APESS - A Service-Oriented Data Mining Platform: Application for Medical Sciences

Mohammed Sabri
University of Sciences and Technology-Mohamed Boudiaf, Computer Sciences Department, Oran, Algeria
E-mail: mohammed.sabri@univ-usto.dz

Sidi Ahmed Rahal
University of Sciences and Technology-Mohamed Boudiaf, Computer Sciences Department, Oran, Algeria
E-mail: rahalsa2001@yahoo.fr

Abstract—The domain medical and public health remains the principal preoccupation of all world population. It makes recourse at several means from various disciplines, including for instance epidemiology, to help clinicians in decision processes. This paper proposes an Assistance Platform for Epidemiological Searches and Surveillance (APESS) for service-oriented data mining in the field of epidemiology. The main aim of the present platform is to build a system that enables extracting predictive rules, flexible and scalable for aid in decision-making by trades' experts. Results showed that the current system provides prediction models of chronic diseases (epidemiological prediction rules), using classification algorithms.

Index Terms—Knowledge Discovery, Services Oriented Architecture (SOA), Web Services, Data Warehouse, Data Mining, Rules Discovery, Medical sciences.

I. INTRODUCTION

The development of information systems and computer technologies has enabled the automation of the activities in every field of the real-world; this has induced a fast increase in the information available, the development of high volume data warehouses and finally, the emergence of Data Mining. The latter corresponds in a set of techniques and methods which from the data (typically stored in a data warehouse) extract usable knowledge in various fields such as environment, public health, pharmacy, biology, etc. However, the growing market draws attention to distributed Data Mining [1] because data and software are geographically distributed over a network instead of being located in a single site. Moreover, the cost is another reason for the distribution: data mining in distributed environments would only increase the cost of licenses because of the need for multiple copies of different types of tools [1]. To optimize investment, users prefer to use components that respond to their specific needs. However, since the arrival of Web and Grid computing, distributed data is now much easier to access. Furthermore, distributed computing in heterogeneous environments has become much more feasible [2]. At the same time, service-

oriented architectures (SOA) are becoming one of the main paradigms for distributed computing [3]. SOA provides solutions for integrating diverse systems that support interoperability, loose coupling and reuse. To full-fill clients need one service invoke another service/services. It is possible that there is some evolution among these external services [4].Through an approach based on services, especially service-oriented architecture (SOA), integrated services can be defined to support the distributed data mining tasks in grids and the Web. Such services can address most aspects taken into account in data mining and knowledge discovery in databases (KDD). Moreover, the most important SOA implementation is represented by web services [5]. The popularity of Web services is mainly due to the fact that they are standardized (adoption of universally accepted technologies, such XML, SOAP, HTTP, WSDL, UDDI, …) [5]. Web Services are simple, flexible and independent from both platforms and languages. Furthermore, their adoption by a number of communities, including the grid community, indicates that the development of data mining applications based on Web services is likely be useful to an important user community [6]. Such Web service is particularly met in business environments where time and data intensive transactions are performed between customers and offered services/products [7]. Regarding our field of interest, namely epidemiology, it is worth noting that data mining has attracted a great deal of attention in the epidemiological information processing, as well as in medical and public health areas as a whole.

As part of our study, we are interesting by the above field by studying health problems currently affecting an important part of the Algerian population: chronic diseases. Indeed, an important issue in the field is to relate a specific chronic disease, for example asthma or diabetes, with physiological and environmental factors. The available data in the field are complex, heterogeneous and uncertain, and it is not easy for epidemiologists to produce predictive rules linking for example asthma or diabetes with these internal and external factors. The main aim of our study will be to provide to experts in the field prediction models for two the above mentioned chronic diseases (asthma and

diabetes), which have been the subject of several studies nationally and represent a major public health problem [8]. Predictive variables will include two physiological parameters (age and gender) and two additional environmental parameters that will give us information about geographical space (province/department) and time (period in the year).

To this end, we set up a service-oriented data mining platform applied to drug sales data. This work was done in collaboration with experts from the public health in the West region of Algeria including pharmacists, doctors, social security, and the laboratory of Biostatistics at University of Oran 1 - Ahmed Ben Bella. An Assistance Platform for Epidemiological Searches and Surveillance (APESS) is proposed based on both: (i) data storage and preprocessing, and (ii) predictive rule extraction in the field of epidemiology.

II. RELATED WORK

There have been many studies published and aimed at applying data mining to service oriented architecture. Shaikh et al. [9] have presented a Web services-based toolkit for supporting distributed data mining called FAEHIM (Federated Analysis Environment for Heterogeneous Intelligent Mining). This toolkit consists of a set of data mining services, a set of tools to interact with these services, and a workflow system used to assemble these services and tools. Chiu and Tsai [10] have introduced a dynamic data mining process (DDMP) system based on service-oriented architecture (SOA). Each activity in data mining process is viewed as a Web service operated on Internet. Furthermore, the Web services are dynamically linked using Business Process Execution Language for Web service (BPEL4WS) to construct a desired data mining process depending on user's requirement. Chen et al. [11] have proposed architecture of metadata service of data mining based on Common Warehouse Metamodel (CWM). These authors divided the service into two types which provide service respectively by analyzing the feature of data service and the services needed by the data mining system. These authors divided the service into two types which provide service respectively by analyzing the feature of data service and the services needed by the data mining system. Xu et al. [12] proposed a service based architecture for data mining applications, including configuration service, service engine, monitor service, analysis service, visualization service, computing service, and data and algorithm provision service. Talia et al. [13] have described the design and the implementation of Weka4WS using the Web Services Resource Framework (WSRF) libraries and services provided by Globus Toolkit. Moreover, the authors presented a performance analysis of Weka4WS for executing distributed data mining tasks in different network scenarios. Ferreira et al. [3] have developed a service-oriented architecture for data mining (Anteater) that relies on Web services to ensure extensibility and interoperability. Their architecture provides simple abstractions for users, and

supports computationally intensive processing on large amounts of data through massive parallelism. Latha et al. [14] proposed a novel method to develop service oriented architecture for a weather information system and forecast weather using data mining techniques. The authors added that the proposal aims at developing a weather information system as a web service that can be used by any type of application and uses the prediction techniques of data mining for weather forecasting. Birant [5] has proposed a flexible service-oriented data mining architecture (SOMiner) that incorporates the main phases of knowledge discovery process. Overall, this architecture is composed of generic and specific web services that provide a large collection of machine learning algorithms conceived for knowledge discovery tasks such as classification, clustering, and association rules, which can be invoked through a common graphical user interfaces (GUI). Zorrilla and Garcia-Saiz [15] proposed a model which joins both facets: data mining and SOA. It describes a data mining service addressed to non-expert data-miners that can be delivered as Software-as-a-Service. The advantage of this model is that the service itself is able to perform all the process by simply indicating where the data file is located. Shelke et al. [16] have proposed an architecture that improves the mobile data mining techniques; in this case, data retrieval for mobile devices should be faster in efficient mobility management using proper web services. Kester and Sam-Aggrey [17] have proposed an Open System Architecture Platform for Big Data based on open standards and with the concepts of service oriented architecture. The authors have claimed that the concepts of SOA with web services did not only reduce the amount of deployed code, but it also reduced the management, maintenance. Also, the services and applications can be created quickly and easily used with a combination of new and old services. Data can easily be mined and pulled for analysis due to the open nature of the architecture engaged and all the open principles with SOA. In paper [18] there is a survey that gives the application of web service for data mining also we build a data mining model based on Web services and going forward it is possible to build a new data mining solution for security according to the prototype of a dynamic web service based data mining process system.

The APESS platform's concept proposed in the present paper focused on service-oriented architecture based on data mining and applied to Epidemiological Searches and Surveillance.

III. PROPOSED APPROACH

Our approach, as illustrated in Fig. 1, is structured into two phases. First, Phase I includes the data warehousing that is part of a process of knowledge discovery in databases (data integration) [19] and the preprocessing preparing exploitable data in the data mining process [20]. Second, Phase II, consists in the realization of the platform APESS, for data mining based on service-oriented architecture, to extract epidemiological predictive models, based on the exploitation of data
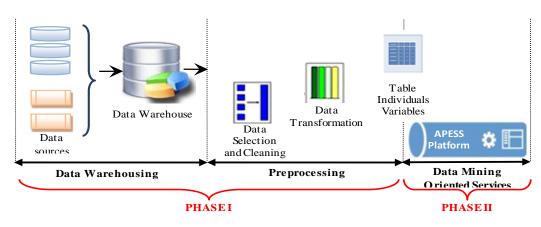
Fig.1. General Architecture of our approach.

issued from the first phase. These two phases are presented respectively in A and B sections.

A. Data Warehousing and preprocessing for APESS platform (PHASE I)

The first phase of our approach is composed by two parts.

Firstly, the design and the realization of the data warehouse, with the objective of obtaining a unique source of data to carry out the data preprocessing tasks for APESS platform. For this, the architecture of the data warehouse (See Fig. 2.) is articulated around three axes: Integration, Building, and Structuring.
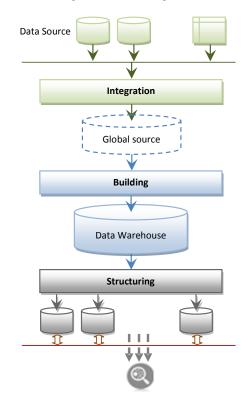

Fig.2. Architecture of data warehouse and preprocessing (PHASE I).

The integration step allows to extract and to group data coming from different data sources [21]. These are recovered, coded, and stored in the global source. The

building step consists in extracting the relevant data and copying it in the warehouse [21] based on the multidimensional concept (star model) [22, 23]. (See Fig. 3).
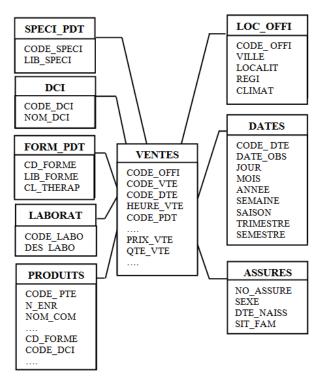

Fig.3. The Star Model of our data warehouse.

Finally, the structuring step is aimed at reorganizing the data in the warehouse in the Data Marts [24]. It is worth noting that the data warehouse consists of a centralized collection of materialized and historical data [21].

Secondly, data from the warehouse is very diverse and heterogeneous and is not necessarily entirely exploitable by the data mining techniques [20]. To circumvent this issue, the preprocessing step is needed to obtain exploitable data, especially in the form of tables of individuals/variables. To this end, the following steps were used: (i) Data selection [19], that applies filters allowing us to select a subset of rows and/or columns, (ii)

Data cleaning [26], in order to process missing data (suppression of records - enrichment by external sources was also carried out during the creation of data warehouse [25]), finally, (iii) Transformation [19] (attributes) and dimension reduction in order to obtain reduced data groups.

B. APESS platform: A Service-Oriented Data mining (PHASE II)

Exploitable data from Phase 1 are integrated in Phase II. (See Fig. 4).

Data mining tasks of the APESS Platform include several basic services. Each of these services is adapted to a specific usage context. The process of services selection and composition is based a priori on predefined rules.

The APESS platform is a runtime environment that is designed and implemented according to a multilayer structure: Data Access Services Layer, Data Mining Services Layer and User Services Layer:

- The Data Access Services Layer (**DASL**) ensures the publication and searching of data, outcome from PHASE I, to be mined in our APESS platform, as well as handling metadata describing available data. In other words, the DASL is responsible for the access interface to data sets and all associated metadata.
- The Data Mining Service Layer (**DMSL**) is a

fundamental layer in our APESS platform. This layer is composed of generic and specific web services that provide a large collection of machine learning algorithms written for knowledge discovery tasks. In DMSL, each web service provides different data mining tasks including among others classification, clustering, association rule mining. They can be published, searched and invoked separately or consecutively through a common GUI.

- The User Services Layer (**USL**) provides the user interaction with the system and offers facilities for visualizing the extracted knowledge models by the visualization services module. In addition, users may want to make specific choices in terms of defining and configuring a data mining process such as algorithm selection, parameter setting, and preference specification for web services used to execute a particular data mining application.
- Finally, the Enterprise Service Bus (**ESB**) is a convenient middleware technology to apply SOA. Also, it is aimed at both: (i) ensuring the interconnection and (ii) managing the mediation of communications and interactions between services. Since the ESB is solving all integration issues, each layer in our APESS platform only focuses on its own functionalities.
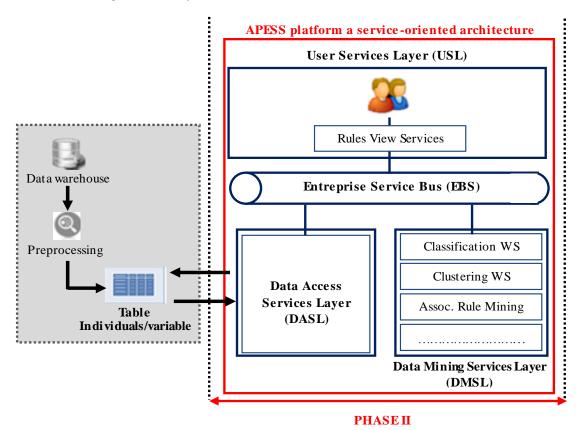

Fig.4. Data Mining Services process in our APESS platform.

IV. EXPERIMENTS

A. Data description (PHASE I)

In our study, we used our APESS platform for extracting epidemiological prediction models for chronic diseases (asthma and diabetes) in the western region of Algeria. To this end, the data used to evaluate our approach came from data records of sales drug in pharmacies (over thirty million records), spread out between January 2003 and September 2014 in 450 pharmacies distributed over 4 departments. As stated above, data warehousing and preprocessing (Phase I) were required to select adequate pertinent data with our approach. Here, more than eight hundred thousand acts of sales concerning the diseases of interest were selected. These acts had an important number of redundancies of sales. This led us to involve experts from the epidemiological field, ending in a total of about 140,000 of prescriptions acts. A descriptive statistical analysis recommended by experts showed different prevalence (See Table 1) for classes' asthma (64.08%), diabetes (34.22%), and a third class obtained by combining these two diseases (asthma-diabetes combination, 1.70%). This last combination with a frequency lower than 2% was not selected because it was considered by experts as insignificant and was not of interest for the current study.

Table 1. Results of the descriptive statistical analysis of data

Diseases	Ratio
Asthma (AST)	64.08%
Diabetes (DBT)	34.22%
Asthma – Diabetes (AST-DBT)	1.70%

Table 2. Description of attributes and the class (Individuals/ Variables table)

Attributes	Signification	Possible values
X1 : MONT	The selected period is the calendar month	**01** (January)....**12** (December)
X2 : DEPART	The province/department where the pharmacy is located	**13** (TLEMCEN), **22** (BELABBES), **31** (ORAN), **46** (TEMOUCHENT)
X3 : GENDER	Patient's gender	M (Male), F(Female)
X4 : AGE	Patient's age - Age intervals	**Chi** (<=16) **Adt1** (between 16 and 40), **Adt2** (between 41 and 65) and **Old** (>65)
Y : THERAP_CLASS	Class of disease to be predicted	**AST** (Asthma), **DBT** (Diabetes)

Before starting the service-oriented data mining (PHASE II), we established the individuals/variables table coming from the data warehouse (See Table 2).

B. Results (PHASE II)

To evaluate our APESS platform and predict chronic diseases, we run as series of experiments on our database,

individuals/variables table with 135,228 individuals, and we rely on induction using APESS's oriented services algorithms. Table 3 presents the results obtained by data mining (APESS) using 2 different classification algorithms, namely ID3 [26] and C4.5 [27].

Table 3. Experiment results by ID3 and C4.5 algorithms

Methods Results	ID3	C4.5
Success rate (%)	88.1881%	88.1881%
Correct Instances	119,255	119.255
Rules	38	33

Although success rates across methods are similar, the C4.5 method produced less predictive rules (33 for C4.5 against 38 for ID3), which renders expert's task earsier. We give the 33 epidemiological prediction rules (conjunctive rules) produced by APESS platform (C4.5 algorithm).

1. *if (AGE = Old and GENDER = M) then AST*
2. *if (AGE = Old and GENDER = F) then DBT*
3. *if (AGE = Adt1) then AST*
4. *if (AGE = Adt2 and DEPART = 13) then DBT*
5. *if (AGE = Adt2 and DEPART = 22 and GENDER = M) then AST*
6. *if (AGE = Adt2 and DEPART = 22 and GENDER = F and MONT = 1) then DBT*
7. *if (AGE = Adt2 and DEPART = 22 and GENDER = F and MONT = 2) then AST*
8. *if (AGE = Adt2 and DEPART = 22 and GENDER = F and MONT = 3) then DBT*
9. *if (AGE = Adt2 and DEPART = 22 and GENDER = F and MONT = 4) then AST*
10. *if (AGE = Adt2 and DEPART = 22 and GENDER = F and MONT = 5) then DBT*
11. *if (AGE = Adt2 and DEPART = 22 and GENDER = F and MONT = 6) then DBT*
12. *if (AGE = Adt2 and DEPART = 22 and GENDER = F and MONT = 7) then DBT*
13. *if (AGE = Adt2 and DEPART = 22 and GENDER = F and MONT = 8) then DBT*
14. *if (AGE = Adt2 and DEPART = 22 and GENDER = F and MONT = 9) then DBT*
15. *if (AGE = Adt2 and DEPART = 22 and GENDER = F and MONT = 10) then DBT*
16. *if (AGE = Adt2 and DEPART = 22 and GENDER = F and MONT = 11) then DBT*
17. *if (AGE = Adt2 and DEPART = 22 and GENDER = F and MONT = 12) then DBT*
18. *if (AGE = Adt2 and DEPART = 31 and MONT = 1) then AST*
19. *if (AGE = Adt2 and DEPART = 31 and MONT = 2) then AST*
20. *if (AGE = Adt2 and DEPART = 31 and MONT = 3) then AST*
21. *if (AGE = Adt2 and DEPART = 31 and MONT = 4) then AST*
22. *if (AGE = Adt2 and DEPART = 31 and MONT = 5) then AST*
23. *if (AGE = Adt2 and DEPART = 31 and MONT = 6 and GENDER = M) then AST*
24. *if (AGE = Adt2 and DEPART = 31 and MONT = 6 and GENDER = F) then DBT*
25. *if (AGE = Adt2 and DEPART = 31 and MONT = 7) then DBT*
26. *if (AGE = Adt2 and DEPART = 31 and MONT = 8) then DBT*
27. *if (AGE = Adt2 and DEPART = 31 and MONT = 9 then DBT*
28. *if (AGE = Adt2 and DEPART = 31 and MONT = 10) then AST*
29. *if (AGE = Adt2 and DEPART = 31 and MONT = 11 and GENDER = M) then AST*
30. *if (AGE = Adt2 and DEPART = 31 and MONT = 11 and GENDER = F) then DBT*
31. *if (AGE = Adt2 and DEPART = 31 and MONT = 12) then AST*
32. *if (AGE = Adt2 and DEPART = 46) then DBT*
33. *if (AGE = Chi) then AST*

C. Discussion

The present findings illustrate an example of epidemiological prediction rules obtained on a dataset collected in the western region of Algeria. It is worth noting that both geographical (DEPART) and AGE descriptors play a prominent role in chronic diseases classification. For instance, we can note that the asthma (AST) is a frequent disease in younger individuals (Chi and Adt1, see lines 3 and 33 in rules above) independently from gender, or even location. For older people (Old), whereas asthma (AST) affects both men and but for diabetes (DBT) affects women (see lines 1 and 2 in rules above). Geographical location could also

be an important factor: in middle age (Adt2), diabetes (DBT) is persistent for departments 13 and 46 (see lines 4 and 32 in rules above).

These findings were confirmed by trades' experts who find them very relevant in the national context [8].

V. CONCLUSION

In this article, we have used a new APESS platform for service-oriented data mining, in the field of public health and in particular epidemiology. Our platform provided prediction models of chronic diseases, using classification algorithms by generating epidemiologic prediction rules.

Practically, the APESS platform allows extracting models for chronic disease surveillance, based on the exploitation of the real data of drug sales in pharmacies. In the context of the analysis of the chronic diseases, the generated epidemiological rules enable us to determine the relations between the disease and the people exposed to it versus the physiognomic characteristics and the environment. It is likely that the generated model with APESS will facilitate the characterization of the diseases by internal and external factors so that the patients are better taken care of.

Taken as a whole, the APESS platform could be of interest to various public health actors involved in the establishment of health policies, particularly in terms of planning of acquisition of pharmaceutical products, by taking into account various factors such age, gender, geographical location and year period and time.

For future work, one could explore how physiological and environmental characteristics can predict other chronic diseases such as hypertension and cardiovascular disease. In addition, it would be of great interest to enrich our database with others factors such as demographic aspect, pollution, socioeconomic status, etc.

ACKNOWLEDGMENT

REFERENCES

[1] N. Chen, N. C. Marques, and N. Bolloju, "A Web Service-based approach for data mining in distributed environments," in *Proceeding of the 1th International Workshop on Web Services: Modeling, Architecture and Infrastructure* (WSMAI-2003), Angers, France, Apr. 2003, pp. 74-81.

[2] W. K Cheung, "*Scalable and privacy preserving distributed data analysis over a service-oriented platform,*" Chapter 7 in the book : Data Mining Techniques in Grid Computing Environments, W. Dubitzky, University of Ulster, UK, 2008, pp. 105-118.

[3] R. A. Ferreira, O. G. Dorgival, and W. Jr. Meira, "*Anteater: A Service-Oriented Data Mining,*" Chapter 11 in the book : Data Mining Techniques in Grid Computing Environments, W. Dubitzky, University of Ulster, UK, 2008, pp. 179-198.

[4] A. Kumar and M. Singh, "An Empirical Study on Testing of SOA based Services," *International Journal of Information Technology and Computer Science*, vol.01, issue 07, pp.54-66, 2015.

[5] D. Birant, "*Service-Oriented Data Mining,*" Chapter 1 in the book: New Fundamental Technologies in Data Mining, K. Funatsu and K. Hasegawa, Published by InTech, Croatia, Jan. 2011, pp.3-18.

[6] A. A Shaikh and O. F Rana, "*FAEHIM: Federated Analysis Environment for Heterogeneous Intelligent Mining,*" Chapter 6 in the book : Data Mining Techniques in Grid Computing Environments, W. Dubitzky, University of Ulster, UK, 2008, pp. 91-104.

[7] T. S. Sobh and M. Fakhry, "Evaluating Web Services Functionality and Performance," *International Journal of Information Technology and Computer Science*, 2014, vol.05, Issue 03, pp. 18-27.

[8] Institut National de Santé Publique, "Transition épidémiologique et système de santé, "Enquête Nationale Santé" Projet TAHINA (Contrat n° ICA3-CT-2002-10011), Ministère de la Santé, de la Population et de la Réforme Hospitalière, Algeria, Nov. 2007.

[9] A. A Shaikh, O. F Rana, and I. J. Taylor, "Web services composition for distributed data mining," in *Proceeding of the 34th International Conference on Parallel Processing Workshops* (ICPP 2005 Workshops), Oslo, Norway, Jun. 2005, pp. 11-18.

[10] C. C. Chiu and M. H. Tsai, "A Dynamic Web Service based Data Mining Process System," in *Proceeding of the 5th International Conference on Computer and Information Technology* (CIT 2005), Shanghai, China, 2005, pp.1033-1039.

[11] P. Chen, B. Wang, L. Xu, B. Wu, and G. Zhou, "The Design of Data Mining Metadata Web Service Architecture Based on JDM in Grid Environment," in *Proceeding of the 1st International Symposium on Pervasive Computing and Applications*, Urumqi, China, 2006, pp. 684-689.

[12] L. Xu, Y. Wang, G. Geng, X. Zhao, and N. Du, "SDMA: A Service based Architecture for Data Mining Applications," *IEEE International Conference on Services Computing*, Jul. 2008, pp. 473-474.

[13] D. Talia, P.Trunfioy, and O. Verta, "The Weka4WS framework for distributed data mining in service-oriented Grids," *Concurrency and Computation: Practice and Experience,* vol. 20, no. 16, 2008, pp.1933-1951.

[14] C. B. C. Latha, P. Sujni, E. Kirubakaran, and S. Narayanan, "A Service Oriented Architecture for Weather Forecasting Using Data Mining," *The International Journal of Advanced Networking and Applications* (IJANA), vol. 02, no. 02, 2010, pp. 608-613.

[15] M. Zorrilla and D. García-Saiz, "A service oriented architecture to provide data mining services for non-expert data miners," *Decision Support Systems*, Elsevier, vol. 55, no. 1, Apr. 2013, pp. 399-411.

[16] R. R. Shelke, R. V. Dharaskar, and V. M. Thakare, "Data mining for mobile devices using web services," in Proceeding of the *International Conference on Industrial Automation And Computing* (ICIAC), Lonara, Nagpur, 12th & 13th April 2014, Published in : International Journal of Engineering Research and Applications (IJERA), vol. 8, no. 2, Apr. 2014, pp. 7-9.

[17] Q-A Kester and W. Sam-Aggrey, "Open System Architecture Platform for Big Data: An Integrated Emergency Disaster Response System Architecture," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, Issue 4, April 2015.

[18] V. Patil and K. Kadam, "A Survey Paper on Implementing Service Oriented Architecture for Data Mining," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol.3, Issue. 7, July 2015, pp. 4815 – 4817.

[19] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, "Advances in Knowledge Discovery and Data Mining," *American Association for Artificial Intelligence Menlo Park*, 1996.

[20] L. Soibelman, M. Asce, and K. Hyunjoo, "Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases," *Journal Of Computing In Civil Engineering*, vol. 16, no. 1, Jan. 2002, pp. 39-48.

[21] F. Ravat, O.Teste, and G. Zurfluh, "Modélisation et extraction de données pour un entrepôt objet," Université Paul Sabatier (Toulouse III), IRIT (Institut de Recherche en informatique de Toulouse), équipe SIG, Toulouse, France 2001.

[22] N. Harbi, O. Boussaid, and F. Bentayeb, " Propriétés d'un modèle conceptuel multidimensionnel pour les données complexes, " in *Proceeding of 8èmes Journées Francophones Extraction et Gestion des Connaissances*, Sophia Antipolis, France, Jan. 2008, pp. 25-36.

[23] R. Kimball, "*The data warehouse toolkit: practical techniques for building dimensional data warehouses*," John Wiley & Sons, Inc., New York, NY, USA, 1996.

[24] N. Selmoune, S. Boukhedouma, and Z. Alimazighi, "Conception d'un outil décisionnel pour la gestion de la relation client dans un site de e-commerce, " in *Proceeding of SETIT 3rd International Conference*, Sousse, Tunisia, Mar. 2005.

[25] J. Han and M. Kamber, "*Data Mining: Concepts and Techniques*," Morgan Kaufmann Publishers, Inc., San Francisco, CA, USA, 2001.

[26] J. R. Quinlan, "*Learning efficient classification procedures and their applications to chess endgames*," Chapter 11 in the book : In Machine learning: An artificial intelligence approach, R. S. Michalski, J. G. Carbonell, & T. M. Mitchell, (Eds.), Tioga Publishing Co., Palo Acto, CA, USA, 1983, pp. 463-482.

[27] J. R. Quinlan, "C4.5: Programs for Machine Learning," Morgan Kaufmann Publishers, Inc., San Mateo, CA, USA, 1993, *Book Review in Machine Learning*, Springer, vol. 16, no. 3, Sep. 1994, pp.235–240.

Dr. Sidi Ahmed Rahal: He is Doctor in computer science since 1989 in Pau University France. He is memberships in professional activities are a member of LSSD (Laboratory Signal, System and Data) and Interest in DataBases, Data Mining, Agent and expert systems.

Authors' Profiles

Mohammed Sabri obtained his Master of Science Degree in Computer Science Department of University of Oran (Algeria) in 2010. He is currently a Ph.D. candidate in the Computer Science Department of science and technology – Mohamed Boudiaf Oran. His research interests include Databases, Data warehousing, Oriented Services Data Mining, and Web Services.