EcoSystem (OCES) which is a set of ontologies, and a toolkit for developing and maintaining ontologies.

The purpose of building all these ontologies is to put structure in an formalized manner. The outcome is usually a knowledge graph. For Industry 4.0 there is one proposal made by Sebastian R. Bader[8], a model that makes use of external data (see Linked Open Data Cloud and BDpedia).

One important aspect that need deep consideration is how these high-level knowledge products able to describe the world will be managed in the future. How entities and the relationships they establish within the confines of an ontology could break the barriers and start to lend themselves to other ontologies getting past one graph. How a semantically charged term is uniquely identified, and what would be the technology powerful enough to break these boundaries. A possible solution might be the emergent technologies called generically Web3.

3. UNIQUE ENTITIES THROUGH DECENTRALIZED WEB TECHNOLOGIES

One important aspect residing with the ontologies is that the terms for classes and properties are uniquely identified through Unique Resource Identifiers, and the best known form of of it is the Universal Resource Locator, the simple link everyone is accustomed with. This is the mechanism necessary to identify uniquely a term, a concept.
Ontologies need management systems and one generic using APIs (Application Programming Interfaces) was proposes by Stephan Bloehdorn et al.[9]. Unfortunately unique identification on long term via URIs is met with inconsistencies the model client-server in a centralized web model is posing. Another major issue is having the same resource duplicated and identified by multiple URIs, and to make matters worse, the same URI might identify a resource that has already been versioned. There is a new class of technologies emerging in the last decade called today Web3.

For the last decade a decentralized web and distributed ledgers began to grow in importance beyond tokenization of different assets like cryptocurrencies and more recently non-fungible tokens (NFTs). Two of the class applications will be taken into account being relevant to a possible context for ontologies and data exploitation: Interplanetary File System (IPFS) and the blockchains. If current access model for web resources is dealing mostly with HyperText Transfer Protocol (HTTP), IPFS is a hypermedia protocol using the peer-to-peer model. The first ensures that data of the ontology is kept as a unique entity, and the second gives way to a valuable secure transaction model based on smart contracts (a small piece of software executing in the environment of a blockchain).

Most of the use cases for distributed ledgers come from developing supply chains or managing IoT (Internet of Things) or FinTech (Financial Technology) systems. It is rare to find ontologies implemented, but there is a proposal for a model involving distributed ledgers and it comes from Ugarte-Rojas Hector, and Chullo-Llave Boris[10] who developed BLONDIE (Blockchain Ontology with Dynamic Extensibility) underpinning a model of implementation for an ontology using OWL (Web Ontology Language) capable of interfacing with Bitcoin and Ethereum networks from where data about transaction would be readily available.

For a possible model of ontologies usage with Web3 technologies, first we need to take care of the data. For this InterPlanetary Linked Data (IPLD) coupled with InterPlanetary Name System (IPNS) are up to the task. IPLD is the "data model of the content-addressable web" (https://ipld.io/) and the IPNS a way to address the data structures (files) using a classical URI.

Ontologies may be tokenized as long as any containing entities could be stored in IPFS as linked data, and doing so becoming identifiable through a Content Identifier - an alternative to URI. The entire ontology could become a smart contract able to define formally the interactions between the entities. An ontology encoded as such could be considered the state of a particular domain. Then, a particular data could be enriched with descriptions, and in the end ready to become the raw material to a Natural Language Processing algorithm or the bases for a Machine Learning Model. The main gain of such a scenario is traceability of each stage in a possible reasoning procedure due to the nature of how distributed ledgers are working as a global distributed trustworthy database.
In a matter of speaking, an ontology will have its classes, subclasses and properties as distinct IPLD data identified by CIDs managed through a smart contract that defines the relations between those entities or defines what happens in case of certain interactions. This could be the bridge with other intelligent algorithms trained to do concept/entity extractions from raw text, and act on the identified concept according to the rules written in the smart contract in case of ontology matching.

The management of an ontology using a smart contract could be seen as far fetched, and for this reason less important at this stage. The model could could be viewed as an upgrade to a distributed storage and a different truly unique persistent identification. An interesting similar approach is debated by Knez et al.[11] which employ the benefit of blockchain in ontology track changing the modifications done using Protégé software for building.

7. CONCLUSION

Ontologies are capable instruments if wielded in an enabling context of connection, common semantics extending reduced appliance to a particular domain.

Most of the ontologies used today in manufacturing are small in coverage, usually extending to only a portion of a particular domain, and not always transparent contrary to the principles of Findable, Accessible, Interoperable, and Reusable (FAIR). There are some examples to counterbalance this picture like The Open Biological and Biomedical Ontology (OBO) Foundry.

One particularly interesting reason for employing small ontologies is that an extensive ontology is producing poor results at the moment reasoners are trying to extract facts. Another reason is directly related to the effort is put into building an ontology, including revisions needed to keep it relevant. There is another aspect concerning ontology building an data enriching which pertains to the issues of personnel qualifications needed to operate such systems. One particular well funded example is the project Big Data to Knowledge (BD2K) developed since 2014 by the U.S. National Institutes of Health. Trainig opportunities abound if one glances at the page https://commonfund.nih.gov/bd2k/resources#Training.

The ontologies of the future have a fertile ground in smart contracts, as this new class of dynamic secure and uniquely identifiable transaction lend themselves to the vast implementation opportunities including industry and manufacturing.

8. REFERENCES

[1] *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS A New Industrial Strategy for Europe*, (2020). https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593086905382&uri=CELEX%3A52020DC0102
[2] Teixeira, J. E., & Tavares-Lehmann, A. T. C. P. (2022). Industry 4.0 in the European union: Policies and national strategies. *Technological Forecasting and Social Change*, *180*, 121664. https://doi.org/10.1016/j.techfore.2022.121664

[3] Directorate-General for Research and Innovation (European Commission), Renda, A., Schwaag Serger, S., Tataj, D., Morlet, A., Isaksson, D., Martins, F., Mir Roca, M., Hidalgo, C., Huang, A., Dixson-Declève, S., Balland, P.-A., Bria, F., Charveriat, C., Dunlop, K., & Giovannini, E. (2021). Industry 5.0, a transformative vision for Europe: governing systemic transformations towards a sustainable industry. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/17322

[4] *RAMI 4.0 - ISA*. (n.d.). Isa.Org. Retrieved September 14, 2022, from https://www.isa.org/intech-home/2019/march-april/features/rami-4-0-reference-architectural-model-for-industr

[5] Bader, S. R., Grangel-Gonzalez, I., Nanjappa, P., Vidal, M.-E., & Maleshkova, M. (2020). A Knowledge Graph for Industry 4.0. *The Semantic Web: 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31–June 4, 2020, Proceedings*, 465–480. https://doi.org/10.1007/978-3-030-49461-2_27

[6] Jaskó, S., Skrop, A., Holczinger, T., Chován, T., & Abonyi, J. (2020). Development of manufacturing execution systems in accordance with Industry 4.0 requirements: A review of standard- and ontology-based methodologies and tools. *Computers in Industry*, *123*, 103300. https://doi.org/10.1016/j.compind.2020.103300

[7] Järvenpää, E., Siltala, N., Hylli, O., & Lanz, M. (2019). The development of an ontology for describing the capabilities of manufacturing resources. *Journal of Intelligent Manufacturing*, *30*(2), 959–978. https://doi.org/10.1007/s10845-018-1427-6

[8] Bader, S. R., Grangel-Gonzalez, I., Nanjappa, P., Vidal, M.-E., & Maleshkova, M. (2020). A Knowledge Graph for Industry 4.0. *The Semantic Web: 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31–June 4, 2020, Proceedings*, 465–480. https://doi.org/10.1007/978-3-030-49461-2_27

[9] Bloehdorn, S., Haase, P., Huang, Z., Sure, Y., Völker, J., van Harmelen, F., & Studer, R. (2009). Ontology Management. In J. Davies, M. Grobelnik, & D. Mladenić (Eds.), *Semantic Knowledge Management: Integrating Ontology Management, Knowledge Discovery, and Human Language Technologies* (pp. 3–20). Springer. https://doi.org/10.1007/978-3-540-88845-1_2

[10] Hector, U.-R., & Boris, C.-L. (2020). *BLONDiE: Blockchain Ontology with Dynamic Extensibility*. arXiv. https://doi.org/10.48550/arXiv.2008.09518

[11] Knez, T., Gašperlin, D., Bajec, M., & Žitnik, S. (2022). Blockchain-Based Transaction Manager for Ontology Databases. *Informatica*, *33*(2), 343–364. https://doi.org/10.15388/22-INFOR490

Exploarea tehnologiilor distribuite în vederea unei posibile integrarea a vocabularelor şi ai agenţilor din domeniul industrial

Numărul tot mai mare de actori industriali şi de cercetare care caută un spaţiu semantic comun se petrece în odată cu transformarea acestora în actori prinşi în paradigma Industriei 4.0. Aceasta implică cloud computing, robotică, Internetul lucrurilor şi fabricarea aditivă, printre alte delimitări conceptuale. Un lucru împiedică extinderea organică a cooperării între industrie şi părţile active ale Spaţiului European de Cercetare reînnoit: sistemele închise de sisteme centralizate care stau în calea realizării Web-ului Semantic, o adevărată viziune Linked Open Data. Aceast articol analizează modelele de comunicare existente şi vocabularele legate de produse, menţinând în atenţie viitoarele lanţuri de date şi surse de adevăr pe care registrele distribuite şi blockchain-urile le oferă. Posibilele căi de exploatare şi practicile existente sunt luate în considerare pentru a privi dincolo de tehnologiile actuale în schimbul de informaţii, date şi cunoştinţe.

Nicolaie Ciubotaru, PhD, Transilvania University of Braşov, Mechatronics and Environment Department, Mechatronics and Environment Department, nicolaie.ciubotaru@unitbv.ro, 0731148005

Angela Repanovici, Prof. Dr. Eng, Transilvania University of Braşov, Mechatronics and Environment Department, Mechatronics and Environment Department, arepanovici@unitbv.ro