Advances in Science, Technology and Engineering Systems Journal Vol. 2, No. 3, 12-23 (2017)

www.astesj.com

Special issue on Recent Advances in Engineering Systems

ASTESJ

ISSN: 2415-6698

Applications of Case Based Organizational Memory Supported by the PAbMM Architecture

Martín, María de los Ángeles[1], Diván, Mario José[2]

[1]Engineering School,Universidad Nacional de La Pampa, General Pico, 6360, Argentina

[2]Law and Economy School, Universidad Nacional de La Pampa, Santa Rosa, 6300, Argentina

ARTICLEINFO

Article history:
Received: 03 March, 2017
Accepted: 25 March, 2017
Online: 02 April, 2017

Keywords:
Organizational Memory
Data Stream Processing
Big Data
Remote Sensing
Measurement & Evaluation

ABSTRACT

In the aim to manage and retrieve the organizational knowledge, in the last years numerous proposals of models and tools for knowledge management and knowledge representation have arisen. However, most of them store knowledge in a non-structured or semi-structured way, hindering the semantic and automatic processing of this knowledge. In this paper we present a more detailed case-based organizational memory ontology, which aims at contributing to the design of an organizational memory based on cases, so that it can be used to learn, reasoning, solve problems, and as support to better decision making as well. The objective of this Organizational Memory is to serve as base for the organizational knowledge exchange in a processing architecture specialized in the measurement and evaluation. In this way, our processing architecture is based on the C-INCAMI framework (Context-Information Need, Concept model, Attribute, Metric and Indicator) for defining the measurement projects. Additionally, the proposal architecture uses a big data repository to make available the data for consumption and to manage the Organizational Memory, which allows a feedback mechanism in relation with online processing. In order to illustrate its utility, two practical cases are explained: A pasture predictor system, using the data of the weather radar (WR) of the Experimental Agricultural Station (EAS) INTA Anguil (La Pampa State, Argentina) and an outpatient monitoring scenario. Future trends and concluding remarks are extended.

1. Introduction

The organizational knowledge management represents a key asset to support decision-making processes by different organizational stakeholders. The main aim of knowledge management systems is to manage, store and retrieve the organizational knowledge, so that it can be used later to learn, share knowledge, solve problems, and ultimately to support better decision-making processes. To ensure an efficient management of organizational knowledge, it is necessary to have technological platforms that support it. In the previous work, we proposed architecture based on data flow processing, for this purpose. Specifically, the Processing Architecture based on Measurement Metadata (PAbMM) [1, 2, 3].

Nowadays, there are processing architectures which allows the real-time data processing through configurable topologies such as Apache Storm [4, 5] and Spark [6]. In this type of architectures, you can dynamically define the processing topology over the data streams and adjust it to different computation necessities, being possible delegating the data structural definition and its meaning inside of the application. In this context, we summarize Processing Architecture based on Measurement Metadata [1, 2], which supported by the framework for measuring and evaluating called C-INCAMI (Context-Information Need, Concept model, Attribute, Metric and Indicator) [7], incorporates metadata to the measurement process, promoting repeatability, comparability and consistency. From the point of view of the semantic and formal support for measurement and evaluation (M&E), the conceptual framework C-INCAMI establishes an ontology that includes the needed concepts and relationships for specifying the data and metadata for any M&E project. Moreover, unlike other strategies for the processing of data streams [8, 9], with the addition of metadata, PAbMM is reliable for guiding the processing of the Measures from heterogeneous data sources, analysing each one in context of origin, as well as its significance within the proposed M&E in which defined.

[1] María de los Ángeles Martín, Street 9 and 110, martinma@ing.unlpam.edu.ar
[2] Mario Diván, Coronel Gil 353, 1st floor, mjdivan@[eco|ing].unlpam.edu.ar

The data streams are structured under C-INCAMI/MIS (Measurement Interchange Schema) [10], which allows gather data and metadata jointly inside the same stream. In this metadata, we can describe the measurement context additionally to the entity under measurement, which permits avoid analysing the measure in isolation way.

The PAbMM evolves the original strategy [1] incorporating support to the big data repositories in contexts of distributed computation. This implies the necessity of gather Big Data and Data Stream Processing technologies; which ensures powerful large data volumes processing, allowing efficient management of knowledge in the Organizational Memory.

The Organizational Memory that integrates the Processing Architecture based on Measurement Metadata serve as base for the organizational knowledge exchange and to be used in recommender systems in decision making processes.

Therefore, by having a well-developed organizational memory that supports the structuring, reusing and processing of organizational knowledge is a primary decision (and likely a success factor) to achieve such an effective management.

Nonaka and Takeuchi have said that an organization cannot create knowledge itself. Conversely, the knowledge creation basis for an organization is the individual's tacit knowledge; and tacit knowledge is shared through interpersonal interactions [11].

Therefore, in order to reach and maintain the organizational effectiveness and competitiveness, an organization needs to learn from past and present experiences and lessons learnt and to formalize organizational memories for enabling to make explicit the individual's tacit knowledge -and why not community's tacit knowledge as well.

One of the main goals of an organizational knowledge management strategy is to make explicit the individuals' implicit knowledge, to try to formalize the informal knowledge in order to allow machine-processable semantic inferences. A way of alleviating this problem from the knowledge representation standpoint is to store the knowledge in a more structured and formal way. We have followed this approach by using the case-based organizational memory strategy. It combines organizational knowledge storage technology with case-based reasoning (CBR) to represent each item of informal knowledge. In general, the organizational memories are intended to store the partial formal and informal knowledge present in an organization with automatic processing capabilities. In particular, by structuring an organizational memory in cases can also facilitate the automatic capture, recovery, transfer and reuse of knowledge for problem solving.

The main goal of this research is the integration of the Organizational Memory and case-based reasoning into the Processing Architecture based on Measurement Metadata as conceptual foundation for any organizational knowledge management. Also, the discussion about the added value of organizational memories; Thus, data, information, and knowledge from heterogeneous and distributed sources can be automatically and semantically processable by web-based applications, for instance, an 'intelligent' recommendation system to support a more effective decision-making process.

This article is organized in six sections. The Section 2 summarizes the conceptual framework C-INCAMI. The section 3 outlines the Processing Architecture based on Measurement Metadata. The Section 4 outlines the case based organizational memory. The Section 5 illustrates the application of the organizational memory to a practical case: a pasture predictor system using the data of the weather radar of INTA EEA Anguil. The Section 6 shows the application of the PAbMM Architecture and organizational memory to an outpatient monitoring and diagnosis case. The Section 7 discusses related work and finally section 8 summarizes the conclusions.

2. C-INCAMI Overview

C-INCAMI is a conceptual framework [12], which defines the concepts and their related components for the M&E area in software organizations. It is an approach in which the requirements specification, M&E, and analysis of results are performed for satisfying a specific information need in a given context. C-INCAMI is structured in the following main components: i) M&E project management, ii) Non-functional Requirements specification, iii) Context specification, iv) Design and implementation of measurement and v) Design and implementation of evaluation. Most components are supported by the ontological terms defined in [13].

The M&E project definition component defines and relates a set of project terms needed for dealing with M&E activities, methods, roles and artefacts.

The Non-functional requirements component allows specifying the Information need of any M&E project. The information need identifies the purpose and the user viewpoint; in turn, it focuses on a Calculable Concept and specifies the Entity Category to evaluate. A calculable concept can be defined as an abstract relationship between attributes of an entity and a given information need. This can be represented by a Concept Model where the leaves of an instantiated model are Attributes. Attributes can be measured by metrics.

Regarding measurement design, a Metric provides a Measurement specification of how quantifying a particular attribute of an entity, using a particular Method (i.e. procedure), and how to represent its values, using a particular Scale. The properties of the measured values in the scale with regard to the allowed mathematical and statistical operations and analysis are given by the scale Type. Two types of metrics are distinguished. Direct Metric is that for which values are obtained directly from measuring the corresponding entity's attribute, by using a Measurement Method. On the other hand, the Indirect Metric value is calculated from other direct metrics' values following a formula specification and a particular Calculation Method.

For measurement implementation, a Measurement specifies the task by using a particular metric description in order to produce a Measure value. Other associated metadata is the data collector name and the timestamp in which the measurement was performed. The Evaluation component includes the concepts and relationships intended to specify the evaluation design and implementation. It is worthy to mention that the selected metrics are useful for a measurement tasks as long as the selected

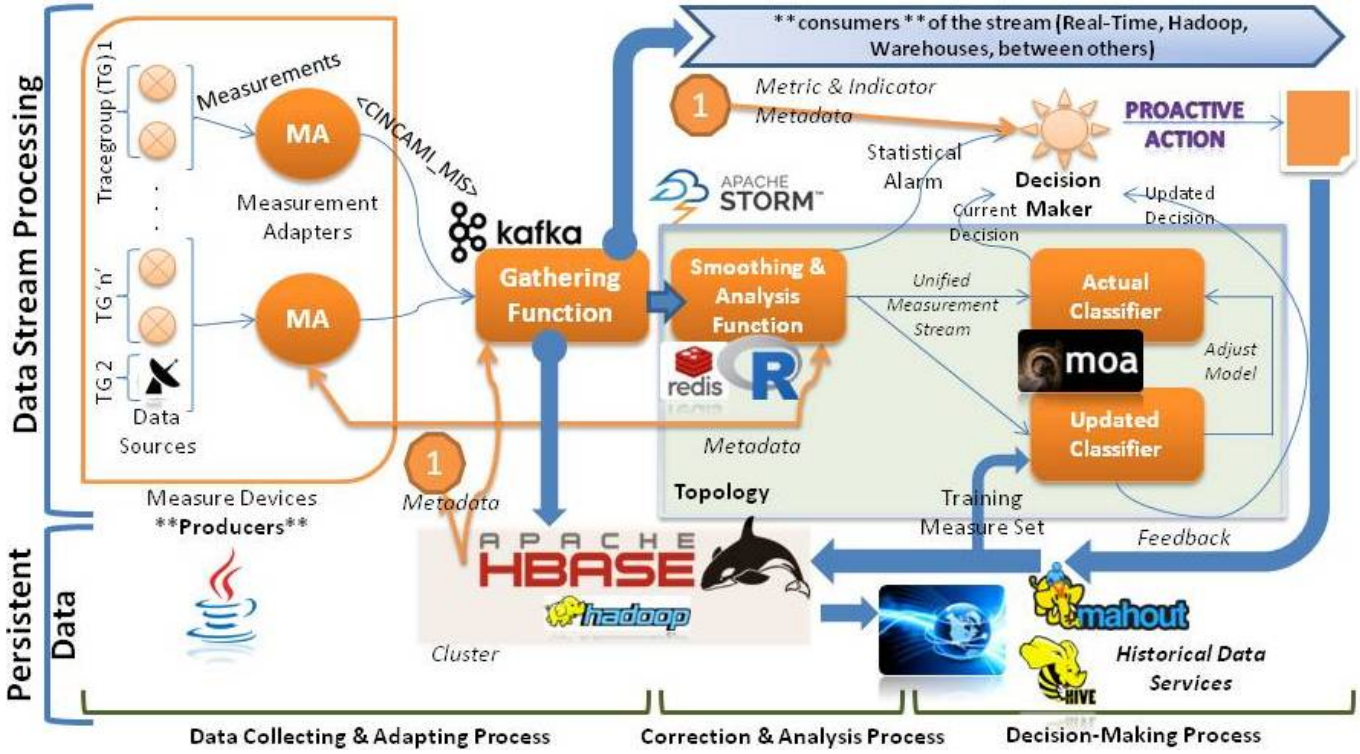

Figure 1. Extensions for the measurement component of C-INCAMI.

indicators are useful for an evaluation tasks in order to interpret the stated information need.

With the formalization of the project M&E based on C-INCAMI, the fact labelling process flow allows structuring the contents of a consistent and aligned with the project objective way. This structuring of the measurements within the PAbMM, maintains the concept that each measure is associated; for example, if a measure of contextual attribute or property.

Inside the stream and jointly with each measure associated with an attribute, the measures associated with the contextual properties are labeled. Thus, it enriches the statistical analysis because it is possible verify the formal and syntactic consistency of each measure against their formal definition, prior to moving forward with more complex statistical techniques.

3. The Processing Architecture Based on Measurement Metadata

The PAbMM is a manager of semi-structured measurements streams, enriched with metadata supported by C-INCAMI, specialized in M&E projects, which incorporates detective and predictive behaviour at online with the ability to manage and provide large volumes of data on demand.

As shown in Figure 1, the conceptual model in terms of stream processing it is as follows. The measurements are generated in the heterogeneous data sources (for example, The INTA weather radar [14]), which supply a module called Measurements Adapter (MA in Figure 1) generally embedded in measurement devices.

MA incorporates together with the measured values, the metadata of the M&E project and reports it to a central gathering function (Gathering Function -GF). GF incorporates the measurements streams in parallel in: a) The big data repository in persistent way, b) The C-INCAMI/MIS stream for the subscribers wishing to process information at the time when it is generated (for example, for INTA weather radar data, a consumer could be the National Meteorological Service), and c) Inside a buffer organized by monitoring groups -dynamic way of grouping data sources defined by the M&E project manager- in order to allow consistent statistical analysis at level of monitoring group or by geographic region where the data sources are located, without incurring in additional processing overhead. Additionally, GF incorporates load shedding techniques, which allows manage the queue services associated with the measurements, mitigating the risks of overflow regardless how they are grouped.

Thus, the C-INCAMI/MIS stream, is incorporated into the big data repository with measurements and metadata, and remains available to meet requests for services associated with data on historical measurement (Big Data Repository and The Historical Data Services in Figure 1). In addition to the measurement stream is sent to the subscribed consumers, a copy of this continues within the data stream processor and applies descriptive, correlation and principal components analysis (Analysis & Smoothing Function -ASF - in Figure 1) guided by their own

metadata, in order to detect inconsistent situations with respect to its formal definition, trends, correlations and / or identify system components that contribute most in terms of variability.

If a situation is detected in ASF, a statistical alert is triggered to the decision maker (DM) for evaluating whether or not to trigger external warning (via email, SMS, etc.) for reporting to the monitoring staff responsible of the situation (for example, in the case of the INTA WR, it could signal a possible device error or WR calibration). In parallel, the new measurements streams are reported to current classifier (Current Classifier -CC-), who must classify the new measurements whether or not correspond to a risk and report its decision to the DM. Simultaneously, the CC is reconstructed online, incorporating new measurements to the training set and producing a new model with them (Updated Classifier -UC).

The UC classifies the new measurements and produce a current decision which will also be communicated to the DM. The DM determines whether the decisions referred by classifiers (CC and UC) correspond to a risk and in which case, what probability of occurrence, act accordingly as defined in the minimum threshold of probability of occurrence defined by the project manager. Finally, regardless of the decisions taken, the UC turns into replacing the previous CC, just if an improvement in their ability for classifying is obtained according to the adjustment model based on ROC curves (Receiver Operating Characteristic) [15].

From a technological point of view, PAbMM evolves the solution proposed in the SDSPbMM, because now it is possible the management of the big data repositories in the distributed computation context, the data provision to thirds by subscription, and maintaining the real-time measurement processing in parallel (See Figure 1). Thus, we prioritize the use of the open source technology with the aim of promoting the extensibility, dynamism and broadcast of the architecture.

In this way and how you can see in the Figure 1, the data sources continue implementing the original interface of SDSPbMM called "*DataSource*", which establishes the responsibilities that a data source must satisfy to provide data in the architecture. Additionally, each MA in PAbMM (See Figure 1) becomes in a producer in terms of Apache Kafka [16, 17]. Thus, the data sent from the producers (for example, the INTA WR) will be processed by a subscription service inside of the processing cluster, under the concept called "Gathering Function" (GF). In this function, the data are collected from the data source, and the measures coming from the same entity under analysis are gathered and organized in a common buffer.

From the common buffer, we use Apache Kafka [17] to generate the unified measurement streams addressed to the consumers (See Figure 1), understanding by this to: a) The real-time subscribers (for example, Argentinian Air Force), b) The real-time processing topology which runs on Apache Storm [4], and c) The big data repository which store the streams in Apache HBase [18, 19] to support the historical data services.

Now, because the PAbMM runs on the Apache Storm and uses Apache Kafka, it is possible to consume and to process the measurement streams from the GF in a continuous way. The latter incorporates flexibility, scalability and dynamism in relation to the configuration of the processing topology, because both the Analysis and Smoothing Function (ASF) and the classifiers are *Bolts* [4], which may be reorganized in agile way and on demand. Moreover, inside the Storm topology, the PAbCMM continues using R [20] for the statistical computing (analysis and smoothing) and Redis [21] as NoSQL database for: *i) Cache Management, ii) Use of intermediate results from R, iii) Storing of snapshot with the last known state of each entity under monitoring.* Finally, the classifiers use the Hoeffding Tree from Massive Online Analysis (MOA) [22] to get a learning incremental strategy which allows to process big streams with limited computation resources.

At persistent level and by a side, PAbMM uses Apache Hadoop and HDFS (Hadoop Distributed File System) [23, 24] to promote the distributed computing, and Apache HBase [19] as columnar database which allows the monolithic scalability and random access to the measurements stored as C-INCAMI/MIS. From this repository, we use Apache Hive [25, 26] to support ad-hoc query on distributed contexts. Moreover, we use Apache Mahout [27, 28] to run classification and clustering analysis according to the needs, which allows one way of feedback the organizational memory.

The organizational memory is case-based and run over the Apache HBase. As the case-based reasoning engine uses MapReduce [24] programs in its recommendation subsystem, each case <fact,solution> is structured in the form of <key,value> to get the advantage from the distributed computing and the parallelism.

Finally and on the one hand, PAbMM is oriented to the distributed computing, scalability and extensibility over mature and open source technologies for supporting big data repositories; and on the other hand, it is possible broadcast the information by Apache Kafka while the stream engine runs over Apache Storm. Thus, PAbMM can be viewed as a simple topology, which allows us to manage the changes, make the adjustments/fixes that being necessaries, and make easier its interoperability.

In the case study that will be presented in section 5, where the architecture is applied for the processing of the data streams from a weather radar, our focus is in: a) Collect the measurements in structured and interoperable way by C-INCAMI/MIS, b) Store all the measures under a single repository, c) Use the stored measures to answer queries or to support a cluster/classification analysis, d) Provide data in real-time and historical on demand, e) Get and maintain an organization memory with the previous experience of the geographical region, and f) Prevent, or at least detect, the potential problematical situations from the documented experience.

4. Case Based Organizational Memory

Once the data (with the associated metadata), are incorporated from the data sources to persistent Big Data repository, it is desirable to structure them in an Organizational Memory, so that can later be exploited and used for recommendation during decision-making processes.

With the aim to manage and retrieve the organizational knowledge, in the last years numerous proposals of models and tools for knowledge management and knowledge representation have arisen. However, most of them store knowledge in a non-structured or semi-structured way, hindering the semantic and

automatic processing of this knowledge. In this section we specify a case-based organizational memory, so that it can be used to learn, reasoning, solve problems, and as support to better decision making as well.

A way of alleviating this problem from the knowledge representation standpoint is to store the knowledge in a more structured and formal way. We have followed this approach by using the case-based organizational memory strategy. It combines organizational knowledge storage technology with case-based reasoning (CBR) to represent each item of informal knowledge. In particular, by structuring an organizational memory in cases can also facilitate the automatic capture, recovery, transfer and reuse of knowledge for problem solving.

4.1. Case Based Organizational Memory Ontology Overview

Although the benefits of applying the knowledge management systems are well known, and the idea of applying case-based reasoning methods to lessons learned and best practices are not new in the knowledge representation area, there is almost no consensus yet on many of the concepts and terminology used in both knowledge management and case-based reasoning areas. In order to reach this aim we have constructed a common conceptualization for case-based organizational memory where concepts, attributes and their relationships should be explicitly specified; such an explicit specification of a conceptualization is one of the core steps for building ontology.

In the following sections, we will describe case-based organizational memory ontology and his application (using the architecture) to the construction of a pasture predictor system using the data of the wheather radar of INTA EEA Anguil.

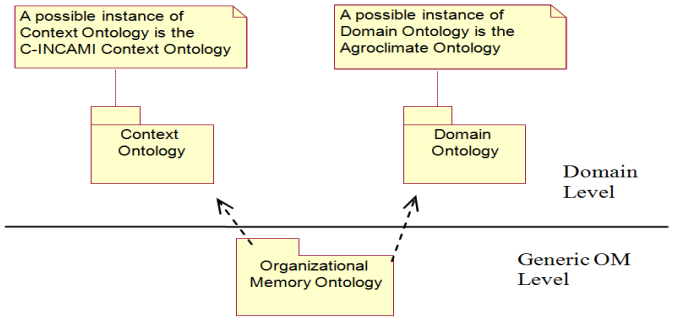

Figure 1. The relationship between ontologies at the specific domain level and at the generic organizational memory level components

The organizational memory ontology aims to be at a generic level from which other representations for specific domain applications can be formulated (see Figure 2). On the one hand, the case-based organizational memory ontology is defined at a generic organizational memory level, and on the other hand, for characterizing the cases according to the specific knowledge domain and its context [29], a domain and context ontologies should also be provided.

The objective of our ontology is to serve as a foundation for the organizational knowledge exchange with semantic power, which in turn facilitates the reuse, the interoperability and the automatic processing by agents [30].

The main concepts of the ontology, which are illustrated in the UML diagram of the Figure 3, are described in the following text, highlighted in italic.

An *organizational memory* is the way in which an organization stores and keeps track of what it knows, i.e., about the past, present and future knowledge. An organizational memory can have one or more *knowledge bases* which are intended to achieve different information needs of an organization –recalling that data, information and knowledge are useful assets for decision making. In addition, an organizational memory may be seen as a repository that stores and retrieves the whole specified, explicit, formal and informal knowledge present in an organization. Thus, a knowledge base is an organized body of related knowledge; taking into account that knowledge is a body of understanding and/or lessons learnt from skills and experiences that is constructed by people.

A type of knowledge base is a *case knowledge base* which stores the acquired knowledge in past experiences, good practices, learned lessons, heuristics, etc. to different domains; that is, it stores cases. A *case* is a contextualized *knowledge item* (i.e., an atomic piece of knowledge) representing an experience by means of a problem and its solution.

The representation of the knowledge through cases facilitates the reuse of the knowledge acquired in past problems to be applied to a new problem in similar situations [31].

A case can be seen as an ordered pair < P, S >, where P is the *problem* space, and S is the *solution* space.

There exists a general description of problems as $P(x_1, x_2, \ldots, x_n)$, where each individual problem is an instance $P(a_1, a_2, \ldots, a_n)$; also a general description of solutions as $S(y_1, y_2, \ldots, y_n)$, and every individual solution $S(b_1, b_2, \ldots, b_n)$ is an instance of that general description. The x_i are variables that characterize the problem (*problem feature*), and the y_i are variables that characterize the solution (*solution feature*), where both are features. A *feature* or attribute is a measurable physical or abstract property of an entity category. Since the stored cases refer to a specific knowledge domain, the features that characterize the problems and solutions are defined by a *domain concept* term; for example, the concepts coming from the meteorology domain ontology (i.e. *Precipitation Accumulation*, *Hail*, *Hail Damage, Environmental Temperature, Environmental Pressure* and so on, as we will illustrate in Section 5.2).

The case-based reasoning process consists in assigning values to problem variables and finding the adequate instances for solution variables. To find the appropriate values for the instances of a solution, the similarity assessment of cases should be performed, so that for each case knowledge base a *similarity assessment model* should be specified.

Greater detail of the functionality of the Organizational Memory can be found in [32], where processes have been formalized by the SPEM metamodel to promote its communicability and extensibility.

4.2. Similarity Assessment Model Representation

Most of the case-based reasoning applications have been focused on problems of specific domains. However, in order to be useful to an organization, a case-based reasoning system should be fitted in with the main knowledge sources that can stem from

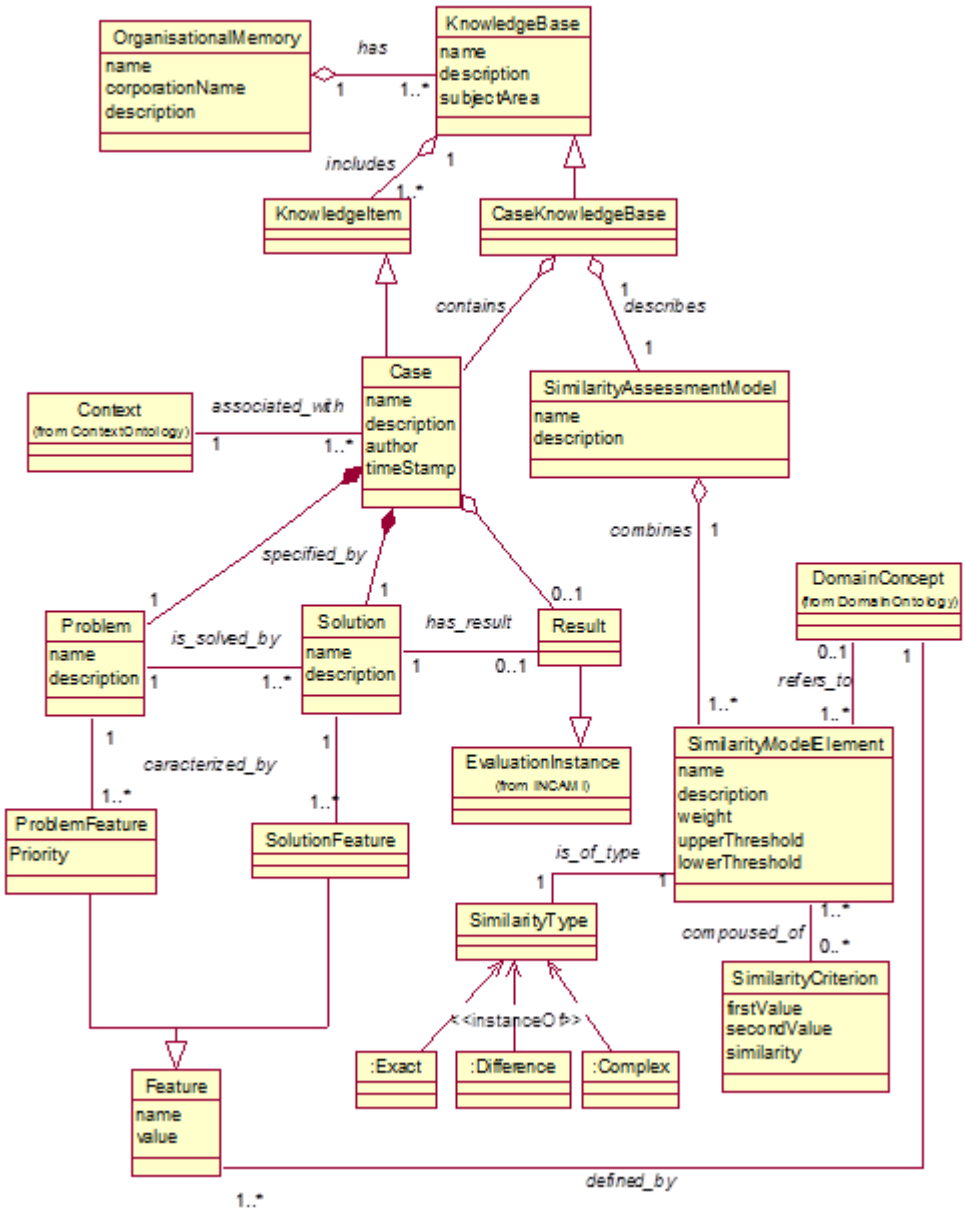

Figure 3. UML diagram that specifies the main terms, attributes and relationships to the Case-based Organizational Memory Ontology.

diverse domains, and so the similarity functions appropriate to each case knowledge base.

Usually, the similarity between a recovered case R and a new case C is defined as the sum of the similarities among its constituent feature values multiplied by their weights, i.e. the so-called Nearest Neighbor formula:

$$Similarity(R, C) = \sum_{f \in F} w_f * sim_f(f_R, f_C) \quad (1)$$

Where w_f is the weight of the feature f, and sim_f is the similarity measurements function to the feature. Therefore, in order to provide an appropriate representation of the similarity model, it is necessary to represent the weights that model the relative relevance, and the similarity function type for each specific feature. The weights are represented as an attribute inside each similarity element (see Figure 3), and the similarity type is restricted to be one out of three general types, namely: Exact, Difference and Complex.

The inclusion of these three types of functions is based on the analysis of numerous investigation works in the case-based

reasoning area, as well as taking into account they cover the similarity representation needs of most cases in the Software and Web Engineering area. Particularly:

- The Exact similarity function returns 1 if two feature values are the same and 0 otherwise.
- The Difference similarity function is inversely proportional to the difference between a feature values. It can only be applied when it is possible to define the value difference; for instance, between numerical values the difference similarity function returns 1 if both features are equals, and return $1/|f_c - f_n|$ in other case.
- The Complex similarity function solves the similarity for all those situations where the two previous functions are not applicable; for example, the semantic difference between two synonymous terms that is neither completely the same nor completely different. If the number of a feature values is finite, it is feasible to have beforehand the similarity measure values for all possible values' pairs. In our model, these parameters are represented in the Similarity Criterion class, which is defined as the assessment pattern used to determine the semantic similarities between two feature values.

Ultimately, an exhaustive glossary of terms, attributes and relationships are shown in [29], where the terminology for the case-based organizational memory ontology is explicitly described.

5. Practical Case 1: A Pasture Predictor

In order to illustrate the above main concepts, attributes and relationships, we will elaborate on an example of case-based knowledge base and its similarity assessment model for a specific domain: a pasture predictor system, using the data of the weather radar of INTA EEA Anguil. This case base stores a body of related knowledge about the growth of pasture based on a range of data, including current weather conditions and forecasts, rainfall events and past climate records, processed by the PAbMM architecture and taking the radar as a data source.

5.1. The Weather Radar of INTA EEA Anguil

Weather radars are active sensors of remote sensing that emit pulses of electromagnetic energy into the atmosphere in the range of microwave frequencies. These sensors are tools to monitor environmental variables, and specifically, the identification, analysis, monitoring, forecasting and evaluation of hydro meteorological phenomena, as well as physical processes that these involve, given the risk that can cause severe events. Its main applications are: a) Weather description, forecasting and nowcasting, b) Forecasting and monitoring of environmental contingencies (hail, torrential rain, severe storms, etc.), c) Security in navigation and air traffic control, d) Studies of atmospheric physics, e) studies of agro climatic risk, f) Provision of basic data for scientific and technological research, and g) Provision of input data for hydrological models (i.e. floods) [14]. The information recorded by the WR is collected through volumetric scans and today, the data are stored in separate files called volumes. The data contains the different variables: reflectivity factor (Z), differential reflectivity (ZDR), polarimetric correlation coefficient (RhoHV), differential phase (PhiDP), specific differential phase (KDP), radial velocity (V) and spectrum width (W).

Two types of data are distinguished: a) raw data and b) some level of data processing or "products". In both cases, the sampling units are 1 km^2 and 1º and each variable are stored in separate files.

Under normal operation, in a full day (00:00h to 23:50h), 8640 files are generated only for the range of 240 km just for one WR. In this sense and for each day, just the WR of the EAS Anguil produces daily a volume of 17GB of data, which represents about 6.2 Tb annually and just for one WR.

From raw data using the proprietary software Rainbow 5, different processing can be obtained, for example, some hydrological products estimating precipitation characteristics as SRI (Surface Rain Intensity), which generates values intensity or PAC (Precipitation Accumulation), which calculates a cumulative rain in a predefined time interval. These products can be formatted in XML or raster image. Also, INTA developed software that can generate more products from raw radar data, for example applications of models to estimate occurrence of hail and hail damage to crops [14].

The users of radar data and radar products, with free and open access are: i) National System of Weather Radar (SiNaRaMe), ii) SMN, iii) Sub secretary of Nation Water Resources (SSRH), iv) National Water Institute (INA), v) Civil Defence vi) Argentinean Air Force, vii) Commercial and General Aviation, viii) Directorate of Agriculture and Climate Contingencies (DACC, Mendoza) ix) Agro climatic Risk Office (ORA), xi) Universities, xii) Research Groups and related product development, xiii) Insurance Companies, xiv) Media, xv) INTA.

In this case, PAbMM define a M&E project which is useful between others things, for detecting decalibration in the WR in each data stream processed. So, and in terms of PAbMM, each WR is a heterogeneous data source that uses a MA. The architecture store the C-INCAMI streams from each WR, provides by subscription the data and allows the monitoring of the stream to make decisions in real time. In this way and from the processed data, you can build knowledge management systems for supporting the decision making in the agricultural production domain, based in the experience stored through the Organizational Memory.

5.2. Knowledge Base for Pasture Production

This case base stores a body of related knowledge about pasture growth in relation to current, past and future weather conditions, so that it serves as the basis to a recommendation system that support the new cycle of pasture production regarding similar past ones.

This knowledge base, collect heterogeneous data from different sources (One source of data is the Weather radar of INTA EEA), as well as manual measurements made by agricultural producers, for example the estimated Daily production of pasture.

The aims of the knowledge base are to support productivity, efficiency and continuing growth in this important industry.

It saves past knowledge about weather conditions (as problems) and their pasture production (as solutions).

By providing 60-day forecasts for the growth of pasture, the aim to help farmers make better decisions in managing their herds, production and costs.

To illustrate the knowledge base and for easier understanding, a simplified model of the Case structure is shown. This case knowledge base characterizes the problem situation through various characteristics including current weather conditions 7-day forecasts, rainfall and hail events, and past climate records. Also the context data of location and time is taken into account. In this example, the case is characterized through four features, namely: Precipitation Accumulation, Hail, Hail Damage, and Rain Forecast that are defined in the meteorology domain ontology.

Table I Example of similarity assessment model for a case base

Feature	Description	Type	Wht
Precipitation Accumulation	Accumulative rain in Last 10 days	Difference	0.40
Hail	Indicates the occurrence of hail in Last 10 days (Possible values are *yes* or *no*)	Exact	0.15
HailDamage	Indicates the occurrence of hail damage in Last 10 days (Possible values are *yes* or *no*)	Exact	0.25
RainForecast	Indicates the next 7 days rain forecast	Difference	0.20

Analogously, the solution will be characterized by the PastureProduction feature; which indicates the daily kg of dry matter produced in one hectare.

For each feature that characterizes a case, we should establish its weight and its similarity function type (see Table 1). These design decisions could be made by an expert taking into account which features are considered more relevant from the similarity point of view to evaluate in the end the global similarity of two cases.

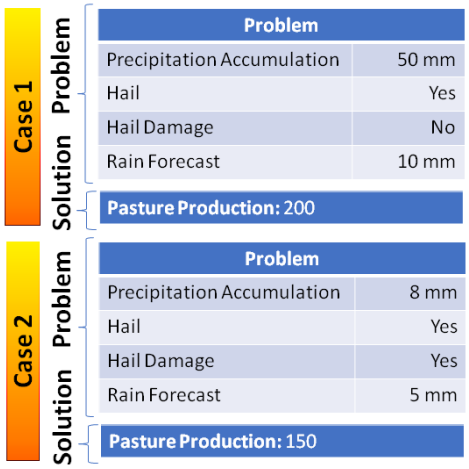

Figure 2. Example of representation of two stored past cases

Once defined the case structure and its similarity assessment model, each case is stored with all the feature values that characterize it and its solution accordingly. Two case examples are shown in Figure 4.

A new decision in herds managing can benefit from the case-based organizational memory by recovering the pasture prediction information of the most similar environmental conditions. Let us suppose that we want to buy animals to fatten and need to know if we have enough grass, and that the case base stores the two cases shown in Figure 4, among others.

In order to minimize risks, we can take advantage of the recorded knowledge by retrieving and reusing the most similar past experience. Therefore, table 2 shows the similarity calculation of each feature of the new case compared to the previous past ones, i.e., "Case 1" and "Case 2". Hence the global similarity calculations give as outcomes:

$$Similarity(Case\ 1, New) = 0.4 * 0.03 + 0.15 * 1 + 0.25 * 1 + 0.2 * 0.2 = \mathbf{0.453}$$

$$Similarity(Case\ 2, New) = 0.4 * 0.083 + 0.15 * 1 + 0.25 * 0 + 0.2 * 0.1 = \mathbf{0.203}$$

Table II. Example of Similarity Assessment between the previous two past cases and the new one

Feature	Case 1	Case 2	New Case	Similarity Case1/New	Similarity Case2/New
Precipitation Accumulation	50	8	20	$\frac{1}{\lvert 50-20\rvert} = 0.03$	$\frac{1}{\lvert 8-20\rvert} = 0.083$
Hail	Yes	Yes	Yes	1.00	1.00
HailDamage	No	Yes	No	1.00	0.00
RainForecast	10	5	15	$\frac{1}{\lvert 10-15\rvert} = 0.2$	$\frac{1}{\lvert 5-15\rvert} = 0.1$

Resulting "Case 1" as the most similar to the new case, and therefore the pasture prediction is 200 daily kg of dry matter produced.

6. Practical Case 2: Outpatient Monitoring and Diagnosis

In this section we show another use of PAbMM and Case Based Organizational Memory oriented to an outpatients monitoring scenario.

Considering a healthcare centre, the doctors need monitor the outpatients for avoiding adverse reactions or major damage in general. In this sense, PAbMM bring to doctors of a mechanism by which they can be informed about unexpected variations and/or inconsistencies in health indicators defined by them (as experts). So, the core idea is: there exists some proactive mechanism based on health metrics and indicators that produces an alarm for each risk situation associated to the outpatient under monitoring.

Considering C-INCAMI, the information need is "to monitor the principal vital signs of an outpatient when he/she is given the

medical discharge from the healthcare centre". The entity under analysis is the outpatient.

According to medical experts, the corporal temperature, the systolic arterial pressure (maximum), the diastolic arterial pressure (minimum) and the cardiac frequency represent the relevant attributes of the outpatient vital signs to monitor. They also consider as necessary monitoring the environmental temperature, environmental pressure, humidity, and patient position (i.e. latitude and longitude) contextual properties.

Indicator Code	ETEMP	
Indicator Name	Level of corporal temperature	
Domain		
(hyperpyrexia, very high fever, fever, normal, hypothermia risk, hypothermia)		
Scale		
Type of Scale	Categorical	
The Value Domain	Ordinal	
Unit	-	
Calculation Method (Elementary Model)		
Name	Temperature Analysis	
Specification	Axillary Temperature (Direct Metric)	Indicator Value
	>= 41.50	Hyperpyrexia
	[38.30 , 41.50)	Very high fever
	[37.50 , 38.30)	Fever
	[36.50 , 37.50)	Normal
	[35.00 , 36.50)	Hypothermia risk
	< 35.00	Hypothermia (While the temperature is descending there is risk of plain encephalogram).
References	1. Loscalzo, J., Fauci, A., Braunwald, E., Dennis L., Hauser, S., Longo, D. (2008). "Harrison's principles of internal medicine". McGraw-Hill Medical. pp. Chapter 17, Fever versus hyperthermia. ISBN 0-07-146633-9.	
	2. Marx, John (2006). "Rosen's emergency medicine: concepts and clinical practice". Mosby/Elsevier. p. 2239. ISBN 9780323028455.	
Decision Criteria		
Indicator Value	Level of Acceptability	
Hyperpyrexia	Not Acceptable	
Very high fever	Not Acceptable	
Fever	Poorly Acceptable	
Normal	Acceptable	
Hypothermia risk	Poorly Acceptable	
Hypothermia	Not Acceptable	

Figure 3. Details of the Level of Corporal Temperature elementary indicator specification

The definition of the information need, the entity, its associated attributes and the context are part of the "*Non-functional requirements specification*" and "*Context specification*" components as discussed in section 2. For monitoring purposes, the metrics that quantify the cited attributes, were selected from the C-INCAMI DB repository (Implemented by Apache HBase in figure 1); likewise, the metrics that quantify the cited contextual properties.

After the set of metrics and contextual properties for outpatient monitoring has been selected, the corresponding elementary indicators for interpretation purposes have also to be selected by experts.

In this way, they have included the following elementary indicators: the level of corporal temperature, the level of pressure, the level of cardiac frequency and the level of difference between the corporal and the environmental temperature. The concepts related to indicators are part of the Evaluation Design and Implementation component. Figure 5 shows the specification of the level of corporal temperature elementary indicator; for example, the different acceptability levels with their interpretations are shown, among other metadata.

Besides, considering that ranges of the acceptability levels (shown in Figure 5) are in an ordinal scale type, then the target variable for the stream mining function (classification) is feasible. So, both classifiers, CC and UC, act relying on the values of the given indicators and their acceptability levels.

6.1. Implementation Issues

Once all the above project information was established, it is necessary for implementation issues to choose a concrete architecture to deploy it. Fig. 6 depicts an abridged deployment view for the outpatient monitoring system considering the PAbMM approach.

Let's suppose we install and set up the MA in a mobile device –the outpatient device-, which will work in conjunction with sensors as shown in Fig. 6. Therefore, while the data collecting and adapting processes are implemented in a mobile device by the MA, the gathering function and other processes can reside in the medical centre computer. The MA component informs the measures (streams) to the gathering function (GF) in an asynchronous and continuous way. MA takes the measures from sensors (in fact, the data sources) and incorporates the associated metric metadata for attributes and contextual properties accordingly. For instance, it incorporates the contextual property ID for the environmental temperature joint to the value to transmit; and so, for every attribute and contextual property. Note that data (values) and metadata are transmitted through the C-INCAMI/MIS schema to the gathering function (GF inside of PAbMM in figure 6), as discussed in section 3.

Although all the values of metrics and contextual properties from monitored outpatients are simultaneously received and analyzed, let's consider for a while, for illustration purpose, that the system only receives data for the axillary temperature attribute and the environmental temperature contextual property from one outpatient, and that also the system visualizes them.

The measures and, ultimately, the acceptability level achieved by the level of corporal temperature elementary indicator, indicate a normal situation for the patient.

Nevertheless, the online decision-making process (Inside of PAbMM, See Section 3), apart from analyzing the level of acceptability met for attributes; also it analyzes the interaction with contextual properties and their values.

At first glance, what seemed to be normal and evident, it was probably not because the processing model can detect in a proactive form a correlation between axillary temperature and environmental temperature. This could cause the triggering of a preventive alarm from the medical centre to doctors, because the increment on the environmental temperature can drag in turn the increment in the corporal temperature, and therefore this situation can be associated to a gradual raise in the risk likelihood for the outpatient.

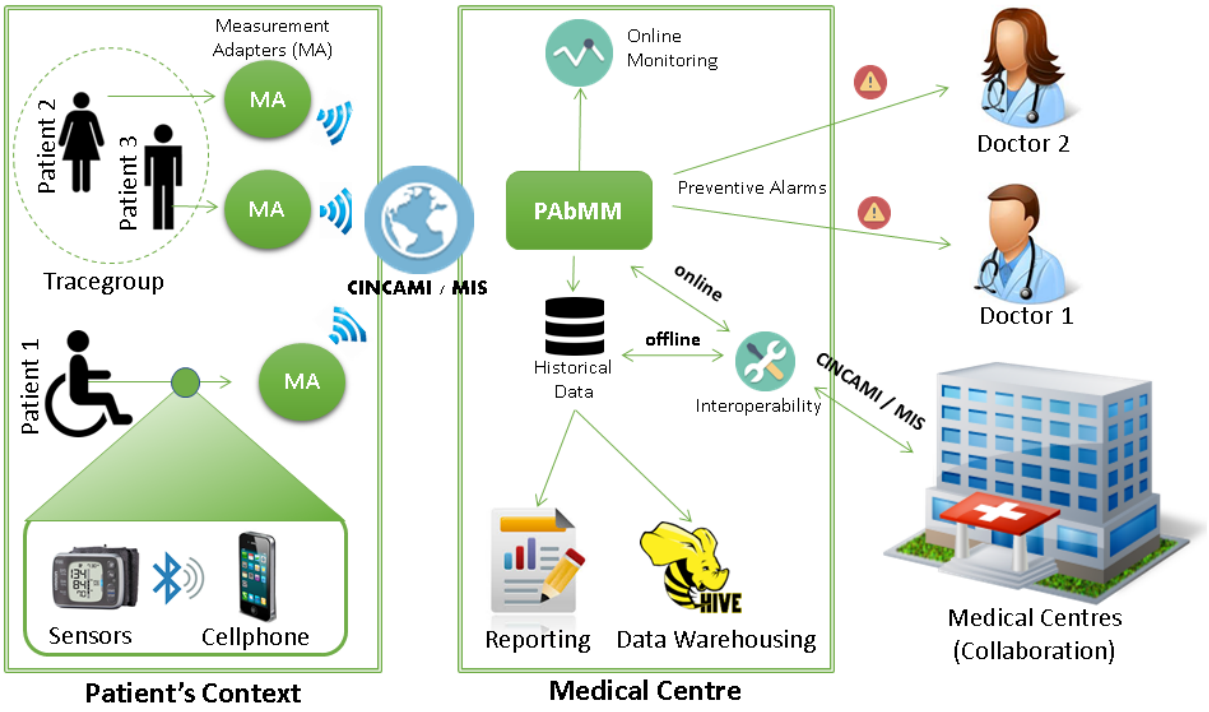

Figure 6. A deployment view for the Outpatient Monitoring System.

6.2. Diagnosis System

Despite of outpatient monitoring task, using the knowledge base of the organizational memory, and the outpatient monitoring data streams, is possible recommend disease diagnosis for a better decision-making. In a similar way to pasture prediction presented in section 5.2, for designing the diagnostic system based on the MOBC, the Organizational Memory must be customized, defining the domain ontology according to the type of knowledge to be stored (See Figure 2).

Case 1

Problem

Problem	
Corporal temperature	38.9°
Systolic arterial pressure	120 mmHg
Diastolic arterial pressure	80 mmHg
Cardiac frequency	95 bpm

Solution

Diagnosis: Influenza **Treatment:**	
Drug	Amoxicillin
Concentration	500 mg
Dose	1 capsule
Daily Frequency	4

Figure 7. Example of a diagnostic case stored in the MOBC.

A problem (possible illness) can be characterized by its symptoms. In order to simplify our example, and for this study, the complete ontology was not developed. We will consider that corporal temperature, the systolic arterial pressure (maximum), the diastolic arterial pressure (minimum) and the cardiac frequency represent the relevant attributes that characterize the problem in the patient, but this list must be completed and adapted to each health institution, according to your needs. Each symptomatic table may have a diagnosis associated with it, and each diagnosis may have different treatments.

Figure 7 shows an example of a case that represents the diagnosis "Influenza", and its respective solution (or corrective action) based on the conceptual model of Figure 2.

For illustrating how the PAbMM classifier can be trained in the arrival of new measures from the monitoring, an example is shown in which the MO stores among its data, two cases: the Influenza case (See Figure 7) and the hypertension case, (See Figure 8). In the case of new data, the classifier will generate a new case (data + metadata using C-INCAMI / MIS) and try to establish an alarm (if applicable) interacting with the MO through the recommendation function.

Measures reported from PAbMM to MO suggest a temperature of 36.9°, a diastolic arterial pressure of 110 mmHg, a systolic arterial pressure of 200 mmHg and a heart rate of 98 bpm. The PAbMM indicates an alarm because it has detected a progressive increase in the pressure, but he does not know how to proceed in terms of medical treatment, and it resorts to organizational memory. The CBR engine of the MO will look for a more similar case. For our example we apply the similarity function Difference, which values the similarity equal to 1 if both characteristics are equal and $1/|f_c - f_n|$ in other case.

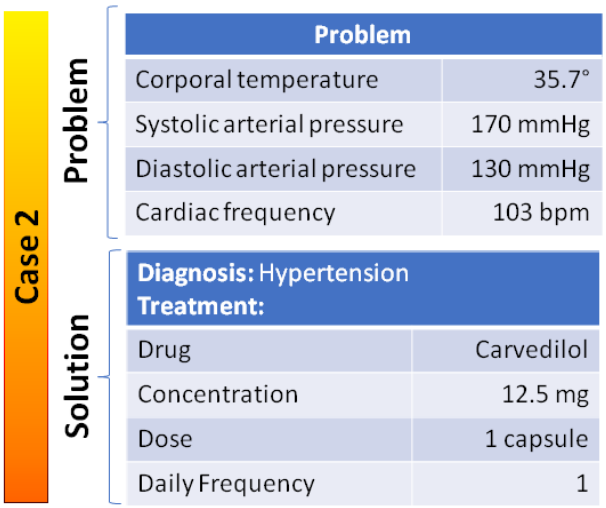

Figure 8. Example 2 of a diagnostic case stored in the MOBC.

Table 3. Medical Recommendation. Example of Similarity Assessment between the previous two past cases and the new one

Feature	Case 1	Case 2	New Case	Similarity Case1/New	Similarity Case2/New
Corporal temperature	38.9	35.7	36.9	0.5	0.83
Systolic corporal pressure	120	170	200	0.0125	0.03
Diastolic arterial pressure	80	130	110	0.033	0.05
Cardiac frequency	95	103	98	0.33	0.20

We also apply equal weight (0.25) to each of the four characteristics. In order to calculate the overall similarity of each pre-existing case with respect to the new case (NC), the formula (1) in section 4.2 is applied, replacing the weights and similarity values for each characteristic (See Figures 7 and 8). As set out below:

$$Similarity(Case\ 1, New) = 0.25 * 0.50 + 0.25 * 0.0125 + 0.25 * 0.033 + 0.25 * 0.33 = \mathbf{0.219}$$

$$Similarity(Case\ 2, New) = 0.25 * 0.83 + 0.25 * 0.033 + 0.25 * 0.05 + 0.25 * 0.2 = \mathbf{0.279}$$

In this way, the most similar case to the new one is the hypertension, which will allow triggering together with the alarm, the recommended treatment (Carvedilol concentration 12.5, 1 capsule per day) for the detected situation.

7. Related Work

In our specific work we have illustrated the use of the case-based reasoning approach to develop a case-based organizational memory; using the advantage of having the PAbMM architecture, which manage large volumes of structured data together with their metadata. The main goal is to exploit the knowledge that con be extracted from organizational memory under a key-value structure (i.e. case-solution structure) stored in the Big Data repository, allowing to incorporate more experience for recommending the courses of actions to the decision-making process.

Firstly, from the organizational memory point of view, there are numerous proposals in the knowledge management area, for example the ones documented in [33, 34]. Most of them capture and store the knowledge in repositories of documents like manuals, memos, and text files systems, etc. where structured or semi-structured storage strategies are seldom used. These approaches usually do not employ powerful mechanisms of semantic and automatic knowledge processing based on ontologies therefore causing very often loss of time and high investment in human resources.

Secondly, from the stream processing architecture point of view, there are recent works which make focus on the data stream processing from a syntactic point of view [8, 9]. In this context, the data model of the stream is based in a key-value structure and incorporates techniques for the adaptive management of high-rate streams [35]. Our architecture incorporates metadata based on a M&E framework, which allows to guide the organization of measures (for example, through snapshots in memory and the last known state of the entities under analysis), facilitating consistent and comparable analysis from the statistical point of view, being able to fire alarms based in the interpretation of the decision criteria of the indicators whose value was got from the data.

8. Conclusions

The organizational knowledge management represents a key asset to support a more effective decision-making process by different stakeholders. In this direction, by having an IT-based organizational memory that supports the structuring, reusing and processing of organizational knowledge is a primary decision to achieve that effective management.

In the previous work [3], we had specified a case-based organizational memory for the stream processing architecture based on measurement metadata. In this paper we deepen the former research and considerably expand the development of the proposed technologies. As a result, we obtained a robust and mature platform for the processing of heterogeneous sources data streams and the management of organizational knowledge.

The knowledge representation through cases facilitates the reuse of knowledge acquired in past problems to be applied to a new problem in similar situations, in addition facilitates the automatic knowledge processing as well.

In this way, the proposed case-based organizational memory can benefit from processing power of the PAbMM architecture, which is supported by distributed Big Data technologies. Also, the PAbMM and the importance about how the measures can be informed jointly with the metadata by C-INCAMI/MIS has been shown. The metadata help guide the processing strategy through the definition of the M&E project and it's possible thanks to the

C-INCAMI framework, which establishes the concepts and relations necessaries to support any M&E project.

Finally, we have illustrated these models and approach with two practical cases. On the one hand a pasture predictor system, using the data of the weather radar of INTA EEA Anguil. This case base stores a body of related knowledge about the growth of pasture based on a range of data, including current weather conditions and forecasts, rainfall events and past climate records, processed by the PAbMM architecture and taking the radar as a data source. On the other side the outpatient monitoring and diagnosis case, shows the processing power of the proposed PAbMM Architecture.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research is supported by the project 09/F068 of the Engineering School and partially through the technical cooperation contract between the Engineering School and the EAS INTA Anguil (Argentina).

References

[1] M Martín and M Diván, "Case Based Organizational Memory for Processing Architecture based on Measurement Metadata," in 5th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), Noida, India, 2016.

[2] M Diván and L Olsina, "Process View for a Data Stream Processing Strategy based on Measurement Metadata," Electronic Journal of Informatics and Operations Research, vol. 13, no. 1, pp. 16-34, June 2014.

[3] M Diván et al., "Towards a Data Processing Architecture for the Weather Radar of the INTA Angui," in International Workshop on Data Mining with Industrial Applications, Asunción, Paraguay, 2015.

[4] A. Jain and A. Nalya, Learning Storm. Create real-time stream processing applications with Apache Storm. Birmingham, United Kingdom: Packt Publishing Ltd., 2014.

[5] Apache Software Foundation. Apache Storm. [Online]. http://storm.apache.org/index.html. Last acces: march 28 of 2017.

[6] M. Frampton, Mastering Apache Spark. Birmingham, United Kingdom: Packt Publishing Ltd, 2015.

[7] F. Papa, H. Molina and L. Olsina, "How to Measure and Evaluate Web Applications in a Consistent Way" in Web Engineering: Springer, 2007, ch. 13, pp. 385–420.

[8] M. Ali, W. Aref, R. Bose, A. Elmagarmid, A. Helal, I. Kamel & M. Mokbel, "NILE-PDT: A Phenomenon Detection and Tracking Framework for Data Stream Management Systems," in VLDB, Trondheim, Norway, 2005, pp. 1295-1298.

[9] D. Abadi, Y. Ahmad, M. Balazinska, U. Cetintemel, M. Cherniack, J. Hwang, W. Lindner, A. Maskey, A. Rasin, E. Ryvkina, N. Tatbul, Y. Xing & S. Zdonik, "The Design of the Borealis Stream Processing Engine," in Conference on Innovative Data Systems Research (CIDR), Asilomar, CA, 2005, pp. 277-289.

[10] M. Diván, "Strategy for Data Stream Processing based on Measurement Metadata", PhD Thesis, Computer Science School. National University of de La Plata (Argentina), 2011.

[11] I. Takeuchi and H. Nonaka, The Knowledge-Creating Company. New York: Oxford University Press Inc., 1995.

[12] H. Molina and L. Olsina, "Towards the Support of Contextual Information to a Measurement and Evaluation Framework," in QUATIC, Lisboa, Portugal, 2007.

[13] L. Olsina and M. Martín, "Ontology for Software Metrics and Indicators," Journal of Web Engineering (JWE), vol. 3, no. 4, pp. 262-281, 2004.

[14] Y. Bellini Saibene, M. Volpaccio, S. Banchero, and R. Mezher, "Development and use of free tools for data exploitation of the INTA Weather Radars (in spanish)," in 43º Jornadas Argentinas de Informática - 6º Congreso Argentino de Agroinformática, Buenos Aires, 2014, pp. 74-86.

[15] C Marrocco, R Duin, and F. Tortorella, "Maximizing the area under the ROC curve by pairwise feature combination," ACM Pattern Recognition, pp. 1961-1974, 2008.

[16] Apache Software Foundation. Apache Kafka. [Online]. http://kafka.apache.org/. Last access: march 28 of 2017.

[17] N. Garg, Apache Kafka. Set up Apache Kafka clusters and develop custom message producers and consumers using practical, hands-on examples. Birmingham, United Kingdom: Packt Publishing Ltd., 2013.

[18] Apache Software Foundation. Apache HBase. [Online]. http://hbase.apache.org/. Last access: march 28 of 2017.

[19] J Spaggiari and K O'Dell, Architecting HBase Applications (Early Release). New York, USA: O'Reilly Media, 2015.

[20] R Core Team, R: A Language and Environment for Statistical Computing. Vienna, Austria: The R Foundation for Statistical Computing, 2016.

[21] M Da Silva and H Tavares, Redis Essentials. Birmingham, United Kingdom: Packt Publishing, 2015.

[22] A. Bifet, G. Holmes, R. Kirkby, and B. Pfahringer, "MOA: Massive Online Analysis," Journal of Machine Learning Research, vol. XI, pp. 1601-1604, 2010.

[23] Apache Software Foundation. Apache Hadoop. [Online]. http://hadoop.apache.org/. Last access: march 28 of 2017.

[24] A Holmes, Hadoop in Practice, 2nd ed. New York, USA: Manning, 2015.

[25] Apache Software Foundation. Apache Hive TM. [Online]. https://hive.apache.org. Last access: march 28 of 2017.

[26] J Rutherglen, D Wampler, and E Capriolo, Programming Hive. California: O'Reilly Media Inc., 2012.

[27] Apache Software Foundation. Apache Mahout: Scalable machine learning and data mining. [Online]. http://mahout.apache.org/. Last access: march 28 of 2017.

[28] A Gupta, Learning Apache Mahout Classification. Birmingham, United Kingdom: Packt Publishing, 2015.

[29] M Martín, "Organizational Memory Based on Ontology and Cases for Recommendation System", PhD Thesis, Computer Science School, National University of de La Plata (Argentina)., 2010.

[30] M. Martín and L. Olsina, "Added Value of Ontologies for Modeling an Organizational Memory," in Building Organizational Memories: Will You Know What You Knew ?: IGI Global, pp. 127-147.

[31] H Chen and Z Wu, "On Case-Based Knowledge Sharing in Semantic Web," in XV International Conference on Tools with Artificial Intelligence, California, 2003, pp. 200-207.

[32] M Diván and M Martín, "Strategy of Data Stream Processing sustained by Big Data and Organizational Memory (In Spanish)," in In proc. of National Conference of Informatics Engineering/ Information Systems Engineering, Buenos Aires, Argentina, 2015.

[33] W Karl-Heinz, "A Case Based Reasoning Approach for Answer Reranking in Question Answering," CoRR, vol. abs/1503.02917, 2015. [Online]. http://arxiv.org/abs/1503.02917

[34] M Lee, M Lee, S Hur, and I Kim, "Load Adaptive and Fault Tolerant Distributed Stream Processing System for Explosive Stream Data", Transactions on Advanced Communications Technology, vol. 5, no. 1, pp. 745-751, 2016.

[35] J Samosir, M Indrawan-Santiago, and P Haghighi, "An evaluation of data stream processing systems for data driven applications," Procedia Computer Science, vol. 80, pp. 439-449, June 2016.