Prototype of assignment intelligent adaptive of service providers inside of ESB with data mining

Andrés Paolo Castaño
[1] *Computer Science Department –University of Caldas, Colombia*

***Abstract* —The SOA philosophy can address new business challenges, become more competitive and provide integrated information systems. In addition, technologies such as BPM, BAM and Web Services are essential complements to SOA. This work aims to use several of these technologies integrated into a single application that will allow in a phase of a previously defined business process, to perform the analysis of input suppliers to the company through the generation of a decision tree using embedded code of the free tool Weka for data mining in order to feedback the business process and evaluate these results to improve the process. For the realization of this prototype we worked with the jBPM suite, the API from Weka to get the J48 algorithm, the postgresql database, the format for data exchange JSON and the web service.**

***Keywords*— Business Process Management (BPM), Data mining, Enterprise service bus (ESB), jBPM, Services Open Architecture (SOA), Web services.**

I. INTRODUCTION

THE evolution of information technologies have had, in recent years, linked to different aspects of the global economy that requires companies to be more competitive and to focus efforts on reducing costs at all fronts in order to maximize utilities, have enabled a constant search and optimization of business processes. This field brought up a revolution in the late 90's, on how to coordinate the internal activities of organizations, has changed dramatically, expanding at a B2B environments that facilitate collaboration among different organizations and today are used in environments designed according to the paradigm of the computing oriented service.

In this context, the business process management systems (BPMS) arise and the workflow management systems thanks to its property of reuse can develop the logic of a business process adaptable to any domain. Although these tools can allow virtually all business process, there are steps in these processes which still remain dependent on human performance; in these scenarios spot out the idea of contributing to the automation of a business process that requires the selection of input suppliers, this task is performed by a human workflow and its result is feed back to the business process for continuation. Here lies the motivation of this work, building a prototype for the allocation of suppliers, using different technologies, on a hand a BPMS, which contains all business process logic and on the other, a data mining application that use a decision tree for this selection, in this process, the data are stored in a database, the data exchange with the different technologies is done with standard formats like XML or JSON. The approach of this solution is made possible by relying on the oriented-service architecture that allows applications developed with different technologies to share data and functionality supporting the creation of new systems with separate applications working together in a decoupled way. The paper looks at related works, indicates the different technologies included for the completion of the prototype and shows the proposed architecture and its implementation with the difficulties encountered and finally, the conclusions and future work.

This article is part of the thesis research project mentored Ph.D. program in Computer Engineering from the University of Salamanca. Director: Dr. Jesús Soto Carrión.

II. CONCEPTUALIZATION AND WORK RELATED

This paper aims to show as possible the application that integrates different technologies and get results in transparent way for the user, i.e., we will indicate as from a suite BPM (jBPM) may involve a process of data mining with a call from a stage of a business process and get the results at the stage of the process that made the call.

There are several works that have been focused on the topic of BPM and data mining, but not as an integrated application but as separate processes that provide services. The work presented by [1] shows how starting from a process model as it is possible to discover by conventional process mining algorithms, they analyze how data attributes influence the choices made in the process based on past process executions, aims at the detection of data dependencies, also, describes how machine learning techniques can be leveraged for this purpose. Likewise, it is the work presented by [2] that describe the design and implementation of a system that manages data mining model assets of an organization that can

support business processes in making real-time decisions and forecasts. This same author presented the work [3] which describes a data mining model management system that addresses the challenges of model aging, management scalability, timely-communication among parties on model changes, semantic gap on interpreting models, and business process integration to support sustainable and operationalized.

There is also the work proposed by [4] which identifies the challenge of case prediction, which for a specific case under the control of a BPMS deals with the estimation of the remaining time until it is completed. An accurate case prediction facility is a valuable tool for the operational control of business processes, as it enables the pre-active monitoring of time violations.

Conceptually, we rely on the Service Oriented Architecture (SOA) and its functioning through services as fundamental elements to integrate and develop applications, in regard to this context we present a general definition developed by [5] "A software architecture as a set of definitions that describe the software components and assigned the functionality of the system components to this components. Describe the technical structure, constraints and characteristics of the components and interfaces between them".

In the same vein, we discuss the concept of Business process modeling and its importance in the development of any industry and as an improvement in a business process positively affects the production and consequently, the products developed, reviewed the work submitted by [6].

Emphasis was placed on the characteristics that differentiate a BPM of a WfMS, at the stage of diagnosis and definition of processes, aspects neglected in a WfMS [7].

The trend today is toward a process-oriented paradigm, where applications should cover the entire business enterprise and the tools are the Business process management systems (BPMS). Establishing the characteristics that must be satisfied the BPMS, will be the centerpiece of development of this project. The BPMS selected must provide:

Process modeling: to capture business requirements in its initial stage and make it available for the rest of the development process.

Running processes: the process execution engine of the BPM system, import the modeling process (defined using BPEL) and then runs and manages instances of processes to meet operational requirements.

Process monitoring: This capability includes the summary view of running processes, of the completed, see states of processes, suspend and resume processes, give warnings and reallocation processes.

Business Activity Monitoring (BAM): analyzes the events generated by business activity and allows give metric.

Afterward, for the integration of different applications was covers the work of [8] about an ESB and the idea of a standard based on the architecture for the integration of heterogeneous systems as proposed by JBI [9], also we discussed the architecture of service components, based on [10] and the use of service data objects (SDO) in accordance with [11].

III. DESCRIPTION OF PROTOTYPE

As stated at the outset of this paper, the idea is to adapt in transparent way an enterprise business process that is modeling in a BPM system, this business process has a stage of vendor selection that used a decision tree generated through free data mining tool Weka, the result is stored in the database and fed back to the stage of the business process where will make the best decision based on the results obtained, will be used in this coupling, universal communication standards and Web services.

Initially different open source applications were evaluated to perform business process modeling, Intalio, Bonita, Enhydra Shark and jBPM, due to jBPM is one of the most robust solutions with a large and growing community, were decided to select this tool. jBPM is a component of large set of business solutions that composing JBoss.

To generate the decision tree was analyzed the API of the data mining tool of the University of Waikato in New Zealand to extract the necessary collection of libraries that make up the J48 algorithm that is an adaptation of the C4.5 and to embed in the application. Weka (Waikato Environment for Knowledge Analysis) is free software released under the GNU-GPL.

The input data for the evaluation of the decision tree are stored in the database postgresql and delivered to the algorithm using the JSON format for data exchange between database and weka. Because weka uses a file structure called ARFF, acronym of Attribute-Relation File Format it became necessary to convert the JSON format to ARFF, so that data with the criteria for the selection of suppliers delivered to the business process are exchanged between different technologies transparently.

Additionally, we included in this prototype a Web service that is responsible for making available the entire business process management to be consulted by a client through an interface created for this purpose.

IV. PROTOTYPE ARCHITECTURE

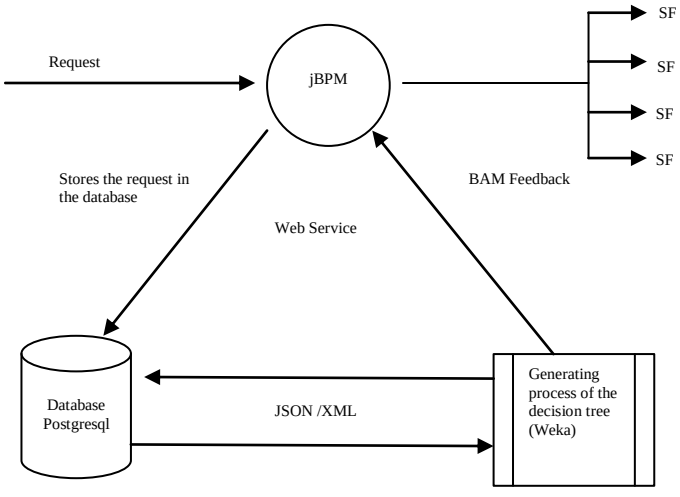

Fig 1. Prototype architecture

A simple business process was designed on the jBPM suite and one of the nodes was associated with an action (action handler) that invokes the process of evaluating different suppliers, in this process were taken the data associated with each potential supplier and stored in the Postgresql database in the table movements. These data are transformed into JSON format in order to make the exchange with the decision tree algorithm. After generating the decision tree results in a confusion matrix was added which contains the number of manufacturers of each class meeting the criteria to be suppliers, the suppliers for each class are taken and carried back to the database, as many tables as suppliers are, tables were created and stored by the supplier and their amount, subsequently, these data are carried to business process node and consulted through a web service.

The following diagram shows the flow between the different components that forming the prototype

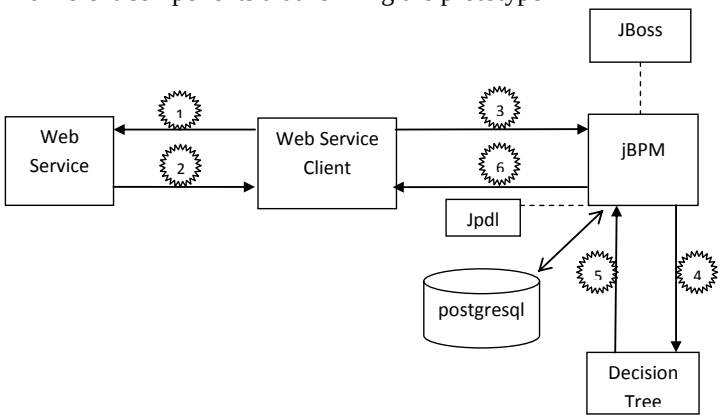

Fig 2. Flow diagram of the prototype

In the user interface is requested read from the Web service, the active business process.

The server executes the request and sends to the user, the defined business process.

The user begins the process by invoking jBPM, with the business process that owns the supplier selection stage.

The information with the criteria for supplier selection is processed in the decision tree using the J48 algorithm.

The resulting confusion matrix is returned to the business process using JSON.

The result is displayed in the user interface and is stored in the database postgresql.

V. PROTOTYPE IMPLEMENTATION

For the integration of the different components, is starting with the module of design of business process of the tool jBPM, here we implemented each of the elements that form the integrated system, using the design platform Eclipse, which is achieved by obtaining a modular system that integrates the components of BPM, Data Mining, JSON and Web services.

The implementation is performed through object-oriented system; this construct consists of a model of classes organized by packages and development projects.

Development projects comprise the following elements:

- jbpmAnalisis
- Mineria
- WebServiceProject

A. jbpmAnalisis

The JbmpAnalisis project is organized in such a way that runs through a previous modeling, a business process. This design is modeled in jpdl, Figure 2 represents the case implemented in this suite the file containing the process is processdefinition.xml.

As general structure a project is proposed having as main element the clsProveedor class, which, by the method activaProceso() starts reading the business process for its further implementation.

As shown in Figure 2, the clsProveedor class uses the definition of jBPM in the MessageActionHandler as the relation of the messages sent between nodes. This is the fundamental basis for the call of the procedures to ensure data mining, which are conducted through clsMineria class.

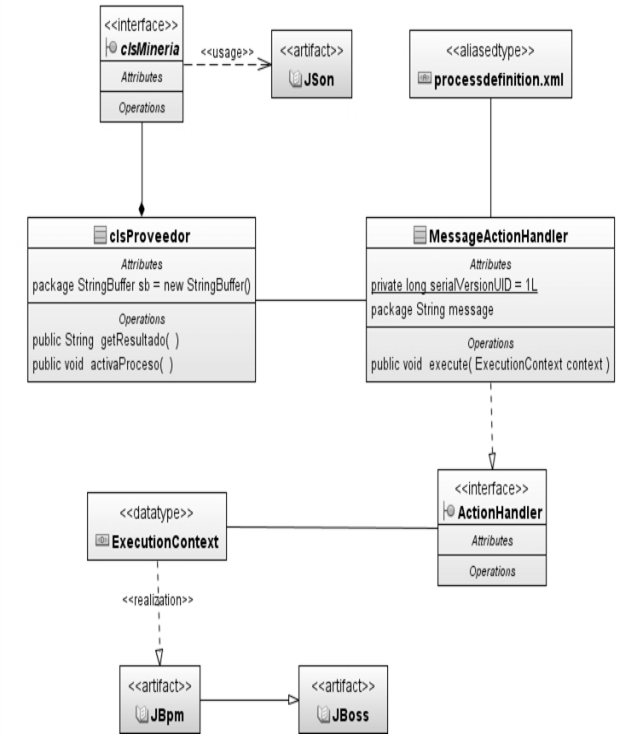

Fig 3. jbpmAnalisis Project

B. Mineria

For the analysis and selection of suppliers, was implemented the J48 algorithm of Weka tool. Its operation is based on the clsMineria class. As shown in Figure 3, the method convierteSB () is used by a web service to publish the results when requested by the client.

For the storage and exchange of data is uses postgreSQL database. The connection to this database is done in the class clsMineria() by the method genera(), which is also responsible

for adapting the Json standard to the ARFF format, which requires weka for its operation.

After generating the decision tree, is obtained a confusion matrix that is stored again in the database, generating a table for each manufacturer. This information will be returned to the node designed in the business process using clsProveedor class.

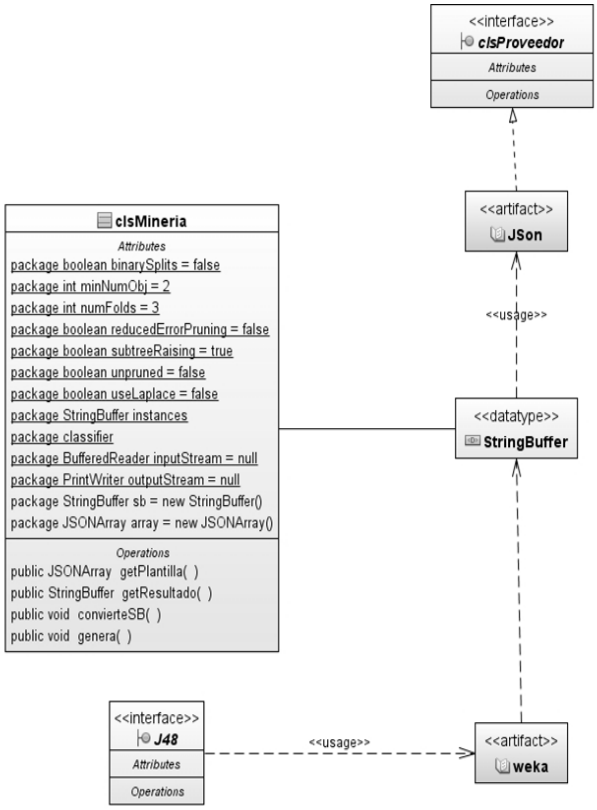

Fig 4. IA Project

C. WebServiceProject

After generating the processing elements of the business process and data mining, must be created a project that offers through a Web service, the results requested by a particular customer.

The publication of the service is generated via file named clsMain.wsdl, its structure is published offering the generaBPM operation, which allows access through the Endpoint clsMain. The access structure to the Endpoint generated as a service is invoked by the client using classes clsMainService () and clsMainServiceLocator ().

VI. RESULTS

When generating the request in the business process that has been diagramed in jBPM, at the initial some data are sent that entering the movements table of the database. This table has the following fields (manufacturer, part, preciomin, preciomax, time) with this data is generated decision tree, based on the criteria preciomin, preciomax and time. The class to evaluate is manufacturer.

The result of the decision tree, delivers the confusion matrix discriminated by the manufacturer with the number of manufacturers that meet the criteria, i.e. we assume the tree is provided with the following information:

('intel', 'processor', '100', '300', '5'),
('amd', 'processor', '100', '300', '7'),
('nvidia', 'processor', '100', '300', '7'),
('via', 'processor', '100', '300', '8'),
('intel', 'mainboard', '200', '400', '7'),
('amd', 'mainboard', '200', '400', '6'),
('nvidia', 'mainboard', '200', '400', '5'), and so on, the confusion matrix will deliver the number of manufacturers that meet the criteria for be potential suppliers.

a b c d <-- classified as
1 7 3 4 | a = intel
2 1 4 6 | b = amd
3 5 3 4 | c = nvidia
4 3 3 4 | d = via

From here we obtain the total per manufacturer, i.e. for the manufacturer intel are 15 potential suppliers, to the manufacturer amd will have 13 potential suppliers, etc. These data are in ARFF format are transformed into JSON format to be stored in the database by creating a table for each manufacturer, the table name is the name of the manufacturer.

These same data are presented in the stage of the business process that generated the request for evaluation with the decision tree using for this purpose a web service with Apache Axis.

VII. CONCLUSIONS AND FUTURE WORK

In this paper we proposed a prototype that would improve the performance of a business process using data mining techniques, as seen through various open technologies, can integrate different applications using standard formats for exchanging data between two different applications into a single integrated system.

Although the integration of different applications is demonstrated, the results obtained returned the number of the suppliers delivered to the decision tree, meet the criteria for a possible selection but does not indicate which of them fits best. Nevertheless it may be advised to use a different technique or supplement along with the decision tree to carry the this evaluation out.

Undoubtedly, the principles of SOA have allowed the development and promotion of this area, however, for this project the technology of Web services has made the system more complicated and slower. This is critical if we think of companies working to support real-time business.

Companies are focusing on designs make their business processes but it is important to determine whether the process conceived and designed fitting adequately to the needs for which it was generated, process mining enables us to discover the process itself, through the recording and analysis of information generated by events and determine their weaknesses, composition, execution and allows us to grow and strengthen the process and its management.

The business world and the world of IT can relate best with the incorporation of a semantic processing, combining semantic web and semantic web services with BPM [12]. SBPM Systems.

REFERENCES

[1] A. Rozinat y W.M.P. van der Aalst, "Decision mining in business processes," BPM Center Report BPM-06-10, BPMcenter. org, 2006.

[2] I. Ari, J. Li, J. Jain, y A. Kozlov, "Management of Data Mining Model Lifecycle to Support Intelligent Business Services". 2008

[3] I. Ari, J. Li, A. Kozlov, y M. Dekhil, "Data Mining Model Management to Support Real-time Business Intelligence in Service-Oriented Architectures". 2008

[4] H.A. Reijers, "Case Prediction in BPM systems: A Research Challenge," Journal of the Korean Institute of Industrial Engineers, vol. 33, 2006, págs. 1–10.

[5] D. Krafzig, K. Banke, y D. Slama, Enterprise SOA: Service-Oriented Architecture Best Practices (The Coad Series), Prentice Hall PTR Upper Saddle River, NJ, USA, 2004.

[6] J.G. Molina, M.J. Ortín, B. Moros, J. Nicolás, y A. Toval, "De los Procesos del Negocio a los Casos de Uso," Jornadas de Ingeniería del Software y Bases de Datos. Pág, 2000, págs. 103-116.

[7] W.M.P. Van der Aalst, A.H.M. Hofstede, y M. Weske, "Business process management: A survey," Lecture Notes in Computer Science, 2003, págs. 1-12.

[8] S. Desmet, B. Volckaert, S. Van Assche, D. Van Der Weken, B. Dhoedt, y F. De Turck, "Throughput evaluation of different enterprise service bus approaches," Proceedings of the 2007 International Conference on Software Engineering Research in Practice (SERP'07), 2007.

[9] S. Vinoski y I. Technologies, "Java business integration," IEEE Internet Computing, vol. 9, 2005, págs. 89-91.

[10] M. Beisiegel, H. Blohm, D. Booz, J.J. Dubray, A. Colyer, M. Edwards, D. Ferguson, B. Flood, M. Greenberg, y D. Kearns, "Service Component Architecture. Building Systems using a Service Oriented Architecture," BEA, IBM, Interface21, IONA, Oracle, SAP, Siebel, Sybase, white paper, version 0.9 edition, 2005.

[11] R. Weaver y S. Specialist, "The Business Value of the Service Component Architecture (SCA) and Service Data Objects (SDO)," IBM Whitepaper, November, 2005.

[12] Z. Laliwala, R. Khosla, P. Majumdar, y S. Chaudhary, "Semantic and Rules Based Event-Driven Dynamic Web Services Composition for Automation of Business Processes," IEEE Services Computing Workshops, 2006. SCW'06, 2006, págs. 175–182.