BIG-INVENTE, TOWARDS AN ADAPTIVE LEARNING APPROACH FOR DISTANCE EDUCATION USING DATA MINING TECHNOLOGIES AND BIG DATA

ANTONIO MAURO BARBOSA DE OLIVEIRA[1], RONALDO FERNANDES RAMOS[2],
CARINA TEIXEIRA OLIVEIRA[1], ELIEZIO GOMES QUEIROZ NETO[1]

[1]Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)
Campus de Aracati
[2]Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)
Campus de Fortaleza
<mauro@ifce.edu.br>, <ronaldo.ramos@gmail.com>,
<carinatoliv@gmail.com>, <egqneto@gmail.com>

Resumo. Observa-se, atualmente, mudanças de paradigmas nos sistemas de ensino, antes centrados na figura do professor, embora muitos dos sistemas de educação existentes, mesmo os de educação a distância, continuem a adotar a forma clássica centrada no aluno. Para se ter um maior nível de personalização no sistema de ensino é necessária a atualização das plataformas de aprendizagem e das teorias de aprendizagem associadas. Neste artigo apresentamos a evolução da plataforma INVENTE desde a sua concepção inicial em 2000 até os dias atuais. A principal mudança que ocorreu na estrutura do INVENTE foi na capacidade de usar técnicas de aprendizado de máquina (Machine Learning) para a modelagem do aluno e permitir a construção de conteúdo dos enormes conjuntos de dados disponíveis na web, mais precisamente no contexto da nova tecnologia Big Data. O desafio do BIG-INVENTE é, portanto, usar essas tecnologias (Data Mining e Big Data), a fim de tornar mais eficiente o processo de aprendizagem.

Palavras-chaves: Aprendizado adaptativo. Educação à distância. Big data. Mineração de dados.

Abstract. Nowadays, paradigm changes in learning systems can be observed, from teacher-centered to learner-centered although many of the existing education systems, including distance education, continue to adopt the classical student-centered model. In order to have a higher level of education system personalization, it is necessary the updating of learning platforms as well as the theories of learning associated. In this article, we show the evolution of the INVENTE platform since from its initial conception in 2000 to the present days. The major change that occurred in its structure was in the ability to work with machine learning techniques for modeling the student and allowing content construction from the huge collections of data available in the web, more exactly in the Big Data context. The challenge of BIG-INVENTE is, therefore, to use these technologies (Data Mining and Big Data), in order to customize and make the learning process more efficient.

Keywords: Adaptive learning. Distance education. Big data. Data mining.

1 INTRODUCTION

Adaptive learning uses computers to adapt educational material according to students' learning needs. This system attempts to transform the learner from passive receptor of information to collaborator in the educational process.

On the other hand, education is increasingly occurring online, resulting in an explosion of data that can be used to improve educational and support adaptive learning. Emerging research groups in educational data mining andadaptive learning are developing methods for mining data available aboutlearners. So, Big Datatechnology make it possible to mine learning information for insights regarding student performance and learning approaches. It refers to the large amount of information that flows through various channels.

In this paper we present the BIG-INVENTE, a framework for distance learning that mines learning information about students, on big data context in order to improve the their adaptive feature. We introduce a novel approach in order to design a pattern classifier aiming at predicting student's profile for the BIG-INVENTE adaptive learning approach. The applied methodology consists of labeling feature vectors based on the events occurred in a time window into a known future.

The remainder of the paper is organized as follows. In Section 2 and 3 we provide a detailed INVENTE project evolution, since their initial approach, on 1999, that used Extended Critical Dimensions concept, until the distributed version based on multi-agent technology. In the section 4 we describe the proposed approach for feature vector building and the predictive-labeling process. Finally, our conclusions are done in Section 5.

2 INVENTE PROJECT

2.1 Extended Critical Dimensions Concept

Existing technologies for distance learning are often based on concepts from learning process practiced under space and time restrictions, i.e. traditional learning process based on a teacher and a number of students together in a classroom. We argue that the employment of such traditional learning model is not well suited to this new environment of distant learning systems. Efficient learning systems must take into account subjective aspects that cannot be measured by traditional dimensioning tools in order to better describe actors' behaviors, according to its inherent Critical Dimensions.

As these subjective aspects can determine actors' behaviors, it is also possible that they could also influence QoS constraints imposed by the learning system and also how network resources will be consumed. Hence, network QoS management can take into account this behavioral approach in order to optimize network resources distribution, for instance. New Critical Dimensions can, thus, be identified.

The objective of the Critical Dimensions described in Hazemi, Hailes e Wilbur (1999) is to help solving pedagogical misunderstanding between traditional and distance learning models in order to better conceive distance learning systems. As consequence of the restrictions imposed by these Critical Dimensions, concepts like Management Behavior, Metaphor, Global Context, and Collaboration play an important role to define distance learning systems (HAZEMI; HAILES; WILBUR, 1999; SOARES, 2001).

Two Critical Dimensions are presented behind in order to illustrate the distance learning system different pedagogical approach:

- **Management Behavior:** In traditional teaching model, learning is, in general, a linear process. Courses have sequential steps, according to an initial planning. In modern teaching model (including distance learning), dynamic navigation offers a great broken in the pedagogical process. Thus, it is necessary to learn how to manager the distance learning to better understand on the learning process.

- **The Metaphor:** In distance learning environments, the use of metaphors is an excellent strategy. In this case, it is very important not to repeat the same language and procedures from traditional model (SOARES, 2001).

The Extended Critical Dimension set adds subjective constraints as the new Critical Dimensions concepts. The cognitive aspects (student's level, teaching's didactic, etc) and the environment variables (weather, traffic etc) are examples of this extension.

In this context, the Figure 1 shows three domains. The internal domain, correspond the network (the information transport). The intermediate domain supports the actors (users) of the system, with theirs cognitive elements. The external domain corresponds to the environmental variables. In reality, the networks resources that an application needs depend of the real time situations. The bandwidth effectively used by an application is a direct function of the personal and external variables around the user. By personal means, for example, the psychological conditions that that has great influence in the cognitive aspects. By external we can consider the independent factors of the will of the user. Both these aspects are here considered in the context of the work proposed.

The main idea proposed in this paper is based on the influence of these new critical dimension concepts (the intermediate and external domains) on the Quality of Service of the network (internal domains). In fact, this domain is already treated in terms of QoS by several mechanisms. So, the new approach consists in using multi-agent modeling based in Extend Critical Dimensions to optimize in real time the resources network allocated to the applications.

2.2 A Distributed e-Learning Platform

In Moura (1999) and Soares (2001) we found a reflection on the network problems to turn effective distance

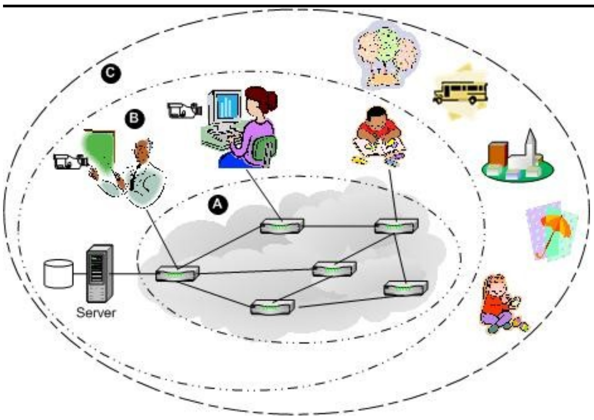

Figura 1: The three levels of the Extended Critical Dimensions set.

education projects. At that time, they emphasize the differences between the technological and the conventional teaching. They also present the technological presuppositions that should be satisfied by the distance learning tools.

In Soares (2001) the e-learning critical dimensions - early studied by Hazemi, Hailes e Wilbur (1999) - are analyzed and the critical dimensions of the Distance Technological Education (DTE) are enumerated. Based on these reflections, Soares has conceived an IBW system (Instruction Based on the Web system), called INVENTE.

The INVENTE architecture proposed by Soares (2001), look at e-learning requisites with a conceptual perspective and not only with a technological vision. In his new conception, this architecture proposes the support to e-learning critical dimensions by means of "easy cultural adaptation", "environment flexibility" and "sense exploration".

Through an open and flexible architecture, that considers important aspects of QoS,INVENTE satisfies presuppositions of an environment for the technological teaching, respecting e-learning critical dimensions. A great advantage of this architecture is the possibility of easy aggregation of new applications to the environment.

In Serra (2001) some improvements have been proposed in the architecture of this system in order to assist important requirements of distribution. A manager of distributed services was introduced in the platform and a distributed videoconferencing application called AVET (Videoconference Application for the Technological Education) was also joined to the environment.

For the third version of INVENTE, now called RE-INVENTE (OLIVEIRA, 2003), we are worried about the QoS aspects at the same time we take into account new constraints imposed by the support of new users based on wireless (IEEE 802.11) or mobile (GPRS, etc.) technologies.RE-INVENTE must now support students or teachers that access distant learning resources by means of a notebook PC or a PDA. This new support will impose new problems to be solved like interface constraints, and QoS limitations.

2.3 RE-INVENTE Distributed Architecture

The distributed architecture has a three-dimensional structure (Figure 2). The *"Management Center"* is based on the Web technology and it is responsible for: the administration of the resources, the management of the database, the administration of the publications and the resources access control.

The *"Web Interface"*, contained in the *"Management Center"*, supplies the necessary interface between the users and the resources. The applications were divided in *"general-purpose-applications"* and *"domain-specific-applications"* with the objective to turn clear the possibility of the architecture to aggregate different kinds of applications, without induction of the use of a pre-defined private technology.

The architecture also highlights the possibility of the applications interact with various network infrastructures (TCP/IP, ATM, etc.).

The *"distance reference layer"* has the function of intermediating the QoS requirements negotiation between the applications and a specific communication system. This service is generic and not depends on a specific communication system.

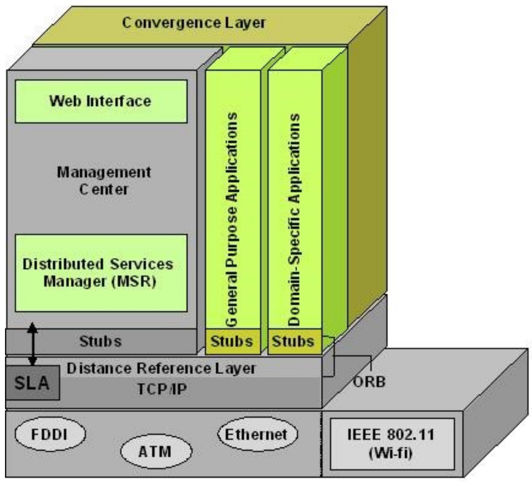

Figura 2: RE-INVENTE Architecture.

The solution of communication adopted was the Remote Call of Procedures (RPC) (TENE; POLO-

NETSKY, 2013). The distributed objects change messages through the distance invocation of a group of methods that compose the communication interface of each object.

To provide this communication solution an ORB (Object Request Broker) was added to the architecture. The *"Stubs"* provide a communication interface to the objects. The *"Distance Reference Layer"* is responsible for solve object requests with transparence and for the call parameters sending and reception in an acceptable format. The *"Distributed Services Manager"* managers of the real time distributed services that are in use.

In the new RE-INVENTE architecture new elements are introduced. We can observe that the infrastructure layer includes Wi-fi technology (IEEE 802.11). The second element in this new version is the inclusion of a new interface to deal with QoS support architectures such as DiffServ and IntServ, by means of traffic marking and SLA/SLS interpretation, represented in the figure by the SLA block, functionally connected to the management center.

2.4 Resources Distribution and Virtual Rooms

The distributed platform proposed by Serra (2001), allows the installation of services in different hosts in a computer network. In some situations, the employment of various equipments will be necessary to provide a specific service. When a service does not demand a lot of computational power, it can be provided by equipment with smaller capacity of processing.

RE-INVENTE *administrator user* is responsible for the resources distribution planning in the network.RE-INVENTE uses the concept of *Virtual Rooms* to become easy the resources distribution. A virtual room, in the context of RE-INVENTE, is a group of distributed resources (hardware and software) that are put together under a single denomination.

To define a *Virtual Room*, the RE-INVENTE *administrator user* informs to the system what hosts and services are available in the platform. He informs the name of the *Virtual Room*, the services that will be made available at this *Virtual Room* and the hosts that will provide those services.

Using this distribution model, a same service can be distributed on several equipments in the network, as well as only one equipment can provide several services. Who decides the distribution of the resources in the network is the RE-INVENTE *administrator user*. He estimates the processing demand of each service and the computational power of each equipment. In the Figure 3, it can be observed two *Virtual Rooms* providing services of audio and video distribution, chat and hypermedia. In this example, two *Virtual Rooms* share a same chatserver, however they use different video and hypermedia servers.

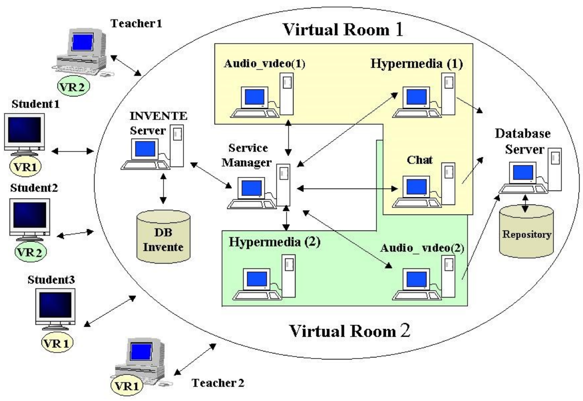

Figura 3: Organization of the resources in virtual rooms.

The grouping of services in *Virtual Rooms* has as objective to turn transparent for the end user (author, teacher, student, etc.) the distribution of the resources. When a teacher will initiate a videoconferencing session, for example, he chooses one of the available *Virtual Rooms* in the environment and does not worry about the resources distribution in the network. This transparency in the distribution of the resources facilitates the cultural adaptation (SOARES, 2001) of the users to the application.

Another advantage of this distribution model is the possibility to limit the use of the resources to the estimated capacity of the installed infrastructure. As each *Virtual Room* expresses a slice of the total capacity of the available resources, the RE-INVENTE *administrator user* creates a number of *Virtual Rooms* that express the total estimated capacity of the available resources.

3 RE-INVENTE AND DYNAMIC LOAD-BALANCING MULTI-AGENT SYSTEM

We believe that to treat aspects that are related to the load-balancing problem in e-learning distributed applications (the user's mobility, wireless connections, electric power consumption, etc.) it is more suitable an application-level approach that involves agents technology. At the same time, we are aware that to conceive a transparent solution for the programmer it is interesting to integrate this solution to a specific middleware.

In the works we have studied, we have not been found equilibrium between these two aspects. While some works prioritize application distribution aspects forgetting important variables (high latency, semantic of objects, etc.) Marron (2000), others approach a large number of variables but ignore important distribution aspects (transparence of resources distribution, integration of load-balancing mechanisms with the middleware, etc.) (RAJAGOPALAN; HARIRI, 2000; ZHOU et al., 2002).

We think that a solution for dynamic load-balancing should take into account all of the available technology (multi-agents, artificial intelligence, etc.) at application level. But at the same time it should also consider the need to integrate this solution to a middleware. This will give transparency to the developer allowing an abstraction of the mechanisms used for the improvement of the end user's QoS (SERRA, 2001). For this reason, in the development of our research, we will try not to prioritize any approach in detriment of others; otherwise we will try to take advantage of both.

We consider that the use of agents and of artificial intelligence techniques it is fundamental in the context of our work for two main reasons. Some works show that the use of agents in the monitoring of the resources and in the maintenance of a resources global knowledge is more effective. This happens because of their collaboration and communication capacities (JUN et al., 1999). At the same time, the number of variables involved in the resources allocation decisions (the local charge, connections situation and characteristics, services behavior, etc.) demands more sophisticated mechanisms that involve artificial intelligence techniques.

A point that we consider of fundamental importance is the object/services semantics. Studies have shown how objects semantics can help in the transparent location of services in distributed environments (FELBER et al., 2001). We know that a number of equipmentsare manufactured especially to better adapt themselves to specific services. For instance, equipments with special audio and video processing power are more appropriated to execute multimedia applications. Following this point of view, we believe that object/service's semantics can be used by the decision mechanisms improving the process efficiency of resources distribution.

3.1 A High-Level Conception

Based on these considerations, two objectives are pursued jointly to improve QoS in RE-INVENTE. The first one is the reduction of the response time to the users (students, teachers, etc.) through an intelligent distribution of the services in the network. The second is the implementation of preventive fault tolerance schemes through the replication and migration of services. To reach these objectives, we propose the implementation of a multi-agent integrated system that will have two main functions:

- Cooperate to maintain a dynamic global knowledge of the resources utilization in the platform through a widely host resources monitoring;
- Take decisions to: reduce the response time of the final users and execute preventive fault actions through dynamic distribution/migration of the services.

To allow the execution of dynamic distribution/redistribution of services on the environment and to hide the load-balance mechanisms of the programmer, we also propose the integration between the multi-agent system and the existent middleware mechanisms of the RE-INVENTE Platform. These mechanisms are inside the *ORB* (Object Request Broker) and the *middleware services*.

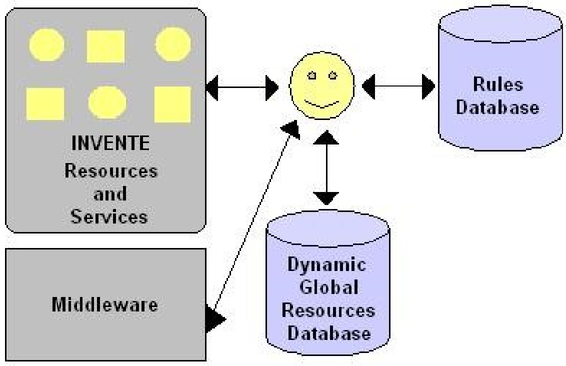

Figura 4: Intelligent Distributed Architecture.

As it can be seen in the Figure 4, the multi-agent system interacts with a group of distributed architecture elements. The multi-agent system will be responsible for monitoring resources, taking decisions and execute actions in the middleware.

To allow the migration and/or the replication of services a *mobile objects technology* Nelson (1999) will be used. This mobility will facilitate the load balancing in the distributed system and the use of preventive fault tolerance mechanisms through the objects replication and migration in the environment.

3.2 Multi-agent System Composition

How it can be seen in Figure 5, the QoS Multi-Agent System maintains a database with information about resources, objects and links of an entire domain. This database is called Dynamic Global Resources Database

(DGRD). This information is dynamically updated through the cooperation between the agents system. Based on the rules database and the resources utilization, the agents can take decisions and execute actions over the distributed architecture in order to accomplish a better QoS to the user.

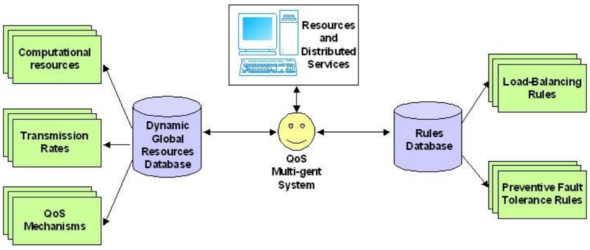

Figura 5: The QoSMultiagent System.

The proposed Multi-agent System (Figure 6) is composed of five different kinds of agents. The "Global Monitoring Agent" (GMA) cooperates with the others system agents in order to maintain the DGRD always updated. The resources information is based on thresholds. The "Local Monitoring Agent" (LMA) monitors the local resources (memory, processor utilization, etc.). When a local threshold is reached, the LMA informs to the GMA. In this way the Domain Manager has a global vision of the system resources.

The Decision Agent (DA) monitors the global utilization of resources in the system. The DA can take decisions based on the "DGRD" and on the "Rules Database" and determine a set of actions into the distributed architecture. A DA decision may imply a set of actions. Each action is delegated to a specific Intervention Agent (IA). The Intervention Agents are responsible for performing the actions in order to reach a better QoS level in the system.

The User Interface Agent (UIA) allows updating the Rules Database. The Rules Database can be updated based on the human knowledge. The monitoring of effects of the RA decisions in the achieved QoS system levels can be made to determine the rules efficacy.

4 BIG-INVENTE, THE PROPOSED APPROACH

Big data technology make possible to mine learning information for insights regarding student performance and learning approaches. It refers to the large amount of information that flows through various channels. Following Doyle (2014), The term "Big Learning Data" encompasses three aspects of learning data: volume, velocity, and variety:

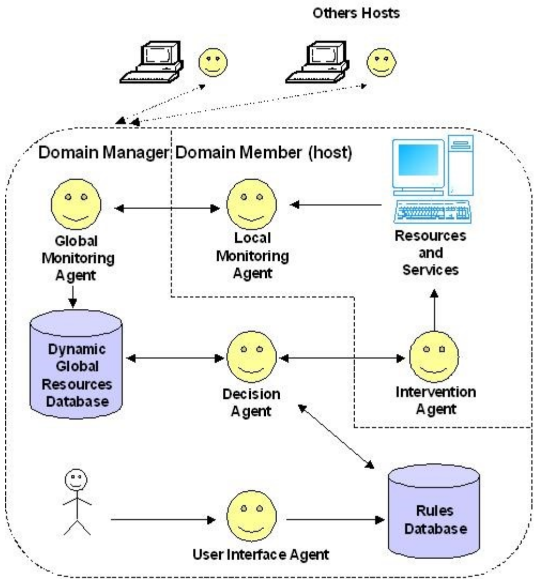

Figura 6: The Multi-agent-based System composition.

- Volume: Big Data can yield information about thousands of learners taking the same course or having the same instructional experience. It can also shed light on multiple data points, over time, about a single learner.
- Velocity: Big Learning Data enables learners and organizations to have rapid access to data even in real time. Velocity instantly would provide her with remedial and enrichment options based on her historical learning patterns and successful strategies from thousands of other learners who also failed that question.
- Variety: Big Learning Data connects the dots, weaving together a wider variety of information from students with different backgrounds. It allows us to see the correlations between performance and environment. Without it, we have traveling expenses and limited representation.

In this section, we describe a machine learning approach to model both the learner and the teacher in order to transform the INVENT in a fully adaptive learning platform. The main idea (figure 7) is that the student model (profile) is constructed from interactions with the system and the model of teaching is constructed from collected and selected web data (big data).

4.1 Pattern Building from Students' Interactions

First of all, it is necessary to build data sets from student actions in the system. According to their performance, they will be grouped together by a clustering non-supervised algorithm. For each group created, the system will build the model of teaching. This model consists in finding necessarily heterogeneous sources and propose to the members of each group a sequence of training and evaluation until the group members decided to close this module and move to another.

The sources of data are obtained from intensive searches to be performed by the search engines specially developed for this purpose.

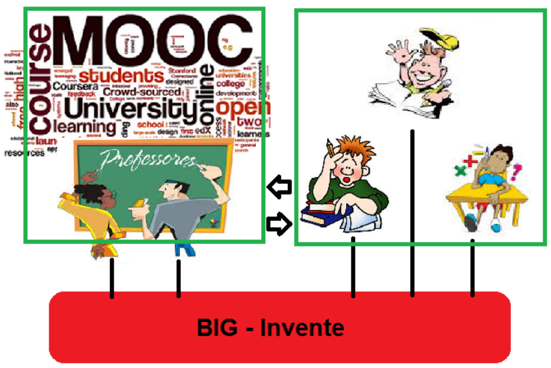

Figura 7: The BIG-INVENTE Proposal.

4.2 Teaching Strategy and Feature Selection

Two learning strategies are available: supervised and unsupervised. In the first case, it will occurs human intervention by teachers in order to label the profile of the students and consequently the group to which they belong and also to accept or to reject the contents found by search engines (web crawlers).

In the second, it will not occur human intervention at all. It is a self-learning system. Users themselves shall terminate the learning process and select learning elements like presentations and assessments they must undergo.

The system has a mechanism for the continuous building of patterns from the actions of group members, as well as evaluations (tests) and the scores obtained.

By this way, it is expected that there will be an increase in the speed and efficiency of the learning process by the use of an automated system, which will be able to classify the student and, in the other side, find the most suitable contents for him.

5 CONCLUSION

This paper presented the INVENTE project evolution. The first version used a Critical Dimension concept that take in account objective and subjective aspects in order to the define network technical parameters. The second version, called RE-INVENTE, had virtual rooms on the distributed system context. The next generation of the project was a oriented Multi-agent System.

More recently, the project makes good use of the Big Data technology. It's the BIG-INVENTE, a framework for distance learning that mines learning information to model both the learner and the teacher in order to transform the INVENT in a fully adaptive learning platform.

The main idea of this new proposal is that the student model (profile) is constructed from interactions with the system and the model of teaching is constructed from collected and selected web data (big data).

Big Data was the great topic on the 2013 World Innovation Summit for Education (WISE) in Quatar. Leading educators, policy makers, and governments from over 100 countries. Among the conclusions: the way learning happens today will not adequately prepare young people for the world of tomorrow and "big data, and the disruptions it can lead to, have led to one of the most creative periods in history in terms of innovation."

The idea with BIG-INVENTE version is to improve theadaptivelearning of the INVENTE framework. For that, we presented in this paper a novel approach aiming at predicting student's profile. A prototype is been implemented and the methodology consisting of labeling feature vectors was already tested with great success.

REFERÊNCIAS

CHEUNG, L.-S.; KWOK, Y.-K. The design and performance of an intelligent load balancing service. *Parallel Processing Workshops*, 2001.

DAMIANI, E. An intelligent load distribution system for corba-compliant distributed environments. *Fuzzy Systems Conference Proceedings*, 1999.

DOYLE, A. C. *DOYLE*: Site. 2014. In: <http://www.innovationexcellence.com/blog/2014/01/29/big-data-in-education-big-potential-or-big-mistake/>. 2014. Disponível em: <http://www.innovationexcellence.com/blog/2014/01/29/big-data-in-education-big-potential-or-big-mistake/>. Acesso em: 2014.

FELBER, P.; JAI, B.; RASTOGI, R.; SMITH, M. Using semantic knowledge of distributed objects to increase reliability and availability. *Object-Oriented Real-Time Dependable Systems*, 2001.

HAZEMI, R.; HAILES, S.; WILBUR, S. The digital university - reinventing the academy. *Springer-Verlag London Limited*, n. 2, 1999.

JUN, k.; BLNI, L.; PALACZ, K.; MARINESCU, C. C. Agent-based resource discovery. *Computer Sciences Department, Purdue University*, 1999.

LESSER, V. R. Multiagent systems: an emerging subdiscipline of a. *ACM Computing Survey*, n. 3, 1995.

MARRON, P. J. Design of a middleware service for scalable wide area network applications. *Distributed Objects and Applications*, 2000.

MOURA, C. O. Concepção e especificação de um sistema ibw para educação a distância. *Dissertação de Mestrado, DEE-UFC*, 1999.

NELSON, J. Programming mobile objects with java. *John Wiley and Sons, Inc.*, 1999.

OLIVEIRA, A. M. B.; GAITI, D.; CASTRO, M. F.; PUJOLLE, G. Re-invente: Un environnement multimédia pour l'enseignement technique à distance. *Proceedings of GRES 2003 (Gestion de Réseaux et Services)*, 2003.

OMG. Common object request broker: architecture and specificqtion revision. *OMG Document*, 1997.

RAJAGOPALAN, A.; HARIRI, S. An agent based dynamic load balancing system. *Autonomous Decentralized Systems*, 2000.

SERRA, A. B. Uma solução de distribuição para aplicações multimídia de tempo real no contexto da educação tecnológica a distância. *Dissertação (Mestrado) - Departamento de Computação da UFC, Universidade Federal do Ceara*, 2001.

SOARES, J. M. Um sistema de gestão para a educação tecnológica a distância, projeto e implementação. *Dissertação (Mestrado) - Departamento de Computação da UFC, Universidade Federal do Ceara*, 2001.

SYCARA, K. P. Multiagent systems. *In AI Magazine*, n. 2, 1998.

TANENBAUM, S. A. *Distributed Operating Systems*. Natick, Massachusetts: Prentice HALL, 1995.

TENE, O.; POLONETSKY, J. Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property*, n. 1, p. 490–495, 2013.

TURGUT, D.; TURGUT, B.; DAS, S. K.; ELMASRI, R. Balancing loads in mobile ad hoc networks telecommunications. *ICT 2003. 10th International Conference on*, 2003.

ZHOU, Y.; AHMAD, H. F.; ARFAOUI, H.; MORI, K. Autonomous information allocation through mobile agents to achieve load balancing in distributed information service system. *Autonomous Decentralized System*, n. 7, p. 35–41, 11 2002. The 2nd International Workshop on.