RESEARCH ARTICLE OPEN ACCESS

Using a Case Based Organizational Memory for Recommendation Systems

Martínez Natalí*, Martín María de los Ángeles**
*(Engineering School National University of La Pampa, Argentina.
** (Engineering School National University of La Pampa, Argentina.
Corresponding Author; Martínez Natalí

ABSTRACT
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 paper we present a summary of an 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. In order to illustrate its utility a practical case using the weather radar (WR) data of the Experimental Agricultural Station (EAS) INTA (La Pampa State, Argentina) is shown. Also related work and concluding remarks are outlined.
***Keywords* -** Organizational Memory, case-based reasoning, Ontology.

Date of Submission: 20-12-2018 Date of Acceptance: 04-01-2019

I. INTRODUCTION

The organizational knowledge management represents a key asset to support decision-making processes by different organizational stakeholders [1]. 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) [2, 3].

The PAbMM evolves the original strategy [2] 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 [4].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 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 five sections. The Section 2 outlines the case based organizational memory. The Section 3 illustrates the application of the organizational memory to a practical case: a pasture predictor system using the data of the weather radar of EAS INTA Anguil. The Section 4 discusses related work and finally the Section 5 summarizes the conclusions.

II. CASE BASED ORGANIZATIONAL MEMORY

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.

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 weather radar of INTA EEA Anguil.

2.2 Case Based Organizational Memory Ontology Overview

The organizational memory ontology aims to be at a generic level from which other representations for specific domain applications can be formulated (see Figure 1). 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 [5], a domain and context ontologies should also be provided.

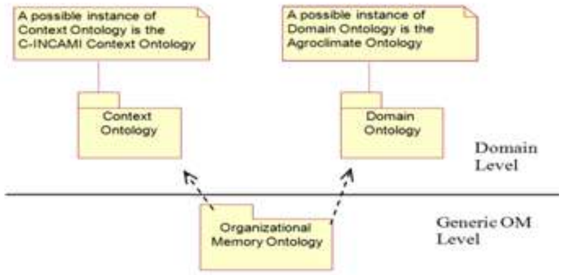

Figure 1. The relationship between ontologies at the specific domain level and at the generic organizational memory level components

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 [6].

The main concepts of the ontology, which are illustrated in the UML diagram of the Figure 2, 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 [7].

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(x1, x2, ... , xn), where each individual problem is an instance P(a1, a2,..., an); also a general description of solutions as S(y1, y2, ... , yn), and every individual solution S(b1, b2, ... , bn) is an instance of that general description. The xi are variables that characterize the problem (problem feature), and the yi 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 3.2).

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 [8], where processes have been formalized by the SPEM metamodel to promote its communicability and extensibility.

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

As we can see in Figure 2, each case knowledge base has a similarity assessment model that specifies it. On the one hand, the similarity assessment model class represents the general description of problems as P(x1, x2 ... xn), i.e. the problem structure, by combining several similarity model elements -one per each problem feature xi.

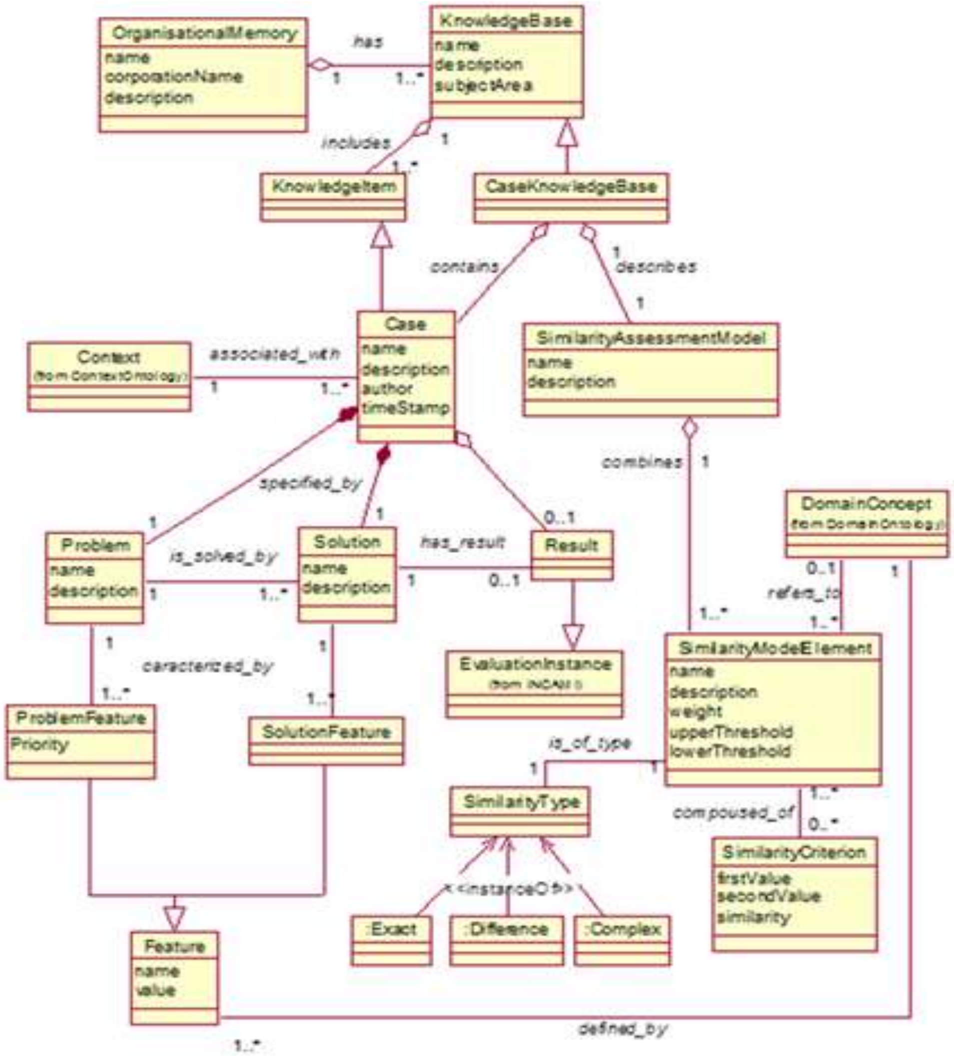

Figure 2. UML diagram that specifies the main terms, attributes and relationships to the Case-based Organizational Memory Ontology.

On the other hand, it is a function with associated similarity elements that models the similarity assessment of cases. In turn the similarity model element is a function with associated similarity criterion that models the similarity assessment of a feature. We propose a model to define the case structure indicating the features that characterize it and the possible similarity models.

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 [5], where the terminology for the case-based organizational memory ontology is explicitly described.

III. A PRACTICAL CASE: 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.

2.2 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) [9]. 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 km2 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 [9].

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, the PAbMM architecture is used to process the data generated by the weather radar. This 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.

2.3 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: PrecipitationAccumulation, Hail, HailDamage, and RainForecast that are defined in the meteorology domain ontology.

TABLE 1
EXAMPLE OF SIMILARITY ASSESSMENT MODEL FOR A CASE BASE

Feature	Description	Type	Wht
Precipitation Accumulatio n	Accumulative rain in Last 10 days	Differ.	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
RainForecas t	Indicates the next 7 days rain forecast	Differ.	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.

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 3.

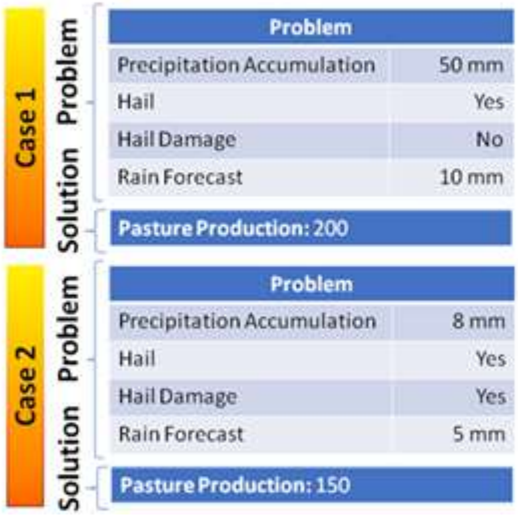

Figure 3. Example of representation of two stored past cases

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 3, 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:

TABLE I. 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	$\frac{1}{0-20|} = 0.03$	$\frac{1}{8-20|} = 0.$
Hail	Yes	Yes	Yes	1.00	1.00
HailDamage	No	Yes	No	1.00	0.00
RainForecast	10	5	15	$\frac{1}{0-15|} = 0.$	$\frac{1}{|5-15|} =$

Similarity(Case 1,New) =0.4*0.03+0.15*1+ 0.25*1+0.2*0.2 = **0.453**
Similarity(Case 2,New) =0.4*0.083+0.15*1+ 0.25*0+0.2*0.1 **= 0.203**

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

IV. 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.

From the organizational memory point of view, there are numerous proposals in the knowledge management area, for example the ones documented in [10,11]. 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.

V. 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 [1], 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.

Finally, we have illustrated these models and approach with a practical case: 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.

REFERENCES

[1]. M. Martín and M. Diván, Case Based Organizational Memory for Processing Architecture based on Measurement Metadata. Proc. 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 Anguil, Proc. International Workshop on Data Mining with Industrial Applications, Asunción, Paraguay, 2015.

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

[5]. 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.

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

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

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

[9]. 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), Proc. 43º Jornadas Argentinas de Informática - 6º Congreso Argentino de Agroinformática, Buenos Aires, 2014, pp. 74-86.

[10]. 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

[11]. 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.

Martínez Natalí" Using a Case Based Organizational Memory for Recommendation Systems' International Journal of Engineering Research and Applications (IJERA) , vol. 8, no.12, 2018, pp 71-78