Automatic Mobile Diagnostic System for Collaborative Patients Based on Data Migration and Users Experimentation

TAITAI Achraf (1), EZZIYYANI Mostafa (1,2), CHERRAT Loubna (2), ELMAMOUNE Soumaya (3)
Database Engineering and knowledge processing Team
Laboratory of Mathematics and Applications
Abdelmalek Essaadi University
Tangier – Morocco
Email : achraf.taitai1@gmail.com, ezziyyani@gmail.com

Abstract

The knowledge extraction is an interactive and iterative process of analyzing a mass of data raw with the aim of extracting distributed knowledge, usable and adaptable to a given situation and profile. In the perspectives of implementing a medical knowledge extraction system, we present in this paper the study, design and development of an incremental knowledge base for medical diagnosis based on decision trees. The knowledge results from shared experiences by connected users to the system produce a base of incremental knowledge which improves gradually for every new experience of various users. We are interested, in the construction process of the knowledge allowing the exchange and the adaptation of the shared results according to the user profile. Our solution includes an automatic verification system to the knowledge base to validate user experiences, it helps to promote the most profitable and most reliable experiences, giving users more choices and more options for a safer operation for its situation. The proposed solution presents a rich model that can be applied in several areas, it is divided into two parts, a server which provides the core of the system developed in J2EE and a mobile part developed with Android.

Keywords : Integration**,** knowledge, profile, decision, adaptive systems.

1. Introduction

The implementation of a solution that supports not only the operation of medical data, but also performs processing and generates dynamic knowledge models to provide suitable and adaptable outcomes for every patient, seems to be difficult to adapt to the constraints required by the sensitivity of medical data and the lack of knowledge to produce the best results, judging by the scarcity of existing programs, that usually, merely exploit an already predefined database.

Our solution which is called MDS-CP: Mobile Diagnostic System for Collaborative Patients, is based on a new methodology for adaptive systems modeling of decision support based on three criteria: The interaction with the environment, the Auto-improvement and the Evaluation.

This new approach presents interesting advantages such as modeling and visualization of system behaviors, services synchronization and resources sharing.

On the other hand, it is necessary to produce classification procedures understandable by the user, we have chosen to work through decision trees that fit this requirement because they graphically represent a set of rules and knowledge also they are more easily interpreted.

In this article, we will describe the various stages that led to the realization of this project, from the preliminary study, passing by the functional and

conceptual analysis, and getting to the realization and implementation of the solution.

What is MDS-CP?

Mobile Diagnostic System for Collaborative Patient (MDS-CP) is a medical diagnostic tool by accumulative knowledge, which allows to its user to generate treatments (medications, tips ...) for illnesses, diagnosed in advance from the application via a decisional system, through the consultation of all the solutions collected from the experiences of other users who have a close profile and have a similar situation to the use case.

The application will also allow the user to consult treatment and even suggest solutions that will be added to the system, the system will perform updates and generate decision trees after each knowledge validation.

2. Targeted objectives

The node object of this project is the design and implementation of a decisional/ Mobile application, dedicated to mobile smartphones and intended for the patients, the application must:

- Contain a dynamic decisional system,
- Have a sharing process, verification, validation and dissemination of knowledge.
- Generate decision trees after each validation of knowledge,
- Have accumulative knowledge through user experiences,
- Be able to find correspondences in terms of profile and use cases,
- Be able to keep a history of use,
- Have a flexible data model and an applicable hierarchy in other topics,
- Contain a system to handle exceptions,
- Ensure the privacy of each user,
- Provide the best reliable results
- Ensure the Inter-Mobile Communication
- Provide simple and representative graphical interfaces.

3. Presentation of the new methodology for adaptive systems modeling

To conceive an adaptive decisional system in an evolutionary and changeable environment is a difficult goal to achieve by the models and the classical approaches because of the unstructured received data, and the sensitivity of the field.

The new methodology is based on three criteria: The modeling and the visualization of the system behaviors, the services synchronization and the resource sharing.

a. The interaction with the environment

The sensitivity of some fields as "medicine" requires the presence of effective and fast special systems which carry out treatments adapted to the needs.

Several studies have been conducted to conceive an adaptive system that meets this need, however, the majority of these studies are based on multi-agent systems that provide results after the cooperation of agents.

The downside of this research is that they eliminate an important option of the adaptive system, it is the interactivity: This option allows to adapt the behavior of the system with the treated environment and check this adaptation at every given moment.

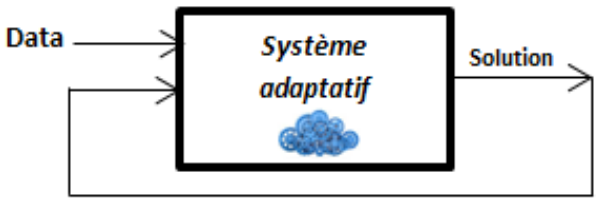

Figure 1: Interaction with the environment via data

The implementation of adaptation involves adapting in each time via the received data and the consideration of results for another adjustment in case of need.

This point was an objective aimed by the new methodology of adaptive systems modeling which aim to take into account the environmental variables for the adaptation in every moment.

b. Self-improvement

In order to improve the system performance, it is necessary to take in consideration the factor of user who must interact at the time of the decision.

Self-improvement, which is based on the system experiment or the choice of user who can change a decision, is the axis which allows in the new methodology to improve the system itself.

After several uses, if the system finds that the decision D2 is not frequented by the user or it has not been chosen as the solution of a given situation, it can trigger a change by another decision as it is shown in the figure.

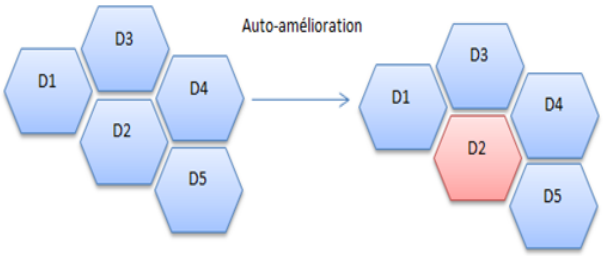

Figure 2: Self-improvement

c. Evaluation

The network construction is done automatically without human intervention, the system creates step by step its own domain ontology based on the accumulative, collected, verified and validated knowledge, allowing to build an evolutionary and robust base.

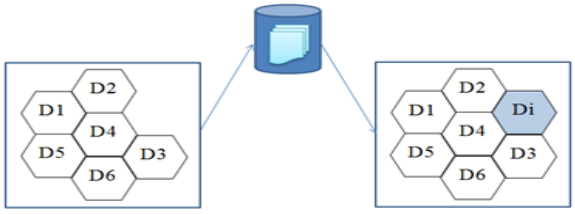

Figure 3: Construction of the decision network

The elementary decisions are created gradually starting from the experiments of the system which supports the formation of an increasingly solid core.

4. Description of the project

The project is composed of three phases, a decisional phase and a contextual research one, as well as a phase of privacy security. The first phase is to integrate a decisional system, the second concerns the processing of data flow, whether those traveling between Smartphones, or those traveling between Smartphones and the server, and the third ensures the security of system data.

a. Decisional Phase

Description of decisional system

The decision-making system that implements our tool is an expert system which is defined as an artificial intelligence system using facts, knowledge and reasoning techniques to solve problems that require some kind of human skills.

Our expert system differs from other conventional systems by its ability to acquire new accumulative knowledge from application users, those users who are in the same network can trigger the change of the knowledge base that fits with the number of views of a given data, thus, the automatic updates of the knowledge base ensures the integrity and reliability of data by fostering in the decision tree the most profitable branches and removing system vulnerabilities.

To each new update is a new knowledge tree more reliable and much richer, all of this by taking into account the views of the users what gives more choices and more options for a safer data exploitation.

The Knowledge Base Construction

The central idea in our approach is to split recursively and efficiently as possible the knowledge set through defined tests using attributes until we get subsets of examples containing almost only the examples that belong to the same class.

This implies that a verification procedure will take place for each node, the knowledge that causes disturbances

will be rejected, and the others will be validated and approved in their position. The resulting tree will still awaiting for shared knowledge to perform data validation procedure.

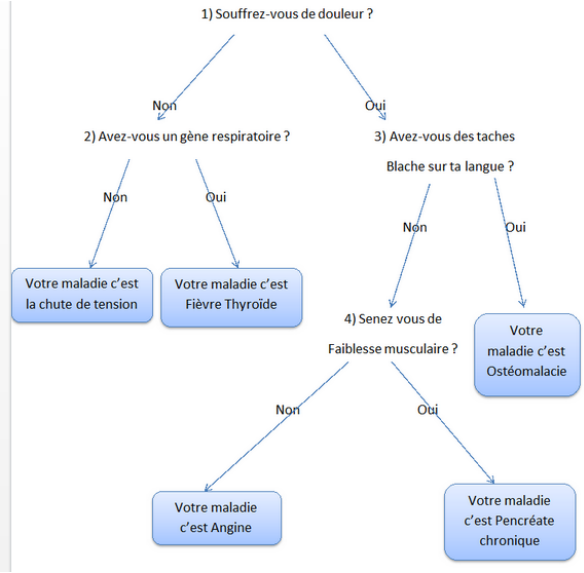

Figure 4: A Part from a knowledge tree

b. Phase of contextual research

The purpose of the contextual research is to meet the needs of the user by providing informations adapted to their specific research context.

This phase is related to the treatment of the data flows and the information exchanges, be it those transiting between users (Ex. Smartphones) via the server, or those traveling between smartphones and decisional system, thus, we are also interested in the knowledge construction process to adapt the shared results according to the profile of their target user.

The figure above shows the transmission and reception of user data to the system base, the data will then be shared to all users.

The Data flows in the system as a social network, except that a shared data goes through a validation process to detect anomalies and ensure the reliability of data flowing in the network.

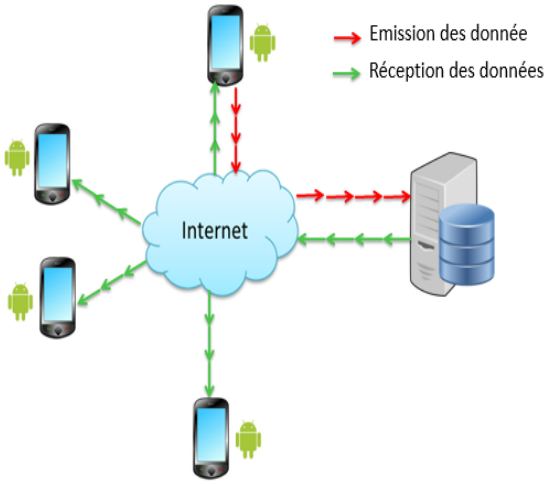

Figure 5: Flow of the data in the system

c. User profile adaptation.

The traditional search engines do not consider too much the context of research, this is why it was necessary to integrate an engine that will adapt the results to the user's profile, in our case,the fundamental need is to offer to a patient the results (treatment, medication ...) of all other patients who shared knowledge and have a profile close to similar to the target user. The Profile here means all the information attached to a patient: gender, location, weight, height, diet...

The relevance of the information provided, its intelligibility and its adaptation to user preferences and usage are a key factors for success or rejection of such information systems. For this, we focused on the modeling of user preferences in information research area in several heterogeneous resources integrated via mediators. So for the definition of an architecture for the user profile [11] taking into account the dimension of focus based on historical interrogation, then the integration of this profile into an information search process to customize the results returned by the mediation system. The use of profile ontologies allows us to:

- Facilitate access to multiple information sources.
- Exchange information to improve, and participate in a collaborative process of distributed resolution.
- Distribute and balance the workload.

Below is a class diagram of the focus that implements relationships between domains and users.

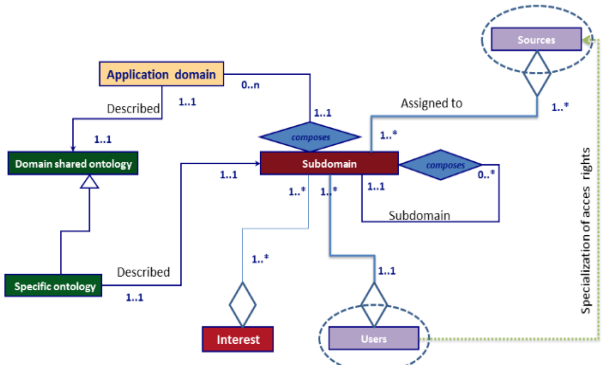

Figure 6: Class diagram of the user interest

i. Hierarchical organization of interests

The public or shared interest is common to all users, but can be customized dynamically as and when users query the mediation system forming appropriate interest centers to different groups of the system. Indeed, the public interest concerns a specific area, provided by experts, and will be used by the system for the construction of users interests classified by categories. These will include knowledge and data shared among a users group of the same mode of operation of the system concerned.

knowledge that make up each area of interest are copied into the user workspace when he selects. The system can then supplement this interest by protected/ private specific knowledge of user.

Generally, the interests provided by the system are not reduced, because it is a prototype, but we plan to create a collection of interests resembling the folders categories of types of users to limit areas of interest in numbers.

In our application context, we are particularly interested in site users to manage online PMD. We identified three groups among those users for which our approach can improve the effectiveness of their work online: administrative, doctors and patients.

- Administrative staff who wish to compose static and newspapers for decision making.
- Doctors who wish to write an adaptive report for their patients.
- Patients who wish to appropriate the contents of the PMD to the changes in his health.

The main common point of these users is ignorance of the accurate purposes of navigation when you start it or the composition of the original document of the PMD. It is therefore necessary to compose the patient's medical document not only in own way to each user, but also as and when the navigation of that user. To do this, the use of cooperative medical ontologies by reusing the techniques from personalization, customization and adaptive query allow optimizing resources and improving the query process of mediation systems.

For example, assume that a doctor needs to monitor a patient consultations via its PMD. In the initial study, the mediation system have to recommend compatible drugs to support the doctor during the medical treatment. In uncertainty, this question requires the use of the ontology of profiles using the same group as the basis for this guide. Therefore, the mediator may recommend medication based on a set of constraints and knowledge of the ontology of profiles (such as disease and health of the person concerned). In addition, this ontology [15][16] can help the doctor to find information about a particular treatment or a particular class of symptoms, it can find the right tools for a medical operation if necessary (such as a particular type of machine suitable for this variety of operation), etc..

ii. Interest dimensions

The interest is to consolidate in a knowledge base the personalized research results in relation with the use profile. This knowledge is used by the monitoring tool for communicating the news of sources on a subject or in a specific area.

The interest is therefore a continuous updating of knowledge of a user or group of users vis-à-vis the data sources integrated by the mediator.

This knowledge is represented by an ontology used in the process of matching queries and in the historical interrogation of sources and in the interaction with the system of mediation. It therefore proposes a cyclical method proceeds in two steps. The first is to extract from the history of interactions a qualified candidate interest of context of use, which reflects a specific need for information. The objective of the second step is then to integrate the context discovered in a knowledge base with respect to the assumption of diversity of interest. This reflects the learning phase of interests which induces changes in the profile.

In order to evaluate the degree of change contexts related to the user interests, we describe a context with

an adaptive ontology where the focus is described by four dimensions. The first represents the history of user interactions with the mediator. The second characterized its recurrent needs as information and is inferred and evolves from the first dimension. The third defines the security exploitation of interfaces and access to data sources. Finally, the fourth includes knowledges in the semantics, data and functionalities offered by the mediation system.

iii. Profile specification of the user queries

For this classification, we were interested also in the consideration of user profiles in the interrogation process whose goal is to improve the relevance in terms of research results in various data sources integrated via the mediator. The solution we provide is based on the use of ontologies to represent knowledge of the research area covered by queries and user profiles on the selected area. In this context, the query ontology can help the users to retrieve the desired information effectively and efficiently from a large collection of data.

In that case, the user is confronted with an information overload in which it is difficult to distinguish pertinent information from secondary information, and even noise. In addition, the evaluation of a query is usually without considering the context and / or specific needs of the user who issued it. The same request made by two different users, produces the same results even if these users have different expectations.

To address this problem, firstly, the use of an evolutionary model of ontology and an adaptation process will allow ontologies to adapt to changing knowledge of an area by a user group. Then, the exploitation of this type of ontology to improve the process of interrogation by an incremental optimization algorithm. The parameters of this algorithm are the different structures and separate data sources, and the degree of importance of these entities to extract pertinent data.

d. Phase of Security and Privacy

This part takes as objectives the improvement of the safety for the second phase, this security have several aspects, the most important are:

- All exchanges between users must be secured,
- The data are shared only after a confirmation from its owner,
- Each profile is invisible if and only if the owner decides to make it visible,
- Each user can select the information to communicate and share with others.

4. Functional and Conceptual Study

This part treats the desugn and the general modeling of the features which define the solution to be developed, this design is carried out with the language of modeling UML.

In this part we present the class diagrams which are considered as the most important in object-oriented modeling, they model the concepts of application domain and internal concepts created of all parts within the application implementation, as well as the use case diagram that expresses the requirement of the system users, so it has a user-oriented vision.

Our first class diagram brings together all the classes which compose our mobile client (see Figure 7), and the relationships between these different classes.

This diagram is linked with a second Diagram (see Figure 8), which has the classes of our core, it is the decisional system which will be implemented at the server.

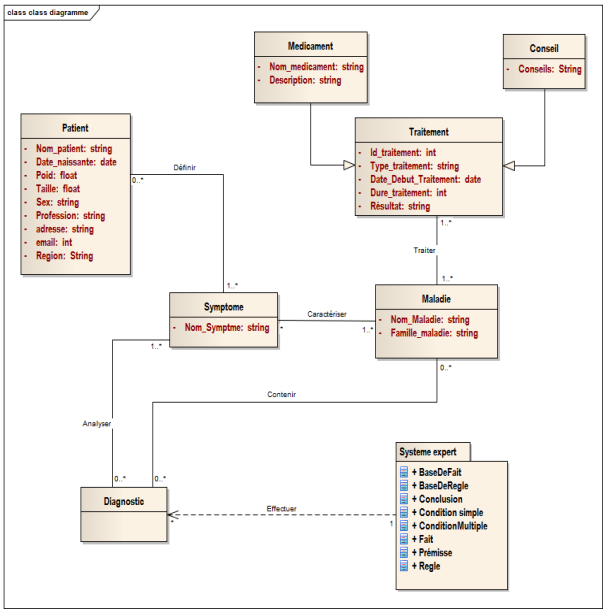

Figure 7: Client-side class diagram

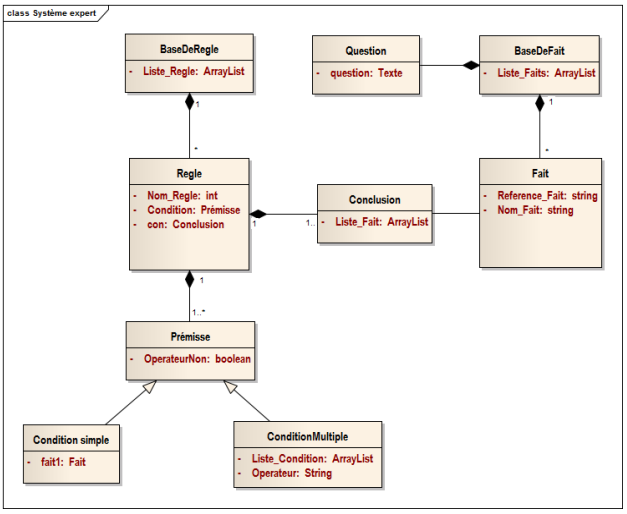

Figure 8: server-side class diagram

Furthermore, our use case diagram (see Figure 9) shows the various actors interacting with our system, we have three actors, the patient who will be at the mobile clients, the expert who manages the integrity and stability of the system and the decisional system who will communicate with users.

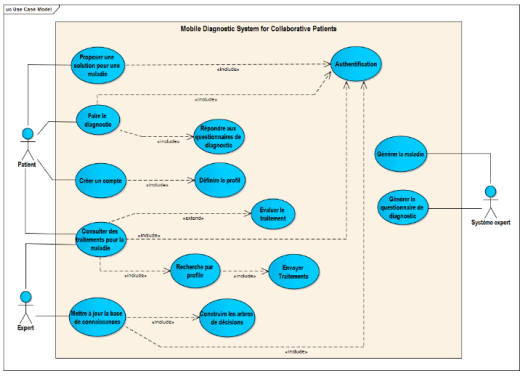

Figure 9: use cases diagram

5. Architecture of the solution

a. Physical Architecture

The figure below shows the physical architecture of the applications that compose the solution.

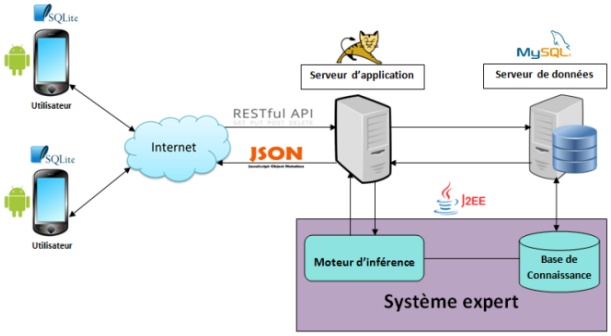

Figure 10: different parts of the system

We have three main parts:

- Client part: It presents the HMI (Human Machine Interface) front office and all visible features. Our HMI will be the Android mobile.
- Business part: This part is in charge of processing mechanisms. It is responsible for the implementation of the project business processes. Our decisional system is developed on this party.
- Data part: deployed on a data server and accessed by client / business parts.

b. Software architecture

For this project we have worked with an architecture Model / View / Controller (MVC), which is a way of organizing a graphical interface of a program. It involves three distinct entities, model, view and controller, each with a specific role in the interface.

The global organization of a graphical interface is often difficult. Although the MVC way of organizing an interface is not the miracle solution, it often provides a first approach that can then be adapted. It also provides a framework for structuring an application.

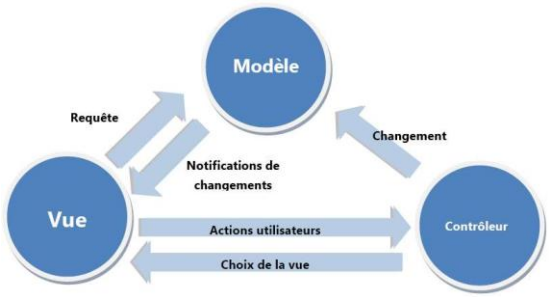

Figure 10: MVC architecture

In the MVC architecture, the roles of the three entities are:

- Data (access and update)
- View: user interface (inputs and outputs)
- Controller: event management and synchronization.

c. Development tools

Mainly we have worked with two tools, the first tool is J2EE with which we have developed the core of our application, and this is the part that performs most of the operations.

This choice was made due to the J2EE application architecture bases on components, which allows a cutting application and thus a separation of roles during development.

For the client part we have worked with the Android system which is free and famous in the world of smartphones.

The communication between the different system components is performed through the Web Services.

6. Implementation of the solution

a. The packages

The decisional system

This application is a J2EE project. The J2EE project contains all the classes that will make treatments and provide results exploited by the Mobile client. The tree structure of this project in Eclipse is divided into four packages respecting the adopted MVC architecture:

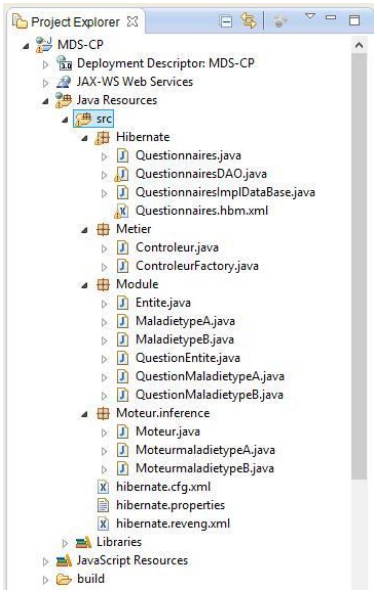

Figure 11: Decisional system development packages

- The "Hibernate" package contains a mapping of the database tables and java classes.
- The "Business" package contains the controller part and the methods of the system.
- The package "Module" Contains the classes instantiation of our decisional systems
- The package "inference engine" Contains the execution engine of the expert system, it often uses the classes of the business layer.

The mobile client

This application is developed with Android, its interest is to provide graphical interfaces required to exploit the results of the decisional system. There is no specific treatment, however it uses the expert systems to perform these functions, it is the "View" layer of MVC module.

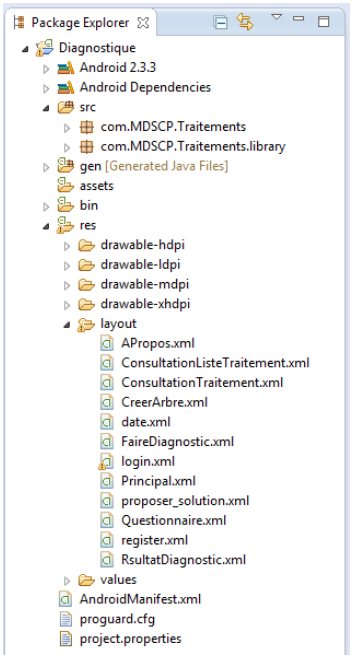

Figure 12: The mobile application packages

The interfaces of our mobile application are of type ".xml" as shown in the figure below.
The SRC folder contains two package:

- The first "com.MDSCP.Traitements" contains Java classes which call on the interfaces.
- The second "com.MDSCP.Traitements.libs" contains classes ensuring communication with the decisional system via the web service REST and exchanges data with the database.

a. The interfaces

The developed tool contains several interfaces; in this article we will show only some of these interfaces.

The main interface (figure 14) of the application offers three possible operations, to make the diagnosis, To consult treatments and to propose a treatment for a given disease.

We present in this article only one simple view of the realized interfaces.

Figure 14 shows the interface corresponding to the operation of diagnosis, the user will have to indicate the painful zone.

Figure 14: Main interface

Figure 13: Medical diagnostic interface

Figure 17 shows the interface of the diagnostic result, this result is displayed after a set of questions (Figure 16) generated by the expent system.

Figure 16: Diagnostic questionnaire

Figure 15: diagnostic result

Conclusion and perspectives

In this paper we propose a fully functional powerful solution that will make life easier for many people. The project is of great technical interest; an extension of this work would be the realization of robustness tests and deployment of the solution. The solution can also functionally evolve by responding to new needs thanks to its flexible architecture carefully designed and the introduction of good programming practices to encourage re-use.

References

[1] L. Rokach and O. Maimon, Data Mining With Decision Trees: Theory and Applications. World Scientific, Singapore, 2008.

[2] A. Knobbe, H. Blockeel, A. Siebes, and D. Van der Wallen, "Multirelational decision tree induction," Principles of Data Mining and Knowledge Discovery, vol. 1704, pp. 378–383, 1999.

[3] A. Knobbe, Multi-relational Data Mining. IOS Press, Netherlands, 2006.

[4] H. Leiva, "A multi-relational decision tree learning algorithm," Master's thesis, department of Computer Science, Iowa State University, 2002.

[5] H. Blockeel, "Top-down induction of first order logical decision trees," Th`ese de doctorat, Katholieke Universiteit Leuven, 1998.

[6] A. Atramentov and H. Leiva, Inductive Logic Programming, vol. 2835, pp. 38–56, 2003.

[7] Y. Arens, C.Y.Chee, C.Hsu, and C.A/Knoblock, " Retreiving and integrating data from multiple information sources", International Journal of Intelligent and Cooperative Information Systems, 2(2): 127-158,1993.

[8] M. Friedman, D. S. Weld , " Efficiently executing information gathering plans", In 15th International Joint Conference on Artificial Intelligence, 1997 Nagoya. Japan, p. 785-791

[9] M. F. Fernandez and D. Suciu.," Optimizing regular path expressions using graph schemas", In Proc. of the 14th IEEE Int. Conf. on Data Engineering (ICDE'98), pages 14–23, 1998.

[10] T. Slimani, B. Ben Yaghlane, K.Mellouli, "Extraction of Relevant Semantic Association with the use of Hyperclique Pattern Method", In proceedings of the second international conference on Web and Information Technologies ICWIT'09, Kerkennah Island, Sfax,Tunisia, June 12- 14- 2009, p.205-216.

[11] A. Cali, D. Calvanese, G. Giacomo, M. Lenzerini, " On the Expressive Power of Data Integration Systems", Proceedings of the 21st International Conference on Conceptual Modeling, 2003, p.338-350.

[12] D. Calvanese, D. Lembo, and M. Lenzerini, " Survey on methods for query rewriting and query answering using views", Technical Report D2I, 2001.

[13] V. Ullman, " Information integration using logical views ", In proceedings of the 6th International Conference on Database Theory ,(ICDT'97), 1997,p. 19-40.

[14] M. Ezziyyani, M. Bennouna, M. Essaaidi, " Mediator of the heterogeneous information systems based on application domains specification : AXMed Advanced XML Mediator", IEEE Journal, International Computer Science an Applications IJCSA, p. 25-45, 2006.

[15] D. Beneventano, S. Bergamaschi, S. Castano, A. Corni, R. Guidetti, G. Malvezzi, M. Melchiori, and M. Vincini, " Information integration: the MOMIS project demon-stration", In Proc. of the 26th Int. Conf. on Very Large Data Bases (VLDB 2000), 2000.

[16] M.-Ch. Rousset , A. Bidault, Froidevaux Ch. Froidevaux , H. Gagliardi , F. Goasdoue , Ch. Reynaud and B. Safar, "Construction de médiateurs pour intégrer des sources d'information multiples et hétérogènes : le projet PICSEL. " Revue I3,2002, Vol.2. n°1. p.5-59.

[17] Amato G., Straccias U., « User profile modeling and applications to digital libraries », Abiteboul, A.-M. Vercoustre (Eds.): ECDL'99, LNCS 1696, p. 184-197, 1999. c Springer-Verlag Berlin Heidelberg 1999.

[18] Bouzeghoub M., Kostadinov D., « Personalisation de l'information: apercu de l'état de l'art et définition d'un modèle flexible de profils », CORIA, 2005, p. 201-218

[19] Kostadinov D., Peralta V., Soukane A ., Xue X., « Intégration de données hétérogènes basée sur la qualité ». INFORSID, 2005, p. 471-486.

[20] Ruzzi M, «Data Integration : state of the art, new issues and research plan » .2004.

[21] Gruber T.R. «Toward principle for the design of ontologies used for knowledge sharing ». Technical report KSL 93-04 Standford University. 1993.

[22] Wache H., Vögle T, Visser U, Stuckenschmidt H, Schuster G, Neumann H and al. « Ontology-based integration of information – a survey of existing approaches ». In proceedings of the international workshop on ontologies and information sharing. Aout, 108-117, 2001.

[23] Automatic construction of query order execution plan for hybrid mediator applied to medical sources

[24] L. Lopes and G. Zaverucha, "Htilde: Scaling up relational decision trees for very large databases," in Proceedings of the ACM symposium on Applied Computing, Honolulu, Hawaii, 2009, pp. 1475–1479.

[25] P. Domingos and G. Hulten, "Mining high-speed data streams," in Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, 2000, pp. 71–80.

[26] X. Yin, J. Han, J. Yang, and P. Yu, "Efficient classification from multiple database relations: A crossmine approach," IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 6, pp. 770–783,2006.

[27] K. Potamias, M. Tsiknakis, V. Moustakis, and S. Orphanoudakis., "Mining distributed and heterogeneous data sources in the medical domain," in Proceedings of Machine Learning in the New Information Age, MLnet workshop, European Conference of Machine Learning, Barcelona, 2000, pp. 27–38.

[28] B. Park, H. Kargupta, E. Johnson, E. Sanseverino, and D. Hershberger, "Distributed, collaborative data analysis from heterogeneous sites using a scalable evolutionary technique," Applied Intelligence, vol. 16, pp. 19–42, 2002.

[29] [16] D. Caragea, A. Silvescu, and V. Honavar, "Decision tree induction from distributed heterogeneous autonomous data sources," in Proceedings of the Third International Conference on Intelligent Systems Design andApplications (ISDA03), Tulsa, USA, 2003, pp. 341–350.

[30] [17] J. Castillo, A. Silvescu, D. Caragea, J. Pathak, and V. Honavar, "Information extraction and integration from heterogeneous, distributed, autonomous information sources - a federated ontology-driven querycentric approach," in IEEE International Conference on Information Reuse and Integration (IRI03), 2003, pp. 183–191.

[31] B. Park and H. Kargupta, The Handbook of Data Mining. Lawrence Erlbaum Associates, London, 2003, ch. Distributed Data Mining, pp. 341–362.