Modeling a Knowledge-Based System for Cyber-physical Systems: Applications in the Context of Learning Analytics

Mamadou Lamine Gueye

Modeling a knowledge-based system for cyber-physical systems: Applications in the context of Learning Analytics

Mamadou Lamine GUEYE

LIUPPA, University of Pau and Pays de l'Adour, Pau, France

mamadou.gueye@etud.univ-pau.fr

Abstract

Knowledge-based systems are major concerns in the field of artificial intelligence for the development of cyber-physical systems capable of self-management and adaptation to their context. The representation and knowledge management of these cyber-physical systems integrating heterogeneous actors must ensure the empowerment and optimization of these systems, as well as their ability to adapt to dynamic and unpredictable changes in their environment. In this document we show how a knowledge-based system based on semantic web technologies and IBM's reference model of Autonomic Computing (AC) can offer intelligent collaboration and coordination between people, data, services, robots and connected objects in the implementation of self-management processes in cyber-physical systems. Our solution consists to design a knowledge base in the field of Learning Analytics (LAs) involving a complex range of knowledge and heterogeneous components. This ontological knowledge base is guided by a functional decomposition approach based on the operating principle of the MAPE-K (Monitor-Analyze-Plan-Execute and Knowledge) autonomous control loop to provide the system with self-management capabilities.

Keywords : Knowledge Base, Ontology, Autonomic Computing, Learning Analytics

1. Introduction

With the fourth industrial revolution, universities underwent pedagogical transformations (University 4.0) induced by the integration of cyber-physical systems (connected objects, robots etc.) into teaching to improve and optimize the learning environment [5]. Thus, educational institutions are seeking to develop self-managing systems to automate mechanisms for understanding, improving and optimizing learning based on knowledge about learning tracking, traces collected through the use of online tools and services [4, 26]. The representation and knowledge management of these systems integrating both connected objects, data, services, robots, people, etc. must ensure the empowerment and optimization of these systems, as well as their ability to adapt to dynamic and unpredictable changes in their environment.

However, syntax-based methods of representing traditional knowledge offer few opportunities to address the semantic management obstacles that allow the orchestration of multiple actors in autonomous systems within the framework of University 4.0. There are also knowledge representation systems that provide empowerment, verification, optimization and adaptation services but are poorly adapted to a functional decomposition approach that allows the levels of maturation of cyber-physical systems to be taken into account (manual, observable, adaptive, self-adaptive, autonomous).

In this article paper, we propose a modeling of a common, scalable, dynamic and adaptive semantic space allowing the generic representation of physical or virtual objects, events, relationships, symptoms, diagnoses and change or action plans for the definition and execution of self-management processes, based on the IBM reference model of the MAPE-K (Monitoring, Analysis, Planning, Execution and Knowledge) control loop for autonomic computing [12].

This document is structured as follows: Section 2 provides a state of the art overview of knowledge-based systems and Learning Analytics, which is our case study. The two approaches we have explored for modeling our proposed solution are presented in Section 3. Before concluding, in section 4 we will present our case study and the proposed overall solution architecture to show a better appreciation of the contributions of ontologies and Autonomic Computing in industry 4.0, more specifically in the context of University 4.0.

2. State of the Art

In this section, we present a review of the literature on the Knowledge-based systems and Learning Analytics.

2.1 Knowledge-based systems

According to GRIMM et al. [1], the representation of knowledge and reasoning is a symbolic branch of artificial intelligence that aims to design computer systems that reason around a representation of the world that can be interpreted by the machine, similar to human reasoning. Knowledge-Based Systems (KBS) have a computer model of the real-world domain of interest in which physical or virtual objects, events, relationships, etc. are represented by symbols. KBS reason on this knowledge and use different inputs to help agents (physical or virtual) solve complex problems or make good decisions [24].

In the field of KBS, several works [2, 4-6] have been done in the conceptualization of data, information, knowledge, and intelligence in different forms using different approaches. According to [1] the most widespread knowledge capture and representation formalities are based on :

- semantic networks that can be found in RDF graphs;
- rules in the form of "if-then", for example in business rules or logical programming formalities;
- logic to achieve a precise semantic interpretation of semantic networks and rules.

Several researches have been conducted and many applications developed in the field of KBS [3], as in other areas of IT. In these systems, real-world facts are represented by simple assertions stored in a knowledge base and then manipulated using certain rules defined according to the objective of the system. We will present some research work carried out as part of the design of a knowledge base and reasoning on this knowledge:

Kitchen [2] presents the AM (Lenat) and Teiresias (Davis) projects in the field of expert systems of artificial intelligence. In these projects, the authors have developed knowledge-based systems capable of imitating the process of human intelligence functioning. The first, AM, applied to elementary mathematics, models an aspect of scientific research concerning the creation of new concepts and conjectures about their behaviour. It is based on the definitions of certain theoretical concepts and a large number of general heuristics to decide on useful research axes and generate examples that form the basis for conjectures. The second Teiresias project is a tool that allows an expert to build and maintain a knowledge base in his field of activity.

Based on the rules provided by an expert in the field, the system generates models that it uses to organize and control the use of the rules by giving the ability to provide a high-level report on the motivations for its actions. For example, for a programmer when a bug is discovered in the knowledge base, Teiresias is able to explain how it came to the knowledge base and even suggests plausible corrections.

According to Soualah Alila et al [4], there are constraints of semantic heterogeneity of resources and heterogeneity of use of learning platforms that do not allow learning content to be adapted to the learner's contextual situation. To help solve this problem, the authors developed an m-learning system based on the learner's contextual constraints to recommend training paths without risk of disruption. The system architecture is based on a formal representation of data and business processes in an ontological knowledge base, as well as the use of metaheuristic algorithms to infer knowledge. According to the authors, them m-learning system offers trainers the possibility to model their know-how taking into account environmental and user constraints.

According to Sanin et al. [5], cyber-physical systems and the Internet of Things can lead to the development of increasingly competent and intelligent systems in industry and academia, based on an interesting and attractive scenario. The authors propose Decision DNA, a knowledge-based system that integrates a complete representation of knowledge for the Internet of Things and cyber-physical systems. Decisional DNA collects explicit knowledge based on the experience of formal decision-making events and uses it to assist in the decision-making process. According to the authors, one of the main advantages of its use is to provide predictive capabilities based on knowledge and experience.

Salayandia et al. [6] propose a system of recommendations and advice for researchers, based on Semantic Web technologies, called MetaShare. It represents the researchers' experience, data collection, and management activities in a form that can be exploited by a machine by connecting the associated data. MetaShare is a decision-making support system for researchers on project planning and implementation based on similar project practices and decisions. Its formal representation of knowledge in the form of ontologies and rules allows for the collection, dissemination, and management of data to facilitate tasks related to the use and exchange of scientific data.

According to Marvis [7], current Semantic Web technologies offer the possibility of representing knowledge explicitly, through tools such as ontologies and rules applied to facts rather than implicitly through procedural programming methods. There is a common, evolving, dynamic and adaptive semantic representation of a problem that allows generic modeling of physical or virtual objects, events, relationships, symptoms, diagnoses and change or action plans for the definition and execution of self-management processes. Based on the above work, [4] and [5] can be used as a support for our approach to design the knowledge base of our system.

2.2 Learning Analytics

Through our case study, we will model a knowledge base in the field of Learning Analytics (LAs). LAs are defined at the first international conference on Learning Analytics and Knowledge (LAK) in 2011 as *"measurement, collection, analysis, and communication of data on learners and their contexts, with the aim of understanding and optimizing learning and the environments in which it occurs"*.

Chatti et al. [19] presented a systematic overview of LAs and its key concepts using a reference model based on four dimensions, namely: 1) data, environments and context (what?), 2) stakeholders (who?), 3) objectives (why?) and 4) methods (how?). We will use this reference model as a basis for presenting different research studies in the field of LAs in relation to each dimension.

Arnold et al. [20] have developed an early intervention solution for college faculty called Course Signals at Purdue University in Indiana. In short, Course Signals uses data from the information system, the course management system and the notebook. This data is then manipulated, transformed into compatible forms and entered into an algorithm generating a level of risk with additional information for each student, represented by a green, yellow or red indicator.

The Signals system is intended to help students understand their progress early enough to allow them to seek help and probably get good grades or change their behaviour. The authors propose to use business intelligence to improve student success at the course level, thereby increasing retention and graduation rates. The Signals prediction algorithm is based on the student's performance, effort, academic background, and characteristics.

Agnihotri et al. [21] propose a model for early intervention with students most at risk of attrition, called the Student At-Risk Model (STAR), applied at the New York Institute of Technology (NYIT). The main risk factors include the student's grades, main subject, and certainty of choice of the main subject, as well as financial data such as the cost of tuition fees in relation to the student's means. STAR allows student support staff to screen and intervene early to improve student retention at university using a dashboard. It provides a binary indicator (risk of attrition or not) for each student and identifies the key factors that make a student unlikely to return the following year. STAR in version 2.0 is built from a dataset contained in a data warehouse where they can be created automatically as soon as a new student registers, then the data mining tools are used to form machine learning models to perform the classification task. These models use variables to predict whether or not a student will return the following year, which is then used to signal the risk of new students.

Niall Sclater et al. [22] have developed a sustainable and flexible LAs service for British universities and colleges, called Jisc's learning analytics. The system collects all kinds of data on student activity in various systems such as LMS such as Moodle, library systems, student record systems or other self-reported data through our student application. The objective of the tool is to provide an alert and intervention system to predict student success.

This solution consists of a mix of commercial and open source solutions consisting of collecting and storing data in Learning Records Warehouses (LRW), analyzing these learning traces with the open platform "Apereo Learning Analytics Initiative" in order to present a dashboard for staff and to alert staff and students and allow them to manage the intervention activity through an alert and intervention system.

Admittedly, the list of LAs applications is not exhaustive, but it can be seen that most solutions consist of collecting, manipulating, transforming data in order to make them compatible before analyzing them using predictive algorithms. We note that the architecture of the proposed solutions for LAs is based on a combination of several technical solutions for data encoding, data mining and storage, and data analysis and processing. This presents many obstacles regarding the syntax and semantics required when integrating multiple data sources and also makes the deployment of LAs complex.

3. Proposal of Our Model

We propose a representation of a knowledge base based on an initial ontology-based approach to provide a common, scalable, dynamic, maintainable and adaptive semantic space that provides decision making adapted to the objectives of the system. The second approach to our modeling is based on the Autonomic Computing operating principle (MAPE-K) which consists of breaking down our knowledge base into autonomous components to provide a self-management process adapted to cyber-physical systems. These two approaches to our knowledge base will be detailed in the following sections.

3.1 Ontology-based modeling approach

Our ontology oriented approach [8] has as its main objective to define conceptual equivalences, which are then used to automatically calculate (reason) relationships between classes, properties, and instances.

Ontologies also allow any concept to be represented in a unique way, and adhere to the open world assumption: everything is permitted until it is prohibited. They provide a model of a body of knowledge in a given field, which can be real or imaginary in order to ensure knowledge management and reasoning on this knowledge, with a view to semantic interoperability between human and/or artificial agents [9]. Ontologies can provide cyber-physical systems with the ability to process and understand data, access a set of structured information and inference rules that they can use to achieve autonomous behaviour.

3.2 MAPE-K oriented design approach

For modeling of our KB to ensure self-management capabilities with a functional decomposition approach, we use the IBM reference model [10] to define autonomous managers, modules that provide autonomous behaviour to system components. It also allows us to better structure knowledge about the measures, symptoms, strategies, objectives, plans or requests for change, etc. in our knowledge base. This model is based, as suggested by Horn [11], on a control loop called the autonomous control loop or MAPE-K loop, presented in the figure below:

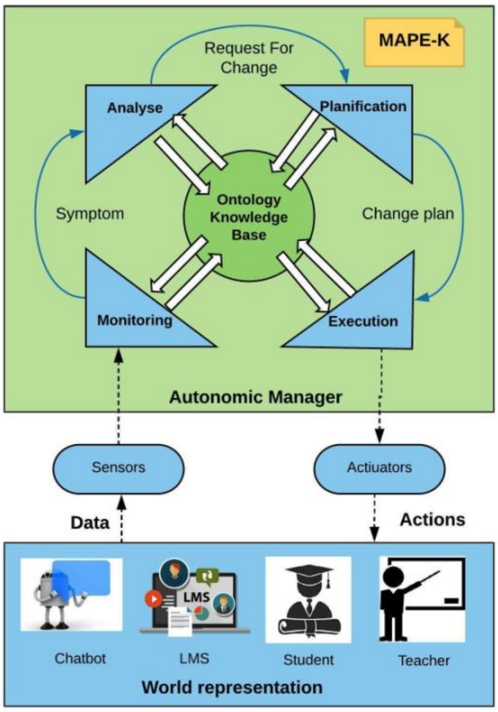

Fig. 1. MAPE-K autonomous control loop

4. Case study

To illustrate the representation of knowledge guided by an ontologically oriented approach in cyber-physical systems, based on the principle of CA [12], we work on scenarios drawn from Learning Analytics [14].

According to the 2012 report of the US Department of Education [15], with the emergence of LAs, "*e-learning systems have the capacity to capture learners' behaviour to provide feedback to a variety of actors to improve teaching, learning and educational decision-making"*. In these scenarios, we will implement knowledge modeling of learning tracking information, traces collected when using online tools and services. This formalization of knowledge in a form that can be interpreted by humans and machines will be used to automate and empower the understanding, improvement, and optimization of learning.

Such a use case is relevant for illustrating the modeling of a KB for cyber-physical systems, as learning systems consist of a variety of interacting physical and/or virtual components (LMS, student, robots, teacher, etc.) and overflowing with a mass of information from different sources.

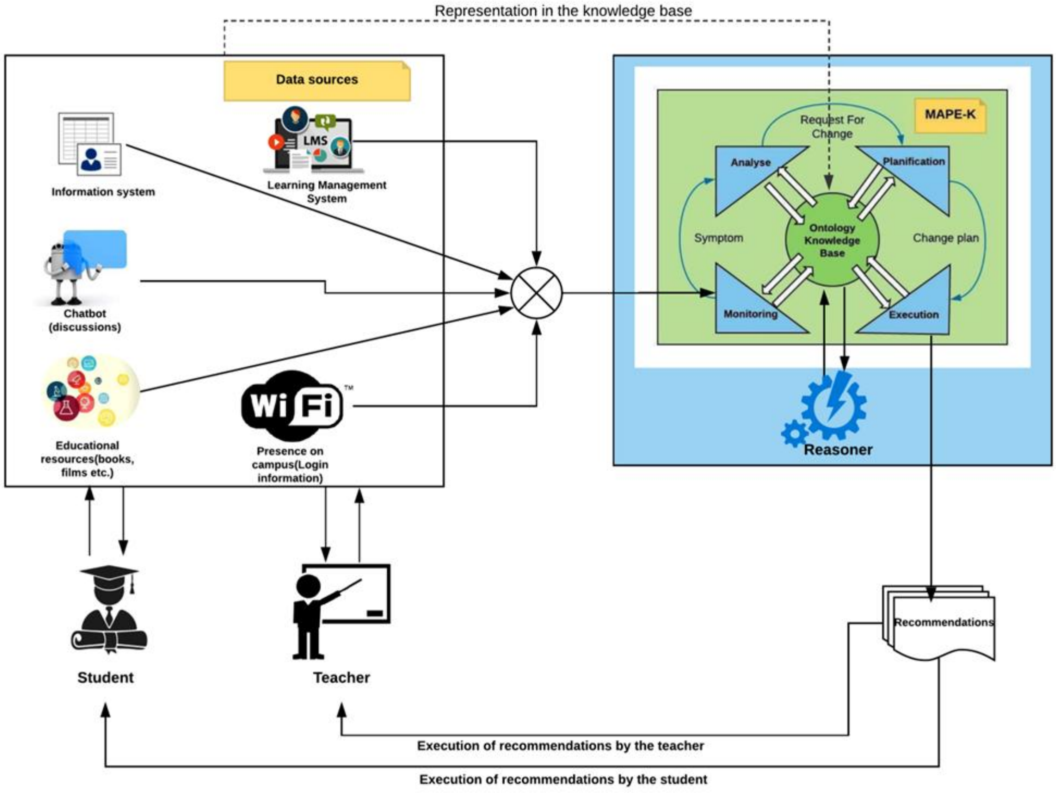

Fig. 2. Global architecture of the solution

The main components of our architecture are as follows:

4.1 Knowledge base (KB)

The knowledge base is the central core element of the MAPE-K loop, it contains all the knowledge relevant to the management of the system. The knowledge represented in our knowledge base concerns the learning management system (LMS), students, teachers, information system, pedagogical resources (books, films, computers, etc.) and chatbot, as well as the concepts, ensuring the functionalities of the MAPE-K autonomous control loop modules providing the system with autonomous behaviour.

We propose a knowledge base with an ontology oriented approach to ensure the heterogeneity constraints of data from different physical and/or virtual components (LMS, student, robots, teacher, etc.) of a learning environment.

This formal representation of knowledge offers us the opportunity to better manage the automation and empowerment of educational decision-making processes for the understanding, improvement, and optimization of learning. We used Protege [23] as an ontology editor to create the ontology of our case study and reason on knowledge using integrated inference engines.

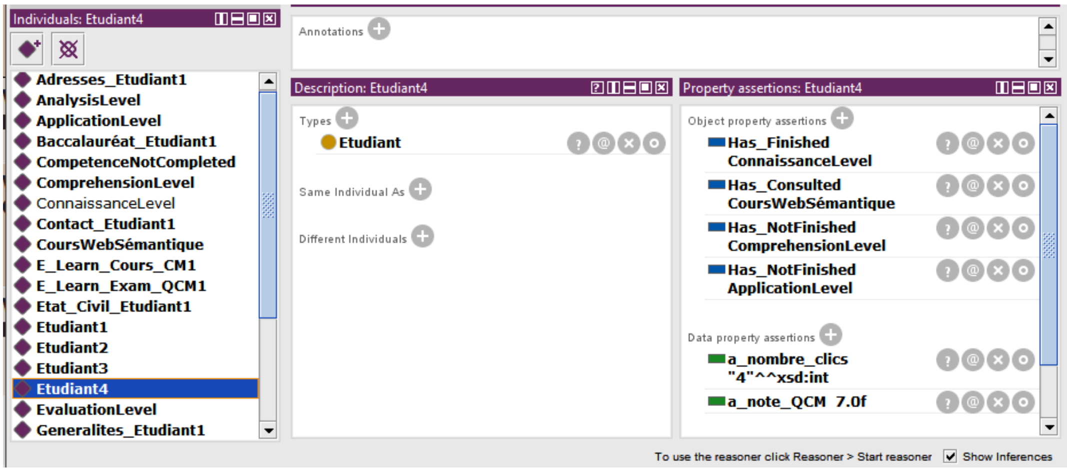

Fig. 3. Representation of a student and his traces on the learning environment

4.2 Reasoner

The reasoner is the trigger for the automatic and autonomous execution of our MAPE-K loop. This inference engine is an important component of the KB, without this software, it would be impossible to reason about knowledge. It applies logical rules of correspondence, selection, and execution [16, 17] to data captured by agents of the learning system and knowledge stored in the knowledge base in order to generate symptoms leading to diagnoses in order to make recommendations to the learner and/or teacher.

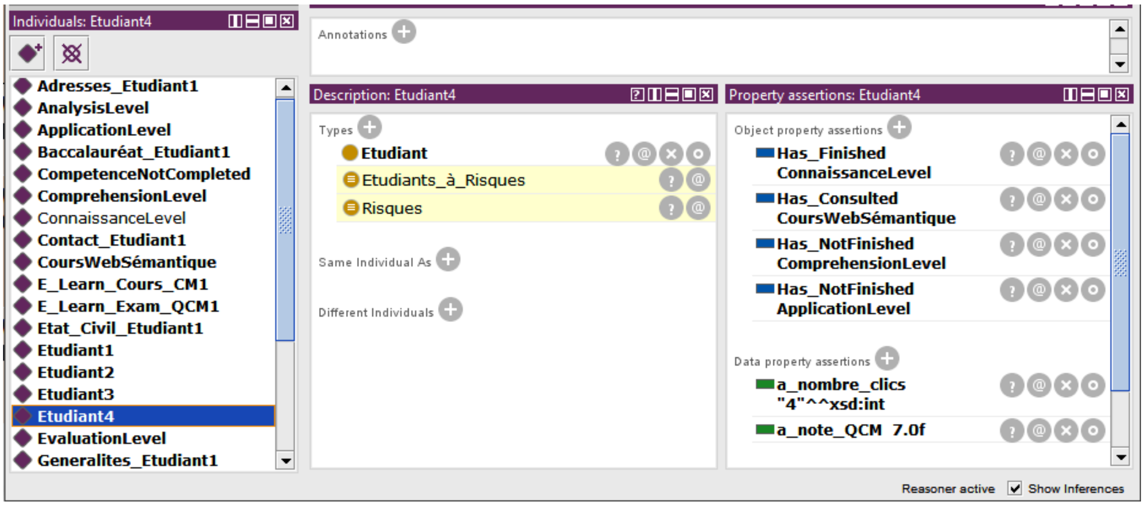

Fig. 4. Classification of the student by the reasoner as being at risk

To ensure the different functions of the system, we defined each function by a set of rules [16, 17] applied to the data captured by the agents of the learning system and the knowledge stored in the knowledge base in order to make recommendations to the learner and/or the teacher.

4.3 Monitoring

The basic function of our system is "Monitoring" which recovers the data captured by the different components of the learning system, presented in the form of facts, figures, etc.; for example:

- *The learner consulted the course on the Semantic Web for 10 minutes*
- *The learner had a score of 7/20 on the Semantic Web exam*

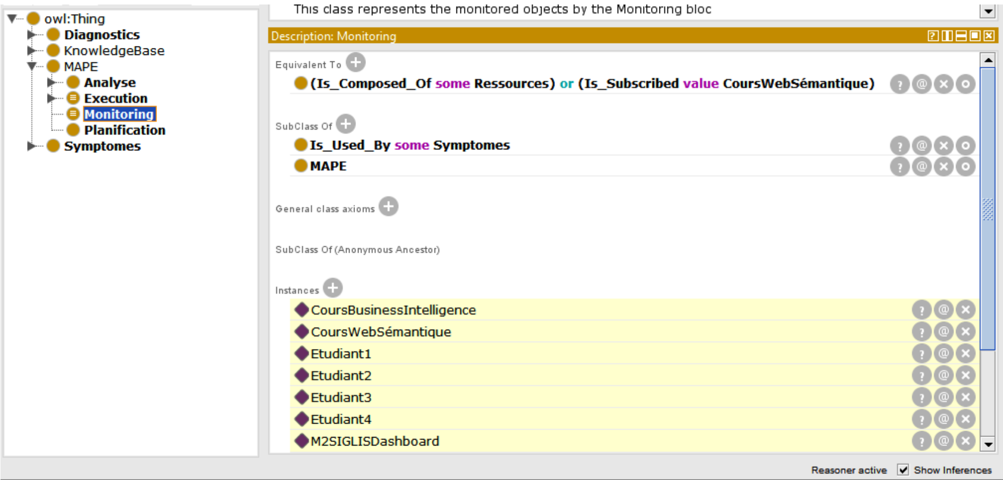

Fig. 5. Monitored components of the system

4.4 Symptoms

The second function is the "Symptoms" which consists in determining the symptoms, i.e. giving meaning to the data, so that they become relevant information for our educational decision-making process. For example, it may generate a "student at risk" symptom if it finds that the student has a score below 7/10.

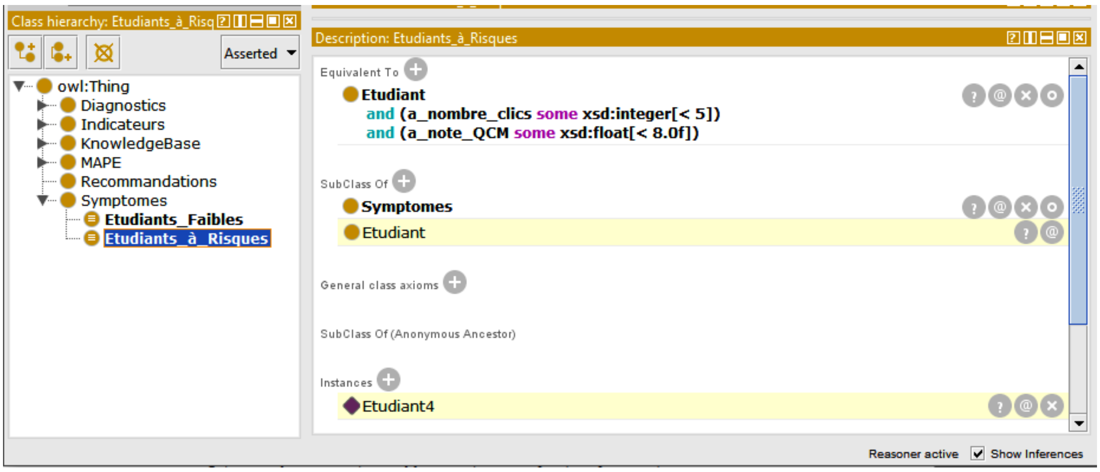

Fig. 6. Symptom of an at-risk student

4.5 Analysis

Analysis, the third function is the subjective interpretation of information in order to determine the causes and the way to act in the process of achieving the objectives of Learning Analytics.

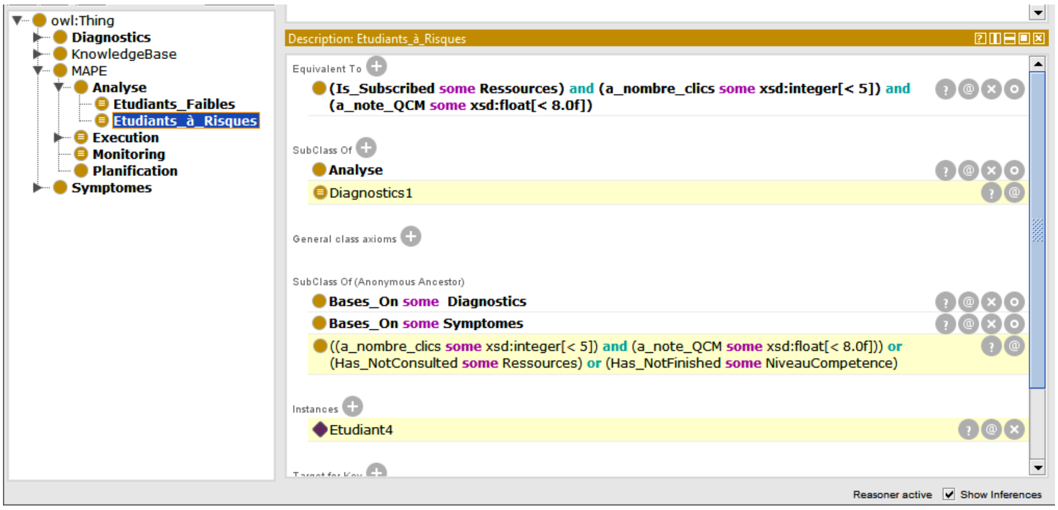

Fig. 7. System analysis results

4.6 Diagnostics

The fourth function is the "Diagnostics", it is based on the symptoms, the results of the "Analysis" function, as well as the knowledge and rules defined in the system leading to the establishment of the diagnosis in a particular way to associate an objective with its diagnosis. This is what generates recommendations (educational decisions) for "Planning" to improve learning. For example, the diagnosis shows that the student has not completed all levels of knowledge acquisition.

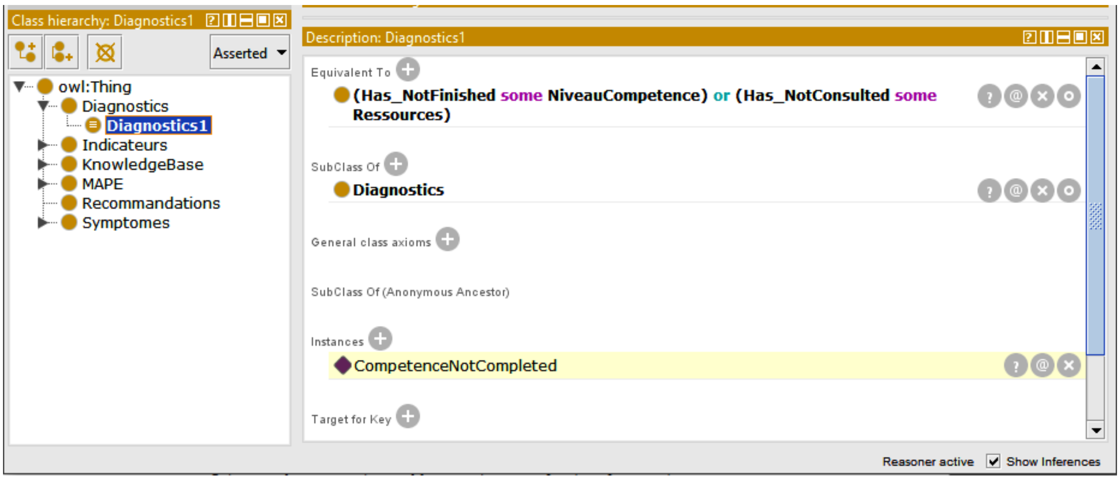

Fig. 8. System diagnostic results

4.7 Planning

The fifth function is "Planification", based on each diagnosis and the reasoning behind each diagnosis, it provides the "Execution" function with a sequential list of possible action plans leading to the implementation of recommendations.

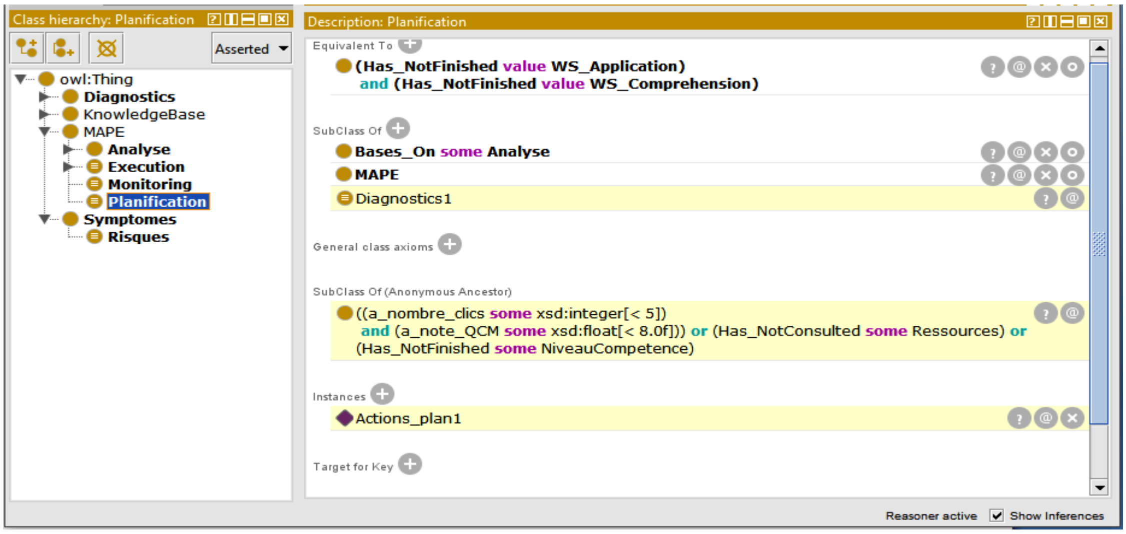

Fig. 9. System planification results

4.8 Execution

Finally, execution generates actions in the form of recommendations to the learner or teacher to improve learning. The latter will be responsible for carrying out these actions in order to implement the necessary changes.

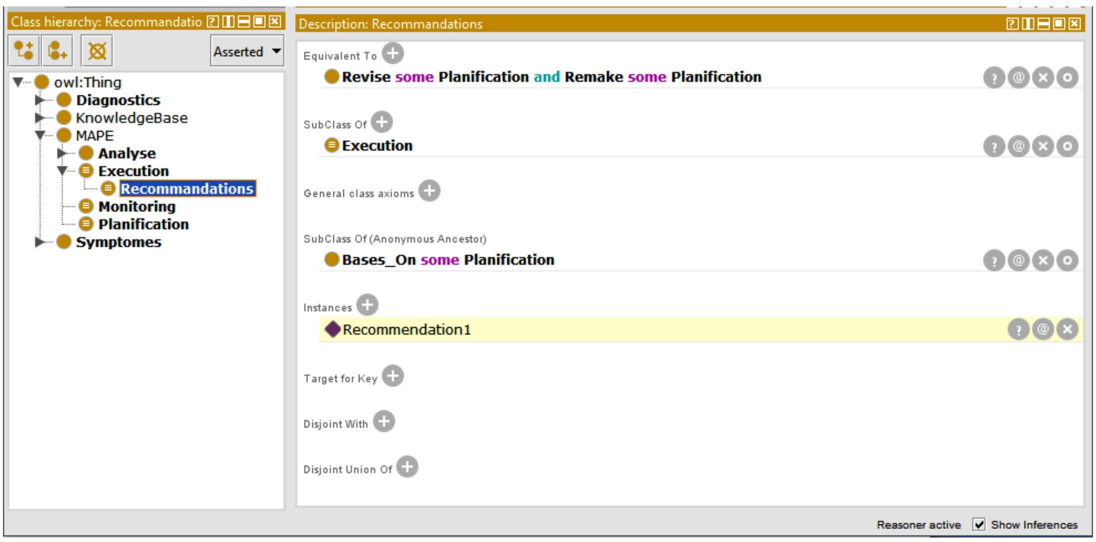

Fig. 10. System execution results

5. Conclusion

In this work, we proposed modeling of an ontology-based knowledge base with a functional decomposition approach, as well as self-management capabilities provided by the CA. In this article, let us implement the continuous improvement of automated decision making and the resolution of problems related to the self-management constraints of cyber-physical systems.

This scalable, dynamic and adaptive system allows the generic representation of physical objects, events, relationships, symptoms, diagnoses and action plans for the definition and execution of self-management processes. It can be further improved to provide better diagnostics for a variety of purposes and results to make its applications more intelligent and efficient.

So, we note that in the LAs, two methods are preferred (xAPI and Caliper) [18] for encoding information on learning follow-up, traces collected when using online tools and services in the form of RDF triplets with the form "subject-predicate-object" + context. The statements are then stored in a Learning Record Store (LRS). We find that ontologies are native with an encoding of information in the form of RDF triplets with the form "subject-predict-object" + context, they also offer a controlled syntax and semantics to ensure interoperability. The knowledge base can be used in data mining and data warehousing.

In terms of perspectives, we intend to improve the work by working on in-depth documentation of the systematic diversification of scenarios and hypotheses in the knowledge base to arrive at a solution that makes it possible to assess the real contributions of a combination of ontologies and knowledge-based systems in LAs.

References

1. GRIMM, S., ABECKER, A.: Knowledge Representation and Ontologies Logic, Ontologies and Semantic Web Languages (2007)
2. KITCHEN, A.: Knowledge based systems in artificial intelligence. Proceedings of the IEEE, 1985, vol. 73, no 1, p. 171-172 (1985)
3. AKERKAR, R., SAJJA, P.: Knowledge-based systems. Jones & Bartlett Publishers (2010)
4. Soualah-Alila, F., Nicolle , C., Mendes , F.: Archive ouverte HAL - Context-Aware Adaptive System For M- Learning Personalization. In : [Online]. [Accessed May 24, 2019]. Available at: https://hal.archives-ouvertes.fr/hal-01238330 (2015)
5. SANIN, C., HAOXI, Z., SHAFIQ, I., et al.: Experience Based Knowledge Representation for Internet of Things and Cyber Physical Systems with Case Studies. Future Generation Computer Systems, vol. 92, p. 604-616 (2019)
6. SALAYANDIA, L., PENNINGTON, D., GATES, A.Q., et al.: MetaShare: From Data Management Plans to Knowledge-Based Systems. In : 2013 AAAI Fall Symposium Series (2013)
7. Marvis U., Shade , K., Atisi, F., Olayiwola, B.: Overview of knowledge-based system, International Journal of Computer &Communication Engineering Research vol. 2, no.3, pp.125-129 (2014)
8. Wikipedia. Ontologie (informatique) [Online].[Accessed March 15, 2019]. Available at: https://fr.wikipedia.org/w/index.php?title=Ontologie_(informatique)&oldid=156460030 (2019)
9. FÜRST, Frédéric and TRICHET, Francky. Raisonner sur des ontologies lourdes à l'aide de Graphes Conceptuels. p. 17 (2006)
10. KEPHART J.O. and CHESS D.M.: The vision of autonomic computing. Computer, 36(1) :41–50 (2003)
11. HORN P.: Autonomic Computing : IBM's Perspective on the State of Information Technology. Technical report, IBM (2001)
12. Aaron Computing.: An architectural blueprint for autonomic computing. Technical report, IBM (2006)

13. Pejman Makhfi. Introduction to Knowledge Modeling. In : [online]. [Accessed March 27, 2019]. Available at: http://www.makhfi.com/KCM_intro.htm.
14. NUMÉRIQUE, France Université (2017).: Learning Analytics, une tendance émergente dans l'éducation - sup-numerique.gouv.fr. In : [online]. [Accessed March 14, 2019]. Website: http://www.sup-numerique.gouv.fr/cid113065/learning-analytics-une-tendance-emergente-dans-l-education.html.
15. U.S. Department of Education (2012). Enhancing learning and training through educational data mining and Learning Analytics [online]. [Accessed March 14, 2019]. Website: https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf.
16. Gong, Y., Janssen, M.:. An interoperable architecture and principles for implementing strategy and policy in operational processes. Computers in Industry, vol. 64, no. 8, pp. 912-924 (2013)
17. Umar, M. M., Mehmood, A., Song, H.: A Survey on State-of-the-Art Knowledge-based System Development and Issues (2015). In : ResearchGate [online]. [Accessed April 1, 2019]. Available at: https://www.researchgate.net/publication/290481752_A_Survey_on_State-of-the-Art_Knowledge-based_System_Development_and_Issues.
18. WIKIPEDIA.: Analyse de l'apprentissage [Online]. [Accessed April 2, 2019]. Available at: https://fr.wikipedia.org/w/index.php?title=Analyse_de_l%E2%80%99apprentissage&oldid=153731273 (2018)
19. Chatti, M.A., Dyckhoff, A.L., Schroeder, U., Thüs, H.: A reference model for learning analytics. International Journal of Technology Enhanced Learning, 4(5/6), 318–331 (2012)
20. Arnold, K.E. and Pistilli, M.D.: Course Signals at Purdue: Using learning analytics to increase student success. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK'12) pp. 267-270 (2012)
21. Agnihotri, L., Ott, A.: Building a student at-risk model: an end-to-end perspective. In Proc. Int. Conference on Educational Data Mining Conference (EDM), 209-212 (2014)
22. Sclater, N., Peasgood , A., Mullan, J.: Learning Analytics in Higher Education. A review of UK and international practice Full report (2016)
23. Musen, M.A. *The Protégé project: A look back and a look forward*. AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4), June 2015. DOI: 10.1145/2557001.25757003.
24. ZARATÉ, P., LIU, S.: A new trend for knowledge-based decision support systems design. International Journal of Information and Decision Sciences, vol. 8, no 3, p. 305-324 (2016)
25. Timothée Duron, Laurent Gallon and Philippe Aniorte – Modelling Learner's Perseverance in Education Software – Information Systems Education Conference (ISECON 2019), April 4-6, 2019, Galveston, TX, USA (2019)