KoMMDia: Dialogue-driven assistance system for fault diagnosis and correction in cyber-physical production systems

Julian Rahm[1], Markus Graube[1], Romy Müller[2], Tilman Klaeger[3], Luise Schegner[1], Andre Schult[3], Rica Bönsel[2], Sebastian Carsch[3] , Lukas Oehm[3] and Leon Urbas[1]

[1]Chair of Process Control Systems, Technische Universität Dresden, Dresden, Germany name.surename@tu-dresden.de

[2]Chair of Engineering Psychology and Applied Cognitive Research, Technische Universität Dresden, Dresden, Germany name.surename@tu-dresden.de

[3]Fraunhofer Institute for Process Engineering and Packaging, Fraunhofer, Dresden, Germany name.surename@ivv-dresden.fraunhofer.de

Abstract—**In complex production systems, the diagnosis and correction of faults requires operators to possess a deep understanding of the specific processes and machines as well as general knowledge about the interactions between different system components. However, in actual work environments operator qualification and experience is quite diverse, which leads to an immense variability in the time required for fault diagnosis and in the quality of corrective actions. Both time and quality of fault diagnosis and correction are vital parameters in the functioning of a plant, because downtimes have severe economic consequences and thus should be kept to a minimum. With the introduction of highly complex and flexible cyber-physical production systems, these problems are aggravated as fault sources vary and diagnosis becomes even more challenging. The paper presents a concept for a self-learning assistance system that supports operators in finding and evaluating solution strategies for complex faults. This concept applies a question-answer approach which allows for an incremental, dialogue-based establishment of common ground between operators and the assistance system.**

I. Introduction

The design and realization of cyber-physical production systems (CPPS) is an important focus of ongoing research and engineering in manufacturing environments. Pursuing the goal of a fully self-organized manufacturing line, CPPS combine disciplined mechanical, electrical and automation engineering with advanced embedded computer science [1]. They are a key feature to achieving the goals of increased flexibility, lower time to market, higher yield, and enhanced product quality [2]. However, despite all automation efforts an aspect that so far has not been taken into account sufficiently is the human: In case of complex faults that cannot be handled by automation, operators need to fix the problem by engaging in appropriate interactions with the CPPS. Such interactions require operators to possess and apply a rich knowledge base about the machine with all its components and the product itself. For instance, in the food processing industry the location of the fault symptoms often does not correspond to the location of the underlying cause [3], and unknown interactions of product characteristics, environmental factors, and process parameters are among the most important contributors to machine operation [4]. On the other hand, the average time during which a machine is running smoothly between consecutive machine stoppages is shorter than four minutes, although 70% of machine stoppages only last two minutes or less [5]. These numbers reflect that in most cases operators only remove the symptoms of a fault as they are unable to determine its actual cause. The result is a cycle of recurring faults and quick fixes. When different products are produced or processed on the same machines as envisioned in CPPS concepts, the interactions between products and machines will get even more complex and more difficult to comprehend, which is expected to increase the already existing problems as operators cannot rely on well-established routines to deal with faults. Accordingly, system knowledge is essential to understanding the complex relations between causes and effects [6]. Therefore, assistance systems (AS) are needed that enhance fault diagnosis and correction by supporting operators, technicians, and maintenance teams in tackling faults by their underlying cause. The present paper proposes an approach to designing a dialogue-based AS for fault diagnosis and correction which elicits and accumulates knowledge via conversational case-based reasoning [7].

The resulting recommender system integrates (1) technical parameters from different information spaces (e.g., machine settings, sensor data), (2) current observations provided by the operator supervising and controlling the machine, and (3) stored knowledge and experience provided by other operators in the past. The result is a cross-domain AS for CPPS [8]. The development of this system is put forward in the interdisciplinary research project KoMMDia. In this project, we tackle the challenge of integrating the digital representation of machines (i.e., engineering documents, sensor data) with knowledge and experience of human agents (i.e., operators, technicians, machine developers). To achieve this integration, we use the benefits of a cyber-physical system incorporating

978-1-5386-7108-5/18/$31.00 ©2018 IEEE

sensor and information networks to attain and provide useful information for fault diagnosis and correction. To illustrate this joint endeavour, the present paper addresses the following research questions:

- How to infer possible reasons for faults based on sensor data and machine behaviour? The high frequency of events and first insights into the stochastic nature of the event-related data suggests to apply machine learning algorithms for detecting recurrent patterns and deducing their underlying causes.
- How to incorporate the knowledge and mental models of humans in the process of fault diagnosis? Humans are an important information source as they provide sensory observations of the current situation that are unavailable to machines, and experience gained from previous fault instances. Both kinds of information contribute to finding the actual cause of a problem, but a major challenge is to elicit, interpret, and use this information in a way that makes it accessible for machines.
- How to combine these two aspects and integrate the digital representation of the production line with information provided by human operators in a semantic representation? This integration can be achieved by using case-based reasoning (CBR) and natural language processing (NLP).

Thus, the present paper focuses on the use of machine data, human knowledge, and their integration in a CBR system. Approaches to combining CBR and NLP have already been proposed in other fields of research such as distributed software development [9]. However, most authors use simple attribute-value similarities and often do not consider changes in the case and knowledge base, but there are some notable exceptions. For instance, a cross-domain recommendation system for personal services has been developed by Moe [8]. However, this system is based on a minimal conception of dialogues and relies on predefined questions to calculate similarities between cases. Moreover, a promising general approach for combining NLP with Linked Data has been proposed by Lopez [10], and a similar concept has successfully been developed for a medical environment by Abacha [11]. Going beyond previous work, our approach is to combine different domains and data sources within our AS, making it possible to adapt and update cases when there are changes in one of these domains.

II. State of the Art

A. Distributed Information Models

To deal with different information sources, a flexible data back-end is needed. There are two candidates that can be used for static and dynamic data. The first one is Linked Data as a part of semantic web technologies [12], a well-known approach to sharing and linking static information in a distributed environment for industrial applications [13]. It denotes the vision of describing, structuring, and connecting knowledge on the World Wide Web in graph-based representations, and making it available to both humans and machines. It uses the Resource Description Framework (RDF) as a data model , and the SPARQL Protocol and RDF Query Language (SPARQL) [14] can be used as a way of querying a triple store.

The second one is OPC UA, a protocol for machine-to-machine communication standardized in IEC 62541, which is quite suitable for handling dynamic machine data and parameters and which can be used to share information between systems in a semantic manner. For the purpose of interoperability between different servers, a uniform data representation is used. Thus, there are specific information models in OPC UA for different application domains, and it is possible to use these models or extend them with domain-specific knowledge. Clients can rely on the fact that all OPC UA servers have the same base model exposing their data [15]. A main benefit of OPC UA is its availability on all layers of shop floor automation, that is from sensors and actuators over cell control to manufacturing execution (MES) and planning systems. This makes it possible to automatically integrate all the distributed but relevant background information into the case base and dialogue components of an AS.

B. Case-based Reasoning

Case-based Reasoning (CBR) is an approach to solving problems based on previous situations with similar configurations of information. The CBR concept originated from cognitive science research on human memory [16]: It operates like a human who selects and adapts his problem solving strategy, based on experience and knowledge.

In CBR, a previous situation is linked with a proper solution so that it is possible to apply and adapt this solution to new problems. A case consists of a given problem and its solution, and all cases are stored in a case base from which they can be retrieved. While alternative approaches like neural networks have received lots of attention due to their high predictability, they lack the capability of explaining their results. In contrast, an advantage of using CBR is that it is comprehensible in terms of how its results are generated and which cases are used for that. Another advantage of CBR is that it can be effective even when the knowledge base or domain theory is incomplete. Certain techniques of automated learning, such as explanation-based learning, work well only if a strong domain theory exists, whereas CBR uses a great variety of examples to overcome knowledge gaps. Furthermore, it can be combined with other methods like rule-based systems [17].

A CBR system performs inference in four phases: *retrieve*, *reuse*, *revise*, and *retain*, which is also called the CBR cycle and was invented by Aamodt and Plaza in 1994 [18], shown in Fig. 1. The first phase is the *retrieve* phase, in which all former relevant cases are retrieved from the case base depending on the current problem description. This can be achieved by different techniques. One approach is to create a complete description, which in most implementations is a simple feature vector or a set of attribute-value pairs [19]. Alternatively, an incremental approach can be taken in which additional discriminating features are obtained by questioning [20]. In the retrieval phase, techniques such as nearest neighbour, decision

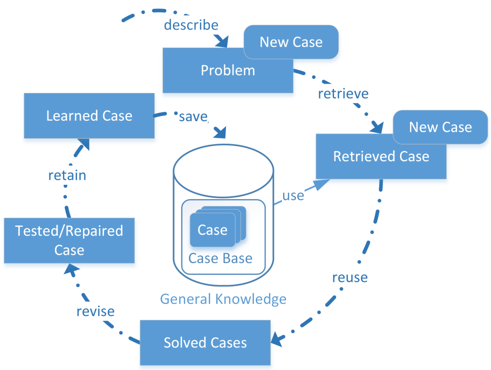

Fig. 1. Standard CBR cycle by by Aamodt and Plaza [18]

trees, Bayesian networks, or genetic algorithms are used to determine the most fitting cases [21]. In the subsequent *reuse* phase, a new solution for the current case is created based on the information provided by the old ones. In the *revise* phase, the given solutions are transformed into a correct solution for the new case. Finally, in the *retain* phase the case and its information is stored in the case base, so it can be used for new problems.

C. Machine Learning

Machine learning (ML) is becoming a standard when working with large amounts of data and high-dimensional datasets. By using supervised learning, the models generated by ML can classify new datasets when trained with known datasets that have already been classified. Numerous algorithms are available for classification tasks, ranging from simple decision trees to deep artificial neural networks [22]. In recent research, it has become quite common to apply ML to sensor data in rotating machines, mostly for condition monitoring and predictive maintenance [23], [24]. However, this approach is not easily adaptable to packaging machines as these machines work intermittently and therefore data analysis is not possible in the frequency domain but only in the time domain [25]. Accordingly, special emphasis needs to be put on adequate feature engineering when the raw sensor data is to be used for detecting anomalies and faults.

D. Human in the Loop

A future trend in CPPS is to create fully self-organized manufacturing lines. Nevertheless, humans remain a vital part of CPPS [1]. As described above, humans are needed to operate the machines, supervise product quality, or perform maintenance tasks. These activities rely on human competence, experience, and a thorough understanding of the current situation. However, a risk associated with the increase in automation is that human skills can decay when they are not being used [26]. When the automation is opaque with regard to its functioning, humans are deprived of the information on which they can base their assessment of the current situation and thus situation awareness is at stake [27]. This can lead to phenomena such as complacency and automation bias, which refer to a lack of vigilant monitoring of the automated system's actions and an uncritical acceptance of the suggestions provided by automated decision aids [28]. Accordingly, a goal is to design automation solutions and AS that keep humans in the loop by providing an interface between CPPS and human decision makers [29]. To this end, much of the research in the Human Factors Community has focused on the design of graphical interfaces that enable human operators to understand the system's functioning [30], [31]. With regard to decision aid concepts, CBR is a suitable candidate as it is based on the human's thorough description of situations [32], can provide different solution alternatives [33], explain its recommendations [34], and ultimately foster the human's conceptualization of the situation in interactive dialogues [7]. This enables a better integration of operators' mental models with the machine's problem representation to establish common ground. Therefore, the system can also be used to teach the operator how the machine works.

E. Assistance Systems in Industrial Environments

In recent years, the terms assistance system, assistive technology, support system, user support, help system, and many other related terms have been gaining importance in the development of human-machine systems. Although the meanings of these terms are quite similar, there is no unified definition that accounts for the similarities and differences between them [35]. We use the term assistance system (AS) to refer to a technical system that supports the human operator in performing his work-related tasks. Such support can take different forms. For instance, the system can collect information, aid in the analysis and representation of information, provide suggestions for decisions, or take over complete action sequences [28]. AS for the process industries have been suggested to encompass the following four stages or combinations thereof: *informing*, *analyzing*, *optimizing*, and *validating* [36]. When determining what stages of human information processing are to be supported by an AS, a note of caution is due: Especially when later stages such as decision making and action implementation are replaced or supported by technology, this can lead to operators overly relying on the AS and refraining from thorough monitoring and cross-checking [37], a problem that has also been reported in the context of fault diagnosis [38]. As thorough information processing is essential in the context for which our AS is developed, our approach focuses on the cooperative retrieval, acquisition, and analysis of information while leaving decisions to the operator. Similar recommendations for CBR systems have been put forward by [32], implying that CBR systems should serve as an extended memory while preserving the human's responsibility for evaluating and integrating the cases.

III. Use Case: Fault Diagnosis in Packaging Machines

To specify and evaluate our approach, the concepts are developed and implemented in cooperation with a company that processes and packages chocolate. We focus on fault diagnosis during the operation of packaging machines that wrap chocolate bars with foil, which currently is performed in the following way: During routine production, one operator is responsible for each machine. In case of a fault, the operator tries to fix it and subsequently restarts the machine. If the same fault reoccurs several times, the operator calls a technician who takes care of the problem based on his thorough understanding of the machine. For the company, a goal that drives the participation in developing the AS is to support the transfer of knowledge between operators as well as between operators and technicians in order to enable operators to fix faults more effectively, to increase the mean time between failures on the one hand and reduce the mean time to repair on the other.

The use case has two main characteristics that make it particularly suitable for an application of CBR: the key role that product-related and environmental influences exert on the production process, and the huge diversity of knowledge among operators. In the packaging of chocolate, the occurrence of faults strongly depends on the product that is being processed, the packaging material, and the specific interactions of different products and packaging materials with the machine. For instance, chocolate bars behave differently depending on their filling, cocoa concentration, or the presence of other ingredients such as nuts. The processing behaviour of the foil used for packaging depends on the quality of the material and the way it was stored. Accordingly, different machine settings are required to process different materials. Moreover, this interaction between product, packaging material, and machine is modulated by environmental factors. For instance, chocolate with cream filling is less robust than bitter chocolate, and at high temperatures chocolate with a high milk content tends to melt and leave smearings on the conveyor belt, which in turn increases the friction between the belt and the chocolate bars. Therefore, knowledge about such interactions is a fundamental requirement for machine operation as well as for the diagnosis and correction of faults. In section IV-E, we take a closer look at the resulting information space that needs to be covered by the CBR system. A requirement for the AS is to be useful for operators with different levels for knowledge. This makes it essential to adapt the representation of information and the user interaction with the AS to the needs of different operators. The usability and acceptance of AS can be increased by enabling operators to attain a better understanding of the situation, for instance by providing further explanations as needed [20]. The combination of CBR and NLP allows for a familiar form of interaction as the system can represent information in a comprehensible way and support dialogues that resemble the way language is used in everyday settings. Accordingly, AS are needed that can deal with the complex interactions characteristic of packaging machines and make them accessible to different operators in an adaptive manner.

IV. Concept for Assistance System

The approach presented in this paper enables a flexible interaction between the three cornerstones of fault diagnosis in CPPS: human, machine, and AS. Fig. 2 illustrates the interplay of these three components. The following concept is pursued: While an operator monitors the machine, he detects a fault. In order to diagnose its underlying cause, he can communicate with the AS in a question-answer dialogue. Based on this dialogue, the operator makes a decision and interacts with the machine in order to correct the fault. Throughout the dialogue, the AS draws on its knowledge base and its capability of recording machine and production data to provide adequate information in response to the operator's queries. It can also pose questions to the operator, depending on its current assessment of the situation. During this dialogue, every input and data storage is processed and evaluated. The following sections describe how our AS meets these requirements by addressing the challenges associated with machine configuration and machine learning, adapting and representing knowledge, and processing natural language.

A. Machine Configuration

The reliability of the AS depends on the amount and quality of the data it can draw upon. For the AS to be effective, an integrated information space is desirable as described in detail in section IV-E. This information space incorporates the machine with all its relevant parts. Technically, the AS is integrated in the entire production line or factory, which is illustrated in the upper part of Fig. 2. As described in section I, in most cases the cause of a fault is not to be found where the symptom occurs but in a previous or subsequent machine or production step. Accordingly, the AS needs interfaces to all parts of the production line, for instance through OPC UA. These interfaces make it possible to evaluate all machine data needed for diagnosing a fault. A basic requirement for the ML to be implemented in our AS (see section IV-C) is that the AS can receive and process sensor data through a real-time protocol. With modern machines, it is comparably easy to provide their engineering and sensor data, a seeded case base, and interfaces to spread machine settings and data in a semantic manner. However, for plants that already exist and are composed at least partly of older machinery, the AS concept also needs to be suitable for brownfield projects. Therefore, an additional communication channel is required which can transport the information necessary for fault diagnosis. Importantly, this channel must neither influence the machine nor provide a possible attack vector. To meet these requirements, the NAMUR[1] Open Architecture (NOA) approach has presented an appropriate concept: the information diode [39]. This approach makes it possible to add additional sensors and distribute their information through an OPC UA interface. Such information can be used by our ML system to improve its results.

B. Knowledge Based System and Representation

Section II-B described a common CBR cycle based on [18]. Going beyond this cycle, a typical feature of conversational CBR systems [40] is their ability to elicit case descriptions

[1]User Association of Automation Technology in Process Industries - https://www.namur.net/

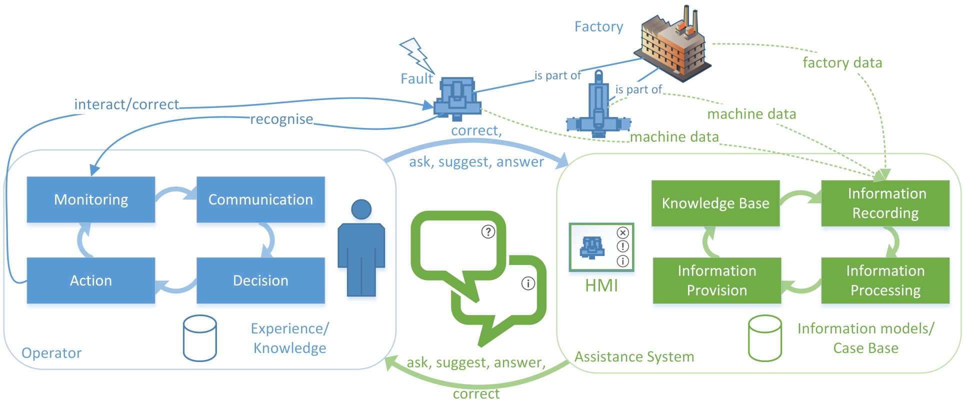

Fig. 2. Relation between the human, machine and assistance system and its components.

incrementally. In our system, we combine both approaches to describing a problem, and thus our system consists of two parts: a static feature vector part that uses data from the machine, MES, and environmental information space, and an incremental question-answer part that engages in dialogues with the operator. The resulting adaptation of the CBR cycle is depicted in Fig. 3. Our AS consists of two subsequent CBR cycles or phases: (1) The first phase is a static one, which receives information about the problem as an input and provides all relevant cases as an output. Based on a problem description provided by the operator (which can be simple and incomplete), the AS retrieves similar cases and presents a first suggestion in its next turn of the dialogue. In addition to answering questions posed by the AS, the operator can also take initiative to specify the problem description at any time, so that the system will retrieve a lower number of possible cases. It should be able to select and formulate its questions to the operator depending on the current set of cases it has retrieved at any given moment. For instance, imagine that the system has retrieved four suitable cases as a consequence of the operator's description. In the next turn, it can pose a question about a fact that best discriminates between two of these cases. The operator can also rule out cases depending on his background knowledge or his assessment of the current situation. The first phase ends when the operator selects the most likely case for reuse. (2) This case enters the second, session-based phase. During this phase, all interactions with the operator are remembered until the case is described sufficiently and the problem can be solved. At this point, the case is stored as a new case. In detail, the second phase that is operating on the selected, reused case presents a diagnosis and a suggestion to solve the problem. The operator can adapt the case and its parameters or the associated solution strategy, which can be seen as a complex revise step. The last part of the second phase creates a new case and retains it in the general knowledge base. However, if the operator realizes during fault correction that the diagnosis or solution suggested by the AS does not fit after all, he can go back and specify the fault more precisely, which will prompt the AS to suggest new cases. Besides the two phases or CBR cycles, a fundamental part of our CBR system is its so-called seed case base. Initially, domain-specific knowledge about the machine and its faults needs to be collected and stored, which provides a starting point for consulting the system in early phases of its application. The appropriateness of the seeds can determine whether and how a CBR system will be used in the future: Depending on their number and quality, dialogues with the CBR system in its early phases will either derive satisfactory solutions or not. If these early dialogues fail, operator's may never use the system again, or if they do use it, it will cost them large amounts of time to insert cases from scratch. Despite the crucial importance of the seed case base, neither a reliable calculation of the required number of seed cases nor a definition of good seed cases is available. Hence, it is useful to create seed cases with the support of domain experts. This can be done via expert interviews and other knowledge elicitation techniques [41], which is a methodologically challenging and time-consuming task. In our project, the seed case base is developed in cooperation with the machine developer and the technicians and operators. Moreover, to increase the validity of the seed case base and reduce the time required for developing it, we use information from virtual commissioning and machine simulation. Beyond that, the entire digital representation of a manufacturing line can automatically be used to transform new seed cases out of a previous version of a machine. This can be done by an interactive roundtrip engineering system [42].

C. Machine Learning

Our AS combines the information gained from human-machine dialogues in the CBR system with the results of ML to find the best matching cases. For the ML part, the machine first needs to be modelled before a suitable feature extraction process is implemented. Relevant features could be typical values on time series such as averages, minima, and maxima. When transporting goods on conveyor belts, the raw data often stems from simple light barriers, and in this case suitable features can be time lapses between the products passing by. To extract those features, a precise model

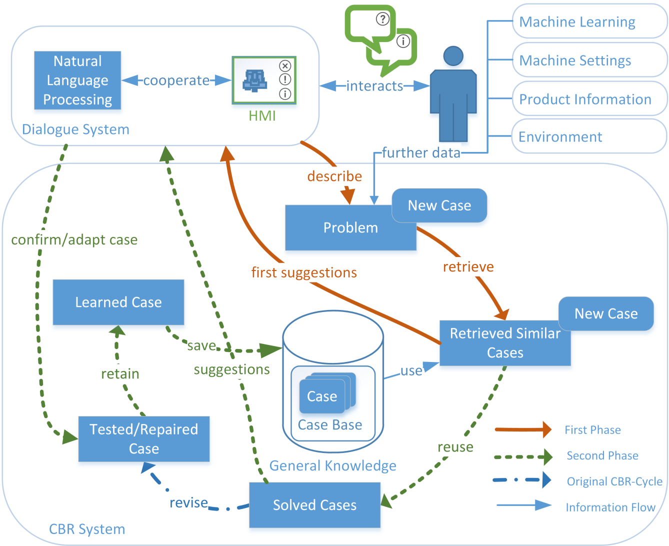

Fig. 3. CBR cycle for our assistance system, adapted from [18]

of the machine and the process is needed, which can be built using the information from the data model. Subsequent to feature extraction, a ML model is generated that provides the likelihoods of different faults. A detailed description of this solution is published in [43]. When these likelihoods generated by the ML are combined with the likelihoods generated by the CBR, the emerging situation model can provide more precise hints for diagnosing and correcting the root cause of faults. After a successful dialogue, the corresponding dataset will be classified, and after a new training this knowledge is available and will result in better likelihood estimations when a similar pattern is detected in the sensor signals later on.

D. Natural Language Processing

Natural language processing (NLP) is a field of artificial intelligence and linguistics that makes it possible to evaluate information expressed by operator in everyday language. NLP is an integral part of our concept and will enable dialogues to be made up of natural question-answer pairs. As the goal of these dialogues is to elicit the operator's concept of the situation, his mental model needs to be represented in our knowledge base. This is an essential prerequisite for the incremental retrieval part of our CBR system. A similar approach was already suggested by Aha et al. [44] and by Abacha et al. [11] who combined an expert system with Linked Data in the medical domain. However, in contrast to our approach these authors used predefined question-answer pairs which were presented to the user. If question-answer pairs are to be based on natural dialogue, text mining is necessary, and a major challenge is to deal with grammar, morphology, and lexical typing errors which are pervasive in written human conversation. Moreover, the system needs to understand the function of words such as "and" and "or" in a particular context. For instance, "and" can precede the description of an additional problem or a specification of the problem that was described in the previous part of an utterance [45]. A major difficulty in NLP is the pervasive ambiguity inherent in human language as described by Allen [46]:

- *Lexical ambiguity* (e.g., the word "duck" might be a noun that refers to an animal or a verb that refers to the action of avoiding an object that is thrown at you)
- *Structural or syntactic ambiguity* (e.g., in "I saw the man with a telescope," the telescope might be used for the viewing or might be held by the man being observed)
- *Semantic* ambiguity (e.g., "go" as a verb has well over 10 distinct meanings in any dictionary)
- *Pragmatic ambiguity* (e.g., "Can you lift that rock?" may be a yes/no question or a request to lift the rock).
- *Referential ambiguity* (e.g., in "Jack met Sam at the station. He was feeling ill." it is not clear who is ill.

To meet these challenges, an ontology with a specific dictionary needs to be created, which can find concepts in the knowledge base according to its evaluation of the operators input. Additionally, a synonym dictionary (e.g., DISCO[2]) should be used to understand the meaning of words in the current context.

E. Distributed Knowledge & Information Management

In existing factory automation systems, most of the information required by our approach exists in a more or less digital manner. Nevertheless, it should be possible for the AS to share its information with other companies or factory locations to gain an increased benefit from distributed information sources. Fig. 4 presents a situation involving two factories and one triplestore. As some cases are related to each other, changing one case should lead to a respective change of the other, related case in order to maintain a consistent case base. Moreover, as described above, the cause of faults often is located in another machine. If we change the parameters of this machine, the case base needs to be updated. Hence, a mechanism is needed that synchronizes our models with interactive roundtrip engineering system concepts [42]. Similarly, if the soft- or hardware of a machine is updated by an engineering office, all cases related to this machine may need to be updated, or new cases may need to be created out of the older ones. For this, our adaptive information space uses five information sources to derive possible cases and their solutions, and accordingly all this information needs to be included in our case descriptions:

a) The dialogue system elicits a natural language description provided by the operator, which is processed to interpret the operator's understanding as described above. From this description, the CBR system gets relevant concepts and key words for classifying the fault.
b) The results of the ML system are used to compare the current situation with the CBR results. The result is used to ask further questions in order to specify the results and to store the results of this resolution process in the case base.

[2]DIStributionally related words using CO-occurrences - http://www.linguatools.de/disco/disco_en.html

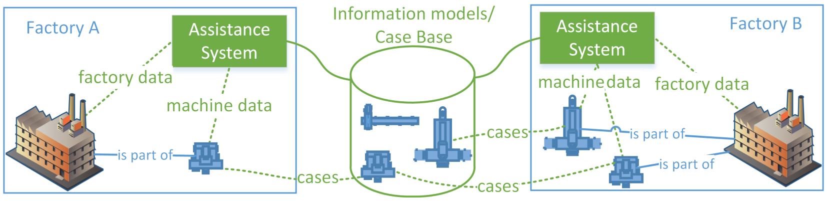

Fig. 4. Using the Linked Data approach, it is possible to share information in a distributed way and use it for the front-end of the assistance system.

c) The current machine settings are compared with product properties, so if the settings have not been adjusted after a product change, we are able to deduce such misalignments. This information can be requested through interfaces such as OPC UA as the most relevant candidate for CPPS.
d) Information about the current product is used to categorize and specify our case base. In our use case, we need information about the product, its recipe, raw material quality, and the packaging material. This information can be requested from existing MES.
e) Information about the environment such as temperature or humidity can be crucial and can provide valuable information about fault causes. Accordingly, this information is also considered in our case descriptions.

An important prerequisite for information retrieval is a semantic description that we can query. Therefore, different ontologies for the five information sources are created or reused from other projects (e.g., the MAnufacturing's Semantics ONtology (MASON) [11] ontology for description of products and processes). Using these ontologies, a case is subdivided into a problem description, a description of its cause, and a suitable step-by-step solution. Additionally, an ontology provides a set of diagnosis tasks to evaluate the correct cause and solution. All faults that have occurred in a factory are collected in a Linked Data storage including a case-based ontology and a collection of all relevant engineering knowledge about the machines.

V. Summary and Further Work

The present article proposed an approach for supporting the diagnosis and correction of faults by enabling cooperative human-machine dialogues. Our AS integrates human knowledge and machine data to derive a joint understanding of the current fault situation. Furthermore, it combines different information sources to determine the most suitable solution. This combination of knowledge representations in Linked Data and a CBR system allows us to establish a flexible and adaptive case base with a descriptive specification of similarity calculation. The current concept can be extended to scenarios involving a distribution of semantic information across factories, making it possible to spread human knowledge across the boundaries of a particular work site.

To orchestrate features of an AS in a flexible manner for a specific use case, a key challenge is to combine and adapt well-known approaches from artificial intelligence (i.e., CBR, ML, NLP) with information space linkage and synchronization (i.e., Linked Data, OPC UA). Moreover, to enhance the functioning of the entire human-machine system, psychological aspects relating to the operator need to be considered. Therefore, a future challenge is the development of suitable interaction patterns between operators and the AS. An important further challenge lies in the development of suitable interaction patterns between operators and the AS. On the one hand, at present many operators lack a thorough understanding of fault causes and thus their solution strategies focus on fixing the symptoms, which can be mitigated by the system providing in-depth information and contributing to ongoing operator qualification. On the other hand, in order to avoid losses in productivity it is important to restart the machine as quickly as possible after a fault, which might call for succinct dialogues that keep the information exchange to a minimum. To find a balance between these conflicting requirements, the system needs to be appropriately integrated into operators' work flows and the production environment as a whole. It also remains an open question for future research whether other types of HMI (e.g., smart watches, AR glasses, spoken language interaction) are promising alternatives, or whether the HMI should be integrated in the machine's stationary HMI, be realized on a separate HMI, or even allow for modern Bring-your-own-device (BYOD) concepts. The integration of the HMI should be implemented under the necessary framework conditions and safety regulations.

Another field for future cross-domain research is to transfer our concept to the process industries. We have conceptualized our AS based on a use case in the food processing industry, but newer approaches like modularization in the process industries lead to a growing number of interesting similarities between the two domains. For instance, the reconfiguration of modules or adaptations of the process may result in faults that go beyond a mechanic's specific process knowledge. Accordingly, it can be helpful to get support from an AS that relies on knowledge gained from similar processes in the past. Further use cases can be found in tasks such as maintenance and commissioning, which are likely to benefit from case-based AS that make domain-specific knowledge available.

Acknowledgment

The underlying project of this contribution was sponsored by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF), under the sponsorship reference number 02K16C072. The authors are responsible for the contents of this contribution.

REFERENCES

[1] L. Wang, M. Trngren, and M. Onori, "Current status and advancement of cyber-physical systems in manufacturing," *Journal of Manufacturing Systems*, vol. 37, pp. 517–527, Oct. 2015.

[2] A. J. C. Trappey, C. V. Trappey, U. H. Govindarajan, J. J. Sun, and A. C. Chuang, "A Review of Technology Standards and Patent Portfolios for Enabling Cyber-Physical Systems in Advanced Manufacturing," *IEEE Access*, vol. 4, pp. 7356–7382, 2016.

[3] P. Tsarouhas, "Reliability, availability and maintainability analysis in food production lines: a review," *International Journal of Food Science & Technology*, vol. 47, no. 11, pp. 2243–2251, Nov. 2012.

[4] J.-P. Majschak, *Grundlagen für Maschinensysteme der Stoffverarbeitung*, 22nd ed., K.-H. Grote and J. Feldhusen, Eds. Berlin Heidelberg: Springer-Verlag, 2014, vol. Dubbel Taschenbuch für den Maschinenbau.

[5] A. Schult, E. Beck, and J.-P. Majschak, "Hin zum Optimum. Steigerung der Effizienz von Verarbeitungs- und Verpackungsanlagen durch Wirkungsgradanalysen," *Pharma+Food*, pp. 66–68, 2015.

[6] R. Müller, "(in press) Cognitive challenges of changeability: Adjustment to system changes and transfer of knowledge in modular chemical plants," *Cognition, Technology &Work*.

[7] D. W. Aha, D. Mcsherry, and Q. Yang, "Advances in conversational case-based reasoning," *The Knowledge Engineering Review*, vol. 20, no. 03, p. 247, Sept. 2005.

[8] H. Hla Moe and W. Thanda Aung, "Cross-domain Recommendations for Personalized Semantic Services," *International Journal of Computer Applications Technology and Research*, vol. 2, pp. 72–76, Feb. 2012.

[9] R. G. Rocha, R. R. Azevedo, Y. C. Sousa, E. d. A. Tavares, and S. Meira, "A Case-based Reasoning System to Support the Global Software Development," *Procedia Computer Science*, vol. 35, pp. 194–202, 2014.

[10] V. Lopez, C. Unger, P. Cimiano, and E. Motta, "Evaluating question answering over linked data," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 21, pp. 3–13, Aug. 2013.

[11] A. Ben Abacha and P. Zweigenbaum, "MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies," *Information Processing & Management*, vol. 51, no. 5, pp. 570–594, Sept. 2015.

[12] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data - The Story So Far:," *International Journal on Semantic Web and Information Systems*, vol. 5, no. 3, pp. 1–22, 2009.

[13] M. Graube, J. Pfeffer, J. Ziegler, and L. Urbas, "Linked Data as Integrating Technology for Industrial Data." in IEEE Network-Based Information Systems (NBiS), Sept. 2011, pp. 162–167.

[14] E. Prud'hommeaux and A. Seaborne, *SPARQL Query Language for RDF*, Jan. 2008. [Online]. Available: http://www.w3.org/TR/rdf-sparql-query/

[15] M. Graube, S. Hensel, C. Iatrou, and L. Urbas, "Information models in OPC UA and their advantages and disadvantages," in *Emerging Technologies and Factory Automation (ETFA), 2017 22nd IEEE International Conference on*. IEEE, 2017, pp. 1–8.

[16] R. C. Schank, *Dynamic memory: a theory of reminding and learning in computers and people*. Cambridge [Cambridgeshire] : New York: Cambridge University Press, 1982.

[17] Y.-J. Park, B.-C. Kim, and S.-H. Chun, "New knowledge extraction technique using probability for case-based reasoning: application to medical diagnosis," *Expert Systems*, vol. 23, no. 1, pp. 2–20, 2006.

[18] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches," *AI communications*, vol. 7, no. 1, pp. 39–59, 1994.

[19] C. Jian, T. Zhe, and L. Zhenxing, "A Review and Analysis of Case-Based Reasoning Research," in *2015 International Conference on Intelligent Transportation, Big Data and Smart City*, Dec. 2015, pp. 51–55.

[20] D. Doyle, A. Tsymbal, and P. Cunningham, "A Review of Explanation and Explanation in Case- Based Reasoning," Department of Computer Science, Trinity College, Dublin, 2003, Technical Report.

[21] Y.-J. Park, S.-H. Chun, and B.-C. Kim, "Cost-sensitive case-based reasoning using a genetic algorithm: Application to medical diagnosis," *Artificial Intelligence in Medicine*, vol. 51, no. 2, pp. 163–200, Feb. 2011.

[22] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, p. 78, Oct. 2012.

[23] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, May 2016.

[24] C. Brecher, M. Obdenbusch, and M. Buchsbaum, "Optimized state estimation by application of machine learning," *Production Engineering*, vol. 11, no. 2, pp. 133–143, Apr. 2017.

[25] A. Lotze, *Steuerungsintegrierte Vorgangsdiagnose in Verarbeitungsmaschinen am Beispiel einer Quersiegeleinheit*, 2008, google-Books-ID: FngkPwAACAAJ.

[26] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, Nov. 1983.

[27] M. Endsley, "Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors Journal 37(1), 32-64," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, pp. 32–64, Mar. 1995.

[28] R. Parasuraman and D. H. Manzey, "Complacency and bias in human use of automation: an attentional integration," *Human Factors*, vol. 52, no. 3, pp. 381–410, June 2010.

[29] R. Müller, S. Narciss, and L. Urbas, "Chapter 10 - Interfacing Cyber-Physical Production Systems With Human Decision Makers," in *Cyber-Physical Systems*, ser. Intelligent Data-Centric Systems, H. Song, D. B. Rawat, S. Jeschke, and C. Brecher, Eds. Boston: Academic Press, 2017, pp. 145–160.

[30] K. J. Vicente and J. Rasmussen, "Ecological interface design: theoretical foundations," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 4, pp. 589–606, July 1992.

[31] C. Borst, J. Flach, and J. Ellerbroek, "Beyond Ecological Interface Design: Lessons From Concerns and Misconceptions," *Human-Machine Systems, IEEE Transactions on*, vol. 45, pp. 164–175, Apr. 2015.

[32] J. L. Kolodner, "Improving Human Decision Making Through Case-based Decision Aiding," *AI Mag.*, vol. 12, no. 2, pp. 52–68, Apr. 1991.

[33] N. Althuizen and B. Wierenga, "Supporting Creative Problem Solving with a Case-Based Reasoning System," *Journal of Management Information Systems*, vol. 31, pp. 309–340, July 2014.

[34] D. McSherry, "Explanation in Recommender Systems," *Artificial Intelligence Review*, vol. 24, no. 2, pp. 179–197, Oct. 2005.

[35] H. Wandke, "Assistance in humanmachine interaction: a conceptual framework and a proposal for a taxonomy," *Theoretical Issues in Ergonomics Science*, vol. 6, no. 2, pp. 129–155, Mar. 2005.

[36] L. Schegner, S. Hensel, J. Wehrstedt, R. Rosen, and L. Urbas, "Architekturentwurf für simulationsbasierte Assistenzsysteme in prozesstechnischen Anlagen," in *Tagungsband Automation 2017*, Baden-Baden, 2017.

[37] L. Onnasch, C. D. Wickens, H. Li, and D. Manzey, "Human performance consequences of stages and levels of automation: an integrated meta-analysis," *Human Factors*, vol. 56, no. 3, pp. 476–488, May 2014.

[38] D. Manzey, J. Reichenbach, and L. Onnasch, "Human Performance Consequences of Automated Decision Aids: The Impact of Degree of Automation and System Experience," *Journal of Cognitive Engineering and Decision Making*, vol. 6, no. 1, pp. 57–87, Mar. 2012.

[39] C. Paul, M. Graube, L. Urbas, T.-P. Henrichs, and S. Erben, "NOA Verification of Request," *atp edition*, vol. 60, no. 01-02, pp. 60–69, Jan. 2018.

[40] D. W. Aha, L. A. Breslow, and H. Muoz-Avila, "Conversational case-based reasoning," *Applied Intelligence*, vol. 14, no. 1, pp. 9–32, 2001.

[41] N. Shadbolt and P. Smart, *Knowledge Elicitation: Methods, Tools and Techniques*, 3rd ed., J. R. Wilson and E. N. Corlett, Eds. Boca Raton, FL: Taylor & Francis, 2015.

[42] J. Rahm, M. Graube, and L. Urbas, "A proposal for an interactive roundtrip engineering system," in *Emerging Technologies and Factory Automation (ETFA), 2017 22nd IEEE International Conference on*. IEEE, Sept. 2017, pp. 1–7.

[43] T. Klaeger, A. Schult, and J.-P. Majschak, "Lernfähige Bedienerassistenz für Verarbeitungsmaschinen," no. 33, pp. 25–28, 2017.

[44] D. W. Aha, T. Maney, and L. A. Breslow, "Supporting dialogue inferencing in conversational case-based reasoning," in *European Workshop on Advances in Case-Based Reasoning*. Springer, 1998, pp. 262–273.

[45] E. D. Liddy, "Natural Language Processing," *Encyclopedia of Library and Information Science*, p. 15, 2001.

[46] J. F. Allen, "Natural Language Processing," in *Encyclopedia of Computer Science*. Chichester, UK: John Wiley and Sons Ltd., pp. 1218–1222.