RECKOn: a REal-world, Context-aware KnOwledge-based lab

(Discussion Paper)

Patrizia **Agnello**[1], Silvia Maria **Ansaldi**[1], Emilia **Lenzi**[2], Alessio **Mongelluzzo**[2], Davide **Piantella**[2], Manuel **Roveri**[2], Fabio A. **Schreiber**[2], Alessandra **Scutti**[2], Mahsa **Shekari**[2] and Letizia **Tanca**[2]

[1]*INAIL - Dipartimento Innovazioni Tecnologiche*
[2]*Politecnico di Milano - Dipartimento di Elettronica, Informazione e Bioingegneria*

Abstract

The RECKON project focuses on interconnection technologies and context-aware data-analytics techniques to improve safety in workplaces, with the ultimate objective of identifying and preventing dangerous situations before accidents occur. In RECKON, prevention is interpreted through the latest monitoring, diagnostics and prognostics techniques from a safety perspective, allowing to detect and use, even in real time, a large amount of data about the entire operational context. Using sensor networks, we are able to collect information that is used in two ways: (i) when a potentially dangerous situation is detected, the system raises an alarm to prevent an accident, and (ii) whenever an accident or a near-miss (i.e., a potential accident that was narrowly averted) occurs, the related useful information is stored in a case report automatically generated and later used to update the accident-prevention politics. This work briefly describes the operational framework of RECKON, along with its modules and their interaction.

Keywords

Workplace safety, Industry 4.0, Context-awareness, Sensor data

1. Introduction

Accidents at work have been a topic of social and economic debate for a long time now. According to the data of INAIL (Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro), 641,638 workplace accidents occurred in 2019, 1089 of them resulting in death [1]. Although the interest in the subject is always high, the evolution of the enterprise context, in which accidents occur, makes it necessary to continuously observe and study these phenomena and to use constantly updated systems for their prevention.

The innovative paradigm introduced by Industry 4.0 [2] has brought about structural, technological, productive and organisational changes in the world of work [3], all based on a high

technological development aimed at increasing the productivity and competitiveness of companies in the market [4]. The use of innovative sensors, for example, makes it possible to analyse large amounts of data extracted directly from the production context. However, while there has been a very strong interest in the digitalization of the working environment, this has been almost entirely channelled into the dimension of machines and their interconnections with the surrounding environment, putting the equally important aspect of human presence in the background [5]. Instead, by putting the operator at the center, the introduction of innovative technological solutions may prove fundamental not only for the company's performance in terms of profits and competitiveness, but also to provide support to the operator in terms of prevention in the field of health and safety at work [6].

It is therefore clear that crucial points of the problem are an accurate and up-to-date analysis and description of the context and how the operator interacts with it, and an efficient use of the data and information derived from this analysis. This is the purpose of the RECKON system, designed to support companies in monitoring and preventing accidents in the workplace.

RECKON exploits the integration of the historical analysis with a conceptualisation and sensitisation of the working context, making it possible to highlight the correspondence between a situation that has already been identified as potentially dangerous and the current working situation. More precisely, the system can be considered as composed of three conceptual parts: (i) sensitization of the companies under consideration (in the current case study, metallurgical enterprises); (ii) integrated, context-aware sensor data stream processing and conceptualization of the context described by those data; (iii) integration of the historical data collected in the critical-events database with other datasets about accidents (both internal and external to INAIL). Thanks to the last two tasks, it will be possible to build a knowledge base for the subsequent analyses useful to improve the context-aware monitoring system.

Sec. 2 of this paper gives an overview of the system, especially focusing on the information workflow (modules two and three); Sec. 3, details module two; and finally, Sec. 4, summarizes the techniques used to perform integration in the third module.

2. System overview

As shown by Fig. 1, part of the system operates directly in the company, to minimise the latency time when a decision is to be made or an alarm is triggered, and another part works, at a higher level (the *RECKON Hub*), in the Cloud, to integrate information from different companies and other data sources.

In fact, the RECKON architecture comprises four levels. In the workplace, we find wearable devices for workers and sensorized machines, and a localisation and imaging system; these two levels allow to localize the workers and machines and to signal critical or abnormal situations via vibration or acoustic signals *(Module 1)*. In the enterprise we also find an edge-computing system to process data acquired from the second level and to generate alarms that will be sent (via second level) to the workers' wearable devices *(Module 2)*. Relevant information from specific installations of the system in the companies are collected and sent to the central Hub. At this level, it will be possible to perform wide-range analyses on alert or hazardous situations,integrate these analyses with external sources of all kinds *(Module 3)*, and visualize
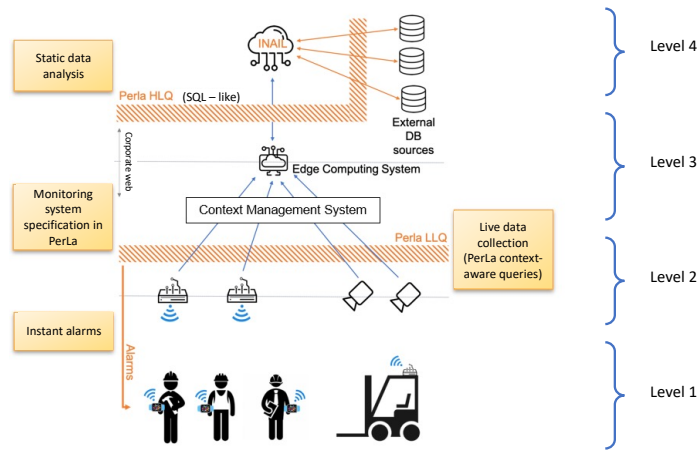
Figure 1: The architecture of RECKON.

data and results appropriately.

In all cases, when dealing with context-aware pervasive systems, two main problems arise: the integration of heterogeneous streams of data coming from distinct data sources and the problem of specifying the system behaviour at various levels, from the low-level support for hardware abstraction to the high-level support for data management, as we can see from the system overview. For this reason, we introduce the PerLa system [7], a middleware for data management and integration that uses the database abstraction for managing the pervasive system, allowing a data-centric view of the pervasive network and providing a homogeneous high-level interface to heterogeneous devices. The query language of PerLa is an SQL-like language that allows access to all the data collected by the network nodes. At the moment, PerLa has been implemented as a prototype, and using it in a real system would need proper engineering. However, being it very clear and expressive, we are using PerLa as a specification language to define the system behaviour and the application constraints. As shown in Fig. 1, there are different types of PerLa queries: the PerLa High Level Queries (HLQ) can be used for specifying the tasks for integrating different datasets; the PerLa Low Level Queries (LLQ) are interposed between level 2 and level 3 and trigger the sampling operations for the main context parameters (e.g., worker's elevation, worker's position, etc.) thus allowing correct identification of context changes; in addition, between the two levels we also find the PerLa Context-aware Queries, that are context-specific and permit to verify particular conditions within the operator-machine-environment situation.

2.1. Context-aware information workflow

In Fig. 2 we show the information workflow of RECKON, describing the context-aware behaviour of the system. Sensor data collection from the field, data processing and integration, and alarm generation, are all context-aware.

The workflow consists of eight main steps, represented in Fig. 2 and detailed below. Let
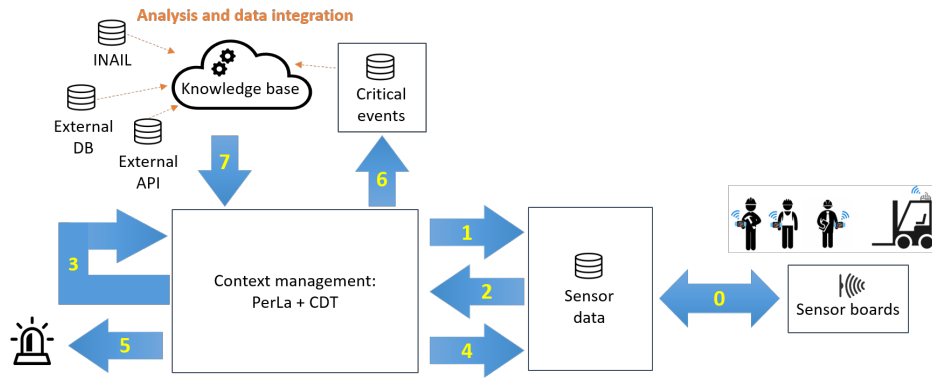
Figure 2: The information workflow in RECKON.

us consider the following use case of the system: *"A worker W enters a delimited area Y of the working facility, containing a fixed machinery X. In order to operate X, workers should use specific Personal Protective Equipment (PPE), currently not worn by W".* We now describe how this scenario is mapped to each step of the workflow.

Step 0: Field data collection This step is responsible for collecting data from the field: without loss of generality, we assume that sensor-specific software can ensure an easy access to sensor data and measurements, in a structured format. For simplicity, we represented these data as part of a relational database.

Step 1: PerLa Continuous Query (LLQ) context monitoring is based on the continuous execution of a set of Low Level Queries on the set of available data sources. For example, to monitor the height and location of workers in real time, a continuous query will be required on all smartwatch devices; or, to update the location of mobile machines, a position parameter sampling operation will be initiated on all devices installed on the machines. In the use case, if we want to detect the usage of machine X by worker W, one continuous query could be: *"What is the distance between the worker W and the machine X?"* (List. 1).

Step 2: Context change detection We detect a context change if the result of the continuous query described in Step 1 (i.e., distance between the worker & machine X) is below the threshold.

Step 3: Critical context activation Updating context parameters may lead to the activation of contexts identified as critical, for which further checks are necessary. In the use case, the usage of machinery X requires a verification on the PPE of W.

Step 4: PerLa Context-aware Query The activation of a critical context determines the execution of Context-aware Queries, possibly generate alarm signals and finally define which are the useful data to be collected in a case report. In the use case, a context-specific query could be: *"Did the worker wear all the PPE required for using machine X?"* (List. 2).

Step 5: PerLa activates context-based alarms The alarms are triggered when critical conditions occur for the context under consideration; e.g., the worker is not wearing the PPE, thus an alarm is triggered and, possibly, some predefined security measures are activated. This closes the context-aware online phase.

Step 6: Accident and near-miss report This step is to automatically generate a case report with the details of the accident or near-miss, based on the Context Dimension Tree model de-

scribed in Sec. 3. The report is stored in a database, available at both corporate and hub level,that grants the privacy of personal and sensitive information. The format of the report will be structured or semi-structured, to ease subsequent automated analysis (Step 7).

Step 7: Refinement and improvement of PerLa queries PerLa queries can be modified or added, leveraging a knowledge base generated from the integration and analysis of sensor data, reports generated in Step 6, historical datasets of accidents, and external data sources. This possibility aims at anticipating the detection of dangerous situations, reducing the number of accidents and near-misses. With reference to the use case, the system could infer that too many non-authorized workers enter area Y to operate machinery X. This new information may suggest to anticipate the control of the PPE, to the moment when workers enter area Y, without waiting for the detection of the use of X. If the PPE is not worn when entering area Y, a pre-alarm is triggered to remind workers to wear the required equipment.

```
1 CREATE OUTPUT STREAM MachineryLocation (id_machinery STRING, ts STRING, x FLOAT, y FLOAT, z
      FLOAT) AS:
2 EVERY 30 seconds
3 SELECT id_machinery, ts, x, y, z
4 SAMPLING EVERY 10 seconds
5 EXECUTE IF EXISTS id_machinery, x, y, z
```

Listing 1: Perla Continous Query (LLQ)

```
1 CREATE CONTEXT Worker_FixedMachinetry_RelativePosition
2 ACTIVE IF SmartwatchProfile EXISTS AND MachEquType = "Fixedmachinery" AND RelativePosition_WM <
      $mimimum_safe
3 ON ENABLE (Worker_FixedMachinetry_RelativePosition):
4    SELECT SmartwatchProfile, Machinery_UsePermission, Workplace_Access,
        PersonalProtectiveEquipment, ProtectionBarriers, SafetyDevices
5    SAMPLING EVERY 1m
6    SET PARAMETER "warning message" = TRUE
7    ACTIVATE ALARM
8 ON DISABLE:
9    DROP Worker_FixedMachinetry_RelativePosition
10   INSERT RECORD INTO Critical_events_db
11   SET PARAMETER "warning message" = FALSE
12 REFRESH EVERY 5m
```

Listing 2: PerLa Context-aware Query

3. Context modelling and representation

A first step in the design of the RECKON context-aware data integration system is a conceptual modeling phase, aiming to understand and model context information. Among the methodologies for designing context-aware systems [8], we use the one based on the Context Dimension Tree (CDT) [9], because it allows to capture the context both at a *conceptual* and a *detail* level and has already been experimented for these purposes [10, 11, 12]. The CDT captures distinct situations in which the users of a system can find themselves and interact with the surrounding environments. The tree's root r represents the most general context; N is the set of nodes; nodes are either black *Dimension Nodes* N_D or white *Concept Nodes* N_C, *a.k.a*, dimension's values; N_D and N_C must alternate along the branches. White triangles are shorthands to represent
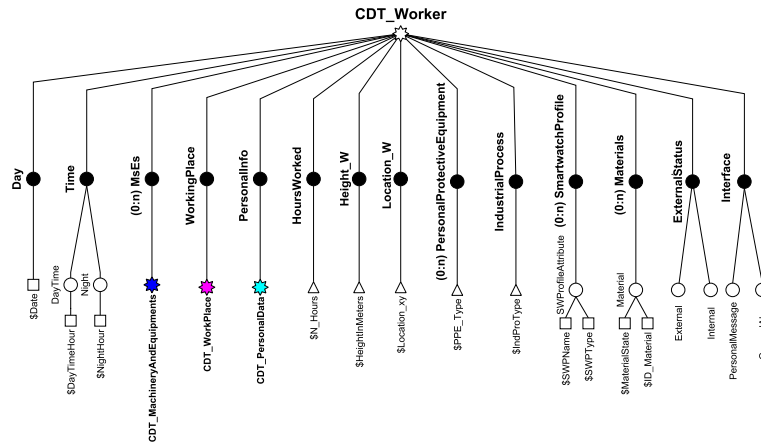
Figure 3: The Worker CDT.

a collection of N_C graphically. Also, each N_C and each leaf N_D without any children can be further detailed through *Parameters*, shown in white squares. The root's children, *a.k.a.*, *top dimensions*, specify the main dimensions of the analysis. Each N_D should have at least one N_C or parameter. Fig. 3 shows the CDT Worker, representing the most relevant dimensions for determining the accident risk exposure of each single worker in the "operator-machine-environment" context in the metallurgic Small and Medium-size Enterprises (SMEs) scenario. The context is understood both as the physical environment (e.g., the environmental parameters, the objects with which the worker interacts such as machinery and equipment) and as the set of characteristics specific to each worker (e.g., the worker's job profile, age, work experience) that help determine the most critical situation for each worker. In the CDT Worker, some N_Ds have cardinality greater than one; the cardinality was introduced to better model the context of the worker who, for example, may interact simultaneously with several machines, or, may wear several Personal Protective Equipments. Note that the dimensions "WorkingPlace", "Personal-Data" and "MachineryAndEquipments" generate sub-trees not shown here that are nested in the CDT Worker. The CDT specification allows us to model all the possible contexts and monitor them during execution, guiding system actions, data collection and interpretation. Through the PerLa Context component [13] it is possible to express formally both the context model (CDT Declaration) and the context-aware system behavior when a critical context occurs (Context Creation clause). Declaring a context also describes the conditions for its activation (ACTIVE IF clause) and defines the consequent action(s) of the system. The ACTIVE IF condition is defined on the context parameters and dimensions by considering some critical thresholds. For example, the context when: (i) a worker is in the proximity of a mobile machine, (ii) the distance between the worker and the machine is less than the critical threshold and (iii) the machine is switched on or moving, is described as: *ACTIVE IF MachEqType = 'MobileMachine' AND Velocity_M > 0 m/s AND RelativePositition_WM < MnimumSafeDistance*. Each of the main accident categories and dangerous situations derived from the analyses defines one or more such critical contexts, that can be modified or added as the system gains "experience", i.e., based on the progressive

analyses of the occurred situations.

4. Extracting knowledge from textual descriptions

A further enrichment of the description of work environments and their possible dangers is the use of ontologies. Since we could not find sufficiently rich ontologies for the Italian language, we developed a new one starting from the CDT terminology and from the analysis of the datasets at our disposal. The ultimate aim of this ontology is to integrate the context-aware knowledge with analysis techniques for Natural Language Processing (NLP), a crucial part of Module 3. To combine the semantic representation with textual data analysis, we divided the ontology into *Grammatical Categories* and *Semantic Classes*. This permits to separate the meaning of each term from its grammatical role, using these two pieces of information either independently or in a combined way, depending on the task. Currently, taking into account the methods used for NLP-based analysis, we have identified four basic *Grammatical Categories*: *Noun, Verb, Adjective and Adverb*. To identify the main *Semantic Classes*, following a bottom-up approach, we started from the most frequent terms of the CDT and the data at our disposal and then generated a hierarchy taking into account their correspondence with the dimensions of the context tree. To embed ontological classes in the textual analyses, a *TAG* is created for each class, and the most frequent terms analyzed are replaced in the datasets with these tags. Here is an example of hierarchy in the ontology with related *TAGs*: *Object(0) –> Work_Object(WO) –> Tool (WOT) –> hammer*. Regarding the combination of *Semantic Classes* and *Grammatical Categories*, at the moment we are working on the formalisation of rules that follow those of the Italian grammatical and logical analysis, being careful not to introduce redundancies. For example, to express the activity "to hammer" we chose to combine the category "Verb" with the class "Tool" in order to group all the accidents involving tools easily without creating a proper class "work activity" or "work activity involving tools".

The third and last module is RECKONition (extensively described in [14]), a NLP-based system for accidens prevention in industry that analyses Italian textual descriptions of previous accidents to build both unsupervised and supervised models. The architecture of this analysis system comprises three models: *Association Rule Generator, Textual Description Clustering* [15] – highlighting groups of accidents at work sharing similarities in their descriptions–, and *Textual Description Inference*, providing next-sentence predictions from the textual description of the accidents. Experimental results are rather satisfactory.

5. Conclusion and future works

We introduced RECKON, a context-aware pervasive system which combines different technologies to efficiently prevent accidents within metallurgical companies. In particular, we focused on describing the context of interest and its embedding within heterogeneous integration and analysis on different levels. To verify the efficiency of the proposed models, we conducted a survey campaign with many manufacturing companies, which resulted in an endorsement of our proposed framework. The next steps will be the actual installation of the sensors in situ and the analysis of the sampled data.

References

[1] INAIL, Andamento degli infortuni sul lavoro e delle malattie professionali, https://www.inail.it/cs/internet/docs/alg-dati-inail-2021-gennaio-pdf.pdf, 2021.

[2] Y. Lu, Industry 4.0: A survey on technologies, applications and open research issues, J. Ind. Inf. Integr. 6 (2017) 1–10. doi:`10.1016/j.jii.2017.04.005`.

[3] J. Lee, B. Bagheri, H.-A. Kao, A cyber-physical systems architecture for industry 4.0-based manufacturing systems, Manufacturing Letters 3 (2015) 18–23.

[4] J. Davis, T. Edgar, J. Porter, J. Bernaden, M. Sarli, Smart manufacturing, manufacturing intelligence and demand-dynamic performance, Comput. Chem. Eng. 47 (2012) 145–156.

[5] P. Fantini, M. Pinzone, M. Taisch, Placing the operator at the centre of industry 4.0 design: Modelling and assessing human activities within cyber-physical systems, Computers & Industrial Engineering 139 (2020) 105058.

[6] V. Paelke, Augmented reality in the smart factory: Supporting workers in an industry 4.0. environment, Proc. of the 2014 IEEE ETFA (2014) 1–4.

[7] F. A. Schreiber, R. Camplani, M. Fortunato, M. Marelli, G. Rota, Perla: A language and middleware architecture for data management and integration in pervasive information systems, IEEE Trans. Softw. Eng. 38 (2011) 478–496.

[8] C. Bolchini, C. A. Curino, E. Quintarelli, F. A. Schreiber, L. Tanca, A data-oriented survey of context models, ACM Sigmod Record 36 (2007) 19–26.

[9] C. Bolchini, C. A. Curino, E. Quintarelli, F. A. Schreiber, L. Tanca, Context information for knowledge reshaping, IJWET 5 (2009) 88–103.

[10] V. Cassani, S. Gianelli, M. Matera, R. Medana, E. Quintarelli, L. Tanca, V. Zaccaria, On the role of context in the design of mobile mashups, in: RMC, 2016, pp. 108–128.

[11] A. Javadian Sabet, M. Rossi, F. A. Schreiber, L. Tanca, Towards learning travelers' preferences in a context-aware fashion, in: Ambient Intelligence – Software and Applications, Springer International Publishing, Cham, 2021, pp. 203–212.

[12] A. J. Sabet, M. Rossi, F. A. Schreiber, L. Tanca, Context awareness in the travel companion of the shift2rail initiative, in: Proc. of the 28th Italian Symposium on Advanced Database Systems, volume 2646 of *CEUR Work. Proc.*, 2020, pp. 202–209.

[13] F. A. Schreiber, E. Panigati, Context-aware self adapting systems: a ground for the cooperation of data, software, and services, IJNGC 8 (2017) 32–61.

[14] P. Agnello, S. M. Ansaldi, E. Lenzi, A. Mongelluzzo, M. Roveri, Reckonition: a nlp-based system for industrial accidents at work prevention, 2021. `arXiv:2104.14150`.

[15] M. J. Zaki, W. Meira, Jr, Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, 2014.