



New directions for applied knowledge-based AI and machine learning

Selected results of the 2022 Dagstuhl Workshop on Applied Machine Intelligence

Bernhard G. Humm¹ · Phil Archer² · Hermann Bense³ · Carolynn Bernier⁴ · Christian Goetz⁵ · Thomas Hoppe^{6,7} · Fabienne Schumann⁸ · Melanie Siegel⁹ · Rigo Wenning¹⁰ · Alexander Zender¹

Accepted: 26 August 2022
© The Author(s) 2022

Abstract

In this article, selected new directions in knowledge-based artificial intelligence (AI) and machine learning (ML) are presented: ontology development methodologies and tools, automated engineering of WordNets, innovations in semantic search, and automated machine learning (AutoML). Knowledge-based AI and ML complement each other ideally, as their strengths compensate for the weaknesses of the other discipline. This is demonstrated via selected corporate use cases: anomaly detection, efficient modeling of supply networks, circular economy, and semantic enrichment of technical information.

Introduction

In 2014 we started a series of annual workshops at the Leibniz Zentrum für Informatik Schloss Dagstuhl, initially focusing on corporate semantic web, and later widening the scope to applied machine intelligence (AMI). In all work-

shops, we focussed on the application of *artificial intelligence (AI)* technologies in corporate and organizational contexts. A number of books [3–5] and journal articles [6–10] resulted from those workshops. Those workshops have an intense atmosphere of interdisciplinarity, collaboration and focus on practical results (see [11]).

✉ Bernhard G. Humm
bernhard.humm@h-da.de

Phil Archer
phil.archer@gs1.org

Hermann Bense
hb@bense.com

Carolynn Bernier
carolynn.bernier@cea.fr

Christian Goetz
christian.goetz@yaskawa.eu

Thomas Hoppe
thomas.hoppe@fokus.fraunhofer.de,
thomas.hoppe@htw-berlin.de

Fabienne Schumann
fabienne.schumann@dictajet.de

Melanie Siegel
melanie.siegel@h-da.de

Rigo Wenning
rigo@w3.org

Alexander Zender
alexander.zender@h-da.de

¹ Hochschule Darmstadt—University of Applied Sciences, Schöfferstr. 3, 64295 Darmstadt, Germany

² Blue Tower, GS1 Global Office, Avenue Louise 326, 1050 Bruxelles, Belgium

³ bense.com GmbH, Schwarze-Brüder-Str. 1, 44137 Dortmund, Germany

⁴ CEA, List, Univ. Grenoble Alpes, 17 rue des Martyrs, 38000 Grenoble, France

⁵ Yaskawa Europe GmbH, Hauptstr. 185, 65760 Eschborn, Germany

⁶ Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin, Germany

⁷ HTW Berlin, Wilhelminenhofstr. 75A, 12459 Berlin, Germany

⁸ dictaJet Ingenieurgesellschaft mbH, Wiesbaden, Germany

⁹ Hochschule Darmstadt—University of Applied Sciences, Max-Planck-Str. 2, 64807 Dieburg, Germany

¹⁰ ERCIM/W3C, 2004, Route des Lucioles, 06902 Sophia-Antipolis, France

With the AMI 2022 Workshop we continued our workshop series at Schloss Dagstuhl after 2 years' interruption caused by the Corona pandemic. In between, we held the AMI workshop online. However, this year's workshop has again shown that nothing can substitute the interactions of in-person meetings with their intense discussions during sessions and the social interactions during coffee breaks, meals and common activities.

In this article, we are happy to present selected results and insights that we gained during this year's AMI workshop. We focus on real-world problems and practical issues raised by the application of semantic technologies, data science and machine learning.

This article is structured as follows. In the following section, we present insights and new directions for knowledge-based AI, and in the section thereafter, innovations for machine learning. Then, selected corporate use cases are shown, namely anomaly detection in manufacturing, efficient modeling of supply networks, circular economy, and semantic enrichment of technical documentation.

Knowledge-based AI

Since the 2010s, AI has experienced an enormous boost, mainly due to spectacular successes in machine learning (ML). Because of this, the term AI is nowadays often equated with ML in media coverage. However, this view is clearly constricted, since *knowledge-based AI* has also been an essential AI approach for many decades and is used in many real-world applications. We show some current and new developments regarding ontologies, WordNets and semantic search in the following sections.

Ontologies

Ontologies are a major approach of knowledge-based AI and are in everyday use. A commonly cited definition of ontology is "an explicit specification of a conceptualization" [12]. Gruber furthermore explains conceptualization as "the objects, concepts, and other entities that are presumed to exist in some area of interest and the relationships that hold them". This definition is broad enough to include formal upper-level ontologies but also knowledge graphs, thesauri and other forms of knowledge representation. (There are entirely different opinions on this, as we point out below.) Examples are domain-specific ontologies like Gene Ontology¹ or the Global Biodiversity Information

Facility² for life sciences, but also Wikidata³ or the Google Knowledge Graph⁴ for general purposes.

Use cases for ontologies are semantic search, decision support systems or interactive systems like knowledge-based chatbots. In addition, conceptualizing a domain of interest is of value per se. Ontology search engines like the Basic Register of Thesauri, Ontologies & Classification (BARTOC)⁵ include more than 3000 ontologies, each with thousands or even hundreds of thousands of entities being specified.

When having a closer look at concrete ontology examples (in the broad meaning of Gruber's definition cited above), different flavors can be distinguished; and with those flavors, entirely different communities are associated.

One flavor includes ontologies, for which the term *knowledge graph* has been used in recent years⁶. For example, for the Google Knowledge Graph, simple standardized ontology schemas like schema.org⁷ are used. schema.org specifies a limited number of classes (currently about 800) like persons, organizations, locations, products, events, etc. Simple standardized modeling languages like RDFa, Microdata and JSON-LD may be used to specify concrete entities. Those simple modeling languages provide restricted semantic expressiveness, including specifying classes, instances, attributes, and relationships, but excluding complex reasoning.

The following example⁸ shows the JSON-LD specification of an organization with its legal name, founding date and email address:

```
{ "@context": "https://schema.org/",
  "@type": "Organization",
  "legalName": "Elite Strategies Llc",
  "foundingDate": "2009",
  "email": "info@elite-strategies.com" }
```

Because of the simplicity of modeling languages and schemas used, we call this flavor of ontologies *lightweight*. Lightweight ontologies are much used in industrial practice, e.g., by all major internet search engine providers and tech companies, including Google, Amazon, eBay, Facebook, IBM, LinkedIn and Microsoft [14]. They are relatively easy to use. Various tutorials and guidelines can be found to access such ontologies or provide data for them,

¹ <http://geneontology.org>.

² <https://www.gbif.org>.

³ <https://www.wikidata.org>.

⁴ <https://developers.google.com/knowledge-graph>.

⁵ <https://bartoc.org>.

⁶ Although the term knowledge graph has been already used since 1972, the current use of the phrase stems from the 2012 announcement of the Google Knowledge Graph [14].

⁷ <https://schema.org>.

⁸ extracted from <https://jsonld.com/organization>.

⁹ <https://jsonld.com>.

e.g., jsonld.com⁹. However, their discussion in the academic community is rather sparse.

Another flavor of ontologies uses modeling languages with much higher semantic expressiveness, such as OWL¹⁰ or F-logic [16]. Such ontology modeling languages allow knowledge engineers to express reasoning, logical operators, quantifiers, etc. Ontology examples are the BFO¹¹ and the Gene Ontology. Consider the following example from [15] formalizing the red color of strawberries using a set relation (`subClassOf`) and a quantifier (`some`):

```
Strawberry subClassOf bearerOf some
RedColourQuality
```

Those ontologies are inherently complex, and, therefore, we call this flavor *heavyweight*. There is much attention being paid to this flavor in the academic community [13, 17–19]. Due to the high expressivity and, thus, inherent complexity of ontology modeling languages, several guidelines for engineering ontologies exist, e.g., [15]. To model the simple statement that strawberries are red in color in the complicated way shown above is, in fact, a recommended modeling pattern from [15].

In those guidelines, lightweight ontologies (including languages for modeling them like JSON-LD) are usually not mentioned at all, or it is stated that this flavor cannot be considered in ontologies at all (e.g., [15], Section 2.5). Even though this viewpoint is rather common, it is still in contrast with Gruber’s broad ontology definition from 1993 cited above.

We notice that around those different ontology flavors, there are different communities dealing with them: rather developer-oriented communities for lightweight ontologies and academic communities for heavyweight ontologies. However, there seems to be minimal overlap and little communication between those communities. It seems like they live on different, disconnected islands. Why is this the case? Maybe there are fundamentally differing goals driving ontology engineering in those different communities.

One goal for developing lightweight ontologies is developing knowledge-based applications like semantic search engines. And the software engineering principle KISS (keep it short and simple) also applies to engineering ontologies to be used in such applications. Consequently, modeling languages, guidelines and tutorials are as simple as possible. The simpler, the better; this is the secret to their success.

One goal for developing heavyweight ontologies is formalizing an application domain as accurately as possible in order to draw valid logical conclusions. And, as reality is complex, the modeling languages and resulting ontolo-

gies reflect this complexity; the more complex, the more accurate, and therefore, the better.

What can we conclude from this observation? We think that it is most important to be absolutely clear about the goals before starting to develop an ontology. To be aware of the different ontology flavors and corresponding communities may help us choose the appropriate modeling languages and tools suiting those goals.

Automated engineering of WordNets¹²

A *Wordnet* is a lexical database of semantic relationships between words in a specific language. The first Wordnet¹³ was created for the English language at Princeton University. As the usefulness of Wordnets as lexical resources became apparent, the Princeton Wordnet was expanded, and WordNets were constructed for other languages.

OMW (Open Multilingual Wordnet)¹⁴ is an open-source project that was launched with the goal to ease the use of Wordnets in multiple languages. OMW has the added benefit of connecting equivalent sets of synonyms (called “synsets” in Wordnet)¹⁵ in different languages by means of an interlingual ID called “ILI”.

OdeNet¹⁶ was constructed from open-source linguistic resources in combination with some manual and semi-automatic corrections. Since OdeNet was constructed independently of existing resources in OMW, it was difficult to connect equivalent synsets in OMW via ILI. As an initial implementation, Google Translate¹⁷ was used in combination with statistical methods as described by Siegel & Bond (2021 [31]). However, this implementation has some shortcomings, including:

- Incorrect ILI classification for some synsets from a semantic perspective.
- Duplicate assignment of some ILIs to multiple synsets.
- Part of speech (POS) for some ILIs being inconsistent between the English Wordnet EWN and OdeNet.

A significant problem in using machine translation to connect equivalent synsets in different languages occurs when translating homographs (words with the same spelling but different meanings) and polysemes (words with the same spelling but different although related meanings). This is particularly noticeable, when a word translated from

¹⁰ <https://www.w3.org/TR/owl2-syntax>.

¹¹ <http://basic-formal-ontology.org>.

¹² The work described here was carried out together with Johann Bergh, Lingolutions.

¹³ <https://wordnet.princeton.edu>.

¹⁴ <https://omwn.org/>.

¹⁵ Synsets are the basic structure of Wordnet: a set of synonyms represents a semantic concept.

¹⁶ <https://github.com/hdaSprachtechnologie/odenet>.

¹⁷ <https://translate.google.com>.

a source language is a homograph or polyseme in the target language. As an example, we take the German word “Unterlegscheibe” from OdeNet. The corresponding English translation is “washer”. Searching for “washer” in EWN, we find three synsets containing the word:

- ILI i94042: someone who washes things for a living.
- ILI i60971: seal consisting of a flat disk placed to prevent leakage.
- ILI i60970: a home appliance for washing clothes and linens automatically.

All synsets in EWN have a short, concise definition. We propose to use this definition to get more context for the disambiguation. First, we combine the word in the synset and the definition and do machine translations. Then, we extract the translated word from the machine translation and look for a corresponding match in OdeNet.

These are the results for the “washer” example:

- ILI i94042: Wäscher: jemand, der beruflich Dinge wäscht.
- ILI i60971: Unterlegscheibe: Dichtung, die aus einer flachen Scheibe besteht, um ein Auslaufen zu verhindern.
- ILI i60970: Waschmaschine: ein Haushaltsgerät zum automatischen Waschen von Kleidung und Wäsche.

As is evident, the machine translation of the second item now enables us to make the correct ILI classification (i60971) for the corresponding OdeNet synset.

Since there still could be multiple candidates in OdeNet synsets for ILIs in EWN synsets, it is necessary to write a classification function to assign weights to each of the candidates, so that the most optimal assignment can be made. Fortunately, OdeNet is very synonym-rich (much more so than other Wordnets), and we can use these synonyms in combination with a German Word2Vec [32] model to do the classification.

The content words in the translation of the definition are added to a vector. All the synonyms in the candidate synset are also added to a vector. For each value in both vectors, a similarity value is computed. These values are summed and normalized to a value between 0 and 1, which is the weighted value for the candidate synset in OdeNet competing for the ILI in a specific EWN synset.

As a result, there are no more duplicated ILIs in OdeNet, and we could correct more than 400 wrong POS tags (of about 36,000 synsets).

Innovations in semantic search

Semantic search is well understood and well documented, see [5]. It exploits the semantic relationship between concepts such as generic term, subordinate term, opposite, synonym, etc. However, relationships to partial terms and

partial meanings of concepts are usually not exploited. How can semantic search be improved further? One idea is to expand the search to include partial terms that are contained in term definitions. Definitions of terms can be created according to the genus-differentia scheme (Gen-Dif) known from Aristotle [18]. In an innovative approach to semantic search [1], we use so-called Basic Linguistic Symbols (BLS) as linguistic identifiers, which can record the nouns of the terms in any language. For example, for >BLS-gauge, in German we have “Meßinstrument”, in English “gauge”, in French “jauge” and in Spanish “calibre” (see Fig. 1). Now we can compose terms using the Gen-Dif pattern as follows. In >BLS-barometer = (>BLS-gauge, >BLS-air_pressure) the left term gauge is the genus (\diamond Gen), and the right term air pressure is the differentia (\diamond Dif). According to the same scheme, we have >BLS-air_pressure = (>BLS-pressure, >BLS-air) with the triples (>BLS-air_pressure, \diamond Gen, >BLS-pressure) and (>BLS-air_pressure, \diamond Dif, >BLS-air). This definition can be inserted into the first: (>BLS-barometer = (>BLS-gauge, (>BLS-pressure, >BLS-air))). Since both the left \diamond Gen and the right \diamond Dif argument of a definition can be resolved by further compound definitions, a so-called *conceptual binary tree* (CBT) is created [1]. The leaves of the CBTs are always indecomposable, atomic concepts like >BLS-air, >BLS-pressure, >BLS-aid, >BLS-thing, and >BLS-measurement. The methodology presented here can also be used to more precisely define and disambiguate the terms used in ontologies and in WordNets. From (>BLS-Barometer, \diamond Gen, >BLS-gauge), (>BLS-Barometer, \diamond SymbolOf, \wedge Barometer) and (>BLS-Gauge, \diamond SymbolOf, \wedge Gauge) e.g. automatically (\wedge Barometer, \diamond subClassOf, \wedge Gauge) can be inferred, i.e., that \wedge Barometer is a subclass of \wedge Gauge.

This allows a *deep semantic search* (DSS) to be implemented. Query processing then, in addition to the usual semantic relations such as superclass, subclass, synonym, etc., also makes use of the terms that occur at any level of the conceptual binary tree. A search radius can be set to the depth to which the partial terms should be included in the search. To find out whether Google also takes compound terms into account when searching, we evaluated Google search against DSS. As an example, various musical instruments were modeled with so-called *Word sense definitions* (WSD).

A Google search for “wind keyboard instrument”¹⁸ returns 24 million search results, most of which refer to pages with *melodicas*. Searching for *pipe organ* only returns the

¹⁸ <https://www.google.com/search?q=wind+keyboard+instrument> (accessed 10/06/2022).

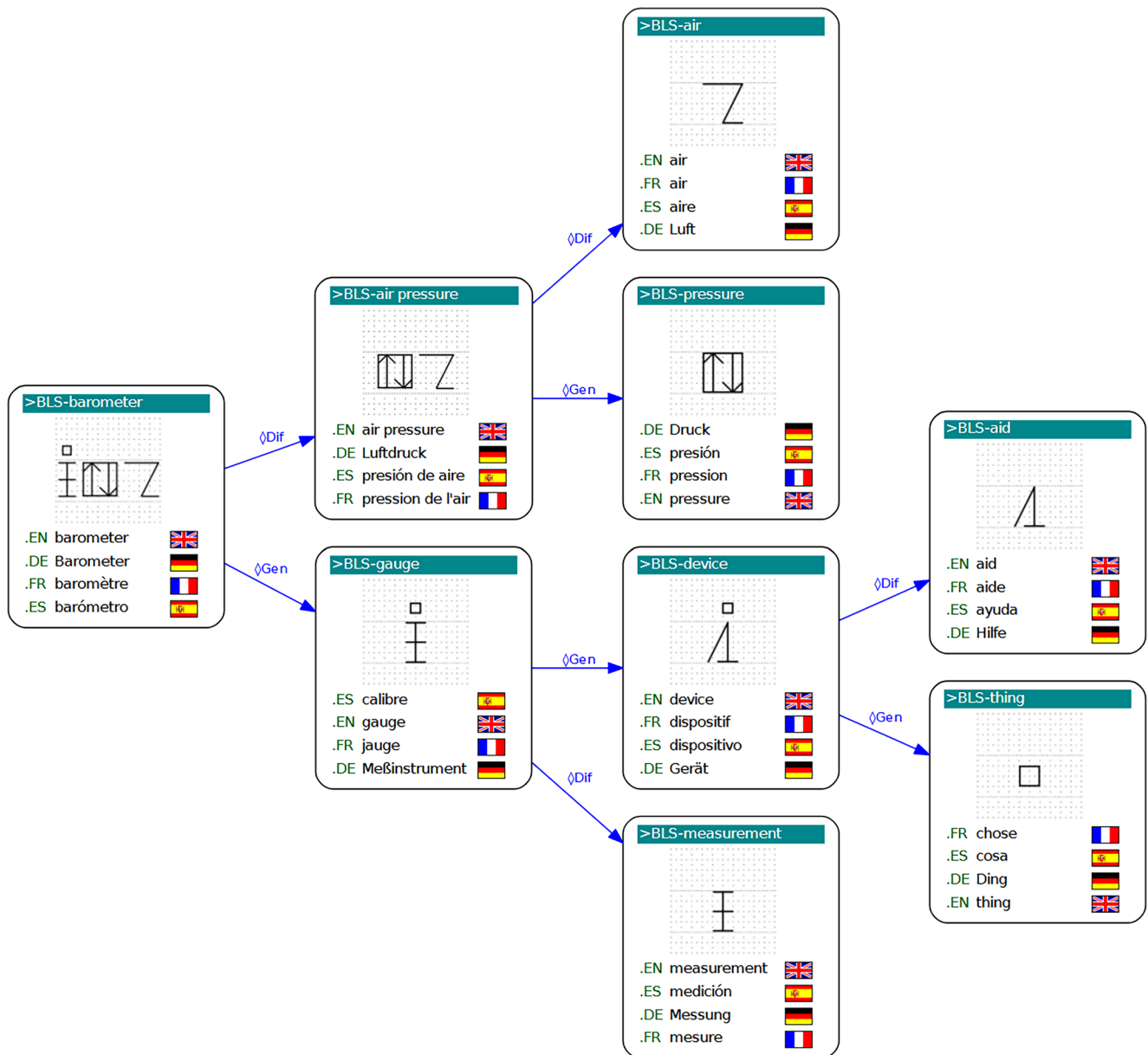


Fig. 1 Composition of the Concept Barometer

quiz question “What is the largest wind keyboard instrument called?” Our deep semantic search finds 67 instruments, 8 concepts for keyboard and 30 concepts for wind. In addition to *melodica* and *pipe organ*, the word *accordion* is also found. The term *accordion* does not even appear on the first Google results page. Thus, DSS can be used to determine additional suitable search terms, which can then be entered in the respective search engine. Also, our sample query “wind keyboard instrument” was evaluated against WordNet Search¹⁹. WordNet finds no result at all,

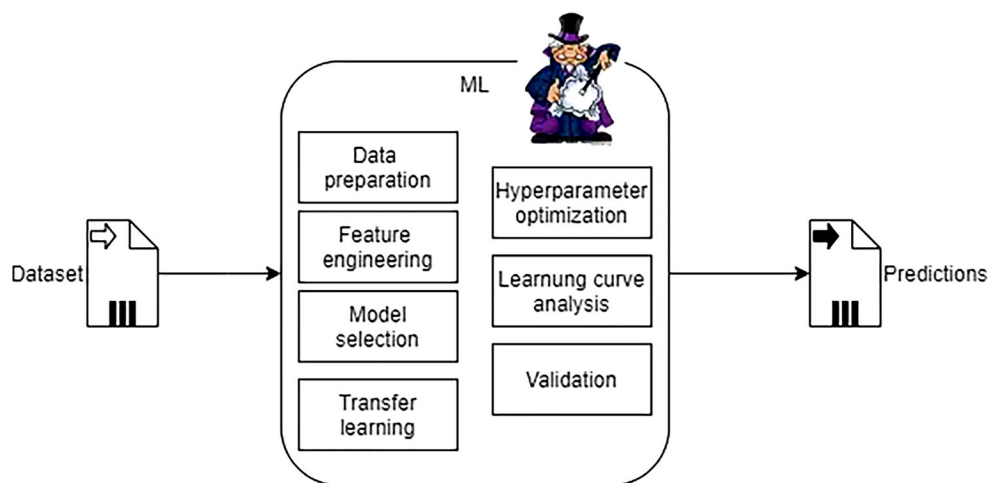
only a search for “wind instrument” finds an entry that describes the term itself, but not the concepts that fall under it.

On the basis of the Longman Defining Vocabulary²⁰, we developed a *Foundation Core Glossary* (FCG) with 450 atomic concepts and approximately 2500 compound terms in eight languages. All new terms can be defined with the concepts of the FCG. This means that queries with any combination of terms from these languages deliver the same results as if they had only been made in one of the languages.

¹⁹ <http://wordnetweb.princeton.edu/perl/webwn?s=wind+instrument> (accessed 10/06/2022).

²⁰ https://www.pu-kumamoto.ac.jp/users_site/rlavin/resources/wordlists/LDV.html.

Fig. 2 Steps for creating ML pipelines



A first version of the deep semantic search has been used in the EnArgus research project for search optimization [2]. The multilingual news app rob.by²¹ also uses this technology. Furthermore, applications are being implemented where the technology is to be used in portal search engines to search for scientific publications. These applications also benefit from the fact that the number of occurrences in documents is noted for each term in the glossary. As a further parameter in the deep semantic search, it can thus be taken into account whether the term is very general or very specific.

Machine learning

Machine learning (ML) is currently the most prominent AI approach, enabling models to make predictions based on previous observations [27]. ML is used to power AI applications, but creating effective ML-based AI applications is complex. It requires high expertise, which experts like computer scientists or data scientists gather over the course of their careers. A great deal of gut feeling is involved. To less experienced scientists, this may look like a “secret art”, making experts operate like magicians (we use the magician icon for illustration in Fig. 2). They apply their expertise and knowledge to analyze a dataset and create an efficient ML pipeline, which can be trained with a dataset to allow predictions for new data. Steps to be performed include data preparation, feature engineering, model selection, hyperparameter optimization, validation, and more. See Fig. 2.

Automated machine learning (AutoML) [37–40] emerged as a research field aiming at generating parts or entire ML pipelines automatically. Presently, a wide variety of different AutoML solutions are available, ranging from open-source to commercial solutions. There is a great variety of

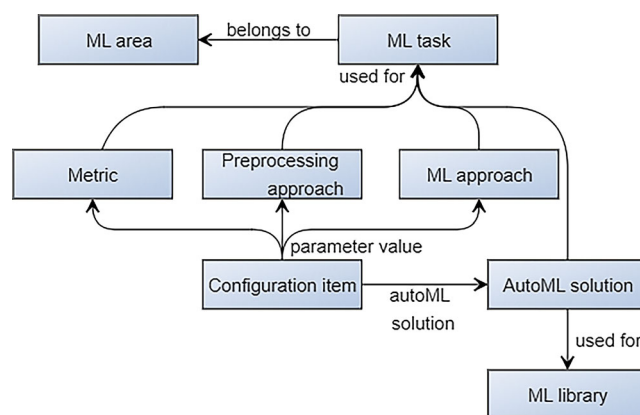


Fig. 3 ML ontology schema [29]

AutoML solutions, differing in functionality, maturity and usability.

However, there is one restriction in almost all AutoML solutions: they focus almost exclusively on one major ML library or ML ecosystem. This leads to some form of vendor lock-in when choosing an AutoML solution.

Meta AutoML [28, 29] is a novel concept aiming at avoiding said vendor lock-in. The concept is to integrate different AutoML solutions in a cloud service. Managed through a metalayer, various AutoML solutions can compete against each other, allowing the user to select the best solution for his use case.

The concept of Meta AutoML led to the development of the *OMA-ML system (Ontology-based Meta AutoML)* [28, 29], combining the idea of Meta AutoML with an ontology for ML. The ML ontology aggregates knowledge about the field of ML and different AutoML solutions to optimize the Meta AutoML process. See Fig. 3 for an overview of the ML ontology schema.

²² <https://www.w3.org/RDF/>.

²³ <https://www.w3.org/2004/02/skos>.

²¹ <https://rob.by/en/App/>.

Fig. 4 OMA-ML leaderboard [29]

LEADERBOARD			
AutoML	ML library	Model	Score
FLAML	sklearn	extra_tree	0.949
Autosklearn	sklearn	ensemble selection	0.946
AutoGluon	pytorch + lightgbm + catboost + sklearn	WeightedEnsemble_L2	0.944
Auto_PyTorch	lightgbm + pytorch	ensemble selection	0.939
MLJAR	LightGBM	Greedy Ensemble	0.938
Auto_Keras	keras	neural network	0.913

The ML ontology is modeled in RDF²² using SKOS²³. It currently consists of over 1800 RDF triples, specifying 3 ML areas (e.g., supervised learning), 42 ML tasks (e.g., classification), 104 ML approaches (e.g., artificial neural networks), 5 pre-processing approaches (e.g., principal component analysis), 55 metrics (e.g., accuracy), 16 ML libraries (e.g., Scikit-learn), 21 AutoML solutions (e.g., Autosklearn), their configuration items, interrelationships and more. The ML ontology is open source under GitHub²⁴.

The ML ontology is the information backbone of OMA-ML and is used in various parts of the application: Firstly, in a configuration wizard, only plausible configuration options are shown to users. Secondly, the ontology guides the automated pre-processing of the dataset by determining if an automated pre-processing is required and which steps. Thirdly, the ontology determines the ML training strategy, guiding the metalayer with instructions on how to execute the training within the AutoML solutions.

OMA-ML is actively being developed as an open-source project and can be accessed as a GitHub repository²⁵. At the time of writing, a minimum viable product is available with an initial set of AutoML solutions integrated, providing classification and regression tasks for tabular datasets. See Fig. 4 for a screenshot of the leaderboard displaying various AutoML results for a dataset training.

²⁴ <https://github.com/hochschule-darmstadt/MetaAutoML/tree/main/controller/managers/ontology>.

²⁵ <https://github.com/hochschule-darmstadt/MetaAutoML>.

Corporate use cases

This section shows four partially new areas of applied machine intelligence: (a) a core problem of automated production processes: the detection of anomalies in time series of sensor data helping to avoid disruptions of industrial processes; (b) the determination of cross-organisational process disruptions and their consequences based on efficient semantic modeling of entire supply networks; (c) semantic product passports describing materials and components attached to goods in order to support repair and recycling and (d) semantic enrichment of technical information.

Anomaly detection

With the rising complexity of modern automated manufacturing processes like production, packaging and quality insurance, the importance of each process step realizing the desired outcome is increasing. Even a single fault in one step could influence the complete production line, resulting in a faulty product, a breakdown of the process or a carryover of the failure through the entire production line. Therefore, it is crucial for modern cyber-physical systems to recognize, react early and prevent such failures. *Anomalies* can be taken as essential failure indicators. With the help of a timely detected anomaly, a fast failure recognition and reaction is possible. Anomaly detection refers to the identification of abnormal system behavior, e.g., behavior that deviates significantly from the regular operation of the system.

Problems with the implementation of anomaly detection in an industrial setup result from the different industrial

Fig. 5 Rotary table dispenser system (Yaskawa Europe GmbH, Eschborn, Germany)

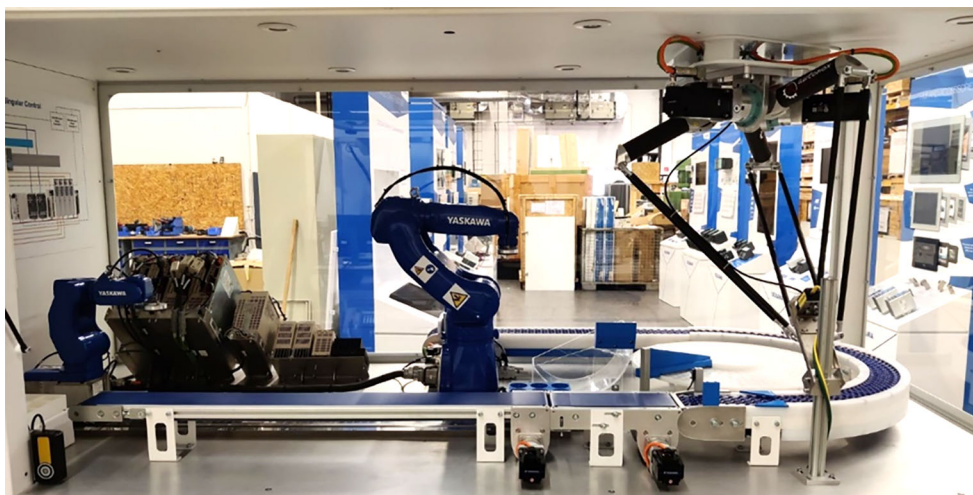
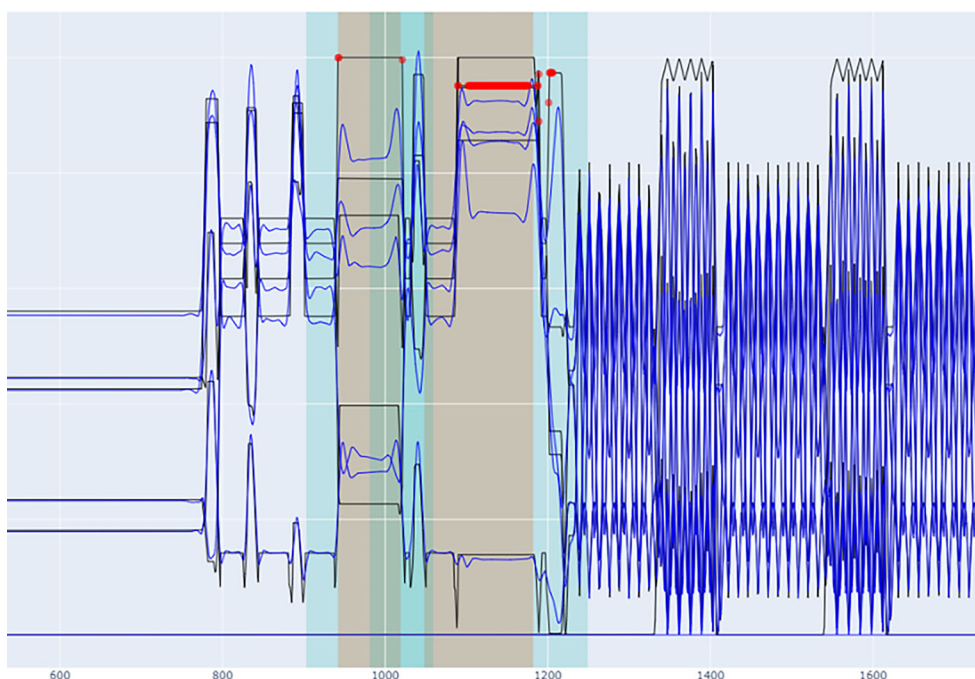


Fig. 6 Anomalous data stream



requirements that must be considered. Among time prerequisites to establish a fast reaction to an anomaly, a reasonable prediction quality and a configurable design to adapt to the dynamics of such a system are essential. Moreover, the communication and processing limitations of those systems must be considered.

In cooperation with Yaskawa²⁶, a leading mechatronics and robotics company, a fully automated, data-driven, and feasible anomaly detection for cyber-physical systems in manufacturing has been developed [30]. The cyber-physical system, shown in Fig. 5, demonstrates different drive units and their collaboration to realize complex processes. The

²⁶ <https://www.yaskawa.de/>.

overall process involves picking small items from a round table in the middle of the unit and placing them into several cups while they move on conveyor belts around the machine.

One example of an anomaly that can appear in the described system is an incorrect establishment of the vacuum generation for picking the cups from the conveyor belt to place them in a delivery position. This can be caused by broken cups, an incorrect picking position, or a worn gasket at the intake. As a result, the system can run into an undefined state, and the normal process can no longer be executed.

To realize anomaly detection while fulfilling industrial requirements, several reconstruction-based, one-dimen-

sional convolutional autoencoders were applied to the cyber-physical systems. The models are trained without the need for domain knowledge. No anomalous data is provided in the training step to meet the data-driven requirements. A sliding window approach achieves the timing requirements and the prediction quality. A shallow structure and the decentralized integration of the separate models address the inherent processing limitations.

An example of a reconstructed data stream and the detection of the abnormal behavior can be seen in Fig. 6; red dots indicate the detected anomalies, and the marked areas display the abnormal sections.

Effective and feasible anomaly detection, considering industrial requirements, can significantly impact automated processes to realize modern and safe cyber-physical systems in production.

Efficient modeling of supply networks

The recent Corona pandemic, the flooding of the German Ahr valley, the accident of the container vessel Evergiven stuck in the Suez Channel, the traffic jams caused by it at container terminals in China and Europe and the current Russian war against Ukraine have shown the vulnerability of our society and economy in the case of unforeseen disruptions. In 2020, the German Federal Ministry for Economics and Climate Protection (Bundesministerium für Wirtschaft und Klimaschutz, BMWK) started the funding of five projects addressing questions of economic resilience in pandemic and other economic crises. Within these projects, the ResKriVer-Project²⁷ addresses—besides other problems—some issues of modeling, documenting, analyzing and simulating supply networks of crises-relevant and potentially substitutable goods and resources, in order to strengthen the resilience of critical supply paths.

Fraunhofer FOKUS²⁸ role within this project—besides its role as consortium leader—is the semantic modeling of supply networks as a knowledge graph, which, in contrast to simple supply chains, considers entire cross-organizational networks in order to perform preventive as well as reactive analysis. The approach chosen by FOKUS is the adoption of W3C's provenance standard PROV-O²⁹ as one of the top-level ontologies of a "ResKriVer common core ontology". This ontology supports the modeling of supply networks on a fine level of granularity and at the same time should allow for (a) the extraction of simulation models for Fraunhofer IML's ODE-Net system for discrete event

simulation and (b) approaches for dependency analysis in order to determine critical nodes within a supply network and their associated probabilities.

During the initial modeling of some examples of supply networks, it soon became clear that this approach will face three major issues. First, the integration of probability information can be addressed by partial adoption of PR-OWL, an ontology for modeling probability information. Second, structural equivalent RDF subgraphs will frequently occur during the modeling of supply networks in order to capture alternative supply chains. Thus, some kind of efficient RDF templating mechanism would be helpful for efficient modeling. Consequently, mechanisms to partially instantiate these subgraphs with available information need to be developed. A basis for such templates could be SHACL³⁰ or—probably better suited—ShEx³¹, augmented by some mechanism to generate new unique anonymous resources and instances during template instantiation. We consider this to be a new research topic. As a third issue, the non-availability of supply chain information, which is usually difficult to obtain for second and higher-order tier suppliers, needs to be addressed. Hence, coping with highly incomplete information about supply networks and acquiring them during the pre-crisis documentation phase and within some current crisis will be a major issue in this applied research project.

Circular economy

On 11 December 2019, the European Commission presented the Green Deal³². The Communication from the Commission also includes a detailed action plan, including an annex with a table of measures within a timeline. In March 2020, the Communication on "A new Circular Economy Action Plan" followed³³. The action plan contains an industrial strategy for a clean and circular economy, aiming at sharing, reusing, repairing and recycling existing products as long as possible. Within the framework of the Multi-Stakeholder Platform for Standardization (MSP)³⁴ W3C and GS1 gave birth to the idea of combining the Linked Data platform of W3C with the work on GS1 Digital Link. A report was handed to the Commission.

The idea contributed to the considerations around the works on the Digital Product Passport (DPP). Good ideas

²⁷ <http://www.reskriver.de>, funded by the German federal ministry for economics and climate protection (Bundesministerium für Wirtschaft und Klimaschutz, BMWK) under research contract 01MK21006A.

²⁸ <http://fokus.fraunhofer.de>.

²⁹ <https://www.w3.org/TR/prov-o/>.

³⁰ <https://www.w3.org/TR/shacl/>.

³¹ <https://shex.io/>.

³² <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52019DC0640> Accessed 2022-06-01.

³³ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0098&qid=1654441498742> accessed 2022-06-01.

³⁴ <https://digital-strategy.ec.europa.eu/en/policies/multi-stakeholder-platform-ict-standardisation> accessed 2022-06-01.

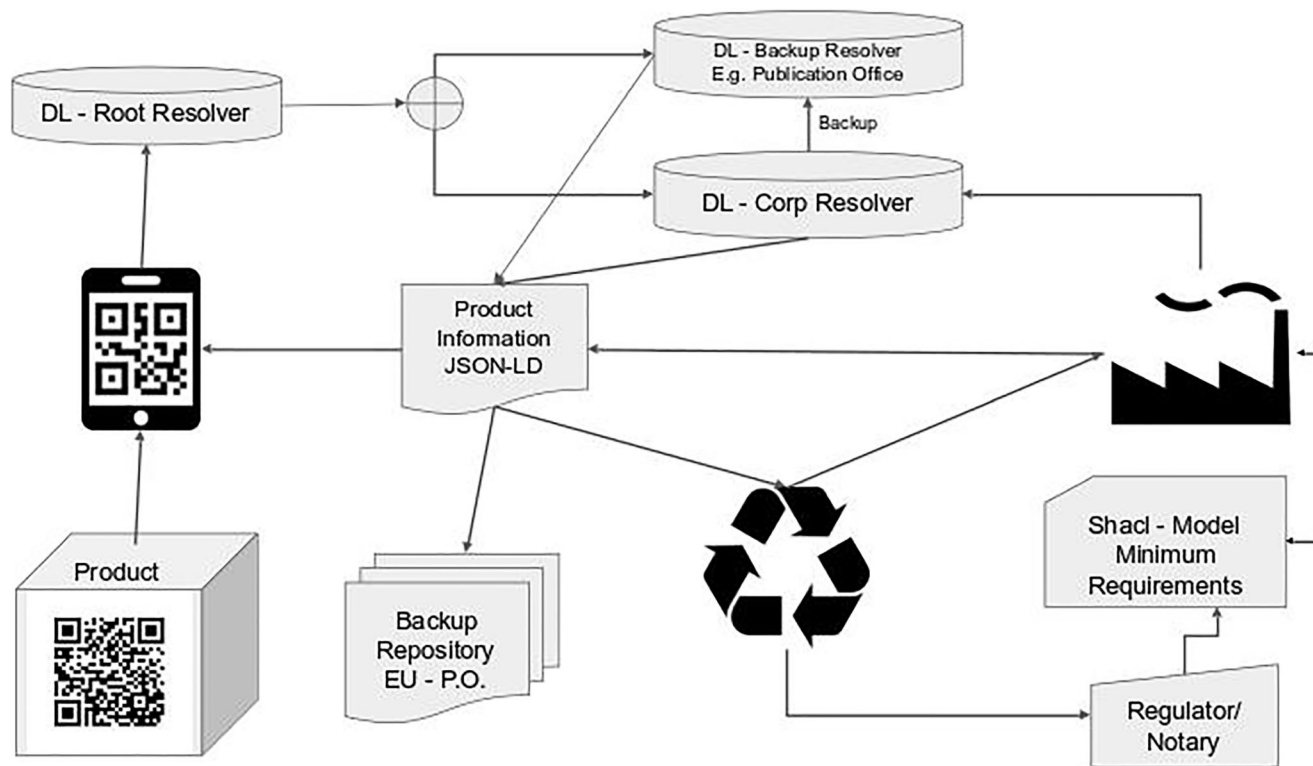


Fig. 7 An architecture for the Digital Product Passport

must be simple, and this is the case here. On a high level, the problem of circularity is seen from an information system perspective. Today, if a product hits the recycling plant, we do not know much about it, and the recycling plant has to do a lot of guesswork and research to find out about the physical object in their hands. This does not really work well and, consequently, recycling can still be improved. But how can we get information to the recycling plant? This question looks simple but has a variety of complex issues of system engineering.

Who can provide useful information? The producer of the tangible goods has lots of information, but not necessarily everything. First, this producer may only assemble components. So the composition of the components is not always known to them. Secondly, the producer has commercial secrets around producing certain goods. So it is not simple and easy to just make information public. The same goes for the producer of the components assembled into a consumer product. It may be produced with secret production techniques not known to the entity assembling the various components. And the producer of the component cannot know about all the products that their component is used for.

Finally, there is the user of the product. Users may be careful with a product or treat it badly. There may be accidents, lightning, repair, etc. in the lifetime of a product that may influence the way the product or its components can

be recycled. This information is with the consumer or end user. But not only with end users, also with their agents and repair services. This is important information that needs to be collected.

While a central database is easier to establish, the large variety of actors or sources of information and the heterogeneity of the information collected makes monolithic systems see their limits quickly.

But what would such a distributed, product-centric system look like? The basic idea is to use GS1's Digital Link Specification [20] together with the Linked Data platform [21] and create a system of annotations to the physical product identified by some IRI. Linked Data annotations [22] have the power to create a knowledge graph around a type of product or the instance of a product. Because the system uses the Linked Data paradigm, data from a variety of sources can be retrieved and easily merged into a full picture. Meanwhile, DL 1.1 makes sure that the annotations are reliably linked to the product instance at hand. A persistent link between the physical and the virtual world is established.

The overall architecture of the system is shown in Fig. 7. We are only at the beginning of a development.

GS1 Digital Link uses the Global Trade Identification Number (GTIN) that is often represented as a barcode, a QR code or an RFID tag [23]. Batch numbers, serial numbers and more can also be included. From the code, a link is

created that is an entry point into the web. The simplest case is that a barcode on a product just points to a web page. The GS1 Digital Link Specification not only uses the simple web but allows for sophisticated operations using so-called “resolvers”, applications identifying that the URI at hand is a GS1 Digital Link URI. These resolvers have a set of typed URIs corresponding to the GS1 Digital Link requested, as indicated in [24]. Although designed for GTINs, the system would work with any other identity scheme (e.g., DNS plus URI) with its own resolver. Standardization work is very likely to begin at ISO shortly to expand the GS1 ideas to encompass any recognized identifier system.

Using the GS1 Digital Link standard, one type of link can now point from the product into an RDF knowledge graph. This graph may receive information from a large variety of sources. Having the product instance at hand, all this information is instantly accessible. If transformed back into HTML, this can be used for consumer information. If given in machine-readable form to a gateway in the recycling plant, it can help sort incoming goods. But the knowledge graph can also help with compliance and regulatory constraints. The minimum requirements for the DPP could be expressed in SHACL [25]. It is then easy to match the regulatory requirements with the existing graph. By including information about access, usage limitations and other constraints, the knowledge graph itself already contains the necessary administrative information. Both can also serve as a proof of compliance [26].

Much remains to be done. Vocabularies and ontologies for the industry will have to be created. The first milestone will be to address the new battery regulation³⁵.

Semantic enrichment of technical information—creating smart technical information

This section deals with the semantic enrichment of technical information that shall ensure the safe operation of machinery put into service in the European Union. The scope of this technical information is defined by the European Machinery Directive 2006/42/EC [33]. Technical information becomes *smart* by being labeled with metadata addressing its issues and concepts as well as the different entities it relates to, for example the machinery and its components (e.g., battery), the actors and their roles (e.g., service technician), and the actions it describes regarding its use and maintenance (e.g., replacing a component). We call these data *context-related metadata*. Thus, smart technical information has been semantically enriched and put into a comprehensive context.

³⁵ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0798> accessed 2022-06-01.

Making legally compliant technical information smart enables it to be retrieved within a semantics-based information system over a faceted or semantic search and to be semantically linked to other pieces of information. Making legally compliant technical information smart is a development that has increasingly taken place over the last few years, due to the change in information consumption, the requirements for context-aware information and the digitalization of information delivery in the context of industry 4.0 [34, 35].

The semantics-based hybrid information system SOLIS-Doc³⁶ intends to include a technological as well as methodologically sound framework for authoring and delivering such smart technical information.

Authoring technical information often occurs within a component content management system (CCMS), enabling technical writers to author and manage technical information in a modular way, meaning that each single issue of the technical information becomes a small information unit, a topic, or a fragment. Putting all the needed topics together, technical writers can, over the CCMS, automatically publish an operating manual, either as a PDF document or as an HTML snippet.

To facilitate the modular authoring of legally compliant technical information within a CCMS, SOLIS-Doc provides the semantic model of an operating manual according to the European Directive as an RDF/RDFS ontology. The semantic model builds a minimally valid, and thus transferable, reference template of the entire mandatory content of such a legally compliant operating manual. The *SOLIS-Doc Reference Template* (see Fig. 8) can be configured and imported into a CCMS and provides technical writers with a reference structure including all the topics and fragments that they can use as a basis within their projects.

Over the SOLIS-Doc reference template, all reference topics have already been pre-labeled with standardized context-related metadata of the new iiRDS³⁷ RDFS-based metadata model [36]. Consequently, technical writers only need to add those project-specific context-related metadata such as product metadata. These proprietary context-related metadata have been modeled within the *SOLIS-Doc Twin of Customer (Product) World*. Thanks to this pre-labeling, the amount of context-related metadata that need to be

³⁶ Developed within the SOLIS-Doc project no. 20_0162_2A, funded by the state of Hesse, Germany, in the funding line no. 2 *Digital Innovation Projects* of the Distr@l program *Strengthening Digitalization—Living Transfer*. <https://www.lidia-hessen.de/projekte/solis-doc-smart-information-for-aftersales-services/> accessed June 21st 2022.

³⁷ iiRDS: Intelligent Information Retrieval and Delivery Standard developed by tekem e.V. Release 1—Nov. 2020 https://iirds.org/fileadmin/iirds_specification/20201103-1.1-release/index.html accessed June 21st 2022.

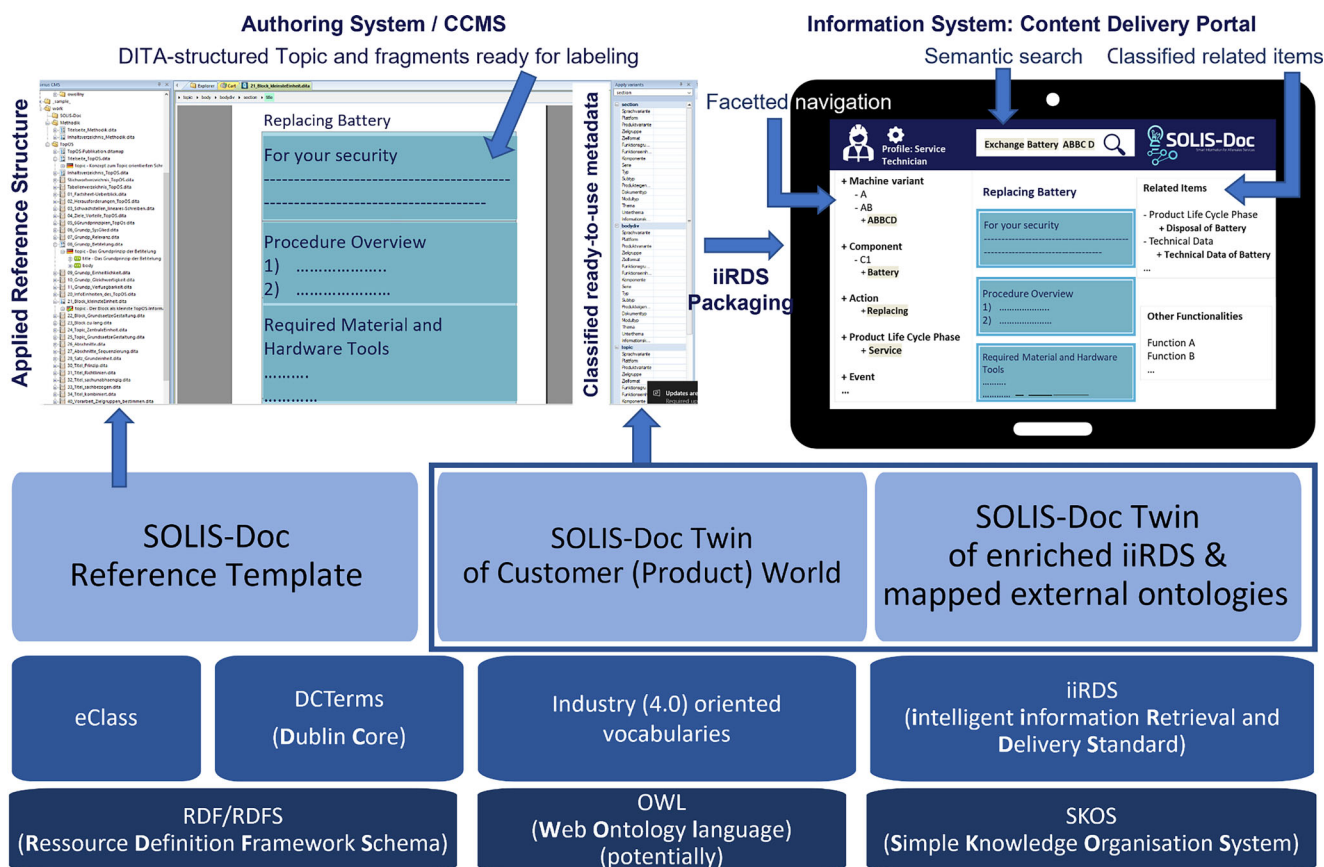


Fig. 8 Technological and methodological framework of SOLIS-Doc

assigned to the information units for creating the intended smart technical information can be reduced.

The SOLIS-Doc metadata model not only uses iiRDS standardized vocabularies for labeling information units but also adds semantics to these vocabularies by modeling relations between them. Thus, SOLIS-Doc becomes an RDF knowledge graph that is a semantically *enriched iiRDS twin*. This enriched iiRDS twin enables an additional reduction of the number of context-related metadata technical writers must assign to information units. For example, being modeled over the iiRDS twin that the action *replacing* (e.g., replacing a battery) is related to the product life cycle phase (PLCP) *service* (*customer:replacing solisdoc:related-PLCP solisdoc:service*), technical writers only need to assign the *action-related* metadata to the topic, while the *PLCP-related* metadata is automatically assigned to it by the CCMS.

As shown in Fig. 8, SOLIS-Doc consists of three semantic models, each of them based on standard frameworks for the development of semantic web applications. It is the aim of SOLIS-Doc to be interoperable with different CCMS as well as different content delivery portals (CDP). Of course, it is a prerequisite that these CCMS and CDP have been (technologically speaking) prepared to integrate semantic

web-based models. That is a development also taking place within the SOLIS-Doc project.

Conclusions

In this article, we presented selected new directions in knowledge-based AI and ML, and their corporate use cases. AI is much more than ML; knowledge-based AI still plays an important role in everyday applications. Knowledge-based AI and machine learning complement each other ideally, as their strengths compensate for the weaknesses of the other discipline. Machine learning approaches deliver good results in areas with little prior knowledge. They usually require a large amount of training data, but at the same time can be scaled to big data and handle noisy data well. However, they are also error-prone and lack the means to explain chosen decisions. This is where the strength of knowledge-based AI methods lies. Since they are based on explicit representations of human expert knowledge, they are applicable in areas where only limited data are available. They can be used for complex logical reasoning, their decisions can be explained to humans, and errors in their knowledge bases can be corrected more easily and quickly. On the

other hand, the development of their knowledge bases is costly.

Hybrid AI approaches combine ML and knowledge-based AI. They have been researched for years and are successfully applied in practice today. The OMA-ML system presented above is a good example. To share experience in developing hybrid AI applications in the community, we are planning a new book about hybrid AI with ML and knowledge graphs to be published by Springer.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bense H (2021) A formal ontology of concepts for deep semantic searching. In: KONT-2021 Knowledge-Ontology-Theory, 8th International Conference
- Bense H, Schade U, Ohrem F, Sikorski L (2014) Recherche-Unterstützung durch Ontologie Visualisierung im EnArgus2-Projekt. In: Proceedings of 3. DGI-Konferenz "Informationsqualität und Wissensgenerierung" Frankfurt a.Main
- Ege B, Humm B, Reibold A (eds) (2015) Corporate Semantic Web – Wie Anwendungen in Unternehmen Nutzen stiften. Springer, Heidelberg. ISBN 978-3-642-54885-7.
- Hoppe T, Humm B, Reibold A (eds) (2018) Semantic applications—methodology, technology, corporate use. Springer, Berlin. ISBN 978-3-662-55432-6.
- Hoppe T (2020) Semantische Suche – Grundlagen und Methoden semantischer Suche von Textdokumenten. Springer Vieweg, Wiesbaden. ISBN 978-3-658-30426-3.
- Humm BG, Bense H, Fuchs M, Gernhardt B, Hemmje M, Hoppe T, Kaupp L, Lothary S, Schäfer K-U, Thull B, Vogel T, Wenning R (2021) Machine intelligence today: applications, methodology, and technology. *Informatik Spektrum* 44:104–114. <https://doi.org/10.1007/s00287-021-01343-1>
- Bense H, Gernhardt B, Haase P, Hoppe T, Hemmje M, Humm B, Paschke A, Schade U, Schäfermeier R, Schmidt M, Siegel M, Vogel T, Wenning R (2016) Emerging trends in corporate semantic web—selected results of the 2016 Dagstuhl workshop on corporate semantic web. *Informatik Spektrum* 39(6):474–480
- Busse J, Humm B, Lübbert C, Moelter F, Reibold A, Rewald M, Schlüter V, Seiler B, Tegtmeier E, Zeh T (2015) Actually, what does "Ontology" mean? A term coined by philosophy in the light of different scientific disciplines. *J Comput Inform Technol* 23(1):29–41. <https://doi.org/10.2498/cit.1002508>
- Hoppe T, Humm B, Schade U, Heuss T, Hemmje M, Vogel T, Gernhardt B (2015) Corporate semantic web—applications, technology, methodology. *Informatik Spektrum* 39(1):57–63. <https://doi.org/10.1007/s00287-015-0939-0>
- Humm BG, Bense H, Geißler S, Hoppe T, Juwig O, Paschke A, Schäfermeier R, Siegel M, Weichhardt F, Wenning R (2019) Current trends in applied machine intelligence. *Informatik Spektrum* 42(1):28–37. <https://doi.org/10.1007/s00287-018-01127-0>
- Schade U, Fillies C, Humm B, Reibold A, Schumann F, Weichhardt F 5 years of semantics workshops in Schloss Dagstuhl: it connects! https://www.researchgate.net/publication/336798165_5_Years_of_Semantics_Work. Accessed 28 Oct 2020
- Gruber TR (1993) A translation approach to portable ontologies. *Knowl Acquis* 5(2):199–220
- Gangemi A, Guarino N, Masolo C, Oltramari A, Schneider L (2002) Sweetening ontologies with DOLCE. In: Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web Siguenza, Spain, 2002. https://doi.org/10.1007/3-540-45810-7_18
- Hogan A, Blomqvist E, Cochez M, d'Amato C, Melo GD, Gutierrez C et al (2021) Knowledge graphs. *ACM Comput Surv*. <https://doi.org/10.1145/3447772>
- Schulz S, Seddig-Raufie D, Grewe N, Röhl J, Schober D, Boeker M, Jansen L (2012) Guideline on developing good ontologies in the biomedical domain with description logics. https://www.iph.uni-rostock.de/storages/uni-rostock/Alle_PHF/IPH/media/GoodOD/GoodOD-Guideline_v1_2012.pdf. Accessed 2022-06-10
- Kifer M, Lausen G (1989pp) F-logic: a higher-order language for reasoning about objects, inheritance, and scheme. *Acm Sigmod Rec* 18(2):134–146. <https://doi.org/10.1145/66926.66939>
- Keet CM (2020) An introduction to ontology engineering. <https://open.umn.edu/opentextbooks/textbooks/590>. Accessed 2022-07-08
- Arp R, Smith B, Spear AD (2015) Building ontologies with basic formal ontology. MIT Press, London. ISBN 978-0-262-52781-1.
- Duque-Ramos A, Boeker M, Jansen L, Schulz S, Iniesta M, Fernandez-Breis JT (2014) Evaluating the good ontology design guideline (goodOD) with the ontology quality requirements and evaluation method and metrics (OQuaRE). https://www.researchgate.net/publication/264990788_Evaluating_the_Good_Ontology_Design_Guideline_GoodOD_with_the_Ontology_Quality_Requirements_and_Evaluation_Method_and_Metrics_OQuaRE. Accessed 2022-07-08
- GS1 (2022) Digital link standard, release 1.2.1. <https://www.gs1.org/standards/gsl-digital-link> (Created 06.2022). Accessed 2022-06-05
- Raimond Y, Schreiber G, Manola F, Miller E, McBride B (2014) RDFa 1.1 primer—third edition, W3C note 2014. <http://www.w3.org/TR/rdf11-primer/>. Accessed 2022-06-01
- Sanderson R, Ciccarese P, Young B (2017) Web annotation data model. W3C recommendation. <https://www.w3.org/TR/annotation-model/>. Accessed 2022-06-01
- GS1 (2022b) Global trade item number. <https://www.gs1.org/standards/id-keys/gtin>. Accessed 2022-06-01
- Nottingham M (2017) "Web linking", RFC 8288. <https://doi.org/10.17487/RFC8288> (Created 10.2017). Accessed 2022-06-01
- Knublauch H, Kontokostas D (2017) Shapes constraint language (SHACL). <https://www.w3.org/TR/shacl/>. Accessed 2022-06-01
- Kirrane S et al (2018) A scalable consent, transparency and compliance architecture. In: Gangemi A et al (ed) The semantic web: ESWC 2018 satellite events. Lecture notes in computer science. Springer, Berlin, pp 131–136
- Russell SJ, Norvig P (2016) Artificial intelligence: a modern approach, 3rd edn. Prentice hall series in artificial intelligence. Pearson, Upper Saddle River
- Humm B, Zender A (2021) An ontology-based concept for meta autoML. In: Maglogiannis I et al (ed) Artificial Intelligence Applications and Innovations (AIAI 2021), IFIP AICT 627. Springer Na-

- ture Switzerland, Cham, pp 117–128 https://doi.org/10.1007/978-3-030-79150-6_10
29. Zender A, Humm BG (2022) Ontology-based Meta AutoML. *Integrat Comp-Aided Engin* 29(4):351–366. <https://doi.org/10.3233/ICA-220684>
30. Goetz C, Humm B (2022) Unsupervised process anomaly detection under industry constraints in cyber-physical systems using convolutional autoencoder. In: *Computational Intelligence for Engineering and Management Applications, Select Proceedings of CIEMA 2022*. LNEE lecture notes in electrical engineering. Springer Nature, Singapore
31. Siegel M, Bond F (2021). Compiling a German Wordnet from other Resources. In: *Proceedings of the 11th International Global Wordnet Conference (GWC2021)*, pp 192–198
32. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: *Advances in neural information processing systems*, pp 3111–3119
33. European Union (2006on) DIRECTIVE 2006/42/EC of the European parliament and of the council of 17 May 2006 on machinery, and amending directive 95/16/EC (recast) in “official journal of the European Union, L 157, 09 June 2006”. <https://op.europa.eu/en/publication-detail/-/publication/9c3d6091-62af-49de-b77a-dceb2716b50/language-en/format-PDF/source-259945913>. Accessed 21 June 2022
34. Schaffner M (2019) Technische Kommunikation im Wandel von Industrie 4.0. In: Stumpf M (ed) *Digitalisierung und Kommunikation. Europäische Kulturen in der Wirtschaftskommunikation*, vol 31. Springer VS, Wiesbaden, pp 143–159 https://doi.org/10.1007/978-3-658-26113-9_9
35. Fritz M (2017) Was ist Intelligente Information? In: Hennig J, Tjarks-Sobhani M (eds) *Intelligente Information*. *tekom Schriften zur technischen Kommunikation*, vol 22. tcworld GmbH, Stuttgart, pp 11–25
36. Parson U (2017) Das Datenmodell der Technischen Dokumentation in iiRDS. In: Hennig J, Tjarks-Sobhani M (eds) *Intelligente Information*. *tekom Schriften zur technischen Kommunikation*, vol 22. tcworld GmbH, Stuttgart, pp 26–39
37. He X, Zhao K, Chu X (2021) AutoML: a survey of the state-of-the-art. *Knowl Based Syst* 212:106622. <https://doi.org/10.1016/j.knosys.2020.106622>
38. Doke A, Gaikwad M (2021) Survey on automated machine learning (autoML) and meta learning. In: *12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp 1–5 <https://doi.org/10.1109/ICCCNT51525.2021.9579526>
39. Ge P (2020) Analysis on approaches and structures of automated machine learning frameworks. In: *2020 International Conference on Communications, Information System and Computer Engineering*. IEEE, Piscataway, pp 474–477 <https://doi.org/10.1109/CISCE50729.2020.00106>
40. Chauhan K, Jani S, Thakkar D, Dave R, Bhatia J, Tanwar S, Obaidat MS (2020) Automated machine learning: the new wave of machine learning. In: *2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020)*. IEEE, Piscataway, pp 205–212 <https://doi.org/10.1109/ICIMIA48430.2020.9074859>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.