Proceedings

International Conference on Software Engineering and New Technologies

Hammamet, Tunisia, 15-17 Décember 2012

Conference Chair: **Mohamed Ridda LAOUAR**

PREFACE

On behalf of the 1st International Conference on Engineering Software and Technologies (ICSENT'12), we would like to welcome the delegates and their guests to the ICSENT'12.

The ICSENT'12 was held in Hammamet, Tunisia during 15-17 December 2012.

ICSENT'12 were initially created by a network of researchers and engineers both from academia and industry in the areas of Design, Analysis and Tools for Engineering Software, ICSENT'12 has been extended to the fields of Software Engineering, Computer Science and Information Technology.

The main target of ICSENT'12 is to bring together software/hardware engineering researchers, computer scientists, practitioners and people from industry to exchange theories, ideas, techniques and experiences related to all aspects of Engineering Software and technologies.

The ICSENT'12 Technical Program Committee (TPC) consists of about 100 experts in the related fields of ICSENT'12 both from academia and industry.

ICSENT'12 is partnered with IAES (Institute of Advanced Engineering and Science).

ICSENT'12 received 75 full-paper submissions from 13 different countries around the world. These submissions were sent to the members of the TPC and additional reviewers for review.

Every submission received at least two in- dependent reviews. We would like to thank the ICSENT'12 TPC members and the reviewers for their hard work and dedication.

The ICSENT'12 Technical Program includes 2 keynotes lectures which are organized into lecture sessions.

So, we would like to extend our heartfelt thanks to our co-sponsors, conference committee, program committee, reviewers, and the many other colleagues who have contributed in difference ways to the success of this event.

Finally, we must thank the many authors who contributed the high-quality papers contained within these proceedings.

We hope you will enjoy your stay here in Hammamet (Tunisia) and we hope you will find this conference rewarding and will continue your great support for the ICSENT'12 in the future.

Mohamed Ridda LAOUAR
ICSENT'12 Chair

ICSENT'12 KEYNOTES

ADVANCED MIDDLEWARE AND ARCHITECTURES

Hamid MCHEICK

University of Québec at Chicoutimi, Canada.

The research area of software architecture and middleware has been developing rapidly. Software architecture is typically defined as the structures of the system that consist mainly of software components and interactions among them. Component is described as a unit of software that is self-contained, self-deployable, and well-defined functionality and can be assembled with other components through its interface. The relationship of components is determined through interaction that can be described as glue to link components together. Thus, interaction is an essential part of architecture and components need to be loosely coupled. Particularly, interaction plays an important role in distributed system.

Middleware is dedicated software which is used to implement the interaction between components without involving knowledge of network protocol and hardware platform. Normally, middleware provides software architecture with mechanisms for delivering data and for supporting transparency. According to the different mechanisms, middleware can be classified as Remote Procedure Call (RPC), Object Request Brokers (including CORBA, RMI) and Message Oriented Middleware (MOM). In practice, they can be applied to different situations according to separate requirements and development environment. We will survey advanced middlewares and enterprise software architectures and discuss the challenges of these area.

Keywords: *software architecture; middleware; Component.*

Biography

Hamid Mcheick is working as an associate professor in Computer Science department at the University of Québec at Chicoutimi, Canada. He has done his PhD in Software Engineering and Distributed System in the University of Montreal, Canada. He has a master degree in computer science from University of Quebec At Montreal (UQAM), Canada. He is currently working on software architecture, Service-oriented computing, Web Applications, software evolution and maintenance, software development methodologies, and aspect-oriented software engineering. He has 20 research papers in international journals and 55 research papers in international/national conference proceedings in his credit.

Dr. Mcheick has given many talk speeches in his research area, particularly in Distributed System and Software Connectors. He has also gotten many awards and congratulations of his quality of research (ICCIT-IEEE2012, ICCIT-IEEE2011, University of Jordan, FASE2006), and quality of teaching (UQAM, UQO). He is a chief in editor, chair, co-chair, reviewer, and member in many organisations (IEEE, ACM), journals (IJCNDS, IJCSI, JTAER, IJWS, etc.) and conferences (CCECE-IEEE, ICISA-Springer, ATNAC-IEEE, ACC-ACM, ICCIT-IEEE, etc.) around the world.

OPTIMIZATION QUERY PROCESS OF MEDIATORS INTERROGATION BY COMBINATORIAL STORAGE OF SOURCES

Mostafa EZZIYYANI

Abdelmalek Essaadi University, Morocco

The challenge posed by the increase and the diversity of information sources on the web, and secondly by the need of organizations to interoperate database systems have created not only the need to use tools for integrating data among multiple users and heterogeneous information sources, but again, these tools must overcome the limitations of current search engines by allowing not only users to ask queries more sophisticated than simple keywords, but also being able to aggregate other elements of answers from different sources to build, in the most optimized possible way by time and space research, the analytical global response to the user query. This need is becoming more and more imposed for medical information, especially with the existence of a multitude of web sources specific to medicine areas and the trend towards computerization of patient medical records.

In this distributed environment where a query involves across several heterogeneous sources, communication cost must be taken into consideration. In his conference we describe a query optimization approach using dynamic programming technique for a given set of integrated heterogeneous sources. The objective of the optimization is to minimize the total processing time including load processing, request rewriting and communication costs, to facilitate communication inter-sites and to optimize the time of data transfer from different sites. Moreover, the ability to store the data more than center site gives more flexibility in terms of Security/Safety and overload the network. In contrast to optimizers which consider a restricted search space, the proposed optimizer searches the subsets of sources and independency relationship which may be deep laniary or hierarchical trees. Especially the execution of the queries can be starting traversal anywhere over any subset and not only from a specific source.

Keywords: *Dynamic Programming Technique; Optimization; Communication.*

Biography

Mostafa EZZIYYANI *is a Professor in Faculty of Science and Technologies of Abdelmalek Essaadi University, Morocco. His research focus was on the Software Engineering and Programming (modelling databases and integration of heterogeneous systems, Multi- System Agents, distributed systems and mediation).*

He is IEEE and ASTF Member, Computer Science Department chair, Responsible in La.S.I.T Laboratory of the Research Direction Information Systems and Technologies. He was an invited professor for several countries in the world (France, Spain Belgium, Holland, USA, and Tunisia). He is member of USA-J1 program for TCI morocco delegation in 2007. He creates strong collaborations with research centers in Databases and Telecommunications for Students' Exchange: LIP6, Valencia, Poitier, Boston, Houston and China. EZZIYYANI has been actively in the research community by serving as reviewer for technical, and as an organizer/co-organizer of various international conferences.

ICSENT'12 PAPERS

International Conference on Software Engineering and New Technologies

CARBON NANOTUBE TRANSISTORS C-CNTFETS AND LOGIC FUNCTIONS WITH INVERTER*

Lobna. I'MSADDAK

University Department, National School of Engineering of Sfax, NSES, BPW 3038, Sfax, Sfax, Sfax, Tunisia
lobna.231177@gmail.com

Hekmet. SAMET

Group, LETI laboratory, Address, Sfax, Sfax, Tunisia
Samet_hekmet@enis.rnu.tn

Mounir. SAMET

Group, LETI laboratory, Address, Sfax, Sfax, Tunisia
Samet_mounir@enis.rnu.tn

Kachouri. ABDENNACEUR

Group, LETI laboratory, Address, Sfax, Sfax, Tunisia
Kachouri_abdennaceur@enis.rnu.tn

Abstract: Carbon nanotube field-effect transistors (CNTFET) modeling approach is being extensively studied as possible successors to CMOS-like. This paper introduced the analysis of the impact of diameter variations, especially logic switching threshold voltage and propagation time. The changing in diameter of CNT which is a critical parameter that can cause large swings in static and dynamic logic gates. SPICE simulations were performed on resistive-load and complementary logic inverter. The transfer characteristics as well as transient behavior were extensively studied. Finally, we have established relationships between the thresholds or propagation times and nanotube diameters.

Keywords : Carbon nanotube field effect transistors (CNTFET); Compact model; Characteristics of CNTFETs; Nanotube diameter; logic gates.

1. Introduction

Carbon nanotube field effect transistors (CNTFETs) have emerged as possible alternatives to silicon transistors in terms of physical modeling, identification of potential applications and understanding of performance limits [Appenzeller 2005], [Javey 2005] and [Guo 2004].

Its interesting structural (high performance), electrostatic properties (e.g., near ballistic transport and low power consumption) and superior characteristics make it attractive for the future integrated circuit applications and seem to be a promising successor for CMOS devices [Avouris 2004] and [Deng 2007].

Therefore, the carbon nanotube field-effect transistors CNTFETs using SWNTs as a channel are expected for low power consumption inverter.

In this paper, CNTFET technology has been employed, to make a very high-performance and ultra-low-power resistive-load and complementary inverter based on conventional CNTFETs from the proposed design.

The theory of CNT transistors is still primitive and the technology is still nascent [Guo 2004] and [Yamada 2000]. However, evaluation of such high-performance transistors in digital circuits is absolutely essential to drive the device design and understand the bottlenecks in multigigahertz processor design [Nanohub]. However, from a circuit designer's point of view, circuit simulation and evaluation using CNTFETs is challenging because most of the developed models are numerical, involving self-consistent equations which circuit solvers like SPICE are not able to handle. This paper presented a novel surface potential-based SPICE compatible modeling technique for single walled, semiconducting, ballistic CNTs in their ballistic limit of performance with one-dimensional (1-D) electrostatics [Guo 2002].

This model is applicable to a wide range of CNTFETs with diameters varying from 0.6 to 3.5 nm and for all chiralities as long as they are semiconducting. This model uses suitable approximations necessary for developing any quasi-analytical, circuit-compatible compact model. Both I-V and C-V characteristics have been modeled. This simple model enables simulation of circuit transfer (DC) characteristics as well as transients. It has been validated against numerical models in [Guo 2004] and [Nanohub] and has been found to be in very close agreement. It has been incorporated in SPICE and has been used to simulate digital logic blocks, functional and processing units.

2. Theory of ballistic CNTFETs

A simple model for ballistic CNTFETs is summarized and the computational procedure is given below [Guo 2002], [Prégaldiny 2006], [Raychowdhury 2004] and [Marani 2011]:

- Assume a V_{DS}, and a beginning-of-the-channel nanotube potential, V_{CNT}.
- Calculate the V_{GS} required to produce the assumed V_{CNT}.

If $V_{GS} \langle \Delta_1$

$$V_{GS} - V_{CNT} = 0 . \qquad (1)$$

If $V_{GS} \geq \Delta_1$

$$V_{GS} - V_{CNT} = \alpha(V_{GS} - \Delta_1). \qquad (2)$$

Where the parameter α is given by

$$\alpha = \alpha_0 + \alpha_1 V_{DS} + \alpha_2 V_{DS}^2 . \qquad (3)$$

- The parameters α_0, α_1 and α_2 are determined as indicated below[Prégaldiny 2006] [Dang 2008]:

$$\alpha_0 = 0.758 - (0.124 \times D_{CNT}) + (0.018 \times D_{CNT}^2). \qquad (4)$$

$$\alpha_1 = -0.325 + (0.181 \times D_{CNT}) - (0.137 \times D_{CNT}^2) + (0.024 \times D_{CNT}^3) . \qquad (5)$$

$$\alpha_2 = \alpha_0 .143 - \left(0.137 \times \exp\left[\frac{(D_{CNT} - 1.1075)}{0.9744}\right]\right). \qquad (6)$$

As the α parameter can not assume negative values, in [Marani 2011] we have demonstrated that the model describes the behavior of CNTFET, having diameter between 1 nm and 3 nm.

This set of functions has been computed for CNTFETs with a diameter between 1 and 3nm [Raychowdhury 2004] and [Marani 2011].

- Calculate Drain current I_{DS} for the first subband is obtained as.

$$I_D = \frac{4eK_BT}{h}\sum_{p=1}^{+\infty}\left[\ln\left(\frac{(1+\exp(-\xi_s))}{(1+\exp(-\xi_D))}\right)\right]. \quad (7)$$

If i = S,D

Where p is the number of subbands, e represents charge of electron is equal to 1.6×10-19 Coulomb, k_B represents Boltzmann constant is equal to 1.38×10-23 JK^{-1},T represents room temperature is equal to 300 K and h represents Planck constant.

$$\zeta_i = \left(\frac{V_{CNT} - \Delta_P - \mu_i}{K_BT}\right). \quad (8)$$

- The gate potential V_{GS}' required to produce the effective gate bias V_{GS} from:

$$V_{GS}^{'} = V_{GS} + V_{FB}. \quad (9)$$

Where V_{FB} is the flat-band voltage.

- The capacitors C_{GS} and C_{GD}:

For $\xi_i \langle 0 and V_{GS} \langle \Delta_1$.

$$C_{Gi} = qN0\left(\frac{AL}{K_BT}\exp(\xi_i)\right) \quad (10)$$

For $\xi_i \langle 0 and V_{GS} \geq \Delta_1$

$$C_{Gi} = qN0\left(\frac{AL}{K_BT}\exp(\xi_i)(1-\alpha)\right) \quad (11)$$

For $\xi_i \geq 0 and V_{GS} \leq \Delta_1$

$$C_{Gi} = qN0\left(\frac{BL}{K_BT}\right) \quad (12)$$

For $\xi_i \geq 0 and V_{GS} \geq \Delta_1$

$$C_{Gi} = qN0\left(\frac{AL}{K_BT}(1-\alpha)\right) \quad (13)$$

$$N_0 = \frac{4KBT}{3\pi V\, b} \quad (14)$$

$$A = -5.3\Delta_1 + 10\Delta_1 + 1 \quad (15)$$

$$B = 0.34\Delta_1 + 1 \quad (16)$$

The threshold voltage is approximated by the following equation:

$$V_{TH} = 0.91 - (0.81 \times D_{CNT}) + (0.27 \times D_{CNT}^2) - (0.033 \times D_{CNT}^3) \quad (17)$$

The maximum frequency of circuit:

$$F_{\max} = \frac{1}{2\pi \times delay} \qquad (18)$$

3. Compact model of C-CNTFET

Figure 3 shows the equivalent circuit of an n-type CNTFET [Raychowdhury 2004] and [Chen 2006].

The voltage generator describes V_{FB} is the flat band voltage, R_D and R_S represent resistances of the doped drain and source regions, while C_{GS} and C_{GD} represent capacitances of gate to drain and gate to source, finally I_D is drain current.

This model is applicable to a range of CNTFETs with diameter between 1 to 3 nm[Raychowdhury 2004] and [Chen 2006]..

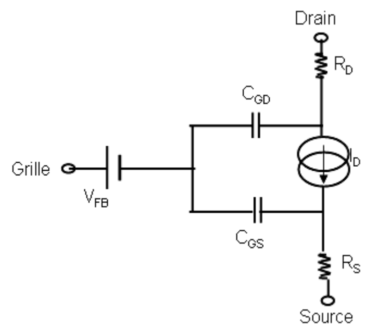


Fig. 1. SPICE compatible circuit model of the ballistic one-dimensional C-CNTFET.

In our study, we have discarded all parasitic capacitance or resistance of interconnects. Yet, we have only included the inherent device characteristic with the device capacitance. In our work we have extracted V_{FB}, R_D, R_S, C_{GS}, C_{GD}, R_L and C_L by best-fit procedure between the measured and simulated values of I-V characteristics of the device [Raychowdhury 2004] and [Marani 2011].

Here, we arbitrarily decided to use 500 mV voltage power supply, and positive logic was used. We took a high level 500mV to represent logical 1 and a LOW level 0V to represent logical 0.

The model uses suitable approximations necessary for developing any circuit-compatible compact model (figure. 1).

Further details regarding the I–V model of CNTFETs can be found in [Marani 2011].

4. CNTFETs inverters

Common types of inverters include resistive-drain, using only one N-type transistor and one resistor; and Complementary, which uses both N- type and P-type transistors per inverter circuit [Chen 2006].

In all simulations of these circuits we arbitrarily decide to use 600 mV power supply and we suppose that the flat-band voltages were equal to +130 mV and −130 mV for n-type and p-type.

1.1. *N-type C-CNTFET with a resistive-load inverter with a resistive load*

An inverter is a digital circuit that inverts the logic state applied at the input. It's a single input-single output device. Digital electronics have two types of logic states, logic '0', a low level and logic '1', a high level. The basic operation of the inverter is to produce an output of logic high for an input of logic low and vice versa.

The inverter is considered as the base building block for digital electronics.

The behavior of the circuit is as follows. For a logic low i.e., 0V DC input, the N-type C-CNTFET transistor is in OFF state and the transistor does not conduct current.

Figure 2 shows a C-CNTFET inverter with a resistive load, which only one N-type transistor and one resistor.

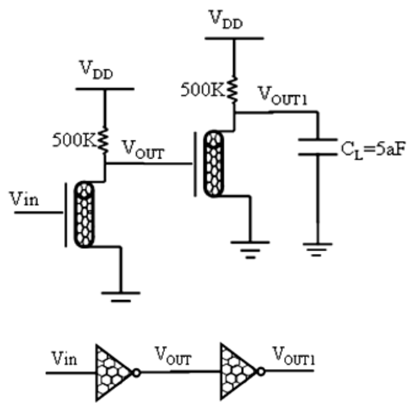


Fig. 2. Schematic and symbol of the N type C-CNTFET inverter with a resistive load for R= 500KΩ and CL = 5 Ff.

Now for a 600mV DC input which is a logic high level, the N-type C-CNTFET transistor is ON.

Figure 3 shows the result of the transient simulation. The waveforms corresponding to the input V_{IN} is observed to be automatically the inverse of output signal V_{OUT}.

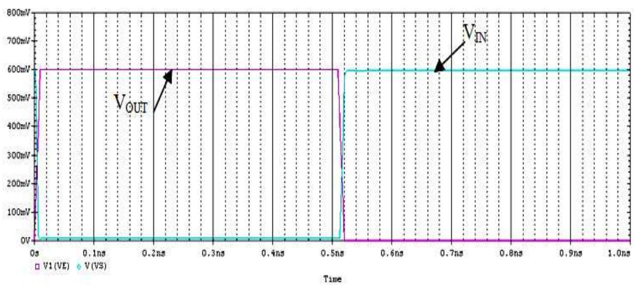


Fig. 3. The N-type C-CNTFET inverter with a resistive-load input/output for D_{CNT} =1.42nm.

The circuit shown in Figure 2 is simulated for 3 D_{CNT} (1nm, 1.42nm and 3 nm) and the DC transfer characteristic of the C-CNTFET inverter with a resistive-load designed with transmission gate is shown in Figure 4, Figure 5 and Figure 6.

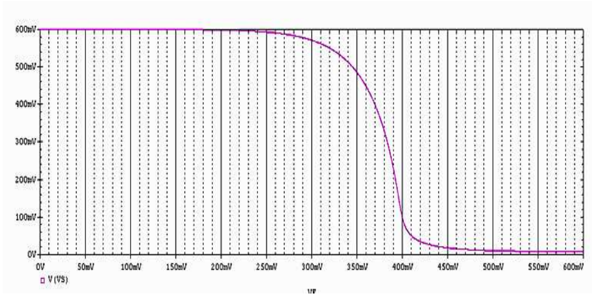


Fig. 4. Transfer characteristics of The N-type C-CNTFET inverter with a resistive-load output for D_{CNT} =1nm.

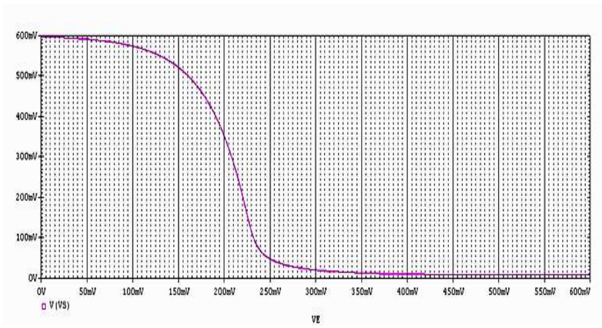


Fig. 5. Transfer characteristics of The N-type C-CNTFET inverter with a resistive-load output for D_{CNT} =1.42nm.

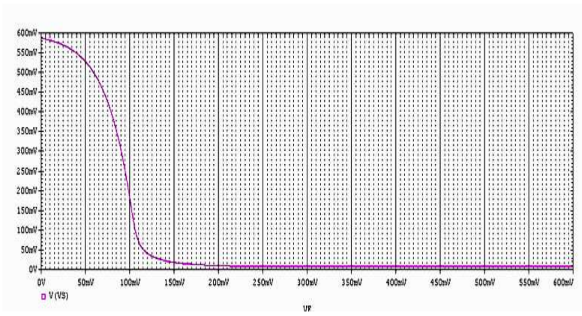


Fig. 6. Transfer characteristics of The N-type C-CNTFET with a resistive-load inverter output for D_{CNT} =3nm.

The rise time and fall time are shown the N-type C-CNTFET with a resistive-load inverter with a virtual 5aF.

1.2. *C-CNTFET complementary inverter*

Figure 7 shows a C-CNTFET complementary inverter, which uses both N- type and P-type transistors per inverter circuit.

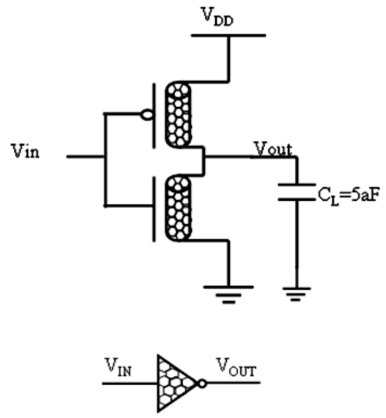


Fig.7. Schematic and symbol of the C-CNTFET complementary inverter.

The same figure consists of two C-CNTFET transistors, the N-type C-CNTFET for D_{CNT} = 1.42 nm and the P-type C-CNTFET for D_{CNT} = 1.42 nm.

The transient characteristic of this inverter designed with transmission gate is shown in Figure 8.

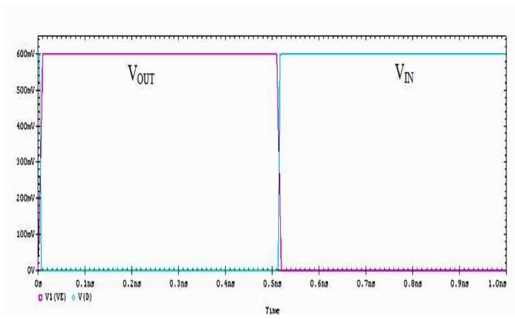


Fig.8. The C-CNTFET inverter input/output.

The DC transfer characteristics of this inverter designed with transmission gate are shown in figure 9, figure 10, figure 11, figure 12 and figure 13.

The DC transfer characteristic of a C-CNTFET complementary inverter (has a N-type C-CNTFET for D_{CNT} =1nm and P-type C-CNTFET for D_{CNT} =1nm) designed with transmission gate is shown figure 9.

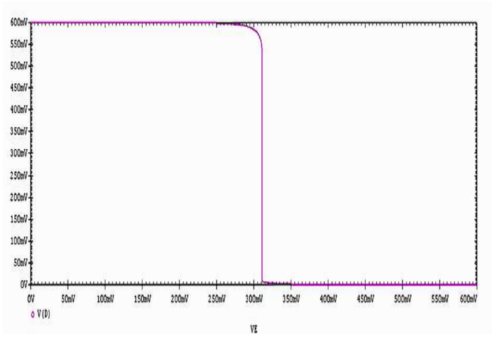


Fig.9. Transfer characteristics of the C-CNTFET complementary inverter output.

The DC transfer characteristic of a C-CNTFET complementary inverter (has a N-type C-CNTFET for D_{CNT} =1.42 nm and P-type C-CNTFET for D_{CNT} =1.42 nm) designed with transmission gate is shown figure 10.

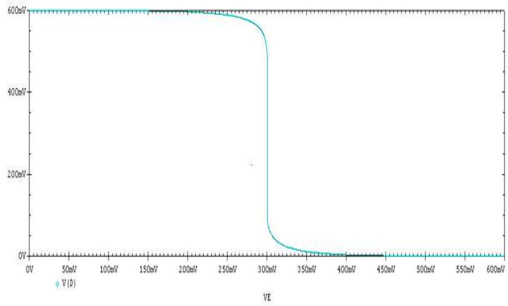


Fig.10. Transfer characteristics of the C-CNTFET complementary inverter output.

The DC transfer characteristic of a C-CNTFET complementary inverter (has a N-type C-CNTFET for D_{CNT} =3 nm and P-type C-CNTFET for D_{CNT} =3 nm) designed with transmission gate is shown figure 11.

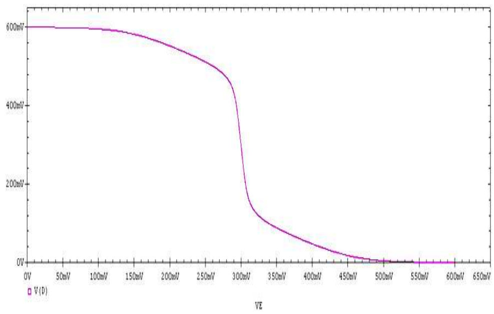


Fig.11. Transfer characteristics of the C-CNTFET complementary inverter output.

The DC transfer characteristic of a C-CNTFET complementary inverter (N-type C-CNTFET for D_{CNT} =1 nm and P-type C-CNTFET for D_{CNT} =1.42 nm) designed with transmission gate is shown figure 12.

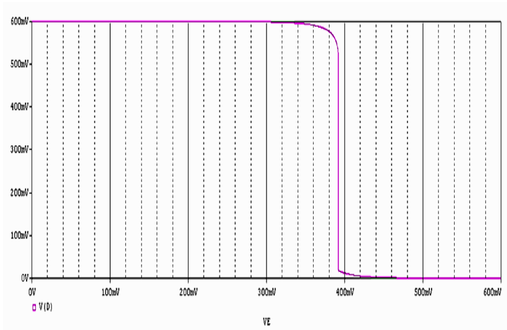


Fig.12. Transfer characteristics of the C-CNTFET complementary inverter output.

The DC transfer characteristic of a C-CNTFET complementary inverter (N-type C-CNTFET for D_{CNT} =1 nm and P-type C-CNTFET for D_{CNT} =3 nm) designed with transmission gate is shown figure 13.

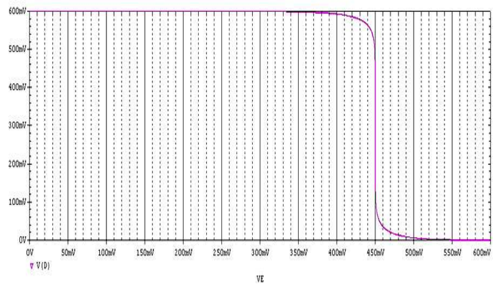


Fig.13. Transfer characteristics of the C-CNTFET complementary inverter output.

The rise time and fall time are shown for the conventional CNTFET complementary inverter with a virtual 5aF.

5. Results

The simulations using SPICE have been performed on two circuits including: resistive-load and complementary logic inverter.

We first have observed that the switching threshold is ideal for C-CNTFET complementary inverter (the N-type C-CNTFET for D_{CNT} = 1.42 nm and the P-type C-CNTFET for D_{CNT} = 1.42 nm) with a high noise margin.

We then have observed that the complementary structure is identified as the most interesting. Therefore, the results are satisfactory in terms of design integrated and speed. Finally, the conventional CNTFET complementary inverter is faster than the CNTFET inverter N-type resistive load.

The inverter circuit shown in Figure 2 is simulated and the results are plotted (figure 3, 4, 5 and 6).

The table 1 below summarizes the results.

Table 1. Results comparison of the N-type C-CNTFET inverter with a resistive load.

Characteristic	SPICE simulation	VHDL-AMS simulation
Rise Time (ps)	1.8	94
Fall Time (ps)	0.15	94
Delay (ps)	1.15	94
VTH (mV)	25	280
Max frequency (GHz)	138.47	1.69

The same table shows the delay, threshold voltage and maximum frequency of N-type C-CNTFET (D_{CNT} =1.42 nm) inverter with a resistive load. As shown in this table we can observe that the VHD-AMS delays and threshold voltage are higher than the SPICE ones. We then have observed that the SPICE maximum frequency is higher than the VHD-AMS one.

The inverter circuit shown in Figure 7 is simulated and the results are plotted (Figure 8, 9, 10, 11, 12 and 13). The table 2 below summarizes the results.

Table 2. C-CNTFET complementary inverter (has a diameter 1.42 nm and the p type C-CNTFET has a diameter 1.42 nm) results.

Characteristic	SPICE simulation	VHDL-AMS simulation
Rise Time (ps)	0.2	24
Fall Time (ps)	0	24
Delay (ps)	0.1	24
VTH (mV)	175	250
Max frequency (GHz)	1592.4	6.634

The same table shows the delay, threshold voltage and maximum frequency of C-CNTFET complementary inverter.

As shown in this table we can observe that the VHD-AMS delays and threshold voltage are higher than the SPICE ones. We then have observed that the SPICE maximum frequency is higher than the VHD-AMS one.

6. Conclusion

In this paper we have presented a CNTFET model, whose structure is to give it directly and easily implementable in circuit simulator, such as SPICE.

The techniques used in our proposed design along with exploiting the relationship between V_{TH} and nanotube diameter (D_{CNT}), have made it an efficient design which uses a minimum number of transistors.

Moreover, we have employed the proposed the model to design some numerical electronic circuits.

The design of basic cells has also been reviewed. As a conclusion, we have described in details both N- type C-CNTFET inverters with a resistive load and complementary inverter types.

In particular we have demonstrated the importance of C_{GD}, C_{GS}, CL and flat-band voltage. However, the effects of the CNT quantum resistances are absolutely relevant in the design of analogue circuits.

We first have verified that the SPICE works well for both N- type C-CNTFET inverters with a resistive load and complementary inverter type. Also, the mathematical model developed has been successfully incorporated into SPICE.

The mathematical models that were developed in VHD-AMS were compared against the models simulated with SPICE.

We then have observed that the conventional CNTFET complementary inverter is faster than the CNTFET inverter N-type resistive load.

Moreover, the channel (with CNT diameter of 1.42 nm) performance is investigated to calculate delay to evaluate the performance of CNTFET which is proposed to replace the standard CMOS technology.

Therefore we have presented a low-power and high-performance logic gate circuit based on carbon nanotube transistors.

Finally, a precise comparison of servel logic structures shows the advantages of the complementary structure of future applications.

7. References

Avouris. P. (2004). “Supertubes: the unique properties of carbon nanotubes may make them the natural successor to silicon microelectronics,” IEEE Spectrum, pp. 40-45

Appenzeller. J et al, (2005). IEEE T-ED, Vol. 52, p. 2568.

Chen. Z; Appenzeller. J; Lin.Y.-M; Sippel-Oakley.J; Rinzler. A.G; Tang. J; Wind; S. J; Solomon. P.M; Avouris. P,(2006). "An Integrated Logic Circuit Assembled on a Single Carbon Nanotube," Science, vol. 311, no.5768, p. 1735.

Dang. T; Anghel. L, (2008). "Portes logiques à base de C-CNTFETs Dispersion des caractéristiques" thèses à l’Ecole Doctorale «Electronique, Electrotechnique, Automatique, Télécommunications, Signal», Institut Polytechnique de Grenoble.

Deng. J; Wong.P, (2007). “A compact SPICE model for carbon-nanotube field-effect transistors including nonidealities and its application—part II: full device model and circuit performance benchmarking,” IEEE Transactions on Electron Devices, vol. 54, no. 12, pp. 3195–3205.

Guo. J; Lundstrom. M, (2002). “Assessment of silicon MOS and carbon nanotubeFET performance limits using a general theory of ballistic transistors,”in Int. Electron Devices Meeting Tech. Dig, pp. 711–715.

Guo. J et al, (2004). IEEE T-ED, Vol. 51, n. 2, p. 172.

Guo. J; Datta. S; Lundstrom. M, (2004). “Assessment of silicon MOS and carbon nanotube FET performance limits using a general theory of ballistic transistors,” in Digest IEDM, pp. 711–715.

Javey. A et al,(2005). Nano Lett., Vol. 5, n. 2, p. 345.

Marani. R; Perri. A. G, (2011). A compact, semi-empirical model of carbon nanotube field effect transistors oriented to simulation software, Current Nanoscience 7 (2) 245–253.

NANOHUB Online Simulations and More, CNT Bands [Online]. Available: http://www.naohub.purdue.edu

Prégaldiny. F; Lallement. C; Kammerer.J-B, (2006). Design-oriented compact models for CNTFETS, in Proc. IEEE Int. Conf. Design Test Integr. Syst. (DTIS) Nano.Technol, pp. 34–39. anded Abstract, 1750-1753.

Prégaldiny.F et al., in IEEE DTIS, 2006, pp. 34–39.

Raychowdhury. A ;Mukhopadhyay.S; Roy.K, (2004).Acircuit-compatible model of ballistic carbon nanotube field-effect transistors, IEEE transactions on Computer-Aided Design of Integrated Circuits and Systems 1411–1420.

Raychowdhury. A et al, (2004). IEEE Trans. Computer-Aided Design, vol. 12, no. 10, pp. 1411–1420.

Yamada.T, (2000). “Analysis of submicon carbon nanotube fiel effect transistors,” Appl. Phys. Lett., vol. 76, p. 628.

E-INCLUSION IS A KEY FACTOR IN YOUTH EMPLOYMENT AND EQUALIZATION

NOURELDIEN ABDELRAHMAN NOURELDIEN

Faculty of Computer Science and Information Technology
University of Science and Technology - Omdurman - Sudan
Noureldien@hotmail.com

Abstract

While the well known term digital divide refers to the gap between those who can effectively use ICT tools, such as the Internet, and those who cannot, the term e-Inclusion refers to the effective participation of individuals and communities in all dimensions of the knowledge-based society and Knowledge-Base economy through their access to ICT.

The effect of ICT in industry and in the creation of new markets for e-services is evolving up, thus the e-skills is becoming a requisite for workplace and for today's lifestyle.

In this paper we show that e-Inclusion polices and strategies have to be in place to widen the youth opportunities in employment and participation in the new knowledge-based economies and societies. In addition, we propose a new e-Inclusion digital gaps model which can provide policy makers with a better understanding to the digital gaps phenomena, and assist them in designing e-Inclusion polices and strategies for bridging these digital gaps.

Keywords: Digital Gaps, E-Inclusion digital gap model, Knowledge-based Society, ICT

1. Introduction

The fast growing diffusion of information and communication technologies (ICT) across all economic sectors is placing new demands on workers' skills. Unlike past decades, in today's job market, gaining basic ICT skills is considered essential for people to enter the workforce, and gaining advance ICT skills is mandatory for many market jobs.

Basic ICT skills are defined as: "the capabilities required for effective application of ICT systems and devices by the individual. ICT users apply systems as tools in support of their work, which is, in most cases, not ICT. Users ICT skills cover generic software tools and specialized tools supporting business functions within industries" [EC-EIDG (2004)].

Most of the employers consider basic ICT skills to include an introductory instruction on: operating systems, productivity applications such as word processing and spreadsheets, email, and the Internet.

Today, ICT skills are not only required for jobs in ICT sector, but across all sectors and job types such as education, business, management, agriculture, construction, and service industries. An OECD study estimated that 20% to 30% of total national employment was ICT-related. ICT specialists1 accounted for only three to four percent of ICT-related jobs. "Basic and advanced" users held the remainder of ICT-related jobs [OECD (2006)].

This changing in skill demands in today's knowledge economy have been the subject of academic research [S. Machin (2001)] [F. Green (2009)]. Accordingly workers in a wide variety of fields have had to adapt and incorporate ICT into their jobs. These changes increase the necessity and importance of ICT skills required by today's workforce and threaten the position of low ICT skilled workers "when they do not succeed in adjusting their skills according to the shifts in the skills demanded in their job or sector of industry" [Welsum and Vickery (2005)].

Thus to adapt youths to face these new skill changes, ICT skills has to be implemented in the educational system and training programs. To insure the adaptation of individuals and local societies, digital gaps have to be bridged through e-inclusion initiatives and programs.

This paper clarifies the necessity of e-inclusion programs to increase employment opportunities of youth people, and propose an e-inclusion model to provide policy makers with a better understanding to the digital divide gaps and their dynamic behavior and to assist policy makers to set e-inclusion policies and strategies.

This paper organized as follows, in section (2) we discuss the growing relationship between employability and ICT-skills, section (3) illustrate the concepts of digital divide and e-Inclusion. Section (4) is dedicated to the role of e-Inclusion in youth employment and equality. In section (5) an e-inclusion gaps model is proposed and conclusions are drawn in section (6).

2. Employment and ICT-skills

Often ICT users apply software systems as tools to support their work, which is, in most cases, not ICT. ICT User skills cover the use of common generic software tools and the use of a specialized systems and tools supporting business functions within industries other than the ICT industry.

Basic ICT skills, such as use of email and search engines and basic use of word processing, power point and spreadsheet applications, have become over the last few years an essential part of a standard set of skills for the majority of jobs in the job market.

Many reasons have contributed to the relatively high level of needs of basic ICT skills in employment. Firstly, the explosion of computer and network uses in private businesses and public institutions. Secondly, the massive amount development of computer applications with their wide application areas. Thirdly, the Internet with its potential applications and low access prices. Added to this has been the strong penetration of mobile phones with their expanded functionalities such as text messaging, browsing the Internet, and sending and receiving audio and video files.

All of the above elements and others have helped to create a firm basic ICT skill level in the workforce, at least in the race for entry-level jobs. A special IDC study commissioned by Microsoft on the demand for information and communication technology (ICT) user skills in European organizations is carried out in 2007 [Marianne and Vladimir (2007)]. A survey on 10 European countries shows that ICT is a key factor in job market. The survey clearly shows that organizations across the board believe ICT skills are critical elements for their future development. Three out of four respondents believe that their employees' ICT user skill levels will influence the organization's competitiveness, its ability to innovate and its growth. And the traditional industries such as agriculture, construction, transportation, wholesale and retail are increasingly investing in ICT and are increasingly looking for employees with ICT skills across the board. These industries and others also face the challenge of attracting technologically minded candidates with a broad spectrum of ICT skills.

Accordingly, it is clear that individuals equipped with ICT skills will certainly have a greater employment opportunity in the job-market. To empower youths with ICT skills, polices that insures embedding of ICT skills in education, tailored ICT training and e-Inclusion programs has to be established.

Due to this exploitation in ICT devices and e-applications, the international, continental, and regional organizations and institutions that are concerns with human and society development have set and deploy ICT policies and action plans.

3. Digital Divide and E-Inclusion

The term digital divide refers to any inequalities between groups, in terms of access to, use of, information and communication technologies [NTIA (1995)] [Chinn and Robert (2004)]. The digital divide measures the gap between those who are empowered to substantially participate in an information and knowledge-based society, and those who are not.

Research suggests a large number of explanations for the digital divide including, but not limited to: education, income age, skills, awareness, ethnic origin, location, gender, political and cultural access; and psychological attitudes to Internet access and usage [Guillen and Suárez (2005)].

Previously, digital divide research has focused on accessibility to the internet and internet consumption. However, with more people accessing the internet, researchers are more interested in how people use the internet and how this usage impacts the cultural and socio-economic behavior.

One problem with the concept of digital divide is its implication of a bipolar societal split. In fact there is no binary division between information "haves and "have-nots", but rather a gradation based on different degrees of access and use to information technology [Cisler, (2000)].

The second problem with digital divide concept is its dynamic nature, that is to say, those who are considered today to be haves, tomorrow can be considered as have-nots. This dynamic nature is due to the rapid change in ICT devices and new developments in applications and services.

Accordingly, the term e-Inclusion is coined to avoid digital divide conceptual consequences and to response to the immergence need of avoiding social exclusion due to ICT gaps.

E-inclusion is a term that encompasses activities leading to social inclusion in societies dominated by the quest for information. eEurope Advisory Group has defined e-

Inclusion as follows [Kaplan (2005)]: “e-Inclusion refers to the effective participation of individuals and communities in all dimensions of the knowledge-based society and economy through their access to ICT”. Further, e-Inclusion refers to the degree to which ICTs contribute to equalizing and promoting participation in society at all levels.

While ICTs can function as exclusion factor and deepen the divides between societal groups, e-Inclusion can supports excluded groups and societies to take part in a wider ICT enabled society. To foster e-Inclusion public roles, most governments, international organizations and private actors develop general e-Inclusion policies, which are then implemented in the form of projects.

4. E-Inclusion and Youth Employment and Equalization

In Europe, by 2020, it is estimated that 35 % of all jobs will require high-level qualifications, combined with a capacity to adapt and innovate, compared to 29 % today [EU (2010)], these percentages are less in US and Japan and more higher in developing countries. If we combine these figures with the fact that too many young people are leaving school early specially in Africa, Asia and Latin America, then this will increase the risk and percentage of unemployed youths in the future and rendering them inactive, living in poverty and causing high economic and social costs.

To face this, many International, continental, regional and national organizations and commissions have set and apply among other strategies and polices an e-Inclusion polices and initiatives. In its minimum benefits to unemployment youths, e-Inclusion programs should enable them to use ICT to chat with each other, and to seek for job. Thus e-Inclusion facilitated an important social function to overcome the isolation of the unemployed. In its maximum, e-Inclusion initiatives and programs aims to qualify youths to handle new ICT market jobs and services, avoid ICT exclusion, insure gender inclusion and equalization.

Many examples for e-Inclusion programs over all the world had suffer from problems that render their value and expected impact socially and economically. Of course any ICT project is complicated, and none can be expected to run smoothly. But the problems with these projects is due to the same reason, a too often focus is given on providing hardware and software and insufficient attention is given to the human and social systems that must also change for technology to make a difference.

5. E-Inclusion Modeling

Digital gaps exist in any society not due only to technical obstacles but there are many other factors that influencing the usage of ICT and causing digital gaps. In order to perform a more detailed analysis of these factors the concept of “e-Inclusion model” was introduced [J. Becker; et al. (2008)]. An e-Inclusion model will provide policy makers with a better understanding to the digital gaps phenomena, and assist them in designing e-inclusion polices and strategies for bridging these digital gaps.

The developed e-Inclusion model considers the dynamic behavior of digital gaps that results from the dynamic and rapid changes in ICT technologies, applications and services, and consequently in skills needed to operate and use these technologies and applications. Therefore, the model implements digital gaps and allow for iteration over these dynamic gaps, moving backwards and forwards, or round and round, as the need arises.

The developed e-Inclusion model is a Non-deterministic Finite Automata (NFA) with λ moves. In the following paragraphs we give the formal mathematical definition for NFA with λ - moves.

An NFA that responds to an empty string λ and moves to the next state is called NFA with λ - moves, or alternatively , a finite automata that is modified to permit transitions without input symbols, and with zero, one or more transitions on input symbols, is an NFA with λ - moves (NFA- λ) [Nasir and Sirmani (2008)].

An NFA with λ - moves exhibits the following five characteristics:

a. Finite set of states (Q)
b. An alphabet Σ of possible input symbols.
c. The transition function (δ) specifies the nature of the transition at a given state due to the input symbols including λ .
d. The initial state q_0.
e. The set of final states (F).

Formally, an NFA with λ - moves is a five- tuple: NFA- $\lambda = (Q , \Sigma , \delta , q_0, F)$

Where

a. Q is a finite set of states
b. Σ is a set of input symbols
c. δ is a transition function such that $\delta : Q \times (\Sigma \cup \lambda) \rightarrow p(Q)$
d. $q_0 \in Q$ is the initial state
e. F subset of Q is a set of final states.

To model e-Inclusion as NFA- λ we define the five tuples as follows:

a. Q is a finite set of states that represents digital gaps and an inclusion state.
b. Σ is a set of input symbols that represent inclusion activities $\{x_1,x_2,x_3,\ldots\ldots x_i\}$
c. δ is a transition function such that $\delta : Q \times (\Sigma \cup \lambda) \rightarrow p(Q)$
d. $q_0 \in Q$ is the initial state
e. F is the inclusion state.

5.1 Model Discussion

To see a working example Fig. 1, shows an e-inclusion gap model that was built around four digital gabs; namely: Infrastructure, Access, Skills and Applications.

Each digital gap was represented as a state; the initial state is a classification state where individuals or societies digital gap/s are identified. The initial state uses λ moves to all other states that represent digital gaps, therefore an individual or a society can be fitted in any digital gap. The inputs into each state represents e-inclusion activities that bridge that digital gap, after which individuals or societies can move either to a next digital gap state or to the accepting state which represent digital inclusion.

To handle the dynamic nature of the digital gaps, values of activities variables $x_{1....}$ x_i are assumed to be changed after each iterative or a cycle corresponding to the nature of the change. The ʎ moves from the accepting state to all digital gaps permits rejoining to e-inclusion activities due to the dynamic change in digital gaps. The arrows joining digital gap states represent the transformation of individuals or societies from one digital gap state to another before reaching the digital inclusion state.

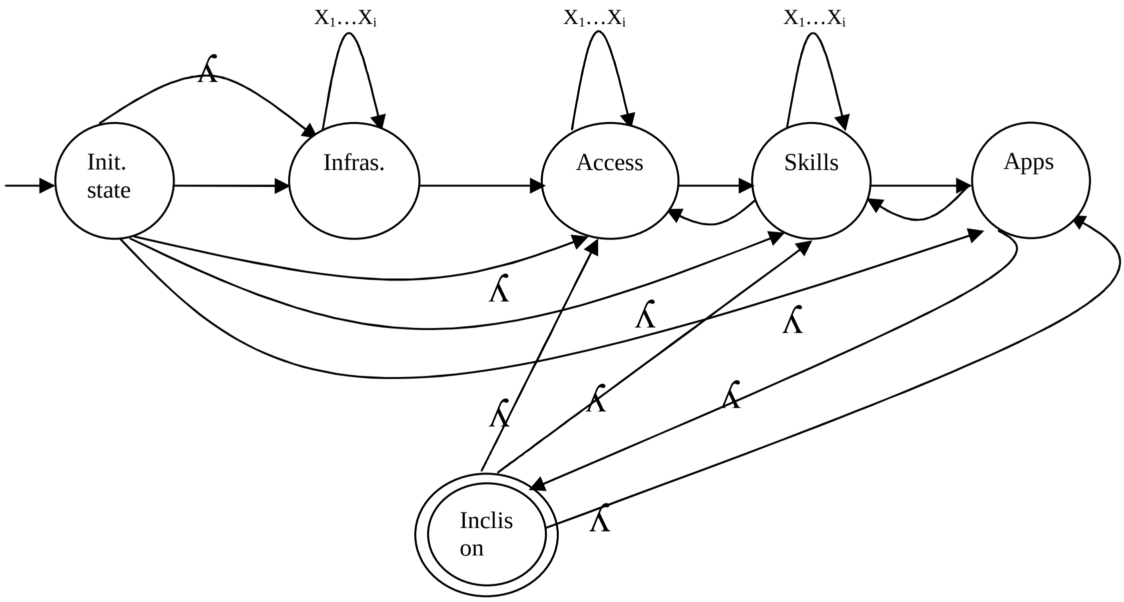


Fig (1): The E-inclusion Model

6. Conclusion

Today's e-Inclusion is central to avoid social exclusion and unemployment. International and continental organizations and national governments have to place e-Inclusion in the heart of their socio-economic development strategies and polices.

This paper provides a generic e-Inclusion gap model which handles the dynamic characteristic of digital gaps. The model provides a better understanding to e-inclusion activities needed to bridge digital gaps.

References

Chinn, Menzie D. and Robert W. Fairlie (2004). The Determinants of the Global Digital Divide: A Cross-Country Analysis of Computer and Internet Penetration. Economic Growth Center.

Retrieved from: http://www.econ.yale.edu/growth_pdf/cdp881.pdf

Cisler Steve. (2000), "Subtract the digital divide," *San Jose Mercury.* Retrieved From:

: http://www0.mercurycenter.com/svtech/news/indepth/docs/soap011600.htm

EC-EIDG, (2004). European Commission, Enterprise and Industry Directorate-General, "e-Skills for Europe: 2010 and Beyond,". Retrieved from http://ec.europa.eu/enterprise/ict/policy/doc/e-skills-forum-2004-09-fsr.pdf

EU (2010). Youth on the Move, Luxembourg: Publications Office of the European Union, ISBN 978-92-79-16673-0, doi:10.2766/69901. Retrieved from: http://www.ec.europa.**eu**/education/pub/pdf/higher/yom_en.pdf

F. Green, (2009). Employee Involvement, Technology, and Job Task, *NIESR Discussion Paper* No. 326.

Guillen, M. F., & Suárez, S. L. (2005). Explaining the global digital divide: Economic, political and sociological drivers of cross-national internet use. *Social Forces*, 84(2), 681-708.

J. Becker et al. (2008). "Digital Divide in eGovernment: The eInclusion Gap Model, LNCS 5184, pp. 231–242, 2008. © Springer-Verlag Berlin Heidelberg 2008

Kaplan (2005). *e*Europe Advisory Group, *E-Inclusion: New challenges and policy recommendations*(July2005), Retrieved from: http://ec.europa.eu/information_society/eeurope/2005/doc/all_about/kaplan_report_einclusion_final_version.pdf

Marianne Kolding and Vladimír Kroa, (2007). e-skills: The Key to Employment and Inclusion in Europe", Retrieved from: http/www.idc.com

NTIA (1995). National Telecommunications and Information Administration. Falling through the net: A survey of the "have not'
s" in rural and urban America. Retrieved from http://www.ntia.doc.gov/ntiahome/fallingthru.html.

Nasir S and P.K.Sirmani, (2008). "Automata Theory", Cambridge University Press India Pvt. Ltd., pp 144-146,2008.

OECD, (2006). The employability of low-skilled workers in the knowledge economy. Unpublished manuscript, Maastricht, UK, Retrieved from: http://rlab.lse.ac.uk/lower/final_papers/grip.pdf 2005

S. Machin, (2001). "The Changing Nature of Labour Market Demand in the New Economy and Skill-Biased Technology Change." *Oxford Bulletin of Economics and Statistics*, 63 (1), pp. 753-776.

W. van Welsum and G. Vickery, (2005), New Perspectives of ICT Skills and Employment," *Organization for Economic Co-operation and Development STI Working Papers*, Paris, France.

International Conference on Software Engineering and New Technologies

A HIGHLY PARALLEL HARDWARE IMPLEMENTATION OF THE DEBLOCKING FILTER USED IN H.264/AVC CODECS

KAMEL MESSAOUDI

Electronic Department, Mentouri University of Constantine,
Route Ain El-Bay, Constantine, Algeria
kamel.messaoud@univ-annaba.org, kamel.messaoudi@u-bourgogne.fr

EL-BAY BOURENNANE

LE2I Laboratory - Burgundy University,
BP 47 870, 21078, Dijon Cedex, France
ebourenn@u-bourgogne.fr

SALAH TOUMI, MAAMAR TOUIZA, AMIRA YAHI

LERICA Laboratory - Annaba University,
Sidi Amar, B.P# 12, Annaba, Algeria
salah.toumi@univ-annaba.org, maamar.touiza@u-bourgogne.fr

In this paper, a new design for the deblocking filter used in H.264 CODECs is proposed. This hardware component is designed to be used as part of a complete implementation of H.264 video coding system. In the proposed implementation we adopt a new strategy for memory management. Several on-chip memories are employed to support efficient parallel access in order to speed up the entire filtering process. Previous implementations of the filter make use of one or two elementary filters, which increases the complexity of control unit and the number of multiplexers required. Therefore, a second contribution of this work is the proposition of a new architecture based on five directional filters. These filters are obtained by subdividing each initial filter into two filters, each for one direction. An additional directional filter is used specifically to filter the left neighborhood blocks. This approach ensures parallel data processing and simplifies the control unit circuitry, reducing as well the number of multiplexers. The proposed architecture is implemented in VHDL and verified to work at 137.85MHz using Xilinx ML501 platform and at 170.95MHz using Xilinx XUPV5 platform, both incorporating Virtex5 FPGA devices. The proposed architecture takes a maximum of 71 clock cycles to process one macroblock, which is about 40% less than the best of the competing proposals.

Keywords: Deblocking filter; H.264/AVC; Hardware implementation; memory management; FPGA.

1. Introduction

H.264/AVC was jointly developed in an open standard process by the world leading experts of the ITU-T Video Coding Experts Group and the ISO/IEC Moving Pictures Experts Group [1]. This new coding standard is well suited for all types of video services including mobile phone applications, broadcast SDTV and HDTV services via satellite, cable or terrestrial transmission, HD-DVD and Digital Cinema. The diagram of the H.264 encoder is shown in Figure 1.

H.264/AVC was proposed to take advantage of the temporal and spatial redundancy occurring in successive visual images [2]. The video compression efficiency achieved in this standard is not the result of any single feature but rather a combination of a number of encoding tools and algorithms, one of these tools is the adaptive deblocking filter (DBF) algorithm [3]. In-fact, H.264/AVC is a block-based coding system; the original frame is partitioned into small blocks; then, the prediction, transformation and quantization are based on them [4]. However, the use of block-based processing sometimes introduces artifacts on the block edges [5]. For that the DBF filter is used to decrease blocking artifacts, which increases its coding efficiency and improves the decoded video quality [6].

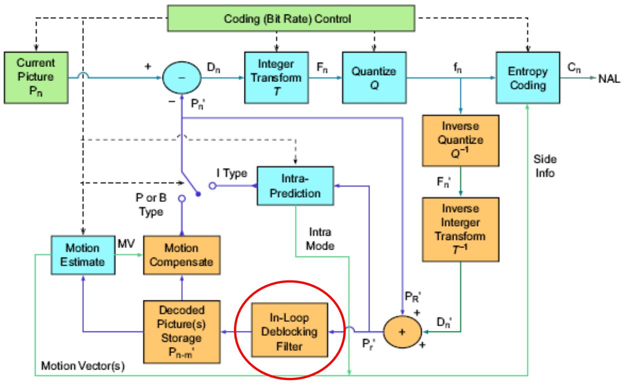


Fig. 1. Basic coding architecture for H.264 encoder.

Previous CODECs utilize a post-filter only in the decoder to improve visual quality at output. At the encoder, specifically at the motion-compensation module, unfiltered decoded frames are used as reference to reconstruct further frames [1]. In H.264, the filter is used in the encoder and in the decoder, in order to manipulate the same reference images [7]. In addition, the DBF algorithm used in H.264 standard is more complex than the DBF algorithms used in previous video compression standards [8]. First of all, the H.264 DBF algorithm is highly adaptive and applied to each edge of all the 4x4 luma and chroma blocks in a macroblock (MB). Second, it can update 3 pixels in each direction, in which the filtering takes place. Third, in order to decide whether the DBF will be applied to an edge, the related pixels in the current and neighboring 4x4 blocks must be read from memory and processed. Because of these complexities, the DBF algorithm can easily account for one-third of the computational complexity of an H.264 video codec [3].
This paper presents a work in progress, focused on the implementation of the H.264 encoder using an FPGA platform, on experiments with embedded operating systems for an embedded processor. To validate our global hardware/software architecture, each hardware accelerator needs to be done and simulated independently. In this paper, we propose new hardware architecture for the DBF filter used in H.264/AVC CODECs. This implementation used the same filtering order in [5], and the two elementary filters are decomposed into four directional filters, each for one direction (left, right, top and

bottom). An additional directional filter (vertical right) is used specifically to filter the left neighborhood blocks. This technique also eliminates the need for a transpose circuit, simplifies the control unit circuitry and allows for pipelined architectures.

This paper is organized as follows: In section 2, the in-loop DBF filter is introduced. In section 3, different existing hardware architectures for the DBF filter and different filtering order are given. The proposed implementation of the DBF filter is verified through simulations, which are shown in section 4. Simulation and synthesis results are provided in section 5. The conclusions are stated in section 6.

2. The deblocking filter

As it can be seen from Figure 1, in an H.264 encoder, the DBF module gets as input the reconstructed MB, from the Inverse Transform/Quant (IT/IQ) module. The IT/IQ module generates the reconstructed MB, one 4x4 block at a time. So, in H.264/AVC, the filtering stage is applied to each 4x4 luminance and chrominance block edge within each MB, in a specific order, as shown in Figure 2.a. Vertical boundary edges (A, B, C and D) are filtered first, followed by the horizontal ones (E, F, G and H) [9]. All filtering steps take place from left to right and from top to bottom. Moreover, MBs are processed in a raster-scan order over the frame.

The deblocking filtering process consists of modifying pixels at the four block edges by an adaptive filtering process. The filtering process is performed by one of the five different standardized filters, selected by means of a Boundary Strength (BS) calculation [1]. Figure 2.b defines graphically some definitions employed in the DBF Filter.

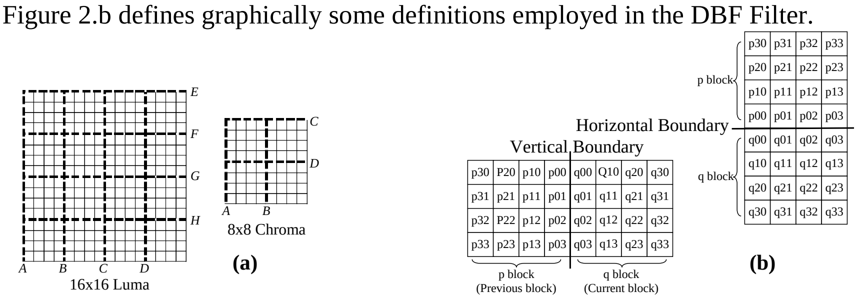


Fig. 2. Edge filtering order in an MB and pixels adjacent to vertical and horizontal boundaries.

2.1. *Boundary Strength*

The boundary strength (BS) is obtained from the block type and some pixel arithmetic calculation to verify if the existing pixel differences along the border are a natural border or an artifact [6]. In this process, it is decided whether the filtering is needed, and how much strength is applied. The choice of filtering outcome depends on the BS and on the gradient of the image samples across the boundary. The BS parameter is chosen according to the rules as shown in Table1 [4]. The result of applying these rules is that the filter is stronger at places where there is likely to be significant blocking distortion, such as the boundary of an intra coded MB or a boundary between blocks that contain coded coefficients [10].

Table 1. Selection of Boundary Strength (BS).

p and/or q is intra coded and boundary is a MB boundary	BS = 4
p and q are intra coded and boundary is not a MB boundary	BS = 3
Neither p or q is intra coded; p and q contain coded coefficients	BS = 2
Neither p or q is intra coded; neither p or q contain coded coefficients; p and q use different reference pictures or a different number of reference pictures or have motion vector values that differ by one luma sample or more	BS = 1
Otherwise	BS = 0

The DBF filter used in H.264/AVC is highly adaptive. When filtering an edge, eight pixels are involved and some of them may be modified according to the filtering boundary strength (BS). BS is set to five different levels, 0-4, in H.264/AVC, and the bigger BS is, the stronger will be the filtering. When BS=0, no filtering is applied and none of the pixels changed; when BS=4, the strongest filter may modify six pixels in the filtering process; and BS=1-3 means some weaker filters, modifying four pixels only [5].

2.2. *Filter Selection*

When the BS has been chosen in the block, the filtering of boundary samples is determined by analyzing each pixel on the block boundary. A Group of samples from the set (p2, p1, p0, q0, q1, q2) is filtered only if BS > 0 and $|p0\text{-}q0| < \alpha$ and $|p1\text{-}p0| < \beta$ and $|q1\text{-}q0| \leq \beta$ [3].

α and β are thresholds defined in the standard [10]. They increase with the average Quantizer Parameter (QP) of the two blocks p and q. The effect of the filter decision is to 'switch off' the filter when there is a significant change across the block boundary in the original image. When QP is small, anything other than a very small gradient across the boundary is likely to be due to image features rather than block effects that should be preserved and so the thresholds α and β are low. When QP is larger, blocking distortion is likely to be more significant and α, β are higher so that more boundary samples are filtered [1].

2.3. *Filter Implementation*

After the boundary strength and filter decision, filtering is applied according to the following rules [1];

Table 2. Filter implementation if BS = 4.

Block	Rule	Input	FIR filter	Output
p	If $|p2\text{-}p0| < \beta$ and $|p0\text{-}q0| <$ round (α / 4) and this is a luma block	p2, p1, p0, q0, q1	5-tap	P0
		p2, p1, p0, q0	4-tap	P1
		p3, p2, p1, p0, q0	5-tap	P2
	else	p1, p0, q1	3-tap	P0
q	If $|q2\text{-}q0| < \beta$ and $|p0\text{-}q0| <$ round (α / 4) and this is a luma block	q2, q1, q0, p0, p1	5-tap	Q0
		q2, q1, q0, p0	4-tap	Q1
		q3, q2, q1, q0, p0	5-tap	Q2
	else	q1, q0, p1	3-tap	Q0

(1) In the case of BS < 4: A 4-tap Finite Impulse Response (FIR) filter is applied with inputs p1, p0, q0 and q1, producing filtered outputs P0 and Q0. If $|p2\text{-}p0|$ is less than

threshold β, another 4-tap filter is applied with inputs p2, p1, p0 and q0, producing filtered output P1. If |q2-q0| is less than threshold β, a 4-tap filter is applied with inputs q2, q1, q0 and p0, producing filtered output Q1.

(2) In the case of BS = 4: Filtering is applied by the rules in Table2.

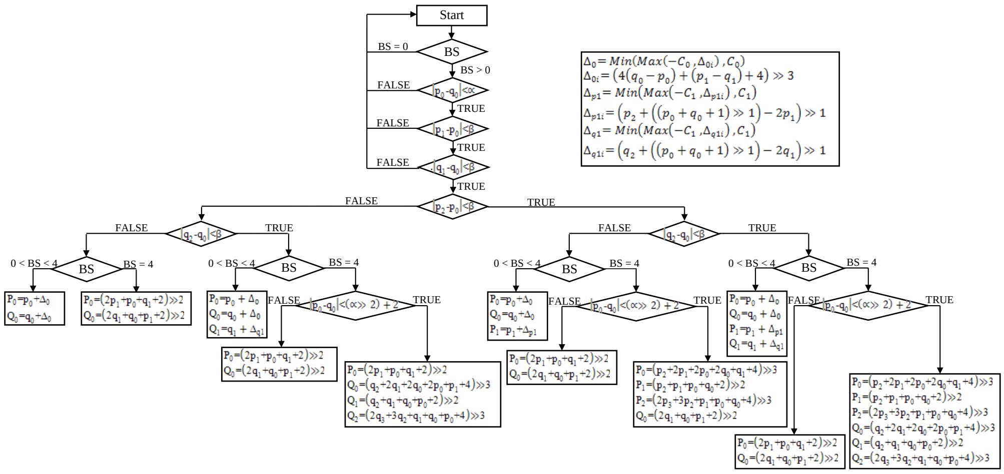


Fig. 3. H.264 Deblocking filter algorithm.

2.4. *H.264 Deblocking filter algorithm*

There are several conditions that determine whether a 4x4 block edge will be filtered or not. There are additional conditions that determine the strength of the filtering for a 4x4 block edge. DBF algorithm can change the values of up to 3 pixels on both sides of an edge depending on the outcomes of these conditions [3]. Figure 3 shows the overall algorithm of the DBF filter.

3. Existing Hardware architecture

The most important restriction imposed by the DBF filter used in H.264 CODECs is the filtering order of pixels. The sequential filtering order proposed by H.264/AVC [10] is shown in Figure 4. The restriction imposed is that if a pixel is involved in vertical and horizontal filtering, then the horizontal filtering should precede the vertical. This is rather loose and offers opportunities for implementation optimization by exploring different filtering schedules, aiming at faster operation through the use of parallelism or at solutions that consume less memory.

Several authors have proposed different filtering orders to limit the access to memory and to give the possibility of parallel processing. Khurana et al. [11] proposed different filtering orders where horizontal and vertical filtering alternate. Sheng et al. [12] proposed an order where a higher frequency of vertical-horizontal filtering direction changes is observed. Li et al. [13] proposed another solution involving a degree of parallelism with vertical and horizontal filtering occurring at the same time, speeding up the filtering at the cost of having to use two filtering units. The scheduling for this solution is presented in Figure 5.a.

Chen et al. [5] proposed a new strategy with only 18 steps instead of 21 steps for the luminance (Figure 5.b). The authors used 4x4 pixel register, one transpose array, one 16x32Bit SRAM and two 1-D filter units (one for filtering the vertical boundary and the other for filtering the horizontal boundary). The use of two filters in pipeline architecture reduces the number of clock cycles required to process a MB to 120.

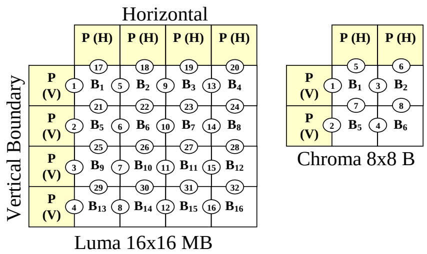


Fig. 4. Filtering order proposed by the standard H.264/AVC.

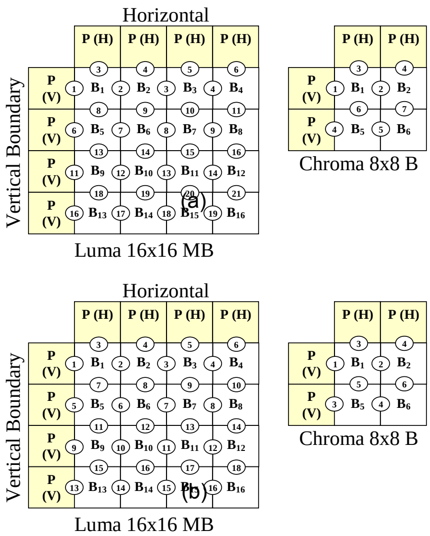


Fig. 5. Filtering order proposed both by Li et al and by Chen et al.

4. Proposed hardware architecture

In this work, we propose a new hardware implementation for the DBF filter, using the last filtering order, based on five units (Figure 8), where the two elementary filters proposed in [5] are decomposed into five directional filters. This technique also eliminates the need for a transpose circuit and allows for pipelined architectures.

4.1. ***Implementation strategy***

For a hardware implementation of the DBF filter and its integration in a complete diagram of the H.264 encoder, three steps are necessary (Figure 6):

(3) Defining a strategy for loading and storing of blocks and that in order to limit the access to external memory (memory management).

(4) Defining a processing strategy using the elementary modules of the filter (global architecture).

(5) Defining a processing strategy in each elementary module in 3 stages: BS selection, filtering decisions and filtering implementation (elementary filter implementation).

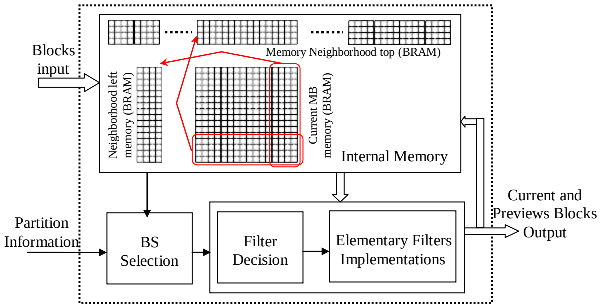


Fig. 6. Deblocking filter implementation strategy.

In a hardware implementation of the DBL filter, two approaches are possible to save the block being processed and their neighboring blocks: either through the use of internal memory or external memory. It is well known that accessing on-chip SRAMs consumes less power than accessing off-chip memory and augments the achievable throughput. In the proposed implementation, we used internal memory to record the 16 blocks being processed, the 4 blocks at left, and near all the blocks near the top in the same row of the image. After every treatment, the four blocks at the right of the treated MB will be recorded in the neighborhood left memory, to serve as left neighborhood for the next MB. The four blocks at bottom in the MB will be stored in the neighborhood memory at the top in the same position than the MB. These blocks will be used to serve as the top neighborhood of the other MBs in the same image (Figure 6). This strategy has the following advantages:

- Limit the number of accesses to external memory, which means we gain perspective of processing time;
- Reuse of same memory to save different successive blocks;
- Reduce power consumption by using internal memory;
- Sets the desired data reading strategy ;
- Sets memories with inputs/outputs of 128 bits using the 36 Kbits memory in the FPGA platform.

BS values for each block are assumed to be provided by the upstream processing modules; these values are calculated based on the type of treatment, the position and type of the block, and the partitioning method. Figure 7 shows the BS selection diagram.

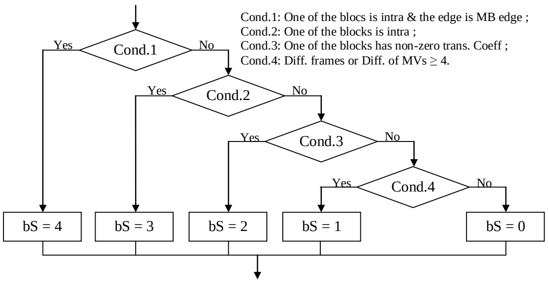


Fig. 7. BS selection block-diagram.

4.2. *Deblocking filter hardware architecture*

In order to implement the DBF filter, we use four cascaded directional filters so that we obtain a parallel processing of blocks as indicated in Figure 5.b. These four directional filters are obtained by subdividing the initial two elementary filters proposed in [5]. An additional directional filter is used in parallel with the first filter (filter left) specifically for left neighborhood blocks. The two horizontal filters are executed in parallel (Figure 8).This approach ensures parallel data processing and simplifies the control unit circuitry, reducing as well the number of multiplexers.

The vertical right filters work alternatively (in ping-pong manner), if the first works the other is set to idle by the selection command line. If the second works, the first is set to idle by forcing the value of BS to zero (no filtering). This is true for the blocks B4, B8, B12 and B16, where the edges on the right are not filtered. Using an additional filter has the following advantages:

- For I/O DATA different elementary modules are connected directly without the need for multiplexers. The conditions will be solely used on the selection inputs (load input).
- The designation of already processed blocks and intermediate blocks pressed easier. Thus for the calculation of memory addresses for these blocks.

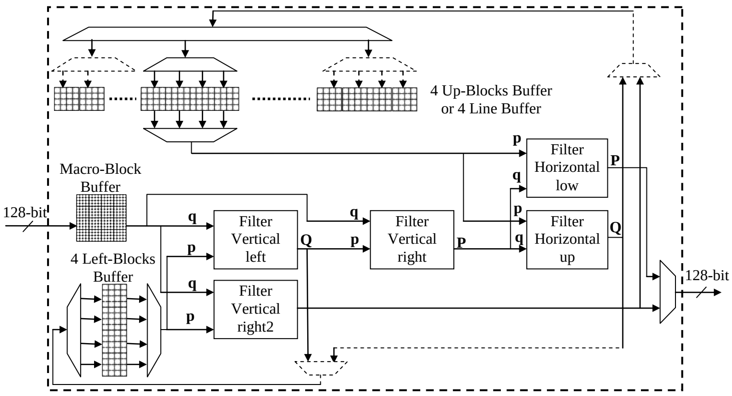


Fig. 8. Proposed hardware architecture for the deblocking filter.

Blocks at image edges are not filtered, because there are no neighbors. This is not the case for other blocks with neighbors in four directions. This causes an irregularity and, consequently, increases the complexity of the control unit of the filter. To avoid this deficiency, we propose to apply the filter for all blocks and to just change the values of BS. We assigned zero for initial neighborhood pixels and zero for the values of BS to avoid filtering the edges without causing an irregularity in the control unit.

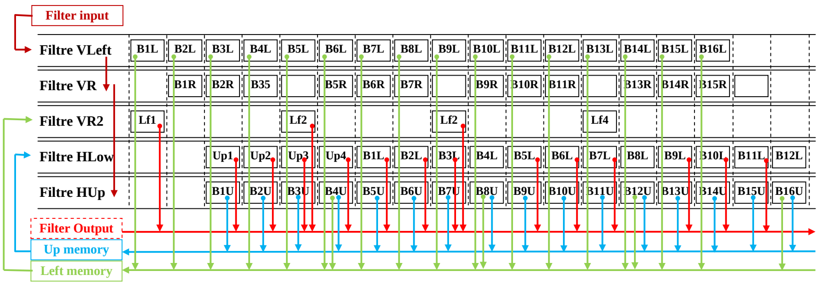


Fig. 9. Timing and order of blocks filtered.

In Figure 9 we show the filter order of different blocks in different elementary filters. We also show the order and timing of output data and intermediate blocks that will be recorded in the neighborhood memories. At the data output of the DBF filter, we observe in order the following blocks: V1 (Lf1), H1 (Up1), H2, H3, H4, V2, B1, B2, B3, V3, B5, B6, B7, B9, B10 and B11 (Figure 10). At the end of treatment, the other blocks will be recorded as follows: Blocks B4, B8, B12 and B16 are stored in the neighborhood-left memory to serve as neighborhood-left to the next block in the same line. Blocks V4, B13, B14, B15 and B16 are stored in neighborhood-high memory to serve as neighborhood-up blocks in the top row of blocks below. At the data output of the DBF filter, it is necessary to add an addressing system to rearrange the different blocks in their exact position in the different MB, and this by calculating their addresses in external memory.

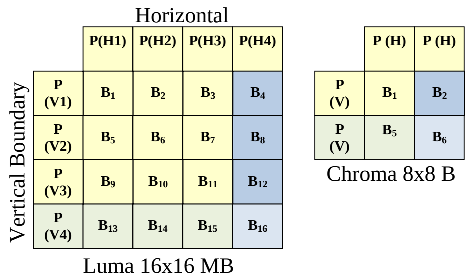


Fig. 10. Deblocking filter DATA output.

4.3. *Elementary filter hardware architecture*

Each directional filter has to process the current blocks of an input (q or p) in one direction, according to the input values (BS, Alpha, Beta and tc0) calculated in the main module. As shown in Figure 2.b, for each filter operation, eight pixels (p3, p2, p1, p0, q0, q1, q2 and q3) on both sides of the edge act as the input of the DBF filter (Figure 11). By using the chart in Figure 3, the internal architecture is almost the same for the five directional filters. The filtering equations are the same; the only differences are theirs outputs which depend on filter direction (vertical-Left, vertical-right, horizontal-up and horizontal-down).

As noted, in each elementary filter, we propose to use different variables for each condition, to avoid using multiplexers on these variables in order to minimize the number of used LUTs. This increases the number of internal registers, but offers the possibility of pipelining in the elementary filters. According to the proposed architecture, the pipeline is needed in both horizontal filters to avoid overlapping blocks. The proposed architecture for the directional filter takes up to 3 clock cycles to process a block boundary.

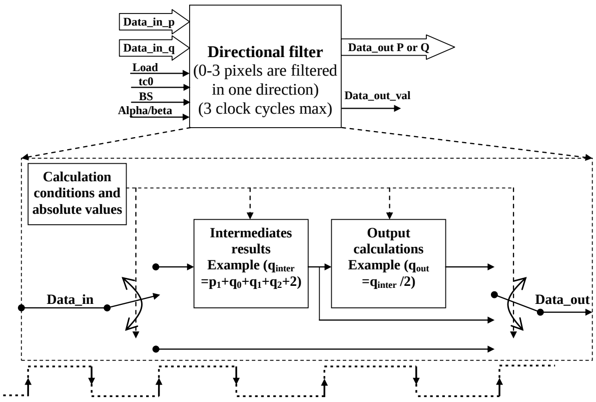


Fig. 11. Elementary filter architecture.

5. Experimental Results

To test the different implementations, we used ISE9.2 tool for project design and synthesis. The design was simulated using ModelSim6.1 environment. The proposed architecture was described using VHDL and synthesized both to ML501 Xilinx Virtex5 and XUPV Xilinx Virtex5 platforms. Three steps are performed in the simulation procedure:

Architecture simulation: This step is to check the performance of the proposed architecture. A testbench file is defined specifically for this step.

Functions simulation: This step involves checking the performance of each elementary module. At this step the hardware architectures of all modules are defined and simulated.

Data simulation: At this level the hardware architecture is simulated using a testbench file which can read images from different files to feed them on the DBF filter module. Simulation conditions are defined for the synchronization of input and output data, specifically to rearrange the processed blocks in their exact position in the filtered image.

The testbench file in this case plays the role of either of a microprocessor if we utilize the implementation into a complete embedded system, or of the intelligent memory controller for the H.264 proposed in previous work [14].
Figure 12 shows the simulation results of the five elementary modules (directional filter). The architecture proposed in this work takes a maximum number of clock cycles equal to 71 cycles to process one MB, which is about 40% less than the best of the competing proposals.

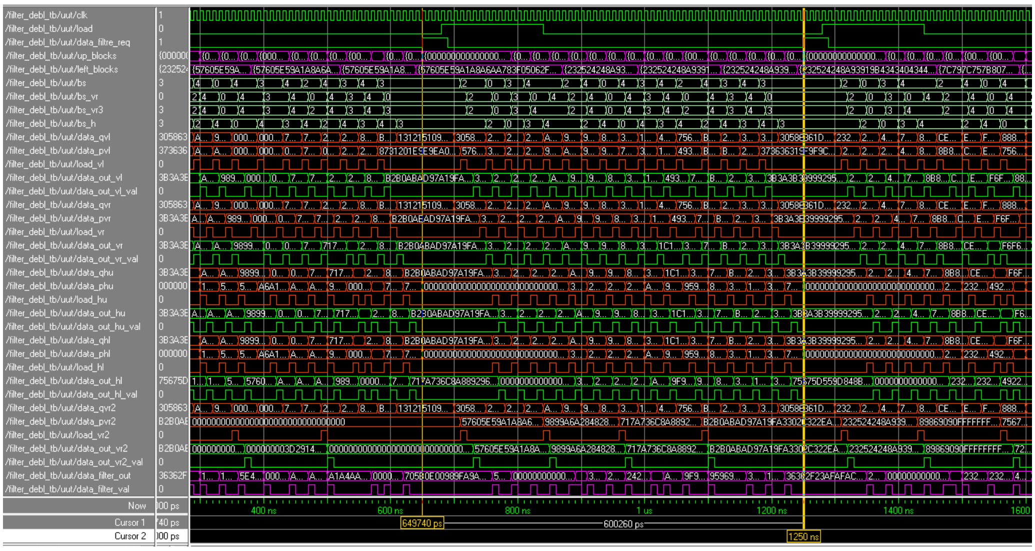


Fig. 12. Simulation results.

Table 3 shows the synthesis results using two FPGA platforms. The Xilinx Virtex5-ML501 with a FPGA devices type LX50, and the Xilinx Virtex5-XUPV with a LX110T FPGA device. The core filter used 5671 LUTs and was able to run at 137.85 MHz which allows the filtering of all blocks of a video with VGA resolution at a rate of 1.173 frames per second considering operations of the five cores. Table 4 shows other proposed solutions concerning number of cycles necessary for filtering one MB and the size of working memory needed.

Original image **Compressed Image (compression ratio 1 / 16)** **Filtered image** **Compressed Image (compression ratio 1 / 4)** **Filtered image**

Fig. 13. Filtering results of compressed-decompressed images of Lena.

Using a testbench file created specifically for the implementation, we can read image files, transfer them on the filter, and finally retrieve and rearrange the result. Using the Matlab functions, we can read and display these files. Figure 13 shows the filtering results of two images of Lena (256×256 pixels). The third image is a filtered image after compression-decompression using the DCT and inverse DCT with a compression ratio of 1/4. Similarly for the forth image with a compression ratio of 1/16. In these figures, we notice that the filter is adaptive; it allows filtering and smoothing surfaces away from image edges. On the natural border of images, the filtering is conditioned with several parameters.

Table 3. Synthesis results.

	Virtex5 LX50		Virtex5 LX110T	
	Used	%	used	%
Number of Slice Registers	3303	11%	3298	4%
Number of Slice LUTs	5671	19%	5671	9%
Number of Block RAM/FIFO	13	27%	14	9%
Maximum operating frequency	137.85MHz		170.95MHz	
clock cycles per MB	55 – 71 Cycles/MB			

Table 4. Comparison with other solutions.

	Cycles per MB	Filter Cores	Memory Bytes	on-chip memory for neighboring blocks
H.264/AVC	192	1	512	No
Khurana et al.	192	1	128	No
Sheng et al.	192	1	80	No
Li et al.	140	1	112	No
Chen et al. [5]	120	2	/	Yes
This work	55-71	5	256	Yes

6. Conclusion

In this work, we have presented a new hardware implementation of the DBF filter used in H.264/AVC CODECs. In the first part of this paper, we study and analyze the computational complexity of the DBF filter. From this analysis, we have concluded that the complexity of the filter lies in the data dependency and in the control module of elementary filters but not in the nature of these filters. Consequently, the memory management for this data is vital, for this reason we have proposed the use of internal memory to save the intermediate results. In the second part of the paper, we have shown the new hardware implementation based on this new strategy of memory management for blocks and MBs. This architecture adopts five elementary filters placed in an innovative manner to provide a parallel processing of different blocks. This architecture presents minimum latency, maximum throughput, full utilization of hardware resources and a combination of both pipelining and parallel processing techniques. The proposed architecture was synthesized for Xilinx ML501 and XUPV platforms. On the first one, the proposed implementation occupies 5671 LUTs, and was able to run at 137.85 MHz. On the second one, the implementation was able to run at 170.95MHz using the same amount of resources. In both cases the proposed architecture takes a maximum of 71 clock cycles to process a MB, which is about 40% less than the best of the competing proposals.

References

[1] E. G. Richardson, "H.264 and MPEG-4 Video Compression," the Robert Gordon University, Aberdeen, UK, 2003.

[2] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, A. Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Trans. on Circuit And Syst. For Video Technology, vol 13, no. 7, pp. 560-576, 2003.

[3] M. Parlak, I. Hamzaoglu, "Low Power H.264 Deblocking Filter Hardware Implementations," IEEE Transactions on Consumer Electronics, vol. 54, no. 2, pp. 808-816, MAY 2008.

[4] S. Shin, D. Oh, Y. Chai, T. Kim, "Performance Improvement of H.264/AVC Deblocking Filter by Using Variable Block Sizes," Springer-Verlag Berlin Heidelberg, pp. 732-743, 2007.

[5] Z. CHEN, W. UAO, U. WANG, M. ZHANG, "A performance optimized architecture of deblocking filter for H .264/AVC," The Jou. of China Univ. of Posts and Telecom., vol.14, pp.83-88, 2007.

[6] K. YANG, C. ZHANG, Z. WANG, "Design of adaptive deblocking filter for H.264/AVC decoder SOC," Elsevier ScienceDirect, vol. 16, no. 1, pp. 91-94, 2009.

[7] S. Wang, S. Yang, H. Chen, C. Yang and J. Wu, "A Multi-core Architecture Based Parallel Framework for H.264/AVC Deblocking Filters," Springer, Sign Process Syst, 2008.

[8] P. List, A. Joch, J. Lainema, G. Bjøntegaard, M. Karczewicz, "Adaptive Deblocking Filter," IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 614-619, 2003.

[9] H.264 Deblocker Core V1.0, http://www.xilinx.com/support/documentation/ip_documentation /h264_deblock_prodbrief_ ds594.pdf.

[10] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG. Draft of Version 4 of H.264/AVC (ITU-T Recom. H.264 and ISO/IEC 14496-10 (MPEG-4 part 10) Advanced Video Coding), 2003.

[11] G. Khurana, A. Kassim, T. Chua e M. Mi, "A pipelined hardware implementation of In-loop Deblocking Filter in H.264/AVC," IEEE Trans. on Cons. Ele., vol. 52, no. 2, pp. 536-540, 2006.

[12] C. Cheng, T. S. Chang, K. B. Lee, "An In-Place Architecture for the Deblocking Filter in H.264/AVC," IEEE Transactions on Circuits and Systems, vol. 53, no. 7, pp. 530-534, 2006.

[13] L. Li, S. Goto, T. Ikenaga, "A highly parallel architecture for deblocking filter in H.264/AVC," IEICE Trans. on Information and Systems, vol. 88, no. 7, pp. 1623-1628, 2005.

[14] K. Messaoudi, E.B. Bourennane, S. Toumi, E. kerkouche et O. Labbani, 'Memory Requirements and Simulation Platform for the Implementation of the H.264 Encoder Modules', IEEE International Conference on Image Processing Theory, Tools and Applications IPTA'10, pp. 133-137, Juillet 2010.

International Conference on Software Engineering and New Technologies

Conception d'un crypto système pour les transmissions de données chiffrées

Mohammed Zakarya BABA-AHMED

Laboratoire de Télécommunications de Tlemcen
Département de Génie Électrique, Faculté de Technologie, Université Abou-Bekr Belkaïd -Tlemcen
BP 230, Pôle Chetouane, 13000 Tlemcen- Algérie.
zaki1300@hotmail.com

Fatima Zohra BENMANSOUR

Laboratoire de Télécommunications de Tlemcen
Département de Génie Électrique, Faculté de Technologie, Université Abou-Bekr Belkaïd -Tlemcen
BP 230, Pôle Chetouane, 13000 Tlemcen- Algérie.
benmansour_fzohra@yahoo.fr

Mohamed ANANE

ESI (École nationale Supérieure d'Informatique),
BP 68M Oued Smar, Alger, Algérie
m_anane@esi.dz

Notre objectif principal était le cryptage d'information qui est devenu obligatoire pour empêcher d'éventuelles attaques ; il peut être réalisé selon deux modèles : La cryptographie à clé secrète et celle a clé publique. Dans cet article, on a opté pour un crypto système hybride combinant la rapidité d'un algorithme symétrique et la sécurité de l'algorithme asymétrique. Celui-ci sera programmé avec un logiciel VHDL est implémenté en hardware sur un circuit FPGA de la famille Virtex-4 permettent ainsi de combiner les techniques cryptographiques symétriques et asymétriques tout en exploitant au mieux les ressources offertes par les circuits FPGA et d'atteindre de bonnes performances en termes de sécurité et répondre aux contraintes des applications en temps réel. Enfin une description matérielle du VHDL de cette architecture ainsi que la simulation par ModelSim pour vérifier les résultats obtenus.

Keywords: cryptographie; clé secrète; clé publique; circuit FPGA; VHDL; ModelSim.

1. Introduction

Le trafic des images numériques est en augmentation continuelle sur les réseaux. La protection des données numériques, devient importante pour de nombreuses raisons telles que la confidentialité et l'intégrité. Le cryptage de ces images répond à ce problème de confidentialité.

Pour empêcher ces attaques, la sécurité doit être omniprésente. Aussi bien lorsque les données ne sont pas utilisées que lorsqu'elles transitent sur le réseau. D'où la nécessité de crypter les données.

La cryptographie est l'étude des méthodes donnant la possibilité d'envoyer des données de manière confidentielle sur un support donné. Elle est divisée en deux grandes catégories : cryptographie à clé secrète et celle à clé publique.

Les algorithmes à clé publique tel que le RSA, qui impose l'utilisation de très grandes clés (1024 bits et plus) pour un niveau de sécurité acceptable, sont incapables de chiffrer une énorme quantité de données dans un environnement en temps réel.

Par conséquent, ces volumes de données ne peuvent être chiffrés que par des algorithmes cryptographiques à clés secrète, néanmoins, ces derniers souffrent du problème de la sécurité de l'échange de clés qui nécessite un canal de transmission très sécurisé mais non disponible.

D'ou la nécessité de combiner les chiffrements symétrique et asymétrique pour assurer d'une part une bonne sécurité et d'autre part une rapidité.

Dans cet article, nous allons définir :

- La cryptographie symétrique et asymétrique et de proposer un crypto-système hybride combinant la rapidité d'un algorithme symétrique, et la sécurité de l'algorithme asymétrique.
- Les Algorithmes ainsi que le logiciel utilisé, se dernier sera implémenté en hardware sur un circuit FPGA de la famille Virtex-4 pour répondre aux contraintes des applications en temps réel.

Enfin nous allons donner les résultats de simulations de l'algorithme AES pour le cryptage et le décryptage des données avec une conclusion et des perspectives.

2. Définition et principe de la cryptographie

La cryptographie concerne la transformation d'un message (texte, image, chiffres) intelligible vers un message codé, incompréhensible à tous sauf pour les détenteurs de la clé de chiffrement.

Décrypter ou casser le code c'est parvenir au texte en clair sans posséder au départ les règles ou documents nécessaires au chiffrement [1].

Une ***clé*** est un paramètre utilisé en entrée d'une opération cryptographique (chiffrement, déchiffrement, scellement, signature numérique, vérification de signature). Elle doit être variable et maintenue en secret (sauf dans certains algorithmes où une part de la clé reste exposée). Elle peut se présenter sous plusieurs formes : mots ou phrases, procédure pour préparer une machine de chiffrement (connexions, câblage, etc.), données codées sous une forme binaire (cryptologie moderne).

La protection apportée par un algorithme de chiffrement est liée à la longueur de la clé, qui peut s'exprimer en bits. Les clés les plus grandes resteront crypto graphiquement sûres pour une plus longue période. Si ce que l'on chiffre doit rester caché pendant de nombreuses années, il faut utiliser une clé très grande [1].

2.1. *Cryptographie symètrique*

La cryptographie symétrique Utilise la même clé pour les processus de codage et de décodage; cette clé est le plus souvent appelée "secrète" (en opposition à "privée") car toute la sécurité de l'ensemble est directement liée au fait que cette clé n'est connue que par l'expéditeur et le destinataire [7].

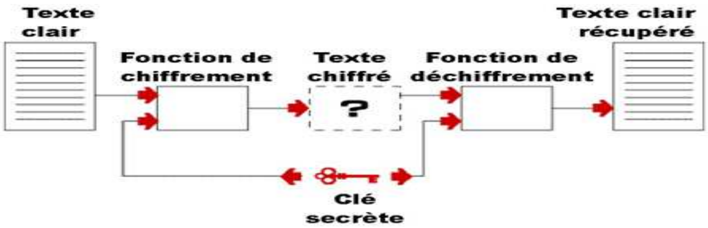


Fig. 1. Schéma de principe de la cryptographie symétrique [1].

La cryptographie symétrique est très utilisée et se caractérise par une grande rapidité et des opérations simples : décalage de certains bits, XOR bit à bit, permutations de certains bits, etc., et par des implémentations aussi bien software que hardware ce qui accélère nettement les débits et autorise son utilisation massive.

De plus, le message (clair) est séparé en blocs successifs, chaque bloc étant chiffré individuellement. Cela permet de disposer de systèmes de chiffrement très performants et très simples. Chiffrement à la volée)

Les deux parties communicantes doivent se mettre d'accord sur la clé secrète utilisée. Le problème est qu'on ne dispose pas d'un canal de communication sûr pour échanger les clés [1].

2.2. *Cryptographie asymètrique*

La cryptographie à clé publique s'attache à résoudre le difficile problème de gestion des clés. Son principe fondamental est de donner à chaque utilisateur deux clés associées, l'une secrète et l'autre rendue publique.

Afin de chiffrer un message à l'intention d'un utilisateur, l'idée consiste à n'utiliser que sa clé publique alors que le déchiffrement doit nécessiter la connaissance de la clé secrète.

Ce concept naturel permet de communiquer de manière confidentielle sans avoir à partager la moindre information secrète initialement [1].

Le schéma de principe de la cryptographie à clé publique est représenté sur la Fig. 2.

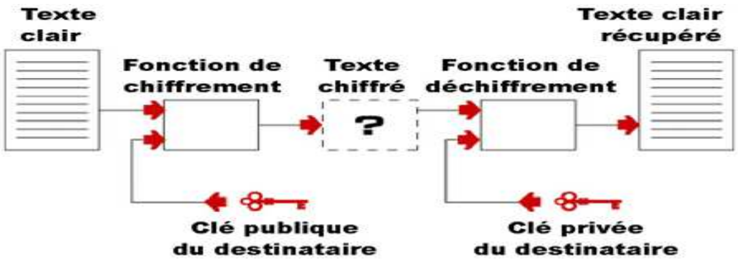


Fig. 2. Schéma de principe de la cryptographie asymétrique [1].

Du fait de l'utilisation de grands nombres premiers, les crypto systèmes asymétriques nécessitent une quantité de calcul importante, ce qui les rend très lents par rapport aux systèmes symétriques [1].

2.3. *Principe de la cryptographie hybride*

Un crypto-système hybride consiste à utiliser les avantages des chiffrements symétrique et asymétrique tels que:

- La rapidité d'un système symétrique qui grâce à une clé secrète valide le temps du transfert de l'information, ou le temps d'une session.
- La possibilité de transmettre la clé secrète par un crypto système asymétrique.

Il utilise le cryptage asymétrique avec une paire de clés publique et privée pour échanger une clé secrète symétrique ou clé de session qui servira à crypter le message.

La plupart des systèmes hybrides procèdent de la manière suivante : Une clé aléatoire est générée pour l'algorithme symétrique, cette clé fait généralement entre 128 et 512 bits selon les algorithmes. L'algorithme de chiffrement symétrique est ensuite utilisé pour chiffrer le message.

La clé aléatoire quant à elle, se voit chiffrée grâce à la clé publique du destinataire, c'est ici qu'intervient la cryptographie asymétrique tels que le RSA. Comme la clé est courte, ce chiffrage prend peu de temps. Chiffrer l'ensemble du message avec un algorithme asymétrique serait bien plus lourd, c'est pourquoi on préfère passer par un algorithme symétrique.

Il suffit ensuite d'envoyer le message chiffré avec l'algorithme symétrique et accompagné de la clé chiffrée correspondante.

Le destinataire déchiffre la clé symétrique avec sa clé privée et via un déchiffrement symétrique, retrouve le message.

3. Algorithmes utilisés

Pour la cryptographie symétrique nous avons opté pour l'algorithme le plus récent l'AES (*ADVANCED ENCRYPTION STANDARD*) et pour la cryptographie asymétrique le RSA (*RIVEST, SHAMIR et ADELMAN*).

3.1. *L'algorithme RSA*

Le RSA est le premier algorithme asymétrique publié en 1977.

La sécurité du RSA, qui est nommé d'après les noms de ses inventeurs (RIVEST, SHAMIR et ADELMAN), vient de la difficulté de factoriser des grands nombres premiers : s'il est facile de multiplier deux grands nombres premiers, il est très difficile de décomposer le très grand nombre obtenu en ses deux facteurs premiers quand on ne les connaît pas.

Les clés publique et privée sont une fonction d'un couple de grands nombres (1024 bits ou plus) premiers. Découvrir le texte en clair à partir de la clé publique et du texte crypté est conjecturé comme équivalent à factoriser le produit des deux grands premiers.

La clé publique contient le produit de deux nombres premiers très grands, et un autre nombre qui lui est propre. L'algorithme de chiffrement utilise ces nombres pour chiffrer le message par blocs. L'algorithme de déchiffrement nécessite quant à lui l'utilisation d'un nombre contenu uniquement dans la clé privée [1].

L'algorithme RSA utilise l'arithmétique de *ęn*, où *n* est le produit de deux nombres premiers distincts *p* et *q*, choisis de façon aléatoire : $n = p \cdot q$ (1)

Ensuite il faut choisir une clé publique e de tel manière que e et Ø(n) soient premiers entre eux avec $Ø(n) = (p - 1)(q - 1)$. (2)

Finalement, à l'aide de l'algorithme d'Euclide étendu, on calcule la clé de décryptage d, tel que : $e \cdot d \equiv [1] \bmod Ø(n)$ (3)

D'une autre façon : $d= [e^{-1}]_{Ø(n)}$ (4)

Il faut noter que d et e sont relativement premiers.

Les nombres e et n forment la clé publique, pendant que les nombres d et n sont la clé privée. Les nombres *p* et *q* ne sont plus alors nécessaires et ne doivent être jamais révélés.

Pour crypter un message m, il est tout d'abord nécessaire de le découper en blocs numériques de taille plus petite que n (normalement la plus grande puissance de 2 plus petite que n), composés de bits binaires. c-a-d, si *p* et *q* sont de 1024-bits chacun, donc le modulo n sera sur 2048-bits et chaque bloc mi doit être légèrement au-dessous de 2048-bits. Le message crypté c sera aussi divisé en blocs ci de la même taille que les mi.

La formule de cryptage est la suivante : $c_i= [m^e_i]_n$ (5)

et le décryptage est obtenu simplement par : $m_i = [c^d_i]n$ (6)

Puisque : $c^d_i = (m^e_i)^d = m^{ed}_i = m_i^{k(p-1)(q-1)+1} = m_i \times 1 = m_i$ (7)

Les clés RSA sont habituellement de longueur comprise entre 1024 et 2048 bits [1].

3.2. *L'algorithme AES*

L'AES (Advanced Encryption Standard) est, comme son nom l'indique, un standard de cryptage symétrique destiné à remplacer le DES (Data Encryption Standard) qui est devenu trop faible au regard des attaques actuelles, il est plus sûr que le 3DES car il présente, entre autres, une plus grande résistance aux attaques par dictionnaires de clés. Les autres attaques ne sont pas applicables dans son cas [2].

Le développement de l'AES a été instigué par le NIST (National Institue of Standards and Technology) le 2 janvier 1997. L'algorithme a été choisi il y a peu de temps : il s'agit de l'algorithme Rijndael (prononcer "Raindal").

A la suite de nombreux tests, c'est finalement Rijndael qui a remporté la médaille, et est ainsi devenu le remplaçant officiel du DES.

Il possède les propriétés suivantes :

– Plusieurs longueurs de clef et de bloc sont possibles :
128, 192, ou 256 bits ;

– Le nombre de cycles (ou "rondes") varie en fonction de la longueur des blocs et des clés (de 10 à 14) ;

– La structure générale ne comprend qu'une série de
transformations/permutations/sélections ;

– Il est beaucoup plus performant que le DES ;
– Il est facilement adaptable à des processeurs de 8 ou de 64 bits ;
– Le parallélisme peut être implémenté

À chaque ronde, quatre transformations sont appliquées :

- Substitution d'octets dans le tableau d'état.
- Décalage de rangées dans le tableau d'état.
- Déplacement de colonnes dans le tableau d'état (sauf à la dernière ronde).
- Addition d'une "clef de ronde" qui varie à chaque ronde [2].

Le Rijndael procède par blocs de 128 bits, avec une clé de 128 bit également. Chaque bloc subit une séquence de 5 transformations répétées 10 fois :

➢ Addition de la clé secrète (par un ou exclusif).

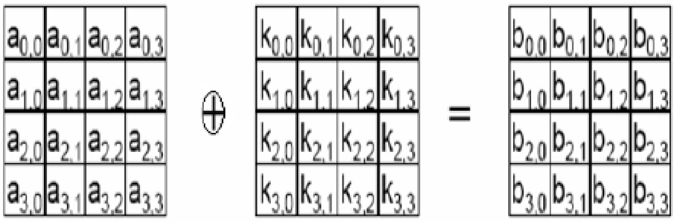


Fig. 3. Addition de la clé secrète [8]

➢ Transformation non linéaire d'octets : les 128 bits sont répartis en 16 blocs de 8 bits (8 bits=un octet), eux-mêmes dispatchés dans un tableau 4×4. Chaque octet est transformé par une fonction non linéaire S [3].

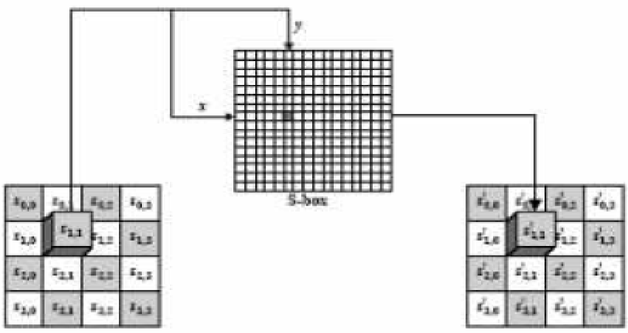


Fig. 4. Substitution d'octets [8]

- Décalage de lignes : les 3 dernières lignes sont décalées cycliquement vers la gauche : la 2ème ligne est décalée d'une colonne, la 3ème ligne de 2 colonnes, et la 4ème ligne de 3 colonnes [3].

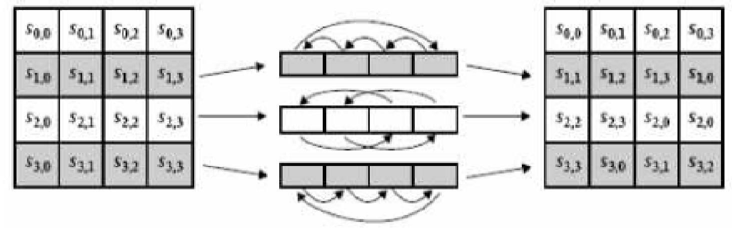


Fig. 5. Schéma de décalage des lignes [8]

- Brouillage des colonnes : Chaque colonne est transformée par combinaisons linéaires des différents éléments de la colonne (ce qui revient à multiplier la matrice 4×4 par une autre matrice 4×4). Les calculs sur les octets de 8 bits sont réalisés dans le corps à 28 éléments [3].

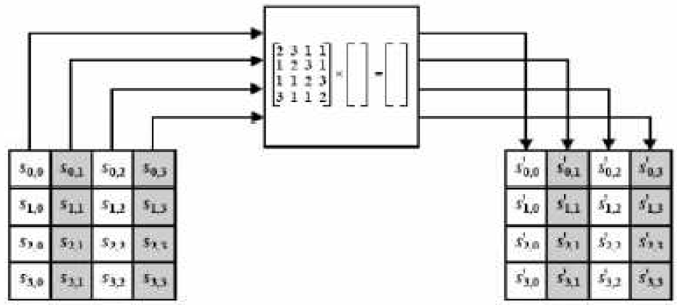


Fig. 6. Étape du brouillage des colonnes [8]

- Addition de la clé de tour : A chaque tour, une clé de tour est générée à partir de la clé secrète par un sous-algorithme (dit de cadencement). Cette clé de tour est ajoutée par un ou exclusif au dernier bloc obtenu [3].

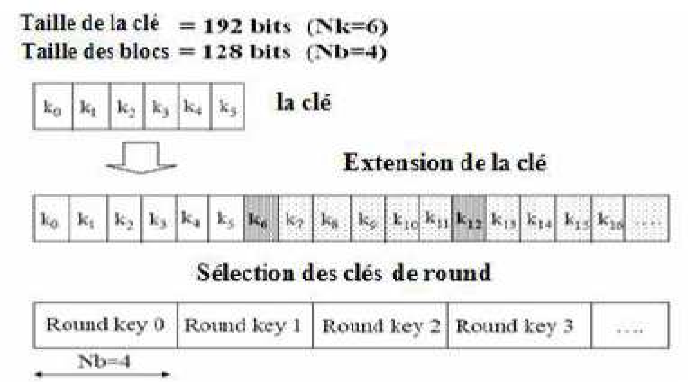


Fig. 7. Schéma des opérations effectuées sur la clé [8]

Remarque : Ici nous avons donné un exemple pour une clé de 192bits.

Cette opération sera répéter 10 fois appart que la clé de tour sera modifié à chaque tour et la clé secrète ne sera utilisé que dans le premier tour. A la fin nous obtenons un bloc crypté à 128 bits dont il sera transmis à la destination et à la destination on fera le décryptage c.à.d. les étapes inverse du cryptage sauf qu'on va commencer par additionner la 10éme sous clé de tour et on termine avec la clé secrète pour récupérer le message clair c.à.d. les données initiales à 128 bits.

La figure 8 nous récapitule les 5 étapes du cryptage des données.

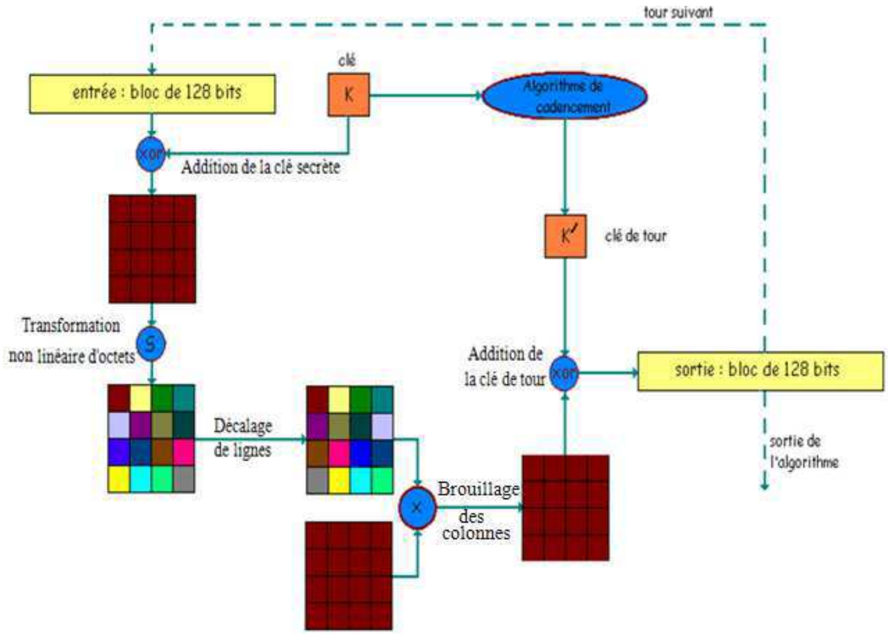


Fig. 8. Schéma détaillé de l'algorithme AES [3]

4. Language de description matériel et matériels utilisés

4.1. *Language de description matériel VHDL*

Le VHDL est un langage informatique qui décrit les circuits. Les programmes de test simulent l'environnement d'un montage, ils bénéficient de la souplesse de l'informatique pour recréer virtuellement toutes les situations expérimentales possibles et imaginables, même celles qui seraient difficiles à réaliser en pratique.

VHDL → VHSIC **H**ardware **D**escription **L**anguage

A ce jour, on utilise le langage VHDL pour :

- Concevoir des ASIC,
- Programmer des composants programmables du type PLD, CPLD et FPGA,
- Concevoir des modèles de simulations numériques ou des bancs de tests.

Une description VHDL se compose de deux parties :

- La description de l'interface du circuit (appelée *entity*) avec le monde qui l'utilise (connecteur, brochage, interface,…). Celle-ci est constituée par la liste des signaux de cette interface, leur sens, leur nature, etc.
- La description de la réalisation du circuit (appelée *architecture*) qui peut contenir trois formes de descriptions :

La description de l'interconnexion de sous circuits (cette forme de description peut être appelée *structurelle*) [4].

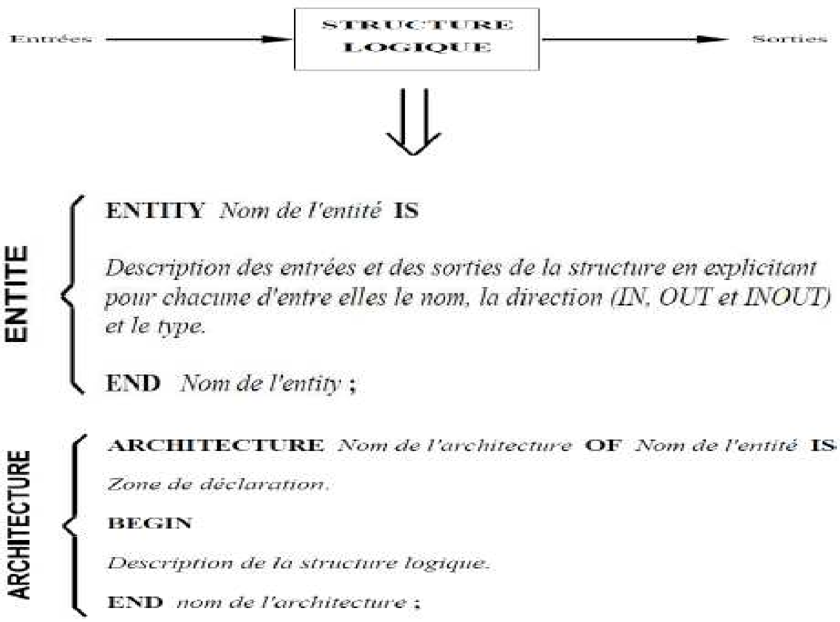


Fig. 9. Structure logique [4]

4.2. *Les circuits FPGA*

Les FPGAs (Field Programmable Gate Arrays ou "réseaux logiques programmables") sont des composants entièrement reconfigurables, ce qui permet de les reprogrammer a volonté afin d'accélérer notablement certaines phases de calcul. Ces circuit ont été inventes par la société XILINX en 1984.

L'avantage de ce genre de circuit est sa grande souplesse qui permet de le réutiliser a volonté dans des algorithmes différents en un temps très court (quelques millisecondes ou parfois moins selon la complexité de l'algorithme et les capacités de l'FPGA).

Un circuit FPGA est composé de nombreuses cellules logiques élémentaires librement assemblables. Celles-ci sont connectées de manière définitive ou réversible par programmation, afin de réaliser la ou les fonctions numériques voulues. L'intérêt est qu'une même puce peut être utilisée dans de nombreux systèmes électroniques différents.

Les FPGA permettent d'atteindre un niveau d'intégration très élevé. En première approximation, un FPGA est un circuit avec un très grand nombre de macro-cellules et une grande souplesse d'interconnexion entre elles [5].

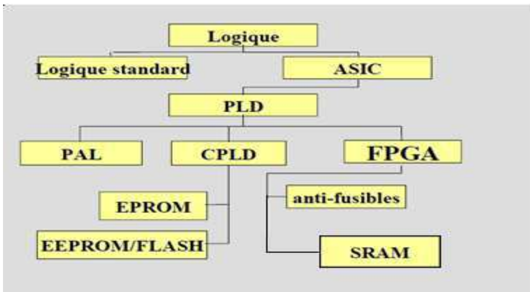


Fig. 10. Les différentes technologies et leurs circuits [5]

- Un FPGA est un réseau (matrice) de blocs combinatoires et séquentiels (CLB).
- Des blocs d''entrée/sortie (IOB) sont associés aux broches du circuit.
- Les CLB et IOB sont interconnectés entre eux par des dispositifs variés.
- Les matrices s'organisent de 8x8 à 128x120.

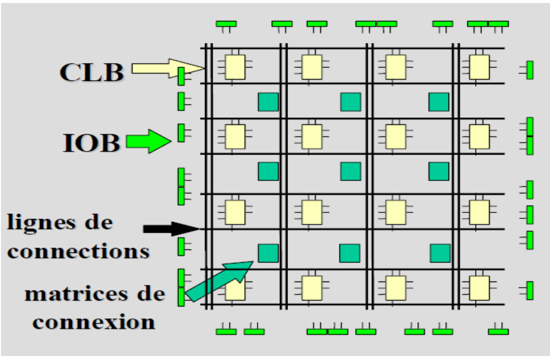


Fig. 11. Architecture générale d'un FPGA [5]

4.3. *Relation entre VHDL et FPGA*

L'implantation d'une ou de plusieurs descriptions VHDL dans un circuit FPGA va dépendre de l'affectation que l'on fera des broches d'entrées/sorties et des structures de base du circuit logique programmable.

Le schéma ci-dessous représente un exemple de descriptions VHDL ou de blocs fonctionnels implantés dans un FPGA. Lors de la phase de synthèse chaque bloc sera matérialisé par des portes et/ou des bascules. La phase suivante sera d'implanter les portes et les bascules à l'intérieur du circuit logique. Cette tâche est réalisée par le logiciel placement/routage, au cours de laquelle les entrées et sorties seront affectées à des numéros de broches.

On peut remarquer sur le schéma la fonction particulière du bloc VHDL N°5. En effet dans la description fonctionnelle d'un FPGA on a souvent besoin d'une fonction qui sert à cadencer le fonctionnement de l'ensemble, celle-ci est très souvent réalisée par une machine d'états synchronisée par une horloge.

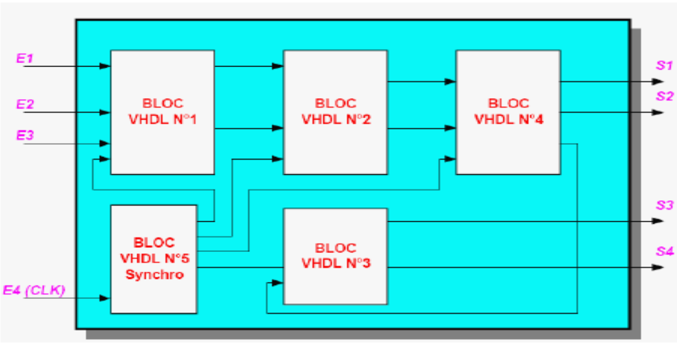


Fig. 12. Schéma fonctionnel d'implantation d'une description VHDL dans un circuit FPGA [6]

5. Résultats de simulation du crypto système AES

Nous allons présenter les résultats de simulation du crypto système hybride que nous avons implémenté sur FPGA. Pour ce faire le langage VHDL a été utilisé pour la description de toutes les cellules qui composent notre schéma qui a été présentée dans la figure précédente. Notre architecture est composée de deux parties qui sont la partie cryptage et la partie décryptage.

Nous allons alors traiter les deux parties séparément et voir le résulta de simulation avec le logiciel Model SIM. Il est à noter aussi que le cryptage de la clé AES est accompli par l'algorithme RSA qui est réalisée par un software basé sur l'outil MAPLE.

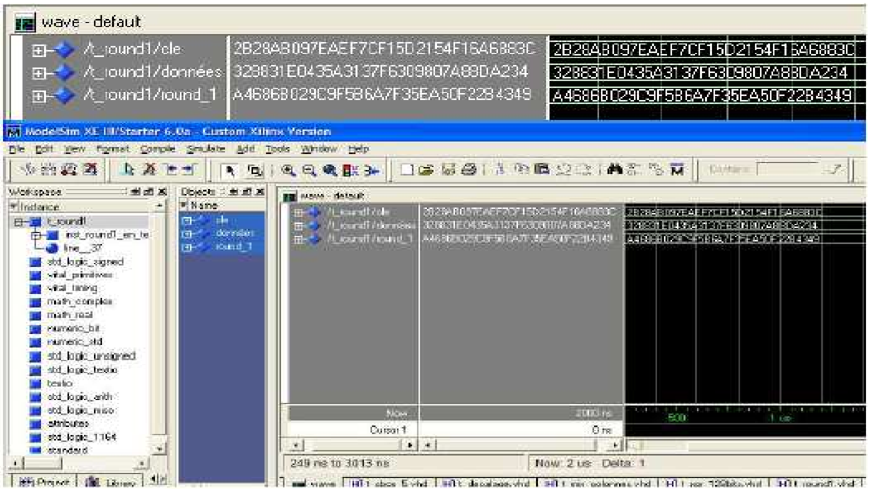


Fig. 13. Résultats de simulation de la cellule 1er Round

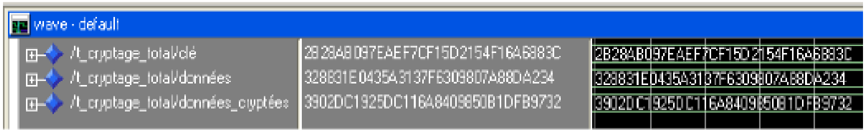


Fig. 14. Résultats de simulation du cryptage des 10 Rounds

Fig. 15. Résultats de simulation de la cellule Calcul de la 1ère sousclé (Round Key)

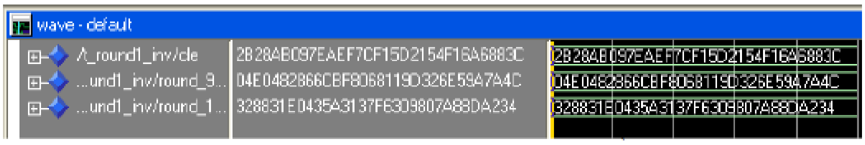


Fig. 16. Résultats de simulation du 10ème Round

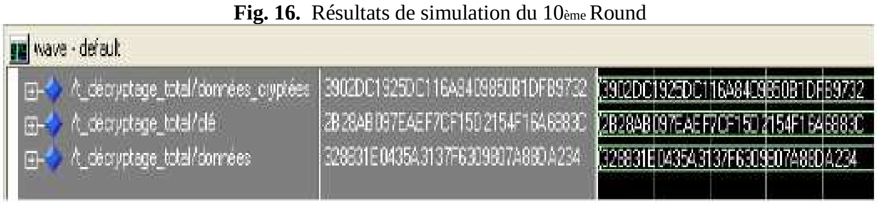


Fig. 17. Résultats de simulation du décryptage des 10 Rounds

Nous allons données les résultats détaillées représentées en hexadécimale sur les deux parties.

5.1. *Partie cryptage*

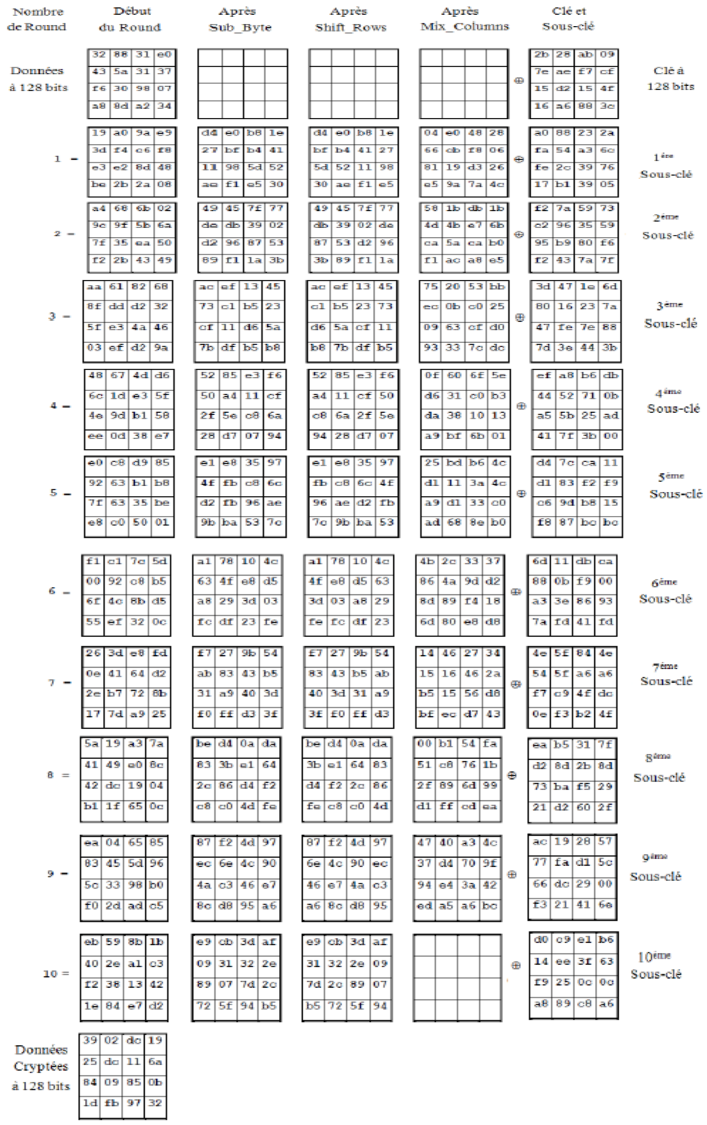


Fig. 18. Résultats de simulation du cryptage

Les résultats de simulation de l'architecture de la partie cryptage a été concluante. Dans cette simulation, nous avons utilisé des données ainsi que la clé sur une taille de 128 bits. La partie cryptage est en fait composée de 10 rounds et chaque round est composé des étapes suivantes.

- XOR 128 bits entre les données et la clé
- Sub_Byte

- Shift_Rows
- Mix_Columns
- XOR 128 bits entre l'étape précédente et une sous-clé.

Sauf pour le dernier Round qui ne contient pas l'étape Mix_Columns.

Après avoir simulé les dix Rounds successivement, on obtient les données cryptées à 128 bits.

5.2. *Partie décryptage*

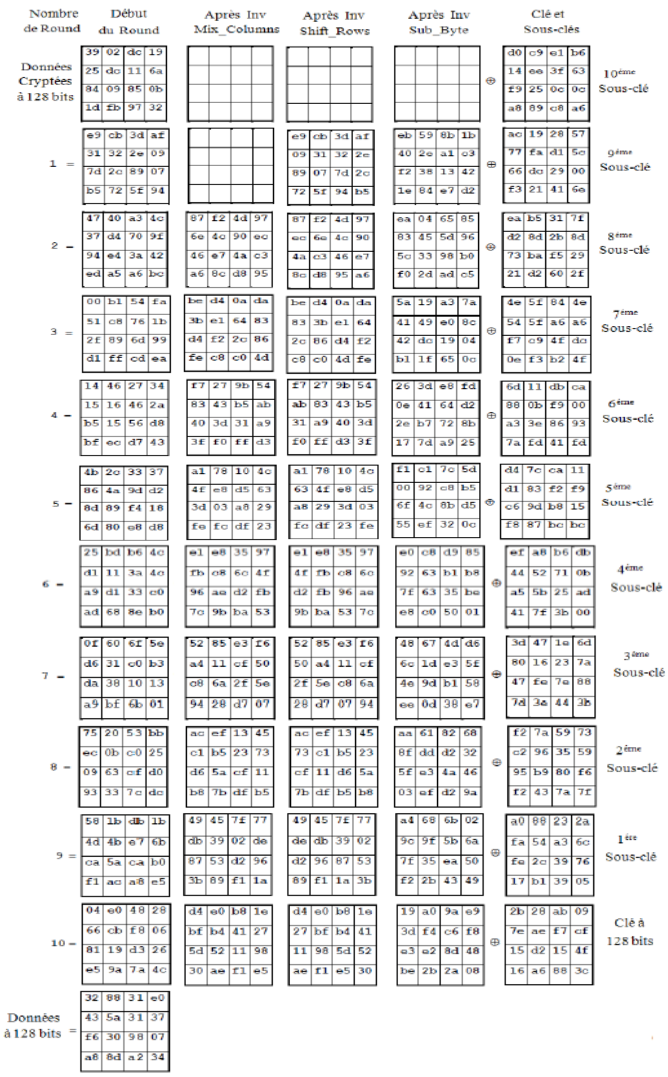


Fig. 19. Résultats de simulation du décryptage

Aux données cryptées, issues de la partie cryptage, on réalise l'opération inverse qui est le décryptage avec la même clé utilisée dans la partie cryptage. Les résultats de simulation de l'architecture qui fait le décryptage sont satisfaisants. De même la partie décryptage est constituée de 10 rounds et chaque round contient les étapes suivantes :

- XOR 128 bits entre les données cryptées et la 10éme Sous-clé
- Inv Mix_Columns
- Inv Shift_Rows
- Inv Sub_Byte
- XOR 128 bits entre l'étape précédente et une sous-clé ou la clé.

Sauf pour le premier Round qui ne contient pas l'étape Inv Mix_Columns.

Après avoir simulé les dix Rounds successivement, on obtient les données décryptées (données initiales) à 128 bits.

6. Conclusion

Nous avons commencé notre travail par une introduction à la cryptographie où les protocoles cryptographiques symétriques et asymétriques ont été abordés pour comprendre le fonctionnement de ce crypto système.

Puis une comparaison entre ces deux systèmes a fait ressortir les avantages et les inconvénients des uns et des autres pour conclure que combiner les chiffrements symétrique et asymétrique permet d'échanger des messages de manière sûre et rapide sans échange préalable de secret. Ensuite, nous avons étudié l'AES dans le but de cerner les opérations mathématiques induites dans ce protocole cryptographique symétrique.

Après que le choix, des méthodes hybride RSA pour la génération de la clé et l'AES pour le cryptage, ait été fixé ; l'architecture réalisant le chiffrage/déchiffrage AES a été conçue puis décrite en utilisant le langage VHDL et enfin implémentée sur un circuit FPGA de la famille Virtex-4 de Xilinx sous l'environnement ISE Foundation 7.1.

Il est à noter que la partie RSA n'a pas été réalisée dans ce travail, vu que celle-ci a été réalisée en software auparavant et que dans ce travail, nous avons alors focalisé sur l'AES. Les performances du système hybride dépendent en grande partie des performances de l'AES et c'est pour cela que ce protocole a fait l'objet d'une implémentation matérielle.qui sera dédiée au cryptage et décryptage d'information volumineuse en temps réel.

Perspectives

Le trafic des images numériques est en augmentation continuelle sur les réseaux. La protection des données numériques, devient importante pour de nombreuses raisons telles que la confidentialité et l'intégrité. Comme perspective Le cryptage de ces images répond à ce problème de confidentialité.

Pour empêcher ces attaques, la sécurité doit être omniprésente. Aussi bien lorsque les images ne sont pas utilisées que lorsqu'elles transitent sur le réseau. D'où la nécessité de crypter les images.

Références

[1] : D. G. MESQUITA, (2006) "Architectures reconfigurables et cryptographie : une analyse de robustesse et contremesures face aux attaques par canaux cachés" thèse *de* Doctorat de l'université de Montpellier II.
[2] : J. P. Gaulier, (1993) "Analyse des algorithmes finalistes concourant pour le futur standard AES" Conservatoire National des Arts et Métiers Centre Régional Associé de Limoges, Mémoire de synthèse.
[3] :http://www.bibmath.net/crypto/moderne/aes.php3
[4] :T. BLOTIN "Le langage de description VHDL" Lycée Paul-Eluard 93206 Saint Denis.
[5] : F. Verdier, (2006) "Les circuits FPGA Concepts de base, architecture et applications " Cours Master Recherche, Université de Cergy-Pontoise.
[6] : AMINE B. and CHOUKRI ADEL C, (2006) "Implémentation sur FPGA des méthodes MPP "P&O" et "floue optimisée par les Algorithmes Génétiques" ; école national polytechnique Alger ; Algérie.
[7] :http://www.securiteinfo.com/cryptographie/symetrique.shtml
[8]: R. Dument, (2006-2007) "introduction à la cryptographie et à la sécurité informatique"; seconde partie du cours de système de conduite des ordinateurs ; notes de cours.

International Conference on Software Engineering and New Technologies

EFFICIENT FEATURE BASED IMAGES REGISTRATION FOR HIGH RESOLUTION SATELLITE IMAGES

BENNOUR AKRAM
University of Tebessa , Algeria
Akram.bennour@gmail.com ,

BORNIA TIGHIOUART
University of Annaba , Algeria
b_tighiouart@yahoo.fr

Abstract

Two decades ago, the satellites gave us global continued images that often separated by several month or years which influence the clarity of the image . The random variation of satellite attitude and the sensor defect, creates numerous special deformations . So it necessary appeared to put automatic or semi-automatic analysis methods for these images to make their comparable, interpretable, and useful. This work is devoted to the process that geometrically aligns at best the common part of multitemporal satellite images, and that by determine the transformation that provides the most accurate match between the images. This process is known by satellite image registration and it is a crucial process in many remote sensing applications. The main algorithm consists of three phases. The first one is the automated generation of control points, and the second one is the robust estimation of mapping functions from control points. Finally, a global linear spatial transformation is applied and the remotely sensed image is efficiently registered. This paper is focused essentially on a comparative study between three novel automatic and efficient feature-based approaches for high resolution satellite images registration, which are those based on Harris corner detector ; scale invariant features named SIFT and speed up robust features called Surf . The features matching phase is carried out using a bidirectional correlation approach leading to a robust set of control points in the multi temporal remote sensing images. We used the random sample consensus (RANSAC) algorithm to handle outliers resulting from the matching process. The methods has been implemented and tested using various high resolution remote sensing imagery including landsat data and spot 5 images over test sites in Kuwait, Copenhagen and "lile de Ré". In experimental results we demonstrate the effectiveness of the proposed approaches and show that they produces accurate results when applied to the registration of high resolution satellite images as well as multitemporal ones, and we notice that SIFT based approach perform the best .

Keyword : satellites image registration feature detection, HRRIS ,SIFT, RANSAC, SURF.

1. Introduction

The automatic processes of the satellite images provide information about the dynamic of the earth. Various fields are concerned by these information, such as map updating, video geo-registration, environmental monitoring, multitemporal classification, change detection, and pattern target localization etc. The most of these applications requires a complex phase of comparison between a couple or several of images. This comparison is only possible if we can assure that the images correspond to the same geographic scene . For this purpose we critically need a process that geometrically aligns at best the common part of multitemporal and/or multisensory satellite images. This process is known by satellite image registration. Because of its importance and its complicated nature, image registration has been the topic of much recent researches . A comprehensive survey of image registration methods was published in 1992 by Brown [BRO 92] and more recently by Zitova and Flusser [NIS 06] and Medha.V and all in [MED 09]. Typically, registration approaches are classified according to their nature, into two broad categories : feature-based and area-based methods [NIS 06], [BRO 92]. Our work belong to the first category of techniques, which are fast and accurate due to the extraction and the use of the useful information only . So it requires in general a primary phase of control points (CPs) detection . A CP is a point whose source image coordinates and reference coordinates are known.

The feature based registration technique via manual selection of CPs is a complex, time intensive , and laborious task [TOT 92], [DAI 99]. Therefore Much of previous researches has devoted to the automatic detection of control points and they improve its efficiency. In general image registration process can be composed to three steps [TON 98]: the first one is control points (CPs) detection, secondly a mapping function estimation between the image to be registered to the reference image, and Finally, a global linear spatial transformation is applied to align source and reference images . In the literature some researchers Used invariant properties of image features [FLU 94], [SHA 02], Others proposed the use of generic algorithm-based matching [JAC 95], the automatic selection of various image features as candidate control points is used in [TON 98], [GOV 99] and [BEN 08], authors in [DJA 93],[MCG 00] used a multi-resolution and wavelet-based approach for matching process. The use of additional information such as digital terrain models (DTMs) have been studied among others [RIG 91].

This paper is focused essentially on a comparative study between three novel and efficient automatic feature-based approaches for high resolution satellite image registration, which are those based on Harris corner detector ; scale invariant features called SIFT and speed up robust features named surf, respectively as features detectors . Firstly, interest points are extracted respectively using Harris corner detector proposed by by Harris and Stephens [HAR 88], scale invariant features called SIFT published by Lowe [LOW 04], and speed up robust features known as SURF published by Bay et al [BAY 06] . which are three of most accurate feature detectors proved in the literature. In the next section we will present more details for each one. Secondly candidates of homologous points are located in the three cases using a similarity measure of correlation coefficients in bidirectional cross-correlation ; then For robust estimation of mapping functions in presence of outliers, we will use the random sample consensus (RANSAC) algorithm [FIS 81]. This algorithm has been reported to handle outliers successfully in many applications [CHE 95], [TOR 01]. We will show that it works too for automated image registration. Finally, a global rigid transformation is generated without outliers and applied to align source and reference images. A comparative study is handled to show the behavior and the influence of the proposed features detectors on the image registration results, which have a crucial underlying effect on the most of satellite images application as we saw above .

The layout of the paper is as follows : In the next two sections the details of the image registration techniques developed in this work are described. Afterwards, experimental results using high-resolution satellite imagery with varying geometric properties are presented. Finally, conclusions are drawn and recommendations for future work are presented.

2. Satellite images registration

The different imaging conditions in satellite imagery provide differences between images such as geometric distortions which can harm the profit of the information given by these images . So we crucially need a process to overcome this impediment. this process is called image registration and it lead to geometrically align at best two or several images of the same scene taken at different times, from different viewpoints, and/or by different sensors, and that by determine the transformation that provides the most accurate match between two images. To obtain an efficient image registration algorithm we must settle some essential criterions [BRO 92], [ESP 02]: the first and crucial one is the research space which is in general a spatial transformation set [MAI 98], [LEE 03], represented in our case by some translations and rotations and called rigid transformations. These transformations can be Global (i.e. applied to the entire image) or local (i.e. image is divided into regions having their own transformations). We have adopted a global and rigid linear registration because of the multi-temporal satellite image characters. Indeed there is may be a time lag of the entire image due to the position of the sensor at the time of taking image which requires a global transformation of the image while keeping its contents as it is. The second one is to define the search strategy to explore the search space in the process of image registration. In our context of feature-based techniques, the research will be direct. we must also chose the similarity measure to be used for matching interest points extracted from the tow images in the features detection step , and that to fined control points which are couples of corresponding points in the two images .The most used are based on the bidirectional correlation research between the images intensity [ESP 02], [LEE 03] , [RAB 05], which is a robust approach inspired by stereo-vision and video analysis algorithms. More exactly, we use a ZNCC (Zero mean Normalized Cross Correlation) based on the following photometric optimization criterion:

$$c_{x,y} = -\sum_{i,j}\left(I_1(x+i,y+j)-\overline{I_1}(x,y)\right)\left(I_2(x+i,y+j)-\overline{I_2}(x,y)\right) / \sqrt{\sum_{i,j}\left(I_1(x+i,y+j)-\overline{I_1}(x,y)\right)^2}\sqrt{\sum_{i,j}\left(I_2(x+i,y+j)-\overline{I_2}(x,y)\right)^2} \quad (1)$$

Where I1 and I2 represent respectively the two images intensity.

Firstly we seek to find a match, in the sensed image, for each interest point of the reference image. Then we proceed an inverse correlation procedure which ensures the robustness of the whole matching scheme. Indeed, only the interest points of the reference image, whose corresponding points (in the sensed image) during the 1st matching step lead themselves to the former original points in the 2nd matching step are considered as successful matches and then retained as control points pairs. But CPs from the first step will always contain false matches, and these will act as outliers and hinder accurate estimation of mapping functions, for this purpose we need a robust estimation to overcome the effects of false matches and achieve correct mapping functions. Therefore we used the powerful estimator called random sample consensus known as RANSAC [FIS 81] which perform well in presence of outliers. It does not require prior assumption of the distribution of outliers. As long as we can postulate boundaries between inliers and outliers [SHA 02], we can use RANSAC to cope with outliers. It works by estimating a model with the minimum required number of control points selected randomly and checking whether other control points support the model. It repeats these procedures for a certain number of times and chooses the best model that has the largest supports. After that, it re-estimates the model using all supporting control points.

The overall approach used to image registration is explained in fig.1. It is based on the extraction of characteristic points, the determination of the control points, the definition of the transformations and the establishment of the transformations in order to obtain the registered image.

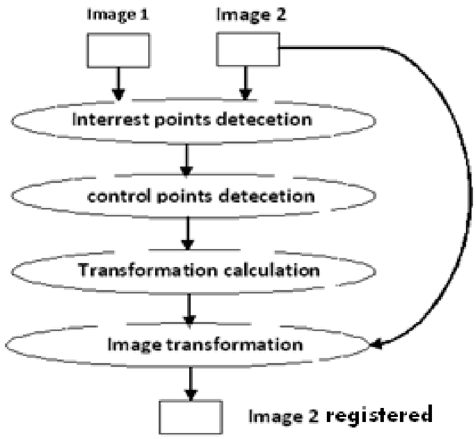


Fig 1 : Approach to Image registration

3. Feature detection

As we saw above, the accuracy and robustness of the feature detector and matching algorithm have a direct impact on the accuracy of image registration process. Ideally for image registration process , feature detectors need to be robust and be able to locate the same features in successive images irrespective of image rotation, translation, scaling or changes in illumination .In satellite images registration context it is not obvious to look for geometric structures such as lines , so The search for interest points would be more interesting. But it is difficult to define them in a formal way; however many studies have been done into the effectiveness of the different feature detection algorithms such as Harris corner detectors and feature descriptors such as SIFT (Scale Invariant Feature Transform) and SURF (Speeded Up Robust Features). A comprehensive overview of the current methods for the feature detection is presented by Schmid and all in [SCH 98] .In [SCH 00] a practical comparison of several feature detectors was accomplished. The operators of Harris [HAR 88] , Forstner [FOS 94], as well as Cottier [COT 94] were evaluated quantitatively. In [ZUL 04] authors presented a mathematical comparison of the Harris, Noble, Kanade-Lucas- Tomasi(KLT) and Kenney point detectors. Authors observed that that the Harris operator perform best and was the most stable of all. In [LOW 04] Lowe described image feature generation with silent invariant feature transform SIFT. Mikolajczyk [MIK 05] compared SIFT with other descriptors such us complex filters , cross-correlation , differential and moment invariants. It was proved that SIFT outperforms other algorithms in terms of invariancy. Recently Bay described in [BAY 06] a new and efficient scale and rotation invariant interest point detector and descriptor. It is inspired by the SIFT descriptor and named speed up robust feature SURF. SURF is claimed by its authors to be more robust and faster against different image transformations than SIFT.

In order to make the paper self-contained, We need first to describe the key elements arising each one :

3.1 HARRIS corner detector

Is one of most widely used and regarded as the toughest and most reliable originally proposed in [MOR 80], and improved later by Harris and Stephens [HAR 88]. It runs on a wide range of images and based on the mean change (in all directions) of the intensity in the neighborhood of a pixel (2).

$$E(x,y) = \sum_{u,v} w(u,v)[\frac{\delta I}{\delta x} + \frac{\delta I}{\delta y}]^2 \qquad (2)$$

when :

• W represents the neighborhood considered (value 1 inside the window and 0 outside);
• I (u, v) is the pixel (u, v) intensity;
• E (x, y) is the average change in the intensity when the window is moved by (x, y);

So we search the local maximum of the minimum values of E in each pixel. Practically we need to research local maximum of the operator HS as in (3,4):

$$HS = det\,(M) - K\,trace2\ (M) \qquad (3)$$

$$M = \begin{pmatrix} \widehat{I_x^2} & \widehat{I_x\ I_y} \\ \widehat{I_x\ I_y} & \widehat{I_y^2} \end{pmatrix} \qquad (4)$$

Such as :

• Î represent the result of the convolution of the picture I by a Gaussian filter.
• Det=determinant(M)
• Trace=trace(M)
• k is a parameter permitting to combine the information of contour given by the trace of M and the information of angularity data by the determinant of M (k=0 .04 according to the authors).

It should be noted that the problem of sensitivity of this approach to the noise that does not appear, as the processed images are of high resolution.

3.2 SIFT

The SIFT algorithm was published by Lowe [LOW 04] .It is used both as a features detector and descriptor. The SIFT algorithm takes an image and transforms it into a collection of local feature vectors. Each of these feature vectors is supposed to be distinctive and invariant to any scaling, rotation or translation of the image.

The steps involved in SIFT are :

• **Scale-space extrema detection:** Key locations are defined as maxima and minima of the result of difference of Gaussians function (DOG) applied in scale-space to a series of smoothed and resampled images :

A DoG image $D(x, y, \sigma)$ is given by :

$$D(x, y, \sigma) = L(x, y, K_i\sigma) - L(x, y, K_J\sigma) \quad (5)$$

Where :

$L(x, y, K\sigma)$ is the convolution of the original image $I(x, y,)$ with the Gaussian blur $G(x, y, K\sigma)$ at scale $k\sigma$, i.e.,

$$L(x,y,K\sigma) = G(x,y,K\sigma) * I(x,y) \qquad (6)$$

where $*$ is the convolution operation in x and y, and

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \qquad (7)$$

• **Keypoint localization:** Scale-space extrema detection produces too many keypoint candidates, some of which are unstable . therefore at each candidate location, a detailed model is fit to determine location and scale. This allows to discarded Low contrast candidate points and edge response points along an edge.
• **Orientation assignment:** Each keypoint is assigned one or more orientations based on local image gradient directions. thereby providing invariance to orientation, scale, andlocation for each feature.
• **Keypoint descriptor:** In This step the algorithm compute highly distinctive descriptors for the features points by creating a histogram of local oriented gradients around the interest point and stores the bins in a 128-dimensional vector .

3.3 Surf features detector

Speeded Up Robust Features is a high-performance scale and rotation-invariant interest point detector and descriptor recently published by Bay et al [BAY 06]. It is inspired by the SIFT descriptor and claimed by its authors to outperform previously proposed schemes against different image transformations. Features points are found by using a so called Fast-Hessian Detector that bases on an approximation of the Hessian matrix (8) for a given image point .

Given a point $\boldsymbol{x} = (x, y)$ in an image $\boldsymbol{I}$, the Hessian Matrix $\boldsymbol{H}(\boldsymbol{x},\boldsymbol{\sigma})$ in x at scale $\boldsymbol{\sigma}$ is defined n as follows

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \qquad (8)$$

Firstly , feature points are selected against different locations in the image, such as T-junctions and blobs. Next, the neighborhood of every interest point is represented by a feature vector. This descriptor has to be robust to geometric and photometric deformations ,noise and detection errors.

4. Experimental results and discussion

As we saw previously the proposed approaches are based respectively on features selection techniques: HARRIS , SIFT and SURFMI . the correlation rations are then computed for interest points of the reference and sensed image resulted from the first phase, this lead to a set of control points. But it remain some bat matches that provide outliers . We judge that RANSAC is the needed estimators for this purpose . finally the algorithm can generate the wanted transformation without outliers and register the images accurately .
In this study, tow datasets were used : In the first one we have experimented our satellite image registration scheme on a sequence of 15 images resulting from an artificial random rigid transformation (rotation and both x and y directions) of a High resolution (HR) multispectral SPOT 5 image (4 spectral bands, 2.5 m resolution) of an urban area. The original images are depicted on fig. 2 as well as one of its transformed copies among the obtained images of the sequence. The second set of images comes from a series of multi temporal High Resolution landsat scenes which differ from the reference by translations and rotations; these are shown in Fig. 2,3,

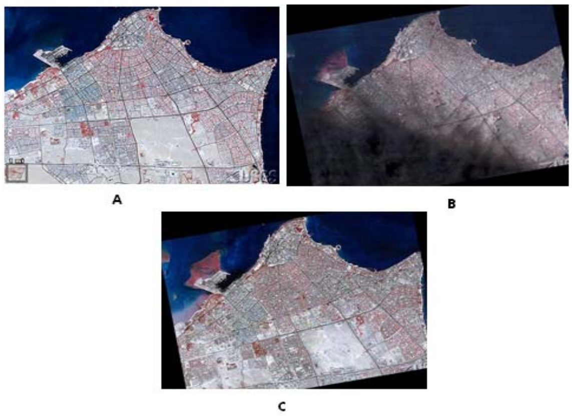


Fig. 2. A. Kuwait, Landsat TM, 31 August 1990
. B. Kuwait, Landsat 4 TM, 23 February 1991
. C. Kuwait, Landsat 5 TM, 14 November 1991

Fig.3. « l'ile de Ré » High Resolution (HR) multispectral satellite images pair. SPOT 5 (2.5m) image before (left) and after rigid transformation (right)

Since no good ground truth is available for these datasets , in a preliminary experiment we evaluate results visually by obtaining the superposition and the mosaics checkerboard of the reference image and the sensed image registered studied approaches as shown in Fig.5.

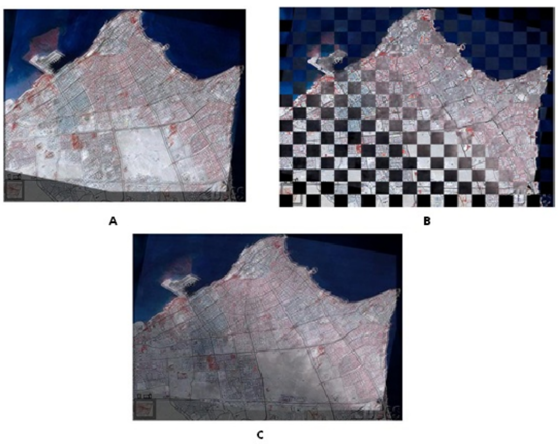


Fig. 4. A. Surposition of images in (fig 2 . A) and (fig 2.C) registered bay Harris based approache.
. B. mosaic checkerboared of (fig 2 . A)and (fig 2 . C)registered bay SIFT based approache.
.C. A. Surposition of images in (fig 2 . A) and (fig 2.B) registered bay SURF based approache.

We can notice that images are well aligned with HARRIS ,SIFT ,and SURF demonstrating the efficiency of these automatic satellite image registration approaches.

But visually experimentations don't allow as to advantage one algorithm to an other . To further demonstrate the reliability of the proposed registration techniques on HR satellite images, we carried out a quantitative evaluation on the first dataset which is a satellite image sequence resulting from random rigid transformations of the reference SPOT 5 image. The result of our registration scheme is exhibited on table.1. Where Mean and standard deviation of the rigid transformations estimation errors are presented such as The real applied rigid transformations being known, we have then calculated the mean and the standard deviation of the errors on the estimated rotation and translation parameters resulting from the registration process as follows:

$$E_i = \left|\Delta\alpha_i - \Delta\alpha'_i\right| .$$

$$\sigma_E = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(E_i - E)^2\right)}.$$

Where $\Delta\alpha_i$ is the real applied rigid transformations parameters on rotation and translations, $\Delta\alpha'_i$ is the obtained results, E is the mean and σ_E is the standard deviation of the results obtained for all images pairs in the dataset.

Used images	***Registration results***								
	SIFT			**SURF**			**HARRIS**		
	ΔX'	**ΔY'**	**Δθ'**	**ΔX'**	**ΔY'**	**Δθ'**	**ΔX'**	**ΔY'**	**Δθ'**
x25y12R12	25.41	12.13	12.16	25.07	11.74	11.84	25.13	11.90	11.88
x12y-15R6	12.50	-14.93	5.99	12.80	-15.07	6.02	12.30	14.96	5.98
x15y10R15	15.03	10.18	14.96	15.74	9.67	14.90	14.63	10.95	15.10
x10y-10R-10	10.14	-9.97	-10.00	10.54	-9.90-	-10.09	10.08	-9.92	-9.94
x-12y20R-15	-11.7	20.16	-14.98	-11.84	20.56	-15.43	-11.83	19.86	-15.30
x18y5R10	18.15	4.98	9.99	17.82	5.03	9.86	18.14	5.12	10.00
x-5y-12R-12	-4.85	-11.61	-11.90	-5.01	-12.01	-11.98	-4.06	-12.17	-12.12
x10y20R15	10.19	19.85	14.90	9.87	19.70	15.91	8.90	20.17	15.6
x15y-10R6	14.94	-9.90	6.03	15.06	-10.20	6.05	14.90	-9.90	6.06
x16y12R-6	16.03	12.79	-5.99	15.90	12.08	-5.87	16.10	12.14	-5.96
x-5y10R12	-4.88	9.92	12.03	-4.96	10.18	12.37	-4.88	10.10	12.12
x-10y20R-10	-9.87	20.15	-9.97	-9.71	19.42	-10.27	-9.96	20.20	-9.96
x-12y20R12	-11.90	19.90	11.90	-12.94	20.35	12.36	-11.35	20.10	12.12
x-12y20R-12	-11.99	20.22	-12.02	-12.17	20.09	-12.13	-11.49	20.15	-12.27
Mean (E)	**0.16**	**0.18**	**0.05**	**0.30**	**0.22**	**0,23**	**0.29**	**0.18**	**0.09**
Standard deviation (σ_E)	**0.02**	**0.04**	**0.01**	**0.11**	**0.04**	**0.06**	**0.10**	**0.05**	**0.01**

Table 1 : Registration results and Mean and standard deviation of the rigid transformations estimation errors

These statistical results confirm the reliability of these registration methods in the context of remote sensing data. In Table .2. we observe that all the algorithms occur well and closely. But we can notice that SIFT based approach perform the best, HARRIS become in the second place, and SURF based one in the last .

5. Conclusion

In this paper we have developed and compared novel and efficient registration algorithms intends to achieve a good quality alignment for two or more multi-temporal high resolution satellite images, in order to be able to compare eventual changes. Treated algorithms are those based respectively on HARRIS corner detector, SIFT and SURF keys-point descriptors. the similarity measure used was the correlation coefficient with combination of RNSAC estimator to dealt with outliers. The process is thoroughly tested using high resolution synthetic test data as well as multitemporal remotely sensed imagery.

The computed results are very encouraging , they show not only the feasibility of the dialed approaches but also its ability to achieve good quality solutions in fast runtime . all of the three algorithms compared in this work perform very well and so closely, and we can clearly notice that SIFT based ones is the best occur one. The following phase will also relate to the comparison of the reference image with the registered one to detect ecologic changes due to human activities in a both dense urban and rural area.

The remaining work is to tackle more complex high resolution satellite multitemporal and multimodalities images with several sites of interest.

References

[BRO 92]. Brown L.G (1992). *A Survey of Image Registration Techniques*. ACM Computing Survey. December 1992

[SCH 98] Schmid.C, Mohr.R, and Bauckhage.C, "*Comparing and evaluating interest points*," in proceedingof international conference on computer vision ICCV, 1998.

[SCH 00] Schmid.C, Mohr.R, and Bauckhage.C, "*Evaluation of interest point detectors,*" in Int. Journal of Computer Vision, vol. 37, no. 2, 2000.

[FOS 94] Forstner.W, "*A framework for low level feature extraction*" in Third Eurpoean conference on Computer Vision, 1994.

[COT 94] Cottier.J.C, "*Extraction et appariements robustes des points d'interet de deux images non etalonnees,*" LIFIA-IMAG-INRIA,Rhone-Alpes, Tech. Rep., 1994.

[HAR 88] Harris.C and Stephens.M.J, "*A combined corner and edge detectors,*" in Alvey Vision Conference, 1988.

[LOW 04] Lowe.D, "*Distinctive image features from scale invariant keypoints*" in International Journal of Computer Vision, 2004.

[MIK 05] Mikolajczyk.K and Schmid.C, "*A performance evaluation of local descriptors,*" *IEEE Transactions on Pattern Analysis & Machine Intelligence*,2005.

[ZUL 04]Zuliani.M, Kennedy.C, and Manjunath.B, "*A mathematical comparison of point detectors,*" in Conference on Computer Vision and Pattern Recognition Workshop, vol. 11, 2004.

[BAY 06] Bay.H, Tuytelaars.T, and.Gool.L.V, "*Surf: Speeded up robust features*" in Proceedings of the ninth Conference on Computer Vision, May 2006.

[MOR 80] Moravec.H, "*Obstacle avoidance and navigation in the real worl by a seeing robot rover,*" Carnegie-Mellon University,Robotic Institue, Tech. Rep., 1980.

[NIS 06]. Nisan A. "*Modèles Numériques de terrain à partir d'images stéréoscopiques*" . Master's thesis. Paul Cezanne University, France.2006.

[MED 09] Medha.V. W , Pradeep.P.M, and Hemant. K. A, *« Image Registration Techniques: An overview ».* International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 2, No.3, September 2009

[TOT 92] Toth.C.K and Schenk .T, "Feature-*based matching for automatic image registration"* , *ITC Journal*, vol. 1, 1992.

[DAI 99] Dai.X and Khorram.S, " *A feature-based image registration algorithm using improved chain-code representation combined with invariant moments", IEEE Trans. Geosci. Remote Sensing*, vol.37, Sept. 1999.

[TON 98]. Ton . J and Jain . A. K, "*Registering Landsat images by point matching".* IEEE Trans. Geosci. Remote Sensing, vol. 27, pp. 642–651. 1998.

[FLU 94]. Flusser . J and Suk .T , "*A moment-based approach to registration of images with affine geometric distortion". IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 382–387.1994.

[SHA 02]. Sharp. G. C, Lee . S. W, and Wehe . D . K," *ICP registration using invariant features".* IEEE Trans. Pattern Anal. Machine Intell., vol. 24, pp. 90–102.2002.

[JAC 95]. Jacp . J.J and Roux . C, "*Registration of 3D images by genetic optimization".* Pattern Recognit. Lett., vol. 16, pp. 823–841.1995.

[GOV 99]. Govindu.V and Shekhar . C ," Alignment *using distributions of local geometricproperties".* IEEE Trans. Pattern Anal. Machine Intell., vol. 21, pp. 1031–1043.1999.

[BEN 08] Bennour.A, Tighiouart.B, Boukir.S, *"Satellite image registration for the follow up of ecologic changes at long-dated".* In proc of *MCSEAI, Oran, Algeria*.2008.

[DJA 93] Djamdji . J.P , Bijaoui . A , and Manière .R . "*Geometrical registration of images: The multiresolution approach".* Photogramm Eng. Remote Sens., vol. 59, no. 5, pp. 645–653.1993.

[MCG 00]. McGuire . M and Stone . H. S," *Techniques for multiresolution image registration in the presence of occlusions". IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 1476–1479.2000.

[RIG 91] Rignot . E. J. M, Kowk . R , Curlander . C.J , and Pang. S.S,(1991). "*Automated multisensor registration: Requirements and techniques".* Photogramm Eng. Remote Sens., vol. 57, no. 8, pp. 1029–1038.1991.

[FIS 81] Fischler.M.A and Bolles.R.C, "*Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography".* Commun. ACM, vol. 24,.1981

[CHE 95] Cheng . Y. C and Lee . S. C," *A new method for quadratic curve detection using K-RANSAC with acceleration techniques"*. Pattern Recognit., vol.28.1995.

[TOR 01] Torr . P , Szeliski . R , and Anandan . P," *An integrated Bayesian approach to layer extraction from image sequence"s*, IEEE Trans. Pattern Anal. vol. 23.2001

[ESP 02] Espiau F.X. *"Métrologie 3D par vision active sur des objets naturels sous-marins"*. PhD thesis. Universityof Nice - Sophia Antipolis, France, February 2002.

[MAI 98] Maintz J. B. A. and Viergever M. A." *A survey of Medical Image Registration",* vol. 2, p. 1-36; Medical Image Analysis. 1998.

[LEE 03] Lee H.Y. et al. "Extraction of *digital elevation models from satellite stereo images through stereo matching based on epipolarity and scene geometry"*. Image and Vision Computing vol 21 . 789–796.2003.

[RAB 05]. Rabaud C. *"La mise en correspondance d'images stéréoscopiques"*. Laboratoire d'Informatique de Robotique et de Micro-électronique, Montpellier, France .2005

[SHA 02]. Sharp. G. C, Lee . S. W, and Wehe . D . K, *" ICP registration using invariant features"*. IEEE Trans. Pattern Anal. Machine Intell., vol. 24, pp. 90–102.2002

International Conference on Software Engineering and New Technologies

EFFECT OF THE POLARIZATION MODE DISPERSION IN A TRANSMISSION LINK BY OPTICAL FIBER

Mohamed CHIKH-BLED

Electronic Department, University Abou bakr belkaid,
Tlemcen / Algeria
mek_chikhbled@yahoo.fr

Amel MOKHDAR

Laboratory of Telecommunication, University Abou bakr belkaid,
Tlemcen /Algeria
amelmokhdar@yahoo.fr

The Polarization Mode Dispersion (PMD) as any other dispersion type is translated by a limitation of the bandwidth or from the product bandwidth- distance. It is related to the vectorial character of the light (most commonly indicated by the term of polarization). In this article we study the effect of polarization as well as the influence of the PMD on the transmission quality in the optical fiber link.

Keywords: Poincare Sphere, Degree of Polarization, polarization mode dispersion.

1. Introduction:

The optical fiber presents certain manufacturing defects (heart of elliptic form) or residual stresses which make anisotropic propagation. The phenomenon of PMD is due to this optical anisotropy making a fast axe and a slow one. Contrary to the other types of dispersion in fiber, the PMD varies very quickly in a dynamic way and results with a limitation of bandwidth. This article is organized as follows: section *2* relates to the Theory and the modeling of PMD in an optical fiber and then aims to explain briefly the PMD phenomenon. In section 3, we represent the simulated connection. Finally in section 4, we discuss these results.

2. Theory and modeling of the PMD in SMF fiber:

Mathematically, two formalisms are used to represent the polarization of a light wave as well as the effect of the optical devices on that: Jones and Stokes formalisms. The former is particularly well adapted to the description of completely polarized light waves, while that of Stokes makes it possible to describe the partially polarized waves [1].

2.1 Jones Formalism:

In the Forties, R.C. Jones introduced a mathematical formalism of description of the polarization effects based on a matrix representation [2]. In the Jones formalism, a polarization state is represented by a bi-dimensional complex vector. The elements of this vector, noted V, specify the amplitude and the phase of the components in x and y of the electric champ

$$V = \begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} E_{ox}\, e^{j\delta x} \\ E_{oy}\, e^{j\delta y} \end{pmatrix} \qquad (1)$$

The form of a polarization state depends only on dephasing δ_{xy} and on report E_{ox}/E_{oy} . This implies that the multiplication of the Jones vector by a constant complex does not modify the corresponding polarization state. Thus, the polarization state can also be defined by the Jones vector standardized (V_n) and defined by:

$$V_n = \frac{V}{\sqrt{E_{ox}^2+E_{oy}^2}} \qquad (2)$$

According to Jones, polarizations at the input S_{in} and the output S_{out} of a fiber length L are bound by a matrix relation represented by a matrix 2× 2 complex U [2]:

$$S_{out} = e^{-j\delta_0}\, u(z,\omega)\, S_{in} \qquad (3)$$

Where $u = \begin{pmatrix} u_1 & u_2^* \\ u_2 & u_1^* \end{pmatrix} \qquad |u_1|^2 + |u_2|^2 = 1$

δ_0 is dephasing between the two components of the electric champ , and U the Jones matrix who can be obtained by integrating the coupling mode equations [3].Once that the Jones matrix is obtained, the PMD can be easily calculated starting from the Jones matrix elements according to the following relation [4]:

$$PMD = \frac{2}{L}\sqrt{\left|\frac{du_1}{d\omega}\right|^2 + \left|\frac{du_2}{d\omega}\right|^2} \qquad (4)$$

2.2 Stokes Formalism:

In the Jones formalism, the polarization states are represented by complex numbers and are characterized by the amplitudes and the phases of the electric champ components. However, it is easier in practice to measure intensities. Moreover, the Jones formalism does not permit to represent the light partially polarized. This is why the Stokes formalism was introduced. With this one, a polarization state is represented by a four dimensions vector S, called Stokes vector:

$$\bar{S} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} \qquad (5)$$

With:

$$S_0 = total\ intensity\ of\ the\ light \qquad (6)$$

$$S_1 = P_0 - P_{\frac{\pi}{2}} \qquad (7)$$

$$S_2 = P_{\frac{\pi}{4}} - P_{-\frac{\pi}{4}} \qquad (8)$$

$$S_3 = P_{CR} - P_{CL} \qquad (9)$$

Where P_q ($q = 0, \frac{\pi}{2}, \frac{\pi}{4}\ ou - \frac{\pi}{4}$) indicate the optical intensity of the light which passes through a linear polarizer forming an angle q (expressed in radians) with axes Ox. P_{CR} and P_{CL} represent the optical intensities of the light which respectively pass through a circular polarizer right and left. In a general way, it is preferable to use the standardized Stokes parameters given by:

$$s_1 = \frac{S_1}{S_0} \qquad s_2 = \frac{S_2}{S_0} \qquad s_3 = \frac{S_3}{S_0} \qquad (10)$$

Then, the vector Stokes standardized is written:

$$\begin{pmatrix} 1 \\ s_1 \\ s_2 \\ s_3 \end{pmatrix} \qquad (11)$$

Consequently, the Stokes parameters vary between -1 and 1.The correspondence between the Jones and Stokes vectors is easy if we develop the expressions of (6) to (9).It comes (for a complete polarized light):

$$S_0 = E_{0x}^2 + E_{0y}^2 \qquad (12)$$
$$S_1 = E_{0x}^2 - E_{0y}^2 \qquad (13)$$
$$S_2 = 2Re\left(E_x^* E_y\right) = 2E_{0x}E_{0y}\cos\delta_{xy} \qquad (14)$$
$$S_3 = 2Im\left(E_x^* E_y\right) = 2E_{0x}E_{xy}\sin\delta_{xy} \qquad (15)$$

This vector can be expressed according to the ellipticity ε and of the azimuth α of the ellipse of polarization in the following way:

$$\bar{s} = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} \cos 2\alpha \cos 2\varepsilon \\ \sin 2\alpha \sin 2\varepsilon \\ \sin 2\varepsilon \end{pmatrix} \qquad (16)$$

This expression of the Stokes vector suggests a geometrical representation associating in each polarization state a point on a sphere of ray unit. The longitude and the latitude of this point are given respectively by 2α 2ε. The sphere obtained is called *the Poincare sphere*. Each polarization state at the input of the fiber for which the polarization state of corresponding output is independent of the frequency to the first order calls *PSP (Principal State of Polarization)*.This means that an optical impulse aligned with a PSP at the input will appear at the output with all its spectral components having the same polarization state [5].

The DOP (*Degree of Polarization*) is the relationship between the power of polarized part of an electromagnetic wave and its total power. This report could be calculated for one period of time for the signals which vary in time. The Stokes formalism offers a tool to study the partially polarized waves. The DOP is defined starting from the Stokes parameters by the following equation [6]:

$$DOP = \sqrt{s_1^2 + s_2^2 + s_3^2} \qquad (17)$$

Because of their perfect circular symmetry the ideal SMF fibers preserve polarization during the propagation of the electromagnetic champ. Conversely when radial symmetry

is broken, accidentally or not, a SMF fiber behaves as an anisotropic medium in which polarization evolves on function to the local birefringence properties. [7].
The birefringence term employed here translated the existence of two indexes of refraction effective n_x and n_y associated the two propagation constants β_x and β_y the modes LP_{01}^{x} and LP_{01}^{y} a fiber with weak guidance.
The physical phenomena inducing of birefringence in SMF fibers are due to the manufacture of the guides and external causes acting on fibers.

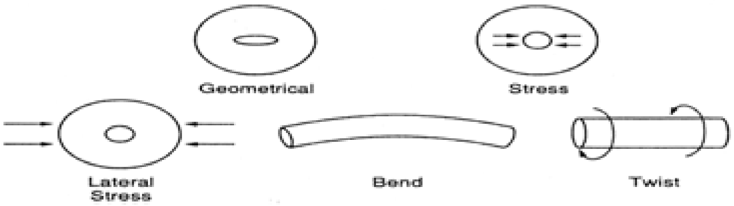


Fig 01. Causes of random birefringence in fiber.

The size characterizing the birefringence of fibers overall is the beat length L_B defined by:

$$L_B = \frac{\lambda}{n_y - n_x} = \frac{2\pi}{\beta_y - \beta_x} \quad (18) \qquad [8]$$

The typical value of L_B is approximately 10 m for the fibers standards of telecommunication [8]. In a real fiber, these properties of birefringence are not defined in a uniform way. The clean polarization modes and their propagation constants can then be defined only locally. The frequential dependence of the propagation constants associated with the clean modes with fiber is at the origin of the Polarization Mode Dispersion noted PMD.

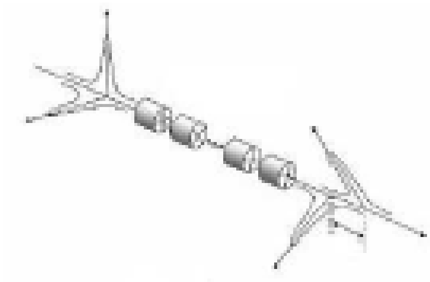


Fig 02. Effects of random birefringence

There are various manifestations of the PMD according to the adopted point of view:

- From the frequential point of view, the PMD appears by a variation of the polarization state in a point Z of fiber when the frequency of the input signal varies. [9]
- From the temporal point of view, the PMD appears by a variation of the travel time of an impulse up to point Z of fiber, i.e. variation of the group time, when the polarization of the input signal varies [9].

In each case, the effects of the PMD are different according to whether the clean polarization states are defined throughout fiber or not. For understanding the PMD in a real fiber well, it is necessary to distinguish two modes from propagation of the polarized light:

- A short distance mode, where it is supposed that there are two clean polarization states is well defined, and where only the local characteristics of birefringence are into account [10].
- A long distance mode, where the distribution of the clean axes and the distribution of birefringence must be taken into account [10].

When we send a signal in a "birefringent" fiber, without being concerned with its polarization, we excite the two modes at the same time. Each one has its own propagation velocity. This group travel times shift causes the unfolding of the signal at the fiber output, and thus a jamming of information (Figure 03) [11] [14-21].

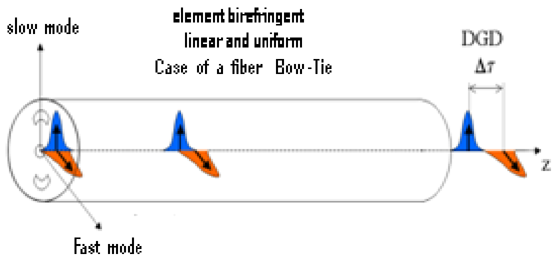


Fig 03. Effect of the PMD on an impulse.

An essential characteristic of this phenomenon lies in its randomness, since it is mainly of extrinsic origin and depends on the quality of the installation of this one. The optical fiber seems a fluctuating medium. The average value of the delay is not thus sufficient to completely describe it and we will thus use statistical data. Principal measurement is the DGD (Differential Group Delay), differential delay between the 2 components corresponding to the major propagation states.

$$DGD = \beta_i \sqrt{L_c}\,\sqrt{L} \qquad (19)$$

With β_i is linear birefringence, L_c and L are respectively the coupling length and the fiber length [12]

Until now days this shift was often neglected because it remains tiny. However, the lengthening of the optical support of transmission increases the value of this temporal shift whereas the breaking value decreases with the increase in the bit rates.

In the Stokes space, each point on the Poincare sphere represents a specific polarization state. The PMD is characterized geometrically by vector PMD $\vec{\Omega}$, this one described at the same time the PSP and the DGD in optical fiber [13]:

$$\vec{\Omega} = \Delta\tau \vec{P_r} \quad |\vec{\Omega}| = \Delta\tau = DGD$$

$\vec{P_r}$ is the Stokes vector associated to the PSP having the highest group velocity .

In the Stokes formalism, Muller matrix R (ω, Z) [4], matrix 3× 3 real describe the evolution of a polarization state to the crossing of a fiber. Polarizations at the input S_{in} and the output S_{out} are bound by the following matrix relation [4]:

$$S_{out} = R(z,\omega)\, S_{in} \qquad (20)$$

By deriving this equation compared to ω we obtain:

$$\frac{d\vec{S}_{out}}{d\omega} = \vec{\Omega}\; x\; \vec{S}_{out} \qquad (21)$$

$$\vec{\Omega}\, x = \frac{R(\omega,z)}{\omega}\; R^{-1}(\omega, z)$$

(x is the usual vector product)

The PMD vector of the first order is characterized by a fixed direction and a constant module. The obtained equation translated the fact that a variation of frequency δω of the input signal led to a uniform rotation of the polarization state of output around PMD vector of an angle ΔΦ = Δτ δω [22].

The PMD vector of the 2nd order is described by the derivative compared to the frequency of $\vec{\Omega}$ as follows [6]:

$$\vec{\Omega} = \Delta\tau_{\omega}\, \vec{P}_{r} + \Delta\tau \vec{P}_{r0} \qquad (22)$$

The index ω indicates derivation compared to ω, the term $\Delta\tau\vec{P}_{r0}$ describes the depolarization of PSPs as for $\Delta\tau_{\omega}$ it is the origin of chromatic dispersion according to the polarization (PCD, Polarization-Dependent chromatic dispersion) which induces a widening or a compression of the impulses.

That can be seen like polarization dependence with respect to chromatic dispersion (DL, Polarization-Dependent changes in Chromatic Dispersion) fiber, described by an effective dispersion [13]:

$$\mathrm{DL_{eff}} = \mathrm{DL} \pm \tau_{\lambda} \qquad (23)$$

$$\tau_{\lambda} = (\pi c/\lambda^2).\Delta\tau_{N} = \frac{1}{2}\frac{d\Delta\tau}{d\lambda}$$

τ_{λ} is the PCD, and the signs plus and minus correspond to alignment with both PSPs [22].

As bit rates increase to 10 Gbps and 40 Gbps, Polarization Mode Dispersion (PMD) becomes one of the leading causes of signal degradation in data transmission. A physical phenomenon in optical fiber that is statistical in nature, PMD causes dispersion, or spreading of pulses in time and distance, causing adjacent signal pulses to overlap and produce bit errors.

The PMD emulator component consists of the PMD channel transfer function considering the first (frequency independent) and second order (frequency dependent) PMD effects.

A linear dispersive fiber can be represented by a 2x2 transfer matrix of the form 24 [15]:

$$T(\omega) = \exp\left(\left(-\alpha - j\beta(\omega)\right).M(\omega)\right. \qquad (24)$$

Where α is the fiber attenuation, β is the mean propagation constant, and M is the unitary matrix that can be written as:

$$M(\omega) = R^{-1}(\omega).D(\omega).R(\omega) \qquad (25)$$

R takes into account the rotation of the principal states of polarization (PSP):

$$R(\omega) = \begin{bmatrix} \cos(k\omega) & \sin(k\omega) \\ -\sin(k\omega) & \cos(k\omega) \end{bmatrix} \qquad (26)$$

Where the coefficient **k** is defined by the depolarization rate

$$2k = \frac{\partial s}{\partial \omega} \qquad (27)$$

The parameter **s** is the direction of one of the two orthogonal eigenvectors.

D takes into account the different propagation speeds on the two PSPs, with the expressions:

$$D(\omega) = \begin{pmatrix} \exp(j\Delta\tau.\,\omega/2) & 0 \\ 0 & \exp(-j\Delta\tau\,\omega/2) \end{pmatrix} \qquad (28)$$

In the second-order approximation the time difference between the two polarizations is given by:

$$\Delta\tau = \Delta\tau_0 + \Lambda\tau'_{\omega} \qquad (29)$$

Where $\Delta\tau_0$ is frequency independent differential group delay, and the differential group delay frequency dependency is represented by the depolarization rate $\Delta\tau'_{\omega}$.

3. Simulation of transmission chains in the presence of the PMD:

The synoptic schema of the connection is represented on Fig 04 below. This chain is consisted of the following elements (from the left to the right):

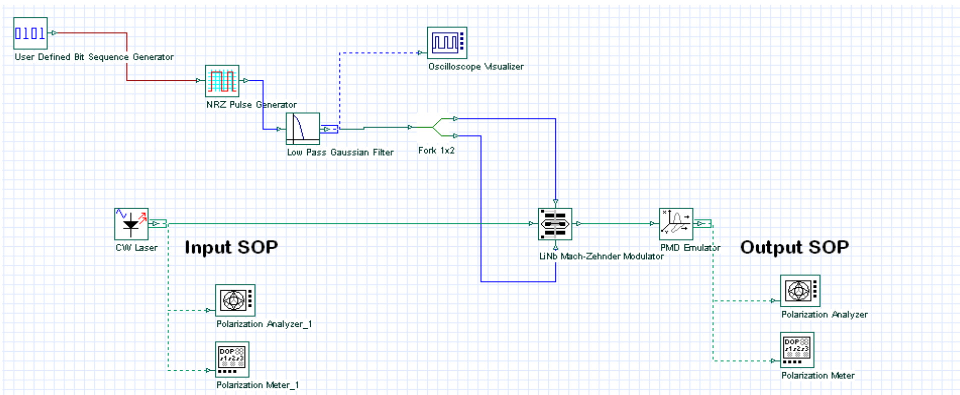


Fig 04. Emulation of the PMD in an optical transmission

This chain is consisted of the following elements (from the left to the right):

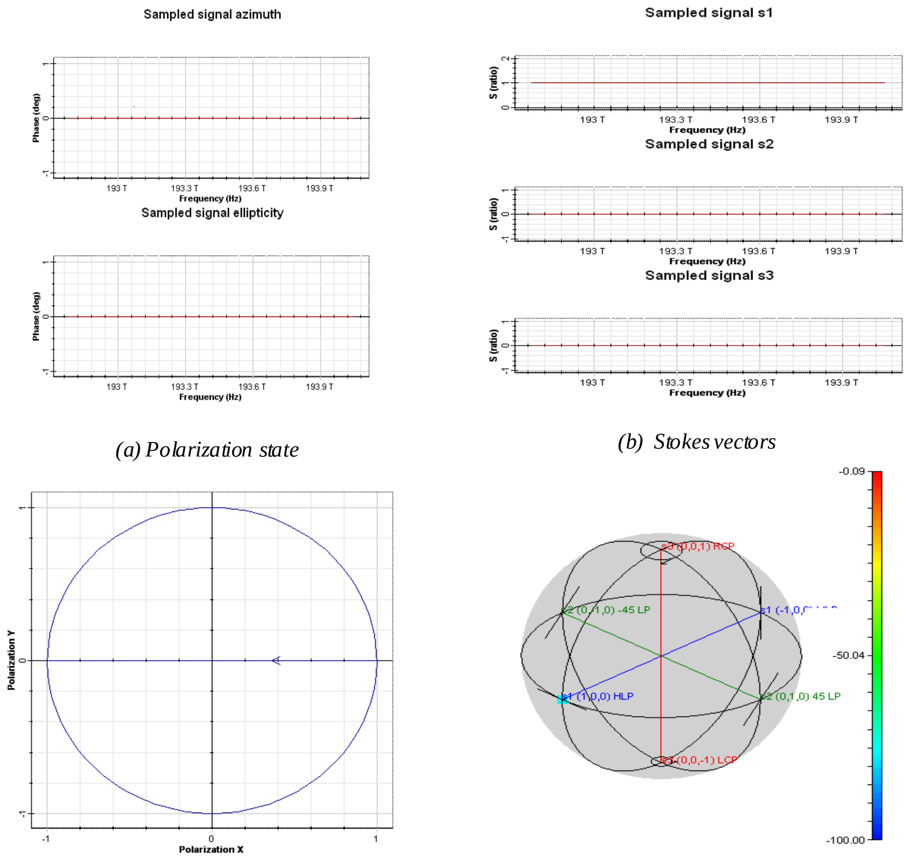
A Generator UDBSG (To Use Definet Bit Generator Sequence) of the bit rate (D) = 10 Gbit/s and binary sequence (11101010010011000011001011101111), a generates a Non Return to Zero (NRZ) coded signal, a filter low passes gaussien , definite according to an Bessel approximation of order 1, and of cuttoff frequency 0,85* bit rate, a generates a continuous wave (CW) optical signal with 1550 nm, a Lithium Niobate Mach-Zehnder modulator based on measured parameters , a fork (1x2) copies the input signal into two output signals. This tool allows to duplicate component output ports and an PMD emulator which simulates the effects of first- and second-order PMD in a linear fiber. The parameters used in our simulations are as follows:

- An optical fiber length L=10 km,
- A Differential Group Delay (DGD) = 71ps,
- A depolarization rate (2k) = 10.8° / GHz,
- A polarization chromatic dispersion (PCD) = 1.3ps/GHz,
- An attenuation and dispersion are taken to zero.

4. Results and discussion:

4.1. *Polarization phenomenon:*

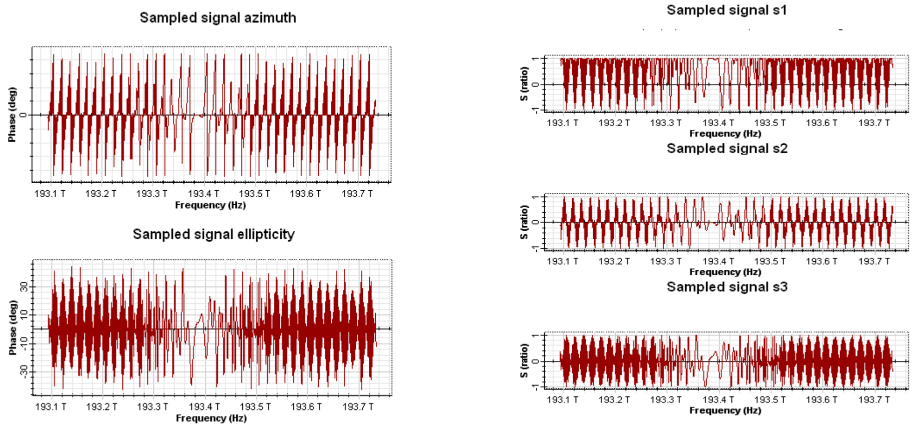
In the first stage, we study the various polarization properties of SOP input / output of the connection. The results are deferred respectively in Fig 05 and Fig 06.



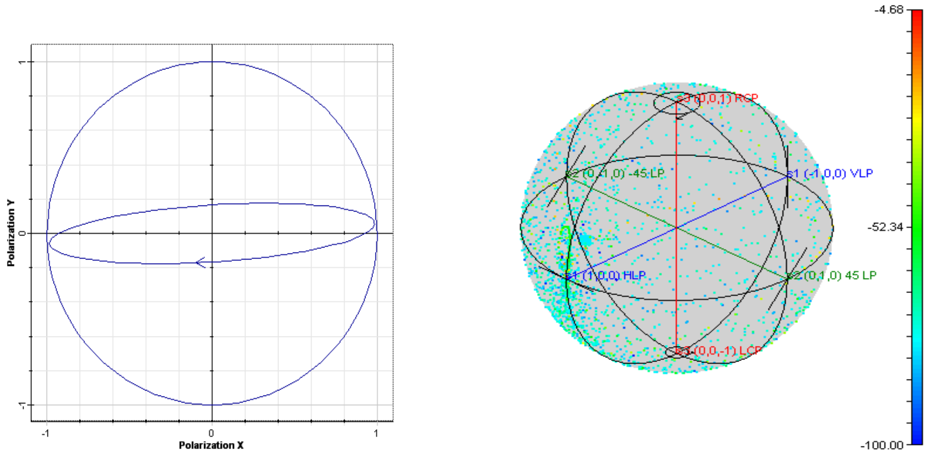
(a) Polarization state *(b) Stokes vectors*

(c) Elliptical display at the reference frequency *(d)Poincare Sphere (completely polarized light)*

Fig 05. Characteristics of polarization of the input Polarization state



(a) Polarisation State *(b) Stokes vectors*



(c) Elliptical display at the reference frequency *(d) Poincaré sphère (partially polarized light)*

Fig 06. Characteristics of polarization of the output Polarization state

The fig (5.a) represents a signal emitted by a linearly polarized laser diode with an azimuth α=0 and one ellipticity $\varepsilon = 0$, this polarization state is aligned according to the fast axis fiber. In this case the birefringence is called linear because the two proper polarization modes are two rectilinear states and it is relatively weak.

The Fig (5.b) represents the Stokes vectors of this polarization state and is defined as follows:

$$\bar{S} = \begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} -75.8134\ dBm \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} S_0 \\ \cos 2\alpha \\ \sin 2\alpha \\ 0 \end{pmatrix}$$

We notice clearly that the total power of the polarized and nonpolarized light at the input S_0 is about -75.8134 dBm , the latter is variable according to the used frequency. This polarization state describes a line in the plan of wave the reference frequency f=193.414 THz (see Fig 5.c). It is located on the equator of the Poincare sphere with a 100% degree of polarization (DOP). That means that the reconcilable laser source is completely polarized. Its state of polarization is represented by only one point on the Poincare sphere (a blue point light) it is shown on the Fig (5.d). The color palette represents the power calculated from S_0 . The marker allows the user to identify the Stokes parameters at the reference frequency.

At the output of the PMD emulator the received signal has an elliptic polarization state (azimuth α = 3.43414 and ellipticity ε = 9.57413) (see Fig 6.a).
The Stokes parameters are represented on the Fig (6.b) with the following values:

$$\bar{S} = \begin{pmatrix} S_0 \\ s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} -77.8637\ dBm \\ 0.93789 \\ 0.11297 \\ 0.32801 \end{pmatrix} = \begin{pmatrix} S_0 \\ \cos 2\varepsilon \cos 2\alpha \\ \cos 2\varepsilon \sin 2\alpha \\ \sin 2\varepsilon \end{pmatrix}$$

It is noticed that the total power of the polarized and nonpolarized light at the input S_0 is about -77.8637 dBm. It is lower compared to the power of input of the laser diode because a part of energy will change on the slow axis of polarization, which will exchange energy with the origin of polarization state.

The dephasing between the components in X and Y of the electric champ varies during the propagation because of the difference in velocity between the two proper modes then the polarization state of the signal varies. These proper modes vary randomly along fiber length L from strong modes coupling which opposes the widening of the signals being propagated in fiber. It is due to really the frequency rotation of the PSPs (depolarization rate), which couples the two orthogonal impulses in entries during the light propagation and also the DGD since the polarization chromatic dispersion (PCD) in simulation is too small to create a substantial deformation of the output signal connection.

The random variation of the polarization states generates a random variation of the polarization characteristics consequence a random change on the level of the stokes parameters, the azimuth and the ellipticity. It is shown on the Fig (6.a) and (6.b).

According to the Fig (6.c) the polarization state describes a left elliptic polarization in the plan of wave (the most general case). It is well located on the northern hemisphere of the Poincare sphere and its DOP is approximately 46.64 % i.e. (DOP=0.4664). We can say that the light is partially polarized, which is represented on the Fig (6.d). In this case the polarization state can break up into the incoherent sum of a completely polarized state (principal polarization state) and of a completely depolarized state. The Stokes vector can be written as follows:

$$\begin{pmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix} = (1 - DOP) \times S_0 \times \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + DOP \times S_0 \begin{pmatrix} 1 \\ S_1 \\ S_2 \\ S_3 \end{pmatrix}$$

For low values of the depolarization rate, we obtain on the outlet side of the PMD emulator the Poincare sphere which is represented on Fig 07

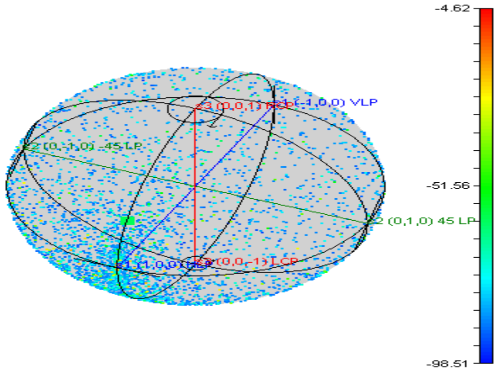


Fig 07. The Poincare sphere for a low depolarization rate (approximately 0.3 deg/GHz)

The degree of polarization (DOP) is approximately 0.2%,(i.e. DOP=0.0020≈ 0), then the light is almost nonpolarized. The DOP is null for a completely depolarized light. In this case, we can consider that the light contains all the polarization states and consequently, all the points of the Poincare sphere are equiprobable. In this simulation the DOP should not be never equal to zero, it is at least equal to 0.002.

4.2. *Second order PMD Effect:*

In the second stage demonstrate the second-order PMD effect on the transmitted signal, it is necessary to add a Photodetector PIN to convert the optic signal into electric signal of bandwidth 50 GHz, Responsivity = 0.55 A/W and dark current = 5nA.

The signal can finally be analyzed after being filtered through a low pass bessel filter order 5 and Cutoff frequency = 0.85*bit rate. The part of reception is represented in Fig 08.

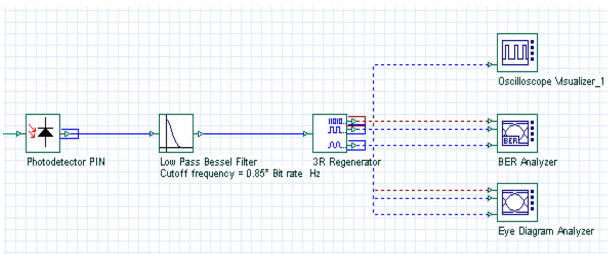


Fig 08. Reception part

The Simulation is performed for the same preceding parameters. The results are reported in Fig 09-12.
The input signal is represented by Fig 09

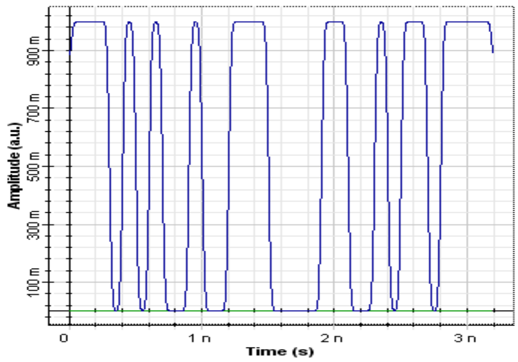


Fig 09. Sequence of the input impulses

The Fig 9, represents the input signal, which is a sequence of 2^5 bits given by the user, this last is a sequence of the type NRZ impulses transmitted in a low pass gaussian filter.

The pace of output impulses sequence is represented in Fig 10.

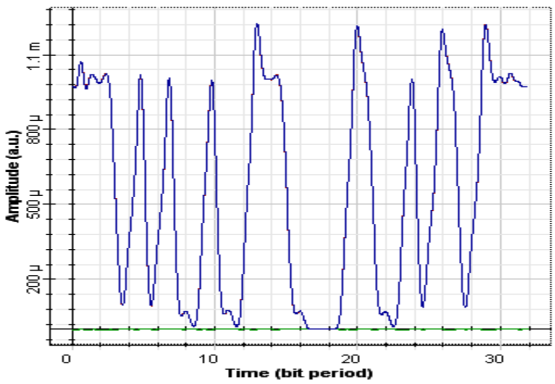


Fig 10. Sequence of the impulses on the output Bessel filter

The Fig 10 represents the effect of the second-order PMD on the impulses NRZ sequence. The presence of power over-shoots on the sequences of "1"s, and the presence of energy on the "0"s are due to the imperfect cancellation of the pulses coupled on the orthogonal axis.

The Fig 11 represents the eye diagram with the variation of the quality factor.

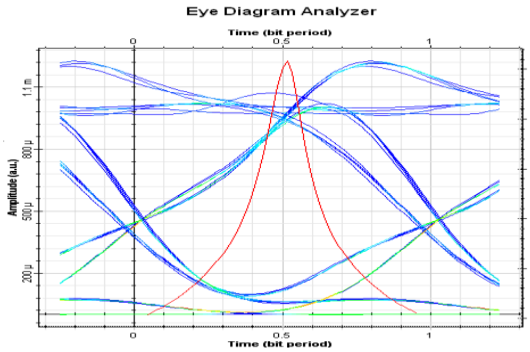


Fig 11. Quality Factor according to time (red) and the eye diagram

According to Fig 11, we notice that the quality factor is to the maximum equal to 11.28. That means that the input signal is received without deformation.

We can represent the effect of the second-order PMD on the Poincare sphere and which appears in Fig 12

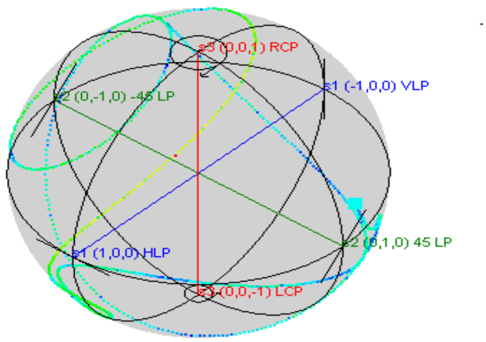


Fig 12. PMD of order 2 on the Poincare sphere

According to Fig 12, we notice that the direction of the rotation axis as well as the swing angle moves with the frequency. Thus the output polarization state carries out a series of rotations on the Poincare sphere.

The PMD causes distortions undergone by the visible signals during the propagation in the systems with high bit rate and which lead to a potential reduction of the performances of the connections. It influences on the transmission quality.

4.3. *Influence of the PMD on the transmission quality:*

In a third stage we study the influence of the PMD on the transmission quality according to the bit rate, the DGD and the input SOP.

4.3.1. Influence of the bit rate and DGD on the PMD:

The Simulation is made for the same preceding parameters, while varying at the same time the DGD and the bit rate. The results are referred in Fig 13 and Fig 14.

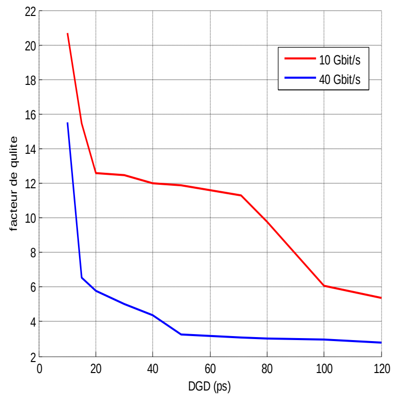


Fig 13. Variation of the quality factor according to the DGD

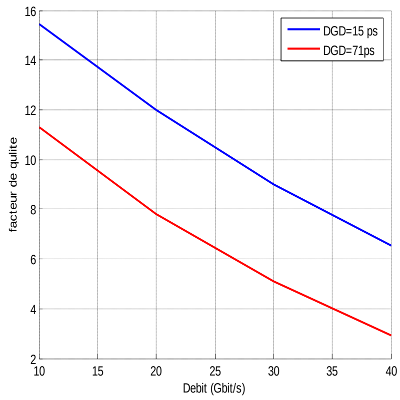


Fig 14. Variation of the quality factor according to the bit rate

The Fig 13 represents the effect of the DGD on the quality factor, we notice that the differential group delay (DGD) will introduce intersymbol interferences (ISI:InterSymbol Interference) between the transmitted impulses which very quickly degrades the quality factor . It is well noticed for a bit rate of 10 GHz/s, the DGD does not exceed the 100 ps while for a bit rate of 40 Gbi/s this last does not exceed the 15 ps to ensure the quality of the output signal. That means that the DGD influences the PMD. The increase in the DGD also generates an increase connection length (see equation 19).

The Fig 14 represents the effect of the bit rate on the quality factor, we notice that for a bit rate of 10 Gbit/s we obtain a quality factor (Q) equals to 15.44 for a DGD of 15ps and 11.28 for a DGD of 71ps. This last decrease to 40 Gbit/s we obtain a quality factor (Q) equals to 6.52 for a DGD of 15ps and 2.91 for a DGD of 71ps. When the bit rate increases, the quality of the signal decreases very quickly. That means that the bit rate limits the PMD.

4.3.2. Influence of the input SOP on the PMD:

The Simulation is carried out with the same parameters while varying the polarization state (the azimuth α of -90° to + 90° and ellipticity ε of – 45° to +45°).

The results of this simulation are reported in Fig 15 and Fig 16.

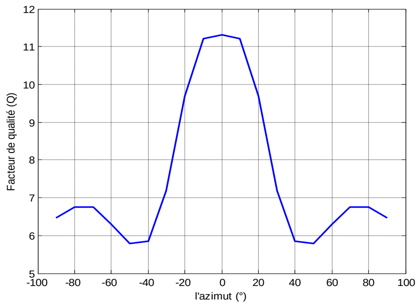


Fig 15.Variation of the quality factor according to the azimuth α

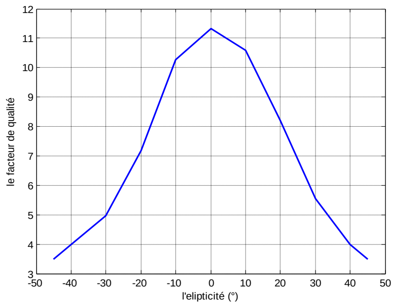


Fig 16. Variation of quality factor according to the ellipticity ε

The Fig 15 represents the effect of the azimuth α on the quality factor, we notice well that for an ellipticity $\varepsilon = 0°$ almost all the values of the azimuth α give a good quality factor (Q > = 6) except for $-40° \leq \alpha \leq -50°$ and $+40° \leq \alpha \leq +50°$ which degrades the transmission quality.

The Fig 16 represents the effect of the ellipticity ε on the quality factor, we sees that to maintain a good transmission quality, it is necessary that the ellipticity ε belongs to the interval [-25°, +25°]. The other values of the ellipticity give a bad quality factor (Q<6).

For a bit rate equal to 10 GHz and a DGD of 71 ps , the input SOP must be linear horizontal or vertical, forming an angle teta with the azimuth or elliptic, never to be circular.

The input SOP influences the signal quality and to obtain a good quality factor, it is preferable to choose a linear polarization along the fast axis (azimuth α=0° and ellipticity ε=0°).

5. Conclusion:

The PMD is a dynamic phenomenon affecting the optical signal during its propagation in fiber. For that, we have presented the formalisms used to study the polarization properties of the optical components.

The parameter making it possible to characterize these polarization properties are the Jones and stokes parameters of the geometrical representation on the Poincare sphere. These sizes permit to apprehend in an easier way, the effects of polarization in the long distance mode.

The PSPs outputs are defined as invariant output polarizations according to the optical frequency owing to the fact that the two modes of polarization can exchange energy between them; this phenomenon is related to the random mode coupling.

To maintain a good transmission quality of an optical signal, whose maximum bit rate is 40 Gbit/s, the DGD should not exceed the 15ps with a input SOP which should never be circular.

References:

[1] S. HUARD, ‘’Polarisation de la lumière’’ Masson, 1994

[2] R. C. Jones, “New calculus for the treatment of optical systems. Description and discussion of the calculus,” Journal of the Optical Society of America, vol. 31, no. 7, pp. 488–493, July 1941.

[3] J. P. Gordon and H. Kogelnik, PMD fundamentals: ‘’Polarization mode dispersion in optical fibers’’, Review PNAS. 97 (2000) 454-4550.

[4] D. A. Nolan, Xin Chen, and Ming-Jun Li, ‘’Fibers with low polarization mode dispersion’’, J.Lightwave. Technol., 22 (2004) 1066-1077.

[5] B. Huttner, C. Geiser and N. Gisin, ‘’Polarization-induced distortions in optical fibers networks with polarization-mode dispersion and polarization-dependent losses’’, Invited paper, IEEE, journal of selected topics in quantum electronics, 6(2000) 317-329.

[6] "Polarization Measurements of Signals and Components", Product Note 8509-1, Agilent Technologies.

[7] M. Fontaine, ” Theoretical and experimental analysis of thermal stress effects on Modal Polarization properties of highly birefringent optical fibers ”, Lightwave Technology, Journal of, Vol.14, N°4, pp 585-591, 1996.

[8] H. Sunnerud, “Polarization-Mode Dispersion in optical fibers: Characterization, Trnasmission Impairements and Compensation”, PHD Thesis of the Chalmers University of Technology, 2001.

[8] A. P. Chimata and C. ‘’Allen, Development of an adaptive polarization mode dispersion compensation systems’’, technical report, university of Kansas, January 2003

[9] L. E. Nelson Leong, ‘’Introduction to polarization mode dispersion in optical systems’’, J. Opt. Fiber. Commun. 1(2004) 312-344.

[10] M. Shtaif and A Mecozzi, ‘’Modeling of polarization mode dispersion in optical communications systems’’, J. Opt. Fiber. Commun. 1(2004) 248-265

[11] Heidemann R., Wedding B. et Veith G., ″10-GB/s transmission and beyond″, Proceedings of the IEEE , November 1993, Vol. 81, n°11.

[12] L. E. Nelson Leong, ‘’Introduction to polarization mode dispersion in optical systems’’, J. Opt. Fiber. Commun. 1(2004) 312-344.

[13] M. Shtaif and A Mecozzi, ‘’Modeling of polarization mode dispersion in optical communications systems’’, J. Opt. Fiber. Commun. 1(2004) 248-265.

[14] Bruyere Frank, ″Impact of First -and Second- Order PMD in Optical Digital Transmission Systems″, Optical Fiber Technology, 1996, Vol.2, pp. 269-280.

[15] Foschini G.F. end Poole C.D., ‘'Statistical Theory of Polarization Dispersion in Single Mode Fibers″, IEEE Journal of Lightwave Technology, November 1991, Vol. 9, n°11, pp. 1439-1456.

[16] Gisin N., Passy R., Bishoff J.C. and Perny B., ″Experimental Investigations of the Statistical Properties of Polarization Mode Dispersion in Single Mode Fibers″, IEEE Photonics Technology Letters, July 1993, Vol. 5, n°7, pp. 819-821.

[17] ″Combat polarization impairments with dynamic polarization controllers″, Site of General Photonics Corporation.
http://generalphotonics.com/techinfocombat.html

[18] ‘’Polarization Mode Dispersion’’, Site of Tektronix.
http://www.tektronix.com/Measurements/commtest/pmd/index.html

[19] ‘’Polarization Mode Dispersion Tutorial’’, Site Web ProForum Tutorials.
http://www.onforum.com/tutorials/polar_mode/index.html

[20] university of michigan, ‘’Polarization Mode Dispersion, Electrical Engineering and Computer Science’’.
http://www.eecs.umich.edu/satin/pmd/

[21] Chromatic Dispersion and Polarization Mode Dispersion (PMD), Light Reading – The Global Site for Optical Networking.
http://www.lightreading.com/document.asp.doc_id=7303.

[22] H. Sunnerud, “Polarization-Mode Dispersion in optical fibers: Characterization, Trnasmission Impairements and Compensation”, PHD Thesis of the Chalmers University of Technology, 2001.

[23] Cristian Francia, Frank Bruyere, Denis Penninckx, and Michel Chbat. " PMD Second-Order Effects on Pulse Propagation in Single-Model Optical Fibers". IEEE Photonics Technology Letters, December 1998.

[24] L. E. Nelson, R. M. Jopson, H. Kogelnik, and G. J. Foschini. "Measurement of Depolarization and Scaling Associated with Second Order Polarization Mode Dispersion in Optical Fibers". IEEE Photonics Technology Letters, December 1999.

International Conference on Software Engineering and New Technologies

SVM classifier for Face Recognition

GHARBI ABDELHAKIM

Tebessa University –Algeria-

Gharbi_hhh@yahoo.fr

MED RIDDA LAOUAR

Tebessa University -Algeria-

Ridda_laouar@yahoo.fr

The feature extraction plays a very important for face recognition. In this paper, a model is proposed for the automatic recognition of faces, by combining two approaches to intelligence Artificial. The images used are represented as pixel intensity, with a front view of the face. Each face will be described by a feature vector representing the three main components of the face are: eyes, mouth and the nose. [Brunelly and Poggio(1993)]. [Chellappa et al (1995)]

Be used for the extraction of characteristic points of the treatments on the histogram of the projections of the image, the distances between these points will construct the set of feature vectors used as input for training and testing SVM multi_classe in end SVM establish its identification and final decision.

Keywords. face recognition, extraction Features ; projection; learning, classifier, localization, support vector machine (SVM).

1. Introduction

In everyday life a person must be identified in multiple situations: a building entrance, access to work, withdrawing cash, payment by credit card, connection to his computer and his e-mail, resulting in the ... memorization required many codes. That is why today many research aims to prove the identity of a person using biological or behavioral characteristics of its own: iris, fingerprint, hand geometry, typing on a keyboard, velocity signature,... biometrics is the science that brings together all of these works and devices. In the spectrum of biometrics, face recognition is special to share the greatnumber of works devoted to it [Hori et al (2006)]. Indeed capture an image of a face, particularly through a 2D camera, is simple and noninvasive. This is a biometric modality easily tolerated by users, and especially since computers are equipped with more and more often "webcam". But the performance of face recognition are still well below what one might hope for such applications. The increase in recognition rate, the decrease in recognition wrongly and faster response time when sought in large biometric databases are the challenges facing the algorithms being developed.

To recognize a person from his image, it must go through certain stages. First you must locate the face in the image. Then it may be necessary to normalize the face back to a standard size. Then we must address the recognition phase itself. In this paper we propose a method which based on the use of support vectors for recognition.

2. Features extraction

2.1. Architecture of the system

The general system architecture is designed by the Fig.1. Because the software is composed of two parts Part learning and part selection (of use).

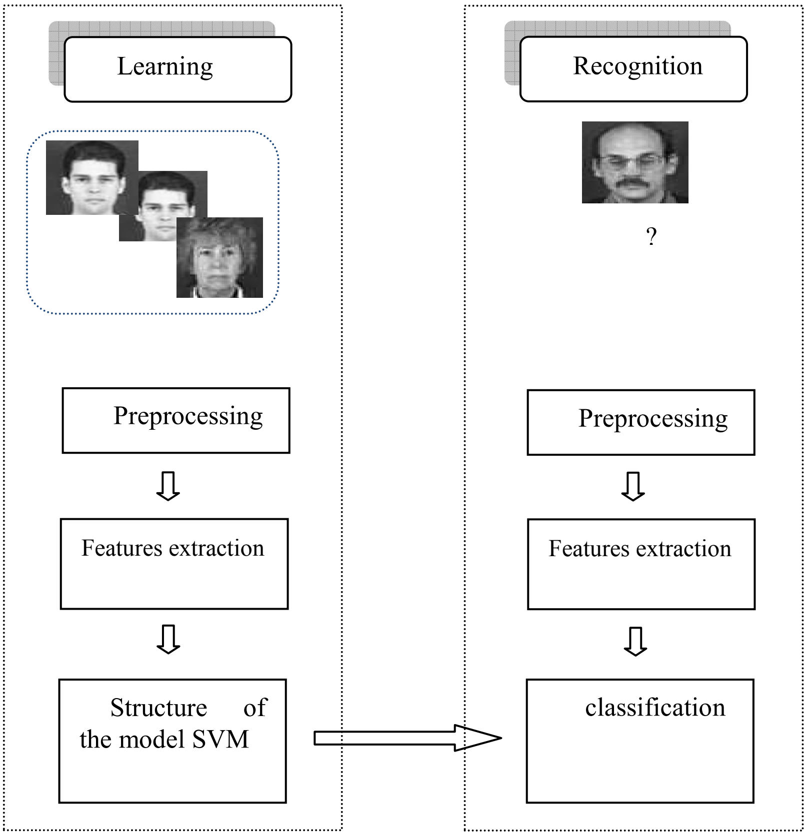


Fig. 1. General architecture of a model of face recognition

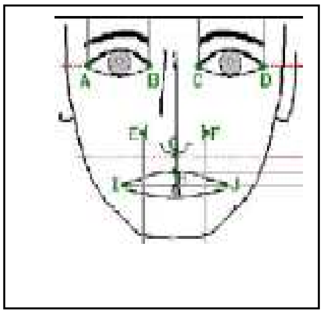
2.2. The pre-processing

We chose the base that contains ORL face images 40 people, with 10 images each. For most subjects, 10 images were taken at different times and with different lighting conditions, but always on a dark background. The image size is 92 x 112.
The purpose of the pretreatment is to improve the image quality by reducing the noise. Because in preprocessing prepares the image that will be subject to future treatments. Face extract is first converted to grayscale, then undergoes a series of preprocessing. Pretreatment consists of modifying the histogram stretching) to enhance contrast, adjust

brightness and reduce the noise coming from the original photographic image applying filters also called smoothing (or sometimes die Foley noise filter). Were chosen based ENT after detection and localization of the face and readjust the image size (48 * 48).

2.3. Proposed geometric model

The analysis of the human face is given by the description of its individual parts and their relationships. This model corresponds to the way in which humans perceive the face, ie, our notions of parts features such as eyes, nose, mouth, etc.. Most of the work focused on the extraction of features from an image of the face and on the definition of an adequate model to represent that face our geometric model used is the one used in [4]



Model [KAMEL et al (1993)] : 12 points and six distances extracts chosen for the vector of Features **V [AD BC EF SG SH SM]**

Fig. 2. The geometric model

2.4. Location of areas of the eyes and nose and mouth

A study of basic and standard ENT given following the extraction by hand rectangles encompassing the eyes, nose and mouth were defined on their spaces for every 400 images from the ORL database.

Fig. 3. Image after locating spaces eyes, nose and mouth

2.5. The integral projection for vertical and horizontal The extraction of the features points

Is the face image I (x, y), the integral projection vertically in the rectangle [x1, x2] * [y1, y2] is defined by:

$$VI(y) = \sum_{x=x1}^{x2} I(x, y) \quad (1)$$

The integral projection in the horizontal rectangle [x1, x2] * [y1, y2] is defined by:

$$HI(x) = \sum_{y=y1}^{y2} I(x, y) \quad (2)$$

These projections generally assist in locating trais, only it is preferable to limit the working window to avoid errors, the two images are then generated horizontal contours (HI) and vertical (VI):

- From (HI), we draw the boundaries of the face.
- From (VI), we obtain the crown of the head, eyes, lower nose and mouth.

2.5.1. Extraction points of eyes

Eye detection is very important in the extraction of characteristic face. To extract the eye corners A, B, C, D we have used the horizontal and vertical projection in the space located above the eyes. The minimum projection Portraits in the rectangle [8.20] * [2.21] gives the line or is the two corners A, B.
With the horizontal projections are the columns of the corners. The ordinate and the points are quads extract

2.5.2. Extraction points of the nose and mouth.

First, it is estimated their position by looking for peaks in the horizontal projection (HI) (nose), and valleys in the horizontal projection (mouth), this evaluation is done according to the prominence of the peak or valley, those whose valuation is the largest are used as the vertical position of the nose and mouth.
The nose is horizontally by defining the search for peaks in the vertical projection (VI), the mouth is located by thresholding the projection of the mean value

3. SVM and recognition of face

In Part learning system takes a set of examples with their classes known in advance and built a decision model using the SVM method. In this phase, we begin by image processing and extraction of their characteristics and their registration with the corresponding classes in a database characteristics

3.1. SVM (Learning)

The SVM method works based on the features and tries to find a decision model. SVM requires that the basis is normalized between -1 and +1, for this it is necessary to convert the values of all vectors in the interval [-1, 1]. We then calculate the maximum and minimum of each attribute of the table and then we apply the following function:

$$f : \mathbb{R} \rightarrow [-1, +1]. \quad (3)$$

SVM module then generates a decision model for each class from the basic characteristics of standard

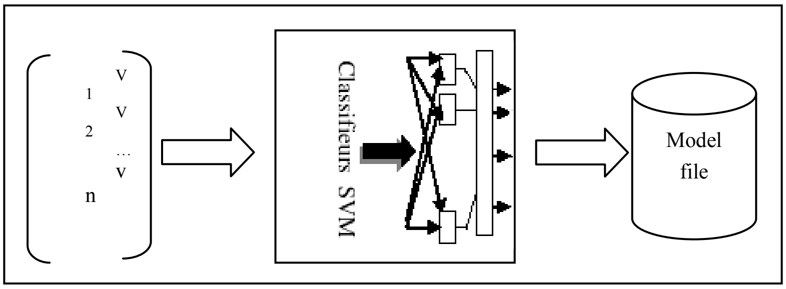


Fig. 4. SVM learning

For each class we take the normalized table and we put the values of this class has one and all other then a -1 is called the module to generate the SVM decision parameters for this class. The parameters generated by SVM for each class are:

- the αi different from 0
- Two features vectors standardized

αi corresponding to different from 0, that is to say the support vector. The overall decision model obtained contains in addition: class parameters:

- the maxima and minima of the nine characteristics.
- the kernel used.
- kernel parameters.
- The labels of classes.

The resulting model can be saved to a file for use in the selection phase

3.2. Classification with SVM

Classification is the actual use phase of the system, it uses a model obtained when learning to decide about the new images.
It takes as input an image face features and extracted in the same way that when learning is then applied to each model class decision function that gives a real value: it is positive if the face belongs to the class, if you face the negative belongs to another class.

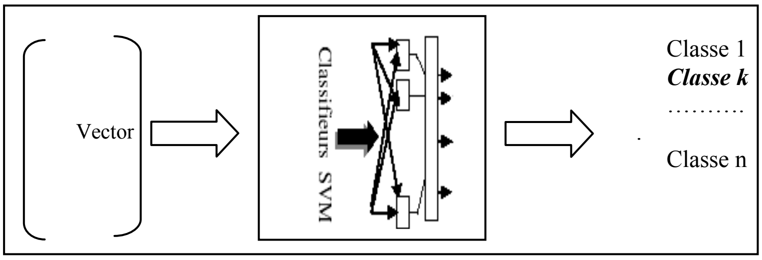


Fig. 5. Classification by SVM

4. Implementation and results

We are currently implementing the verification system in face of the MatLab programming environment. This system is divided into two modules.
Module 1 performs location, extracting feature vectors from the input image faces, calculates
components of the feature vector of each facial image of face. The extraction of feature points based on the integral of the vertical projection and horizontal.
Module 2 is a call for multi-class SVM which consists of a learning module svm_multiclass_learn) and a classification module (svm_multiclass_classify). After the training phase, SVM provides the answer to the recognition of face images. This system is being tested.

References

Brunelly R, Poggio T, " Face Recognition: Feature Versus emplates", IEEE Transaction on Pattern Analysis and Machine Intelligence, 15(10), p.1042-1052, 1993

Chellappa R, Wilson c, Sirohey S, "Human and Machine Recognition of faces: A Survey", PIEEE, 83, pp.705-740, 1995.

Hori Y , M. Kusaka, and T. Kuroda. "A Real-Time Face Detection Core". Symposium on VLSI Circuits Digest of Technical Papers, pp. 188–189, June 2006

KAMEL M. S., SHEN H. C., WONG A. K. C. and CAMPEANU R. I.: "System for the recognition of human faces", IBM Systems Journal, Vol. 32, No 2, p. 307-320, 1993.

International Conference on Software Engineering and New Technologies

DESIGNING EMBEDDED WEB SYSTEMS FOR HEALTHCARE

SIDI MOHAMED O. MOULAYE ABDELLAHI

Department of computer science and mathematics, Laboratory of computer science
Cheikh Anta Diop University of Dakar, Senegal
sidimed_makhtour@yahoo.fr

MOHAMED TAHAR KIMOUR

Department of computer science, Laboratory of Research on Embedded Systems (LASE)
Badji Mokhtar-Annaba University, Algeria
kimour@yahoo.com

MBAYE SENE

Department of computer science and mathematics, Laboratory of computer science
Cheikh Anta Diop University of Dakar, Senegal
ngalagne@yahoo.com

During recent decades embedded systems have become more widely used in modern society of today. These systems incorporate processors, which have become more powerful and more cost efficient. Now, embedded systems are of interest in areas where not used before, allowing for the development of ever more distributed and complex control systems. To handle the increasing complexity of such systems, industry is constantly looking for new software development tools and strategies. In this paper, we present an approach to design such systems while using an instance of a healthcare control application. Such application would support both patients and their healthcare providers in the process of treatment. It deals with the management of the conditions of patients at home while making available information on such conditions at the remote clinical center. The designed system is built around networked embedded devices and integrates the web technology.

Keywords: Embedded systems, Healthcare, Product lines.

1. Introduction

During recent decades, Information and Communication Technology (ICT) have known important advances, offering significant processing power and communicability in almost every device that we encounter in our daily life. Today's embedded systems are mobile, pervasive, and may be permanently connected [Patrick et al. (2008)].

On the other hand, advances in web technology allow ensuring remotely and constantly access to patient's information and enables new directions in development of e-health [Al-hakim (2007)][Kirovski et al. (2007)][Ko et al. (2008)]. In the process of monitoring and treatment of patients, it seems to be self-evident to design a system that would support both patients and their healthcare providers. However, little work in integrating embedded devices with internet for the support of patients have been done to date [Chenhui et al. (2008)][Lee et al. (2008)][Spasov and Petrova (2008)].

In this paper, we present an approach to design a Home Healthcare System (HHS) for monitoring and control of patient's conditions at home. This system incorporates a local network at the patient residence that wirelessly connects various sensors that deliver continuously some important measurements on the body of the patient to a web portal database server, situated at the monitoring center, which is accessible by clinicians such as doctors, nurses, etc.

With an HHS, individual citizen will have the provisions and handles to be self-responsible for his own health and to manage the health delivery process [Lee et al. (2008)][Eslami et al. (2009)]. On the other side, clinicians will be provided with more accurate data from continuous monitoring of patients in their natural environment and will be able to give more accurate prescribes.

A home healthcare system (HHS) should adapt to changes regarding the patient needs and to variations in the residential environment.

In the development of our HHS, we have adopted the product line strategy and an architecture oriented process; inspired from the most notable web developments processes such as webML [Ceri et al. (2000)] and UWE [Koch et al. (2002)], but also from methods dedicated to embedded distributed systems.

A product line (PL), also called system family, is a set of software systems sharing a common, managed set of features that satisfy the specific needs of a particular market segment or mission and that are developed from a common set of core assets in a prescribed way [Clements and Northrop (2002)][Gomaa (2004)][Kang et al. (2002)][Krueger (2006)]. A product line holds commonalities for all the members of the family, and offers variabilities for each single member of the family.

The rest of the article is structured as follows: In Section 2 we present an overview of recent development in healthcare systems and possible applications and different challenges. Our proposal for the architectural design of the system is discussed in Section 3. Section 4 describes related works and compares them with our proposal. Finally, the paper is rounded off in Section 5 by a conclusion and directions for future research.

2. Embedded Healthcare Systems

Embedded systems are not always standalone devices. Many embedded systems consist of small, computerized parts within a larger device that serves a more general purpose. Embedded systems are designed to do some specific task, rather than be a general-purpose computer for multiple tasks. Some also have real-time performance constraints that must be met, for reasons such as safety and usability.

Embedded systems contain processing cores that are typically either microcontrollers or digital signal processors (DSP). They have evolved a lot in aspects such as, mobility, wireless communication, integration with internet, increasingly computing performance, widely enhanced security, etc. They offer new applications in various domains. One of the most important of them is the healthcare system.

For instance, as an embedded system, the mobile phone is more and more used as a valuable tool for patient monitoring. In fact, the mobile telephony has known many advances especially in useful resources, such as serial ports and Internet connections.

Therefore, the phones can interact with electromedic devices– like patient biosensors, for example – and transmit vital signals through Internet protocols, such as TCP/IP and UDP. With this technology, patients may be oriented and assisted to perform a wide variety of medical tasks that have only been performed by trained medical personnel in the past. Such tasks can range from a diabetic patient who checks his/her blood glucose level 2-4 times a day and adjusts the insulin dose appropriately, to the sensing of crucial medical parameters such as blood pressure, glucose level for diabetes and sending them over the cell phone to the local physician’s office or to a distant monitoring center.

A care plan guides the patient regarding measurements that must be done (e.g., blood pressure), medications that should be taken, and other personalized recommendations according to the treatment [Loques and Sztajnberg (2010)][Sztajnberg et al. (2009)].

3. Home healthcare system design

Home healthcare systems are increasingly complex and dynamic. The design activity provides the architectural description of the system compliant with the used component model. In the design of such systems, the current best practices dictate the use of hierarchical distributed architectural models. In doing so, we employ two distinct levels of structural decomposition: the subsystem level and the component level.

At the first level, the system is designed as a collection of concurrent subsystems, possibly communicating with each others. Subsystems are the system parts design or implementation units that can be developed independently, with complex functionality, stored in a repository and reused in multiple applications [Sentilles et al. (2008)]. The subsystem level comprises decomposition according to logical cohesion among offered functionality. The second level concerns the component model, used for modeling small parts of control functionality.

Using two granularity level for designing the system, our design flow is composed of four steps: overview of the system, context model, subsystem model, component model. Subsystems and component models are provided with their static and behavioral views.

At the first step, we build an overview of the system, where we explicit in natural language the various functionalities of the system and its interactions oh the users. Secondly, we model the context of the system, where we depict the system as black box interacting with external entities of its environment. At this step, end-to-end requirements are depicted especially those related to timing requirements. Thirdly, the subsystem model shows subsystems and their interactions. Fourthly, for each subsystem we define its components (either hardware or software). To promote reuse, we design the HHS by adopting a combination strategy between component-orientation and product line concept.

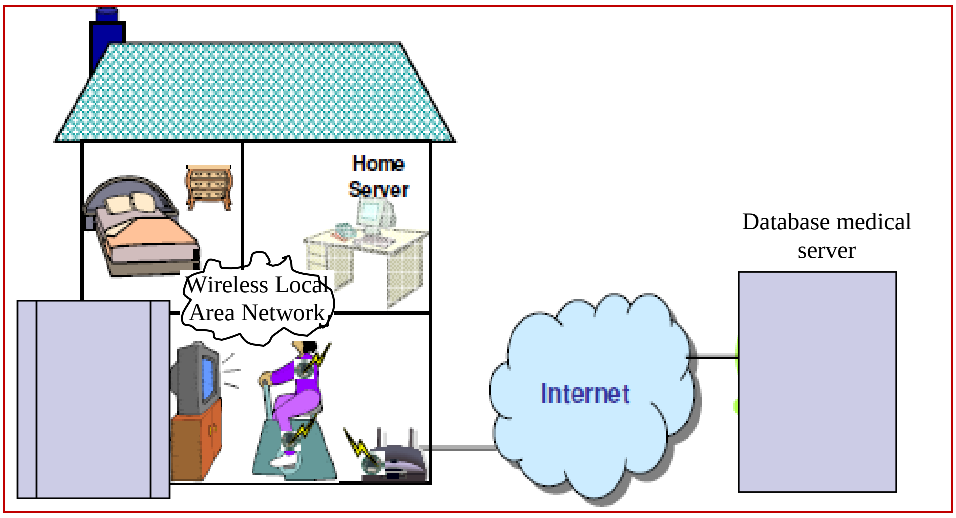


Fig. 1. A typical architecture for a home health monitoring system.

3.1. *Overview of the System*

The proposed Home Healthcare System (HHS) should constantly monitors patient' states and inform him to perform needed tasks such as blood pressure and glucose level measurements. It should communicate such measurements to a medical tracking center and eventually remotely notify the doctor about critical situation. It should help patients at home (especially the elderly ones) when they have to manage specific medical tasks. It has to keep track of their regime of medications and doctor appointments and has to check parameters such as blood sugar or blood pressure. Medical staff (doctors, nurses, and physicians) has the opportunity to remotely monitor the patients' daily activities and adjust his/her care plan based on the treatment progress. For the patient, this can mean a smaller number of visits to the doctor and shorter periods of hospitalization.

Thus, the HHS should be composed of two main parts, connected by internet: one at the patient's home side, for local patient monitoring (LPM), and the other at the healthcare center side, called Healthcare Center (HC). At the LPM, using appropriate mobile devices (sensors), patient's measurements such as glucose level, weight, blood pressure are constantly registered to the home gateway. A PC, a PDA, or even a smart phone may be such gateway that automatically uploads (via internet infrastructure) patient's measurements to a server (healthcare center) for access by healthcare providers (clinicians). To this end, a home wireless local area network (WLAN) is needed in a home healthcare system (Fig. 1).

The communication is usually based on one of the wireless standard from IEEE 802.15 wireless network family [Park and Jayaraman (2003)][Kolitsi and Umpierrez (2007)][Istepanian (2004)]. The warning messages to the patient should be shown in one or more display devices such as TV, cellular phone, tablet, PC, display of the PHS, etc.

The gateway is equipped with a radio module that receives data from the patient's networked sensors, runs the monitoring algorithms, activates the alarms if the medical parameters are above the limits and sends the results to the database server (at the HC). It is connected to the HC through an Internet, GSM or dial up connection.

The gateway runs the monitoring algorithms and sends the results to the HC system through the internet connection. In the case when the patient is in a bad condition and he needs an emergency medical assistance, the gateway can send to the HC, and from here to ambulance service, the precise coordinates of the patient.

The HC system includes a healthcare facility to remotely monitor a large number of patients, and interfaces for use by doctors, caregivers, and relatives of the patients. Collected at the LPM and uploaded to the HC system, the relevant patient's measurements interpreted in real-time. In the case of the identification of an abnormal patient condition, the HC system may activate a LPM's local device (a TV or a smart phone, for example), increase the frequency of monitoring, or, depending on the condition severity, send an alarm.

The HC system, which should be an internet portal, tracks readings and eventually enables the patient to adjust insulin usage in consultation with their doctor, analyze over time behavior, and reminders for tests and treatments, and alerts in case of potential hazardous conditions.

For increased interoperability and common interface to medical services the front-end to the HCS could be realized as Medical Web Portal. This could be implemented using proven best-practices for design of Web applications [Ceri (2000)][Koch and Kraus (2002)]. The function of this system includes control of data exchange with the home side, management and support of database with electronic medical records, interface to various medical and non-medical services, global reasoning over data from BSN and feedback to the patient.

3.2. *Context model*

It is very important to understand the scope of the overall system. We model the system context using a context diagram, which is a UML class diagram with stereotypes. At this step, we explicitly show external interactions between the system (hardware and software), modeled as a black box, and the environment. At this step, we identify the interacting actors that are outside the system.

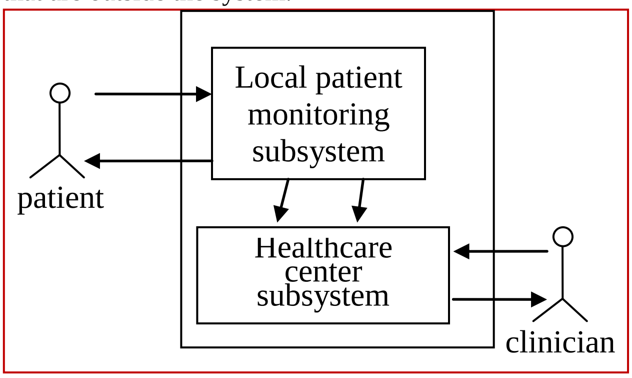


Fig. 2. The context diagram of the HHS.

A system context diagram (Fig. 2) explicitly presents the boundary between the system and the external environment. To differentiate between different kinds of external entities, stereotypes are used. Using the UML class diagram, the system context is depicted showing the system as an aggregate class with the stereotype «system», and the external environment is depicted as external classes to which the system has to interface [Gomaa (2000)].

In order to facilitate system decomposition, our context diagram will draw not only the external actors and the overall system, but also the networked parts of the system. For instance, Fig. 2 illustrates two networked parts of our HHS, namely: the local patient monitoring subsystem and the central healthcare subsystem.

3.3. *Subsystem model*

The subsystem view specifies a component model used for modeling independent distributed components with complex functionality [Koch and Kraus (2002)][Kimour et al. (2012)]. It is based on two important concepts: separation of concerns and distribution. Separation of concerns promotes system comprehension and reuse; hence reduce cost and development delay.

The care plan subsystem is a key element of the HHS. The objective is to guide the patient based on local measurements that should be done at the LPM system (e.g., blood pressure), medications that should be taken, and other personalized recommendations according to the treatment [Loques et al. (2010)][Sergio et al (2008)]. It can notify the patient to remind him of some event or to send a message to the doctor. This feature is interesting to increase the patient's treatment adherence, warning him of daily tasks such as taking a medication, or remembering that it is time to measure the blood pressure. Components are represented by rectangles and connectors are represented by lines. Dashed elements correspond to optional (Fig.3).

One decomposes a problem into sub-problems until one arrives to the basic problem for which basic solution elements can be chosen [Kimour et al. (2012)]. To be able to build software systems out of reusable components, we need a component model [Sentilles (2008)].

A component model is mainly based on the concepts of component, port, interface, and connector. For each subsystem we identify components with three types [Gomaa H. (2000)] data components, control components, and interface components.

A data component contains persistent information recording patient status of situation. A control component models cohesive application logics (algorithmic). An interface component is defined for each input or output device such as sensor, actuator, display device, etc. Moreover, to enable adaptability at run-time, we allow customization through the use of product line strategy, in terms of both the patient and the residence (layout of the residence, distribution and type of the sensors and devices, etc.). Changes in the patient's health situation during the system operation may lead to modifications in resource configuration, such as medical sensors, or in the care plan.

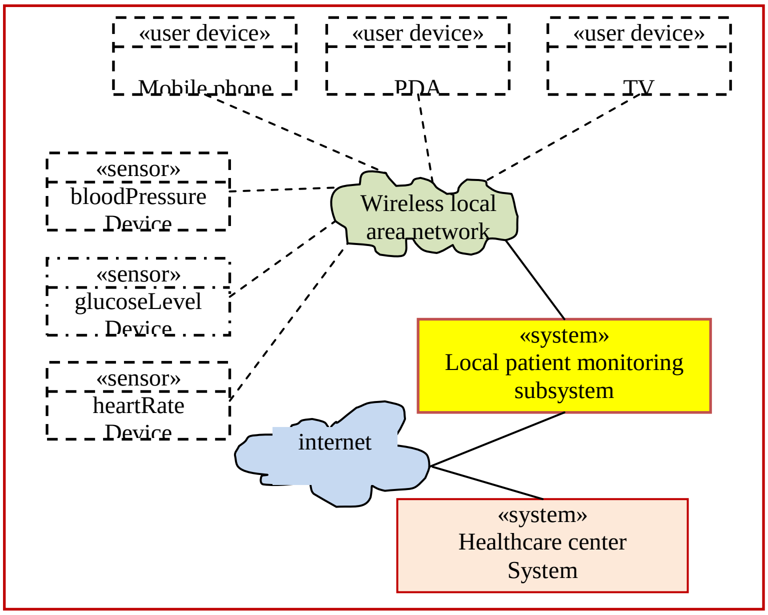


Fig. 3. A product line architecture of the LPM subsystem.

In this case, a reconfiguration of the product architecture at runtime is needed. In case of the availability of various devices at home (TV, PC, iPad, etc.), they could be selected and bound at runtime. This selection is based on context rules that should be defined at deployment time.

To do this, we represent variability at the architectural level that we associate with a decision table to formalize context rules of selection at run-time. Therefore, in terms of the product line architecture, these devices can be defined as variabilities whose architectural elements can be bound at deployment time or at runtime.

The dynamic variabilities should be clearly and explicitly described at the architectural level, specifying the adaptations and the rules under which such adaptations should take place at runtime [Jaring and Bosch (2002)]. In our approach, variabilities are described as architectural variation elements and conditions, which determine the architectural adaptation actions

Due to the space limitation, we only decompose the subsystem of the local patient monitoring (LPM). The remainder ones could be developed at the same manner. Indeed, the LPM subsystem (Fig. 4) will be decomposed according to the separation of concerns principles, leading to the above-mentioned three types of components in embedded systems. Indeed, we decompose such subsystem into a database component (persistent information about patient status and historical states), an interface component for handling patient's body sensors, an interface component for user input and display, and a local monitoring component for handling the application logic at the LPM subsystem.

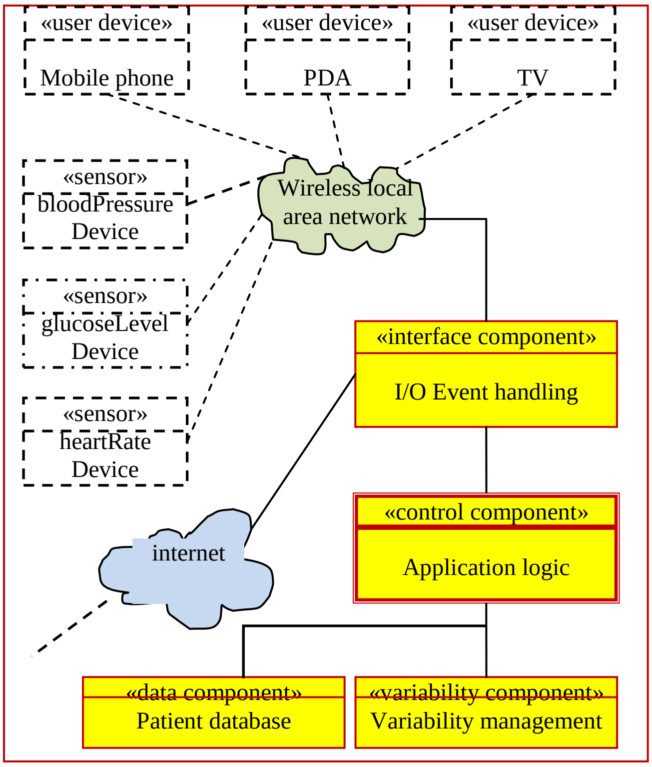


Fig. 4. The component model of the local patient monitoring subsystem.

To be able to derive product at run-time, we define the variability component for dealing with variability description and the decision table of product derivation. In the variability component, the decision table specifies for example, that the system selects the device only if the patient, located by a location service of the local monitoring component is in the same room of the device.

4. Related works

A variety of existing approaches propose the incorporation of variability in software systems in a way to product line engineering in embedded systems. Feature modeling has been used in most of these works for modeling variability. However, as features have no semantics [Czarnecki and Antkiewicz (2005)], their use in automatic adaptation is not guaranteeing correct behavior. One of the important proposals, the SPLIT approach [Coriat (2000)], considers variation points to have attributes and therefore uses the UML classifier, to represent a variation point. However, using this technique systematically requires development of specific scripts and programs to manage it, since it is not integrated in UML, nor in UML design tools [Capilla and Topaloglu (2003)]. In [Sergio et al (2008)], authors have presented a software architecture model for home care systems. Both kinds of variabilities are described by means of contracts. However, it is unclear whether the adaptive behavior specification can be made to deal with changes related to availability and quality of resources, and to the user needs, as proposed in our approach. This feature facilitates the independent evolution of both the architecture and the adaptation issues of the dynamic variabilities.

5. Conclusions and future work

Domains like healthcare have taken advantage of the recent advances in information and communication technology, especially, in the wirelessly connected devices and the web.This has lead to an increase in software complexity, with extensive variations in both requirements and resource constraints. Built around the emerging technology, healthcare applications require a higher degree of adaptability and dependability. Our designed architecture of healthcare system allows the patient at his home to be self-responsible for his own health, while managing the health delivery process. On the other side, physicians will be provided with more accurate data from continuous monitoring of patients in their natural environment and will be able to give more accurate prescribes. There is still room for improvement and completing the work. In particular we plan to focus on developing a flexible, model driven, product derivation from the domain model of healthcare for completing the product-line based process.

References

Al-hakim L (2007): Web and Mobile-Based Applications for Healthcare, Management, Hershey, PA: IRM Press.

Capilla R. and Topaloglu N.Y. (2003): Representing Variability Issues in Web Applications: A Pattern Approach. In Computer and Information Sciences, ISCIS 2003, LNCS, 2869, pp. 1035–1042.

Carvalho S.T., Murta L., Loques O. (2012): Variabilities as First-Class Elements in Product Line Architectures of Homecare Systems, SEHC, June.

Ceri P. Fraternali and Bongio A. (2000): Web Modeling Language (WebML): a Modeling Language for Designing Web sites, Computer Networks, 33(1–6), pp. 137–157.

Chenhui Z., Huilong D., Xudong L. (2008): An integration Approach of Healthcare Information System, Proc. IEEE Conf. on Biomedical Engineering and Informatics, pp.606-609,

Clements P., Northrop L. (2002): Software Product Lines: Practices and Patterns, SEI Series in Software Engineering, Addison-Wesley.

Coriat M., Jourdan J., Boisbourdin F. (2000): The SPLIT Method, SPLC1, Kluwer Academic, pp. 147-166.

Czarnecki K. and Antkiewicz M. (2005):Mapping Features to Models: A Template Approach Based on Superimposed Variants. 4th International Generative Programming Conference, Springer, LNCS, 3676, Tallinn, Estonia, October,pp.422–437.

Eslami M. and van Sinderen M. (2009): Flexible Home Care Automation Adapting to the Personal and Evolving Needs and Situations of the Patient, in 3rd Pervasive Health, Los Alamitos, pp. 1–2.

Gomaa H. (2000): Designing Concurrent, Distributed, and Real-Time Applications with UML, Addison Wesley, 2000.

Gomaa H. (2004): Designing Software Product Lines with UML: From Use Cases to Pattern-based Software Architectures, The Addison-Wesley Object Technology Series.

Istepanian R., Jovanov, E., and Zhang Y.T.,(2004): Beyond Seamless Mobility and Global Wireless Healthcare Connectivity, IEEE Transactions on Information Technology in Biomedicine, 8(4), Dec., pp. 405-441.

Jaring M. and Bosch J. (2002), Representing variability in software product lines: A case study, In Software Product Lines: Proceedings of the Second Software Product Line Conference (SPLC2), LNCS 2379, August, pages 15–36, San Diego, CA.

Kang K., Lee J., and Donohoe P. (2002): Feature-Oriented Product Line Engineering, IEEE Software, 19(4), pp. 58-65.

Kimour M.T., Boudour R., and Ghanemi S. (2012): Closing the gap between the requirements specification and the design model for embedded systems, Second International Symposium on Modeling and Implementation of Complex Systems (MISC 2012), Constantine, Algeria, may.

Kirovski D.; Oliver N.; Sinclair M.; Tan D. (2007): Health-OS: A Position Paper, In Proceedings of the 1st ACM SIGMOBILE International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environments, San Juan, Puerto Rico, June 11, pp. 1–3.

Ko E.J.; Lee H.J.; Lee L.W. (2008): Ontology and CDSS based Intelligent Health Data Management in HealthCare Server, In Proceedings of World Academy of Science Engineering and Technology, Vienna, Austria, August(13–15), pp. 119–123.

Koch N.and Kraus A.(2002): The expressive power of uml-based web engineering, In IWWOST, . Springer, LNCS- 2548, pp. 105–119.

Kolitsi, Z. and Umpierrez, M. (2007): Personal Health Systems: Deployment opportunities and ICT research challenges, Conf. report, February(12-13), Brussels, http://ec.europa.eu/information_society/ehealth.

Krueger W. (2006): New methods in software product line practice, com. Of the ACM, 49(12), pp. 37–40,

Lee H., Lee S., Ha K., Jang H., Chung W., Kim J., Chang Y., and Yoo D. (2008): Ubiquitous healthcare service using Zigbee and mobile phone for elderly patients, Int.Journal of Medical Informatics, issue 78, pp.193-198,.

Loques O. and Sztajnberg A. (2010): Adaptation Issues in Software Architectures of Remote Healthcare Systems, in 2nd SEHC, Cape Town, South Africa, pp. 24–28.

Park, S. and Jayaraman, S. (2002):Enhancing the Quality of Life Through Wearable Technology, IEEE Engineering in Medicine and Biology, 22(3), ,May/June, pp. 41-48.

Patrick K., Griswold W. G., Raab F ., and Intille S. S. (2008): Health and the mobile phone, Amer. J. Prev. Med., 35(2), pp. 177–181.

Sentilles S., Vulgarakis A., Bures T. (2010) : A Component Model for Control-Intensive Distributed Embedded Systems , Proceedings of the 11th International Symposium on Component-Based Software Engineering pp. 310 – 317

Sergio T., Carvalho y., M.urta L., Loques O. (2012): Variabilities as First-Class Elements in Product Line Architectures of Homecare Systems, SEHC, Jun 2012.

Spasov G. and Petrova G. (2008): Web-based Personal Health Systems – Models and Specifics, Proc of Conf. on Computer Science, Kavala, Greece, Nov (4-6).

Sztajnberg A., Rodrigues A., Bezerra L., Loques O., Copetti A., and Carvalho S. T. (2009): Applying Context-aware Techniques to Design Remote Assisted Living Applications, Int. Journal of Functional Informatics and Personalized Medicine, 2(4), pp. 358–378.

WUMA: ANALYSEUR D'USAGE DU WEB

Boubetra Ayoub, Attia Abdelouahab, Boubetra Abdelhak, Benabid Sonia

Dept. of computer science
University of Bordj Bou Arréridj
Bordj Bou Arréridj, Algeria
bobayoub5@yahoo.fr

L'objectif de ce travail est de concevoir et réaliser un outil logiciel nommé WUMA (Web Usage Mining Analyser) qui permettra au « Webmaster », qui peut être considéré comme un QoS Manager, d'avoir l'ensemble des connaissances sur le site web qu'il gère en vue de son amélioration. Cette dernière peut concerner l'adaptation du site web aux comportements de ses utilisateurs et cibler la bonne organisation de la structure du site en prenant les bonnes décisions concernant ce qu'il faut mettre comme contenu dans le site et sur le choix de la bonne forme et le meilleur agencement des pages web. Il s'agit en fait d'extraire de l'information à partir du serveur hébergeant le site Web et prendre les décisions d'amélioration de la qualité de service (QoS). Cet intérêt nous a poussés à créer et mettre en exploitation un site éducatif appartenant à l'université de la formation continue de Bordj Bou Arreridj en Algérie (www.ufcbba.net) pour mener notre étude d'amélioration de la qualité de service.

Keywords: web services, QoS, Web usage Mining; Web Mining.

1. Introduction

L'explosion du nombre de sites web connectés sur Internet et la croissance accélérée du nombre d'internautes confirment de plus en plus la position du web comme un média de masse. Par conséquent, le problème de « l'audience » des sites web revêt de plus en plus d'importance. L'audience d'un site web est un subtil mélange entre le nombre de visiteurs de ce site (le quantitatif) et l'intérêt qu'ont les internautes à visiter le site(le qualitatif).

Face à cette croissance continue de l'utilisation du web et la prolifération du commerce électronique, des services Web et des systèmes d'information basés sur le Web, il est devenu très nécessaire d'améliorer la qualité des services (QoS) des sites web destinés à cet effet [1]. Cette amélioration qualitative exige et nécessite l'étude des données relatives aux accès historiques effectués par les usagers afin de faire sortir les traces (patterns) des comportements de ces derniers. D'un autre côté, pour mener à bien cette

étude, la grande masse de données relative à ces accès nécessite le recours aux techniques de fouille de données [2] chose que nous allons investiguer dans ce papier. L'ensemble d'applications utilisées pour fouiller ces données web dans le but d'extraire des connaissances est connu sous le nom de fouille de Web (Web Mining).
La caractérisation où le profilage des internautes fréquentant un site Web et l'identification de leurs motifs de navigation est un problème incontournable pour les concepteurs de sites Web [3] qui visent à assister l'internaute, prédire son comportement et personnaliser la consultation. Ces trois considérations ont motivé d'importants efforts dans l'analyse des traces des internautes sur les sites Web et l'adaptation des méthodes de classification aux données du Web durant ces dernières années.
Au cours de ces dernières années l'activité sur le Web et les données résultantes ont connu une croissance très rapide, vu la croissance exponentielle du nombre des documents mis en ligne. Selon une évaluation de l'Internet World Stats au 31décembre 2011, il y avait (2, 267, 233,742) d'internautes dans le monde [4], et le nombre de sites Web a atteint les 580 millions de sites, dont plus de 175 millions considérés comme actifs en janvier 2012[5]. Aujourd'hui, le traitement et l'analyse des données sur ce nombre d'internautes et de sites forment le domaine du Web Usage Mining (WUM).
Dans ce travail, nous nous intéressons aux problématiques du WUM dans un but d'amélioration de la qualité de service des sites web en général et plus particulièrement ceux destinés aux communautés éducatives et universitaires.
Dans ce papier la section 1 est une introduction sur l'intérêt du web mining. La section 2 est une spécification des services attendus du WUMA. La section 3 contient le processus de web mining. La section 4 explique l'approche proposée. La section 5 montre la conception des services du WUMA. La section 6 explique l'implémentation des services du WUMA et on termine par une conclusion.

.

2. SPECIFICATION DES SERVICES ATTENDUS DU WUMA

Ce travail apporte des connaissances et des informations, principalement, sur les éléments suivants :

2.1. *Connaissances sur les visiteurs du www.ufcbba.net*

- Les connaissances sur les utilisateurs d'un site sont obtenues directement auprès des utilisateurs eux-mêmes dans l'approche panéliste (âge, sexe, ancienneté sur le Web). Dans le cas des sites à base d'inscription, ces connaissances sont recueillies directement à partir du login et du profil de l'utilisateur donné par l'internaute au moment de l'inscription. Ces données dites explicites, fournies directement par les internautes sont très souvent erronées. Il est également possible d'acquérir des connaissances sur les utilisateurs du site en reconstituant leurs profils en fonction de leurs activités passées sur le Web.
- Détenir le pourcentage des visiteurs par semaine, par mois et par an.
- Avoir une visibilité géo-spaciale: d'où proviennent les visiteurs ?

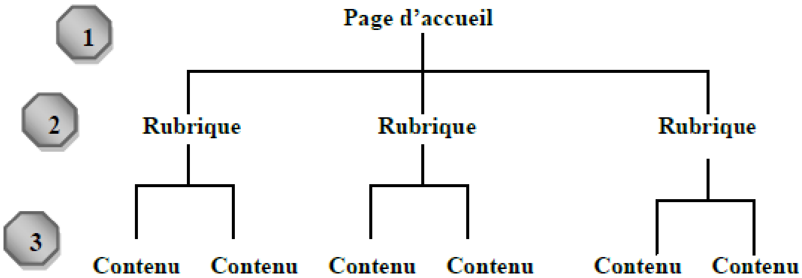


**Figure 1. - La représentation d'un site web par un arbre.
- Exemple d'un chemin de navigation.**

2.2. *Connaissances sur les pages du www.ufcbba.net*

- Reconnaître les pages Web les plus et les moins consultées (pages populaires et pages impopulaires).
- Connaître les combinaisons des pages consultées.
- Savoir quels sont les liens les plus utilisés.

2.3. *Connaissances sur les navigateurs et les systèmes d'exploitation*

- Connaître le pourcentage des navigateurs les plus utilisés.
- Connaître également le pourcentage des systèmes d'exploitation les plus utilisés.

2.4. *Connaissances sur les formes d'exploitation du site www.ufcbba.net*

Les pages d'un site sont matérialisées et accessibles par une adresse Internet spécifique appelée adresse d'allocation de la ressource (URL: Uniform Resource Locator). La structure d'un site Internet simple peut être représentée par un arbre dont la racine correspond à la page d'accueil du site comme il est indiqué dans la Figure 1. Chaque point (ou nœud) présente l'adresse d'une page particulière et les segments reliant ces points indiquent la présence d'un lien hypertexte amenant aux sous-branches immédiates de l'arbre. D'après le schéma ci-dessous de la Figure 1, il est possible de retracer le chemin de navigation de l'internaute sur le site ; par exemple le chemin (1, 2, 3). Cependant, il n'est pas toujours aisé de représenter l'architecture d'un site, en particulier les sites complexes.

3. PROCESSUS DU WEB MINING

Le WUM (Figure 2) consiste en l'application des techniques de fouille des données pour découvrir des patrons d'utilisation à partir des données du Web dans le but de mieux

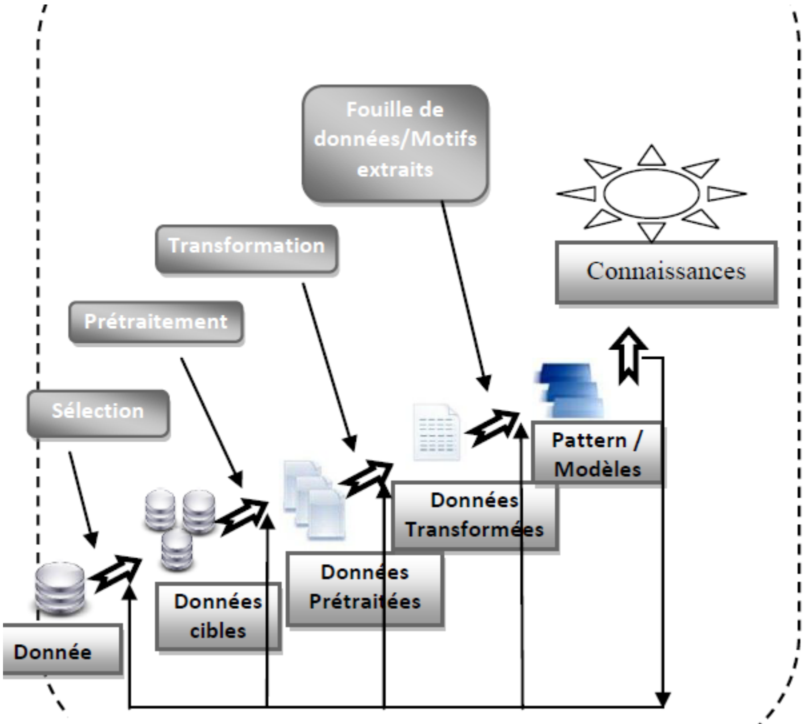


Figure 2. Un processus global et itératif du WUM.

comprendre et servir les besoins des applications Web [6], de découvrir des constantes, schémas ou modèles, dans les ressources d'internet ou les données le concernant.

- La première étape dans le processus de WUM, une fois les données collectées, est le prétraitement des fichiers Logs qui consiste à nettoyer et transformer les données.
- La deuxième étape est la fouille des données permettant de découvrir des règles d'association, un enchainement de pages Web apparaissant souvent dans les visites et des clusters d'utilisateurs ayant des comportements similaires en termes de contenu visité.
- L'étape d'analyse et d'interprétation clôt le processus du WUM, elle a comme objectif de filtrer les modèles inintéressants de l'ensemble trouvé dans la phase d'extraction des modèles. Ce filtrage dépend de l'application finale que l'on souhaite faire du Web Usage Mining (adaptation des sites web, système de recommandation, pré chargement des pages, etc. . .).

4. APPROCHE PROPOSÉE

Notre approche est répartie sur deux phases, la phase 1 consiste à l'application de l'algorithme Apriori sur les données web pour rechercher les itemsets fréquents (les pages consultées ensemble) et au niveau de la phase 2 on génère les règles d'association valides à partir des Itemsets générés en phase1.

Phase 1. Rechercher les itemsets fréquents (les pages consultées ensemble)

- Cette règle suggère qu'il existe une relation forte entre la visite de la page C.html et D.html parce de nombreux utilisateurs qui visitent la page C.html visitent aussi la page D.html.
- Pour commencer on fait la transformation de la table du fichier log vers une table binaire.

Le support de l'ensemble {B.html, C.html, D.html} est égal à 2.

Pour que cette règle soit générée, son support doit satisfaire la relation suivante :

Le support est >= le seuil donnée minsup.

Pour calculer le support de la relation :

On a une règle d'association X -> Y où X et Y sont des ensembles d'items disjoints et N le nombre des internautes. Le support de la règle (X->Y) est égale à :

Support,s(X->Y) = Sigma (xUy) / n

Exemple :

Considérons la règle {B.html, C.html} -> {D.html}. Le support de l'ensemble {B.html, C.html, D.html} étant égal à 2 et le nombre total de transactions étant égal à 5, le support de la règle est 2/5 =0,4. Donc 40% des utilisateurs qui visitent la page B et C visitent aussi la page D.

5. CONCEPTION DES SERVICES DU WUMA

Vu que notre système logiciel utilise des traitements basés sur des flots de données en entrée, nous avons pensé à l'utilisation de la méthode SA (Structured Analysis) pour mener la spécification et la conception de notre logiciel.

6. IMPLEMENTATION DES SERVICES WUMA

Dans cette partie on suivra les différentes étapes des techniques et méthodes à appliquer sur le fichier Log de la période étendue du mois de Mars au mois de Juin de l'année 2012 du site www.ufcbba.net qu'on a créé afin de bien mener notre étude dans le but de l'extraction des connaissances. Il s'agit d'exploiter les connaissances pour la bonne gestion du site web et prendre les décisions d'amélioration de la qualité de service.

Etape1 :

Pendant l'importation du fichier Log indiqué dans la Figure 6, il sera transféré à notre base de données (PostgreSQL) dans une table composée de plusieurs colonnes et chaque colonne correspond à un champ spécifique du fichier Log afin de pouvoir la manipuler et appliquer les différents traitements nécessaires pour les étapes suivantes.

7. CONCLUSION

Avec la croissance explosive des sites sur le WWW, les mesures appropriées qui non seulement évaluent la qualité d'un site web mais aussi de fournir aux Managers de sites un aperçu sur les problèmes potentiels qui sont d'une importance considérable.

Ce travail a traité les mesures multidimensionnelles relatives à l'usage du web dans un contexte appartenant à notre environnement éducationnel qui est celui de notre université de la formation continue. Les résultats obtenus, par rapport à l'échelle de l'étude, suggèrent des caractéristiques psychométriques satisfaisantes d'utilisation du site web en question.

8. References

[1] WebCT: http://www.webct.com/.

[2] C. Groeneboer, D. Stockley, T. Calvert, Virtual-U: A collaborative model for online learning environments, Proceedings Second International Conference on Computer Support for Collaborative Learning, Toronto, Ontario, December, 1997.

[3] M. Spiliopoulou, L. C. Faulstich, K. Winkler, A Data Miner analyzing the Navigational Behaviour of Web Users, Proceedings of workshop on Machine Learning in.

[4] http://www.internetworldstats.com/stats.htm, «world internet usage and population statistics».

[5] Statistiques prises du site Web Server Survey January 2012.

[6] R. Cooley. Web Usage Mining: Discovery and Application of Interesting Patterns from Web Data. PhD thesis, University of Minnesota, May 2000.

CLOUD AS A MAJOR CATALYST OF CHANGE IN CONTEMPORARY BUSINESS ENVIRONMENT

Atif Farid Mohammad

Department of Computer Science, University of North Dakota,

Grand Forks, ND 58202 USA
atif.mohammad@my.und.edu

Hamid Mcheick

Computer Science Department, University of Quebec at Chicoutimi
Chicoutimi, Quebec Canada
hamid_mcheick@uqac.ca

The business environment around us is a on a continuous change on a very fast pace as Cloud computing is taking over as new business environment. To cope up with the issues of process changes, such as adaptation of new or different rules and regulations at the time of an expansion of a business extends a need of its business processes to be reengineered. This article discusses in detail about three fundamental concepts of Business Process Reengineering or BPR, Systems Reengineering and Service-Oriented Architecture or also widely known as SOA. This study provides background of these concepts as well as provides the artifacts researched and work by several researchers and business analysts.

Keywords: Business Environment, Business Process Reengineering, Cloud Computing, Service Oriented Computing

1. Introduction

The term business process reengineering or BPR was coined by Michael Hammer and James Champy in 1990 [1]. The two pioneers define business process reengineering as “the fundamental rethinking and radical redesign of business process to achieve dramatic improvements in critical, contemporary measures of performance, such as cost, quality, service, and speed” [1]. However, achieving dramatic improvements through business process reengineering is easier said than done. According to Abdolvand et al, [2] 60-80% of BPR have been unsuccessful. The high failure rate coupled with huge costs; make BPR a very risky operation. We will also discuss when organizations need to embark on BPR and when they need not go through that path.

System reengineering can be described as focusing on improving existing or legacy systems for the new need or reducing operation cost. In some articles [18] [19], reengineering is closely related to maintenance, which is defined by the modification of a software product after delivery to correct faults, to improve performance or other attributes, or to adapt the product to a changed environment. By placing reengineering on a continuum with maintenance and new development, the true nature of reengineering becomes apparent. Maintenance entails making corrective, perfective, and adaptive changes to software, while development focuses on implementing new capabilities, adding functionality, or making substantial improvements typically by using new computer resources and new software technologies.

The use of Service-Oriented Architecture or SOA in the Cloud provides services data abstraction. This abstraction can be understood as services messaging metadata, which can be in the shape of XML, XSD or other set industry standards providing interoperability to a user's request. The messaging between services is encapsulated for information-hiding purposes. A SOA solution for a bank as shown in Figure 1.1 is designed to maintain customer service by providing customer care as well as web self-serve options for consumers.

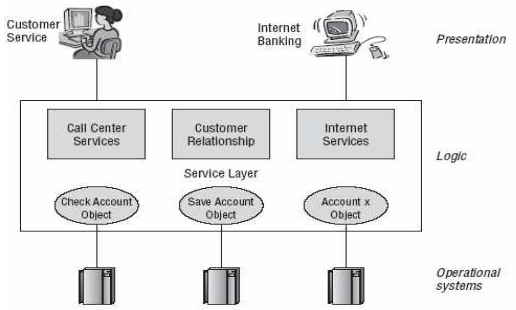


Figure 1. Bank customer service solution. [3]

SOA eases the development of ever-changing applications that compare data with stationary applications while maintaining a decoupled relationship between these applications on a simultaneous basis. Cloud is a loose union of services and SOA has brought in a fresh approach for business information technology departments, making it easy to assemble and configure IT components like building blocks that can be combined to provide easy and fast solutions. For example, a bank provides a line of credit by checking a consumer's credit; an automotive spare-parts seller checks the inventory; Fed-Ex maintains the shipping status for delivery to consumers, etc.

This article sheds light on the use of SOA to design the services deployable in the Cloud using business process reengineering process to make applications simple and easy to use by both internal and external users of the organizations, which are interested to move to Cloud computing.

2. Business Process Reengineering

This section of the article presents detailed discussion on the concepts worked by several researchers.

2.1. *Success Factors of a BPR*

Abdolvand et al [2] came up with five success factors which they called positive readiness indicators as shown in Figure 2.1, and they are: egalitarian leadership, collaborative working environment, top management commitment, supportive management, and the use of information technology. The authors based their research on two Iranian companies in different industries as well as research done by other experts. Each main factor has sub-factors. The combination of the factors and their sub-factors cover all the important factors [2].

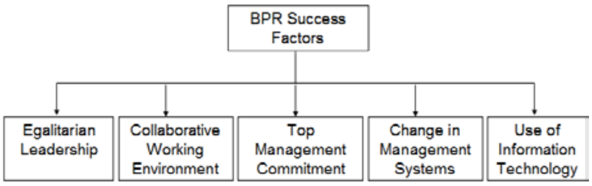


Figure 2. Business Process Factors

A critical success factor is what Abdolvand et, al, [2] called egalitarian leadership. Top management should drive the changes by providing vision and employees should become more responsive. However, while communication is very important in the success of any venture, nonetheless, culture plays an important role in how people communicate. For example, in low power distance cultures such as the United States and some West European cultures, relations between leaders and the average people is much closer than other cultures that have high power distance. Such cultures (high power distance cultures) always expect ideas to come from the top in the form of commands and the rest are expected to obey without questioning the wisdom of such ideas. Would this egalitarian leadership approach work universally? We believe it is highly unlikely to work across cultures.

The third and fourth success factors are top management commitment and supportive management. Management should have a clear strategic vision, or, in other words, they must know where the organization is and where it is headed [2]. The authors were short on description on this topic. However, we believe without managerial commitment and involvement, the organization could easily fail in realizing its objectives. Management needs to guide the path and offer support to their employees in order to transition seamlessly into their new environment. For example, one way management could support its employees is to provide training.

The fifth success factor is the use of information technology. As per Abdolvand el al, [2] “IT is introduced as a critical component and even a natural partner of BPR, which has a continuous and important role in BPR projects”. We agree with the authors’ findings. In fact, IT is a natural “ally” of BPR. It is intrinsically an integral part of the process. However, organizations need to make due diligence before undertaking the proposed change. They must calculate their ROI before investing on IT. If you can’t recoup your investments in a reasonable time, don’t invest in the proposed project.

2.2. *Failure Factors*

Abdolvand et al [2] cited resistance to change as the most common factor that causes the failure of BPR. The authors called this factor the “unreadiness factor”. Humans resist change, particularly, when they don’t expect to benefit from the change. We believe since humans are self-interested, and some employees may even go to great lengths to derail the process, management must reassure its employees that they won’t be worse off when the process gets completed. The fear factor is a legitimate concern which needs to be addressed before the change starts.

Another factor that could contribute BPR to fail is having unreasonable expectations about the project. According to Herzog et al [4] “The first consists of training about the importance and role of BPR, about the benefits of BPR, the role of cooperation, the commitment to employee training, and the resource availability for training”. Herzog et al

[4] argue that management must be knowledgeable and have realistic expectations about the benefits of BPR and what it can actually contribute to the bottom-line.

One often overlooked factor in BPR failure is culture. We already mentioned some success factors that may not work across all cultures. According to Newman et al, [5] “The tenets of Confucianism suggest that a BPR effort will challenge Chinese ideology and the workforce will resist radical change including changes to the management hierarchy. Confucius considered it important to maintain a state of harmonious equilibrium in society”. The authors found that Chinese companies considered cultural issues as the most important factor when undertaking BPR. Newman et al mention that Chinese culture prefers gradual change, and BPR advocates a dramatic change; hence, most companies would not adopt BPR. Moreover, due to the risk-averse nature of Chinese society, Newman et al, [5] also have found that “most Chinese managers advocate stability and want to maintain power and control. Indeed, many employees are satisfied with getting explicit work instructions and are not willing to initiate actions that are often associated with risks”. We agree with the authors. Usually, uncertainty avoidance cultures, which include Chinese culture, try to avoid taking chances. Such cultures do not like ambiguities. Since BPR is fraught with uncertainty, it may not work in those cultures.

2.3. *Benefit of BPR*

Organizations need to make a thorough research before they decide on implementing BPR. Having metrics in place ensures that an organization knows what it is getting into. However, in order for reliable metrics to be established, management must be knowledgeable [5]. There is no black and white answer to this question and experts disagree on this. S, some suggest if the organization is doing ok, it should not rush into BPR; making changes gradually might be enough. However, prominent experts on the field such as Michael Hammer advocate no less a change than a complete (100%), fundamental, and a radical change. Hammer says don’t do changes piecemeal [1].

We suggest that BPR may not be the best approach for some organizations. The factors that could influence an organization’s decision on whether to implement BPR or not, could include on evaluating where the organization stands versus the competition.

3. System Rengineering

For quite a while in last four to five decades systems were being created for industries requirements, design decisions process and business rules successful implementation. In order to effectively use current assets of an organization, it is important to develop a systematic strategy for the continued evolution of currently elder systems to meet changing mission, technology and user needs. Moreover, knowledge of the business rules and technical decisions is often embedded in the code. Such knowledge is difficult to recover after many years of operation, evolution, and personnel change. The software portion of a legacy system may have been written 10 or 20 years ago, developed using archaic and ad-hoc methodologies, and subjected to prolonged and sometimes need costly maintenance. The result is a legacy system that lacks the ability to evolve and develop for meeting changing demands. Systems reengineering is an esteem requirement to fill this gap. Years ago, legacy systems are created for customer and business environment needs and used by companies. Moreover, legacy system may have been written 10 or 25 years ago, developed using archaic and ad-hoc methodologies, and subjected to prolonged and sometimes need costly maintenance. The problems to legacy system are lacking the ability to evolve and develop for meeting changing demands.

3.1. *The Goals of System Reengineering*

The following figure 3.1 shows the goals of system reengineering:

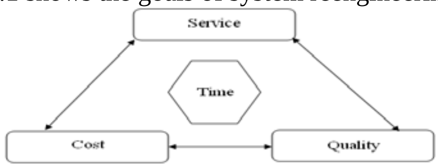


Figure 3. An overview of system reengineering [6]

Reengineering is the transformation in a systematic way of an existing system into a new form in order to improving of an operation, system capability, new function and high performance at a lower cost, less time , high quality service to the customers. The word system is a combination of business rules, processes, and operations of user applications. Software is usually used to elaborate a system in the world of Information Systems.

3.2. *System Reengineering Approaches*

There are three main approaches [6] utilized to reengineer any system, following is given the detail of each to understand the goals of each approach, the first one is "Big Bang" approach, replaces the entire system at one time. This approach main advantage is often to be used by projects that need to solve an immediate problem, such as migration to different technology, totally different user interface as well as the system architecture. The disadvantage with this approach is the result tends to be on the bigger software projects or ERP software, such as SAP, may not always be suitable. For large systems, this approach may consume too many resources or require large amounts of time before the target system is produced. The risk with this approach is high the system must be functionally intact and work in parallel with the legacy system to assure functionality.

The second approach is "Incremental" approach to a software system re-engineering is also known as "Phase-out" approach. In this approach, system fragments are re-engineered and added incrementally as new versions of the system are needed to satisfy new goals that contain the training of staff on an individual error free module one by one. The project is broken into re-engineering fragments based on the existing system's fragments. Every module of the software has similar with a little enhanced functionality on the new technology is developed and tested on the basis of the legacy software module. The main advantage of using this approach as it has less risk than the “Big Bang” approach as the risks can be identified and monitored at the time of development and beta testing of each module. The disadvantage of using this approach is the use of much longer timeframe to reach to the goal of reengineered software.

In the "Evolutionary" approach, as in the Incremental approach, fragments of the original system are replaced with newly reengineered system fragments. The components of the legacy system need to be broken by the use and reengineered into new modules. The main advantage is converting of a legacy system to latest technology. There is one disadvantage and it is that similar functions must be identified throughout the legacy system then refined as a single functional module.

There are quite a few factors involved in the “Migration” process of a legacy system to be converted in to a WEB UI system, such as the coding language, base technology that contains RDBMS, operating system the new networking technology involving wireless networking as well as advanced hardware infrastructure. System reengineering activities in general require techniques and tools that help to dig out valuable knowledge about the structures and inner mechanism of software systems from the source code using reverse engineering, since other information sources such as system documentation and developer's knowledge are often not available or consistent. A key dilemma during reverse engineering tasks is that the information extracted from the source code has to be condensed and abstracted towards an abstraction level which can support software engineers to perform essential analyses or modifications of the system and preventing them from an information overflow that makes software reengineering goals complicated. These activities have to be supported by tools. An important consideration in building tool support for reengineering is what information must be provided, and how this information is modeled.

3.3. *Using Data Flow to Explain System Reengineering: A Case Study*

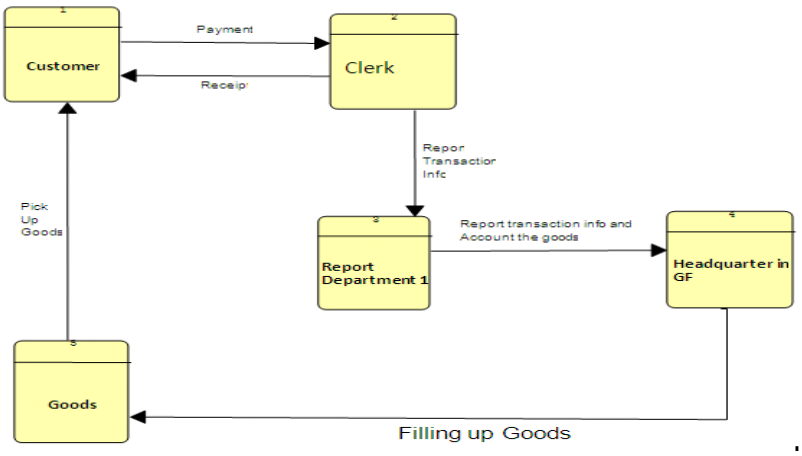


Figure 4. Legacy system

Let us assume that there is a branch of retail company found in Grand Forks .ND. The legacy system of this company is that: First, customer picks up goods in the branch company. After that, customer pays the bill and serviced by a clerk at front desk. Moreover, Clerk hand in transaction information to another clerk who works in report department and, after accounting the goods, reports transaction information to headquarter in grand forks after. While headquarter received information from branches, will send goods to the branch for the need. Figure 3.2 shows the data flow of the business process of this branch of the retail company.

Reengineered System in Retail Company in Grand Forks

After analysis, the system reengineering team translated the legacy system to a Cloud System by using accounting report system which is sold by an information technique company and operated on using a computer. The accounting report system sends transaction information immediately through the website when transaction is occurred. Installed AR system, the report department is dismissed automatically, and transaction information would be sent by clerk who works at the front desk at first time. While receiving report transaction information, headquarter is going to fill up the lacks of the goods. This reengineering brought the retail company benefits as described:

a) The service time is reduced, as the clerk would report to headquarter immediately after customer's purchase(s) of the goods for initiating restocking purpose.

b) The cost of good could be reduced significantly per branch.

c) System reengineering introduced an improved business operations cycle.

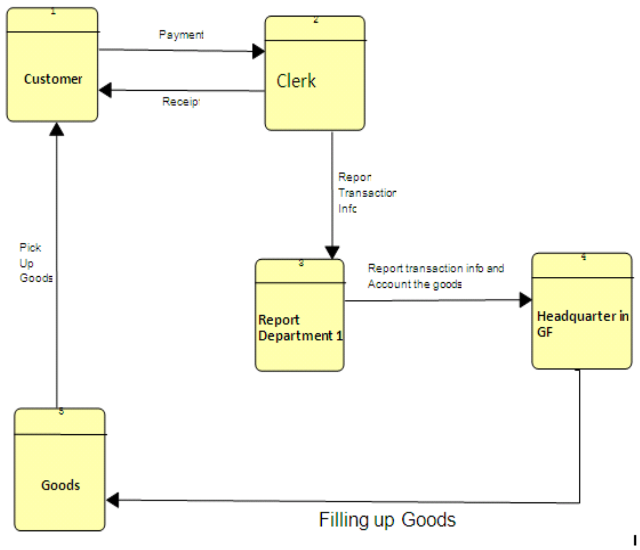


Figure 5. Reengineered Cloud System

Therefore, a basic assignment of system reengineering is to develop a brief model as a basis for such tools, which can enable the software development team(s) to characterize several associated software applications artifacts needed during the reengineering process and to build high-level abstractions from low level legacy program elements. Problems to be addressed in this project can be as follows:

a) Specification of a model to represent design information of a mini or main-frame system.
b) Specify operations to query the design information stored in the model.
c) Specify operations to abstract low level program entities into more abstract high-level entities using ERD and DFD mechanisms.

4. Service Oriented Architecture

SOA provides the framework to work modularly for the Cloud. Due to this flexibility, any business can adapt SOA and assemble its services as needed. A bank's services illustrate this in Figure 1. After changing a consumer base or adding new products or services for current customers or to attract new customers, these functionality components can be reassembled. This reconfiguration of service components is invaluable for saving money as well as time. New components can also be added to provide better or more services on the base functionality components. In the case of a merger or cooperative partnership with other related businesses, the components can be included as an additional service to the consumers at one platform. SOA provides the flexibility for businesses to improve and change with the pace of time.

The challenges of the present financial climate can be resolved by software engineers and decision makers, such as CTOs or MIS managers, by aligning business needs, using service components to improve service to consumers within a tight budget constraint. SOA is a boon to an enterprise looking to create service components in an agile fashion and reuse an existing components infrastructure. According to Schreiner and Lamb [7], systems of the future will be based on the concepts of SOA. Service applications will be composed of a number of individual services running at several servers. As illustrated by Erl [8], service component application logic can be divided into two levels: a service interface, where loosely coupled services are available with their implementation and technology platform; and a service-using application level in which service application logic is developed and deployed on different technology platforms. These services communicate via open protocols.

The purchase planning scenario of a commercial or residential property is an example of a system where the property-selling agency can be provided a flexible approach using SOA. The levels discussed above can be adapted as application and service interface levels. We now take a closer look at how SOA helps to plan a purchase, instantiate, and adapt the composed property purchase service. Initially, three services can be identified as:

i) "Property for sale" search service
ii) Offer submission to initiate a purchase service
iii) Credit check and mortgage service

In this example, partner service implementations are selected as follows. The "property for sale" service can be a local server provided by the real estate agency. The offer submission service can be the web services provided by another agency, or it can be assumed by law firm service providers. Such legal purchase services may provide the possibility of applying web service policies, but apart from that, the services might not be configurable. The credit check service is a web service provided by the financial department of a bank or a mortgage provider. Considering the strict security requirements of financial transactions, these services typically have fixed functional and nonfunctional requirements—i.e., security policies—that cannot be altered.

Duo to SOA, these services can be combined at one front-end platform on a website to provide the "property for sale" information and other services as mentioned above. To make this clear, we can take the example of Amazon's [9] store-front. Customers use a browser to get the displays on Amazon.ca. The front-end website infers the signals or the customer's intent triggers and calls upon services that do things like acquiring the data for the current on sale products, or getting the customer's order. The business sector is searching for ways to convert its Information Technology or Management Information System departments to swiftly adapt to changing business needs. SOA facilitates service orientation.

4.1. *Challenges of SOA*

As the world has become a global village, there is a significant need for businesses to fulfill consumer needs by providing services globally. This pursuit presents some challenges as given below:

- Flexibility is required and increasing demand of standardized services with seamless experience for consumers

- Reduction in operational cost of these services by getting satisfactory results due to improved efficiency, which means users control their business, rather than technology
- Services are mostly distributed, getting regulatory approvals
- Getting rigid information systems wrappers developed to get the services combined at one framework
- Efficient use of existing resources, services are heterogeneous.
- The patterns of services interaction are unpredictable
- The service(s) “front end” is less useful for testing purposes due to decoupled implementation on the backend.

SOA characterizes and provides the composite framework of the services based on open standards. Multiple services can be combined to serve a consumer at one platform; for example, a travel agent can provide car rental, hotel reservations or can book an entire vacation. This is a convenient way for the consumer to get several services with one phone call or one website utilization. Expedia [10] can be taken here as an example.

4.2. *Related work on SOA*

SOA provides the methodology to form web services as functional building blocks to build an accessible infrastructure for consumers, which is autonomous and not platform-dependent. Platform independence makes it easier to develop web services, such as operating systems, programming languages, etc. Quality of Service or QoS is an important factor in SOA as well. We live in an information age, managed by electronic devices. These devices communicate with each other using data communication networks, the best example here being Cloud. Most of the services we use in our daily life are communicated over these networks and are delivered to us on our digital gadgets. Quality of Service (QoS) for networks is an industry-wide set of standards and mechanisms for ensuring high-quality performance for critical applications. By using QoS mechanisms, network administrators can use existing resources efficiently and ensure the required level of service without reactively expanding or over-provisioning their networks several researchers, Korostelev et. al., [11] Borcoci1 et. al., [12] Zeng et. al., [13] have focused on the web services domain, while working on QoSaware service composition and management. During this ongoing research, the composition of QoS-aware services in P2P networks has engrossed much of the researchers’ interest. For automatic maintenance of QoS levels research in a P2P infrastructure, Huynh et. al. [14] have proposed self-organized and QoS-aware Virtual Service Communities. Gu et. al. [15] examined large-scale P2P systems in detail for their QoS-aware service compositions. They introduced SpiderNet. This framework is capable of doing:

- Service path selection
- Service path instantiation
- Benefit-driven peer clustering

This framework model produced results with a higher success rate of service provision than other problem-solving techniques in which the most appropriate solution of several was found by alternative methods in successive approaches. Let us take an example of an insurance company, which sells the following business services:

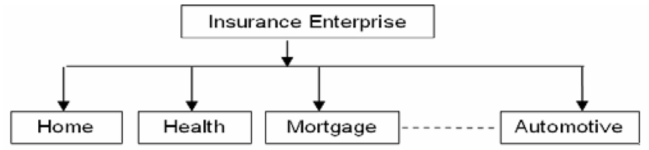


Figure 6. A Horizontal Framework of Services

1. Home insurance
2. Health insurance
3. Mortgage insurance
4. Employment insurance
5. Critical illness insurance
6. Automotive insurance

We can see these services as the following framework. Let us explore this horizontal framework of business services, as depicted in Figure 4. All of the mentioned business services are directly related to one enterprise. When a consumer needs one insurance product, the insurance enterprise must evaluate all of the service provisions and locate the appropriate service(s) from the listed services. From the researchers' point of view, a structure for service management is required to maintain QoS in a SOA solution. The aim of this research is to construct a framework that supports QoS for a SOA solution. SOA permits business organizations to interrelate their information systems using predefined rules for mutual dealings within and with consumers [16] [17]. SOA is a methodology that interconnects businesses and computational assets, such as software in use, current system applications and human resources of the organizations involved in to design, develop, and deploy a SOA solution.

5. Conclusion

Business process reengineering or BPR is a relatively new business idea. Since its inception in 1990, numerous organizations have tried to adopt BPR. However, the high failure rate indicates that BPR is not beneficial for everyone. Moreover, the cost of failure is very high both in terms of money and morale. Before they try adopting BPR, organizations must take into account on where they stand versus the competition. If they are way behind, then they have no choice but to adopt BPR. In addition, since culture plays an important role in determining the success or failure of BPR, organizations that are located in uncertainty avoidance cultures and high power distance cultures must think twice before adopting BPR.

Furthermore, even if all the factors that necessitate the adoption of BPR exist, it would still hard be for most organizations to succeed without a knowledgeable and hands-on management. A knowledgeable management helps set realistic goals and expectations and reduces employee anxiety about the changes to come through effective communication. We would like to conclude that the risks outweigh the benefits, and most organizations must not try to adopt BPR unless they feel they have nothing to lose.

Software architecture is a vital part of software engineering, and the use of Service-Oriented Architectures (SOA) for Cloud deployable services presents an advanced architectural concept with significance for a system's transition from legacy state to a Cloud system state. With SOA-provided flexibility, the new tailorable systems can be produced, and platform independence can also be achieved. Services designed using SOA are formulated software applications, and this formulation is closer to business domains. Research has also indicated that the call for additional metadata of service descriptions is growing quickly, and the amount of data collected from use experiences with a service needs to be stored for analysis of service and its future use. This data handling in terms of storage locations and synchronization raises issues and serious concerns about service performance. The service can have performance issues in terms of message communication to and from the user to the service provider due to this additional contextual information.

References

[1] Hammer, Michael. & Champy, James. Reengineering the Corporation. HarperCollins Publishers, Inc., New York, NY 10022. 1993.

[2] Abdolvand, Neda. Albadvi, Amir. & Ferdowsi, Zahra. Assessing Readiness for Business Process Reengineering. IT Department, Engineering Faculty, Tarbiat Modares University, Tehran, Iran. P. 497-510. 2007

[3] M. Rosen, B. Lublinsky, K.T. Smith, M. J. Balcer. Pulished. Applied SOA: Service-Oriented Architecture Design and Strategies. Wiley Publishing Ltd. 2008.

[4] Herzog, N. V., Polajnar, Adrej, & Tonchia, Stefano. Development and Validation of Business Process Reengineering (BPR) Variables: A Survey Research in Slovenian Companies. International Journal of Production Research. Vol. 45, No. 24, 5811-5834.

[5] Newman, Mike. & Zhao, Yu. The process of Enterprise Resource Planning Implementation and Business Process Reengineering: Tales from Two Chinese Small and Medium-Sized Enterprises. Info Systems J 18, 405-426. 2008

[6] Atif Farid Mohammad; The path from a Legacy System to GUI System. IEEE Conference on Systems, Computing Sciences and Software Engineering. Bridgeport, CT. December 2008.

[7] R. Schreiner, U. Lang; Protection of complex distributed systems. Proceedings of the 2008 workshop on Middleware security. pp. 7-12. 2008.

[8] T. Erl. Service-Oriented Architecture: Concepts, Technology, and Design. Prentice Hall PTR, 2005.

[9] http://www.amazon.ca/; Accessed on November 18, 2009.

[10] [http://www.expedia.com; Accessed on November 18, 2009.

[11] http://technet.microsoft.com/en-us/library/cc757120.aspx; Accessed on November 18, 2009.

[12] A. Korostelev, J. Lukkien, J. Nesvadba, Y. Qian, "QoS management in distributed service oriented systems", Proceedings of the 25th IASTED International Multi-Conference, pp. 345-352. 2007

[13] E. Borcoci1, A. Asgari, N. Butler, T. Ahmed, A.Mehaoua, G. Kourmentzas, S. Eccles, "Service Management for End-to-End QoS Multimedia Content Delivery in Heterogeneous Environment", Proceeding of Advanced Industrial Conference on Telecommunications/Service Assurance with Partial and Intermittent Resources Conference/ E-Learning on Telecommunications Workshop, pp. 46-52. 2005

[14] L. Zeng, B. Benatallah, A. Ngu, M. Dumas, J.Kalagnanam, H. Chang, "QoSAware Middleware for Web Services Composition", IEEE Transactions On Software Engineering, Vol. 30, No. 5, pp. 311. 2004

[15] H. Huynh, E. Lavinal, N. Simoni, "A Dynamic QoS Management for Next Generation of Services", Third International Conference on Autonomic and Autonomous Systems, pp. 11-11. 2007

[16] X. Gu, K. Nahrstedt, "QoS-Aware Service Composition for Large-Scale Peer-to- Peer Systems", Book Chapter: Grid Resource Management: State of the Art and Future Trends, Chapter 24, pp. 395 - 410, Kluwer Academic Press, October 2003

[17] W. Dostal et al.: Service-orientierte Architekturen mit Web Services: Konzepte – Standards – Praxis. Heidelberg, Elsevier, 2005

[18] Peter H. Feiler. Reengineering: An engineering problem. Technical Report CMU/SEI-93-SR-5, Software Engineering Institute, Carnegie Mellon University, July 1993

[19] Jean-Marc DeBaud, Bijith M. Moopen, and Spencer Rugaber. Domain analysis and reverse engineering. In Proceedings of the International Conference on Software Main-tenance (Victoria, BC, Canada; September 19-23, 1994), pages 326-335, 1994.

International Conference on Software Engineering and New Technologies

A NEGOTIATION APPROACH FOR REQUIREMENTS ENGINEERING PROCESS IN COOPERATIVE INFORMATION SYSTEMS

Chaouki Boulekdam

LIRE Laboratory, University Mentouri,
Constantine, Algeria
boulekdam.chaouki@ umc.edu.dz

Nacer eddine Zarour

LIRE Laboratory, University Mentouri
Constantine, Algeria
nasro-zarour@umc.edu.dz

The requirement engineering process aims to provide a requirements specification document that defines the system to be developed. In this paper we focus our work on the negotiation phase for the requirements engineering process in cooperative information systems (CIS). The negotiation includes the discussion of requirements conflicts and the search of a compromise approved by all stakeholders. In order to improve this phase we propose a new negotiation approach which overcomes the following problems: requirements could be in conflicts, and the existing conflicts cannot be explicitly well expressed i.e. conflicts are often unknown. In this paper we detail the different phases of our approach with their models and finally we validate this approach by a case study.

Keywords: Requirements conflict; Requirements Engineering (RE); Negotiation; Cooperative Information Systems.

1. Introduction

The cooperation is a concept and an old practice which has become currently the object of active research by the integration of modern technologies of information. However, it appears normal to raise the question: why so much of references to the concept of cooperation currently? Indeed, the current economic context encourages the companies with the fusion-acquisition and the externalization of many activities. Consequently, the companies are obliged to better communicate with their customers, partners, suppliers and subsidiary companies, to produce the best products and services more quickly and at least cost. Taking into account complexity, the diversity of knowledge and competences which the projects require, the organizations adopt cooperation solutions. The importance of the cooperation increases according to the complexity of the system to develop and the number of stakeholders involved. Several studies revealed the importance of cooperation in the organizations. A study of the organization [Parametric Technology Corporation (2005)] shows that the engineers pass two thirds of their time to cooperate. Another study of the organization [Frost and Sullivan (2006)] shows that more than 36% of the performance of a company is due to its index of cooperation, which is more than twice the impact of its strategic orientation (16%) and more than five times the impact of the market and influences of technological turbulence (7%). For that, they had to interconnect their information systems (IS), to share their competences, to share resources (information, knowledge, process, etc.) and thus to constitute CIS, i.e., to gather distributed IS, possibly heterogeneous and should be coordinated to meet jointly and in a homogeneous way a common total aim.

The principal quality of CIS is to interoperate physical sites by respecting their autonomy in this perspective, the paradigm multi-agents proposes concepts particularly interesting for the development of CIS: such as the dynamic organization, autonomy of control, decentralization, the interaction which includes according to [Chaib-draa et al. (2001)] the following types: cooperation, coordination, and negotiation. Autonomous agents and multi-agent systems represent a new approach for the analysis, the design and the implementation of complex computer systems and other CIS [Zarour (2004)].

In a multi-agent system agents interact in order to perform tasks or achieve goals. The interaction takes place, usually in a common environment where agents have various zones of influence, notably various parts of the environment on which they can act. These zones can be disjoined but, in the majority of the cases, they are superimposed (the environment is shared by the agents). While interacting in a shared environment, the agents must coordinate their actions and have mechanisms for conflicts resolution.

The majority of works in the field of CIS is devoted to the two phases: design and implementation justified by the great number of software tools which supports them. However, very little attention is given to RE process whose importance is very recognized today [Zarour (2004)].

Many studies [Jorge (2007)]; [Lubares et al. (1993)]; [McGraw and Haribson (1997)]; [Nuseibeh and Easterbrook (1999)] show that the failures in the implementation and the use of computer systems are the results of a bad comprehension of the needs to which these systems try to answer. It is well known that these failures can involve high development costs, delays in the delivery of product and consequently the loss of reputation. However, the correction of errors made during requirements analysis is less costly if they are detected early in the development cycle.

RE characterizes the process which would lead to the production of a coherent set of specifications on system design. According to [Nuseibeh and Easterbrook (2000)], a RE process includes the phases of elicitation, modeling, analysis, specification, validation and management of requirements. Other authors propose different divisions of the RE process as [Kotonya and Sommerville (1998)] which highlighted all the above mentioned phases except that of modeling. These differences do not necessarily explain by an omission of phases, but can be regarded as different ways of seeing the process for example Kotonya and Sommerville which include modeling in the analysis phase.

So we propose the definitions hereafter coming from one or the other of the two visions knowing that beyond these apparent differences, they share a common ground:

- The requirements elicitation is a task which consists in gathering information concerning the needs for the users to try to understand the problem and the field of application; this requires the involvement of all stakeholders.

- The requirements analysis focuses on examining, understanding and modeling results of the elicitation phase in order to clarify the requirements, to remove inconsistencies and to ensure completeness and non-redundancy.
- The requirements specification consists in transcribing in a document all the requirements identified during the elicitation in a coherent form.
- The requirements negotiation is to explain the disagreements on the system requirements are inevitable when multiple stakeholders are involved in the process. So we cannot consider that disagreements are failures but reflect the divergences in the requirements and the dissimilarities of the origins and the responsibilities for the people involved.
- The requirements validation according to EIA-632 is focused on the verification of the final requirements document to detect the conflicts, the omissions and the deviations from standards. [EIA (1999)]
- The requirements documentation consists with documented the requirements to allow the communication and the approval of the requirements as well as the traceability of the products of other work.
- The requirements management is performed throughout the RE process. This step consists in following the evolution or the change of the requirements, making the traceability and the control of the various versions of these requirements [Wiegers (2006)] during their life cycle of requirement, i.e. during all the RE process [Fiksel and Dunkel (1991)].

Since we are talking about cooperation between enterprises in economic matter, while in computational point of view each company with it IS which was conceived in isolated way to meet the specific needs of the company. For that these systems will be incompatible and heterogeneous. These ISs should be coordinated to answer jointly and in a homogeneous way a common global objective. This generates conflicts and contradictions between the objectives of the different stakeholders.

The negotiation plays a fundamental part in the activities of cooperation by enabling people to solve conflicts which could put in danger cooperative behaviors [Chaib-draa et al. (2001)]. Then we can consider negotiation as a process of improvement of the agreements by reducing inconsistencies and the uncertainty on common points of view or action plans by the structured exchange of relevant information. This exchange is done by well-defined protocols according to the model of negotiation applied.

The conflict is difficult to define because it takes many forms and occurs in different settings. It seems that the conflict is, in essence, a disagreement, a contradiction or inconsistency. The term applies to any conflict situation where there are individuals or groups whose goals, cognitions or emotions are incompatible and lead them to oppose.

There is no universal model that can be applied to any negotiation situation. The main reason is due to the existence of a large number of parameters [Zarour (2004)]. These parameters depend on the negotiation techniques used, the language that agents use to negotiate and the types of agents used.

The main objective of our work is twofold: (i) proposing a new RE process (ii) the proposal for solutions to the problems related to negotiation in RE process in CIS.

This article is organized into six sections. In the following section we will first discuss studies and work relating to this field. Section 3 presents our proposed approach and discusses the choices motivated. Section 4 shows the prototype developed to test the proposed approach. Section 5 presents a discussion on our work. Finally, the carried out perspectives of the work are presented in the conclusion of the paper.

2. State of the Art

In this section we give a brief overview of the literature on RE processes and techniques of detection of the existing conflicts.

The RE guarantees the process of transformation of the needs expressed by the customers into system requirements that are technically exploitable. In systems that must be integrated in complex environments, the challenge is to satisfy all the needs and constraints of stakeholders directly or indirectly from the use and operation of the product. The satisfaction of needs is to meet the expectations and constraints of stakeholders involved in the life cycle of the product. It also aims to give the system acceptability and meet the expectations or the concerns of all those who are or will be affected by its objectives.

2.1. Approaches of RE process

Kotonya and Sommerville suggest a RE process that shows linear iterations between activities (fig.1). They state that the activities in the overlap model are often performed iteratively.

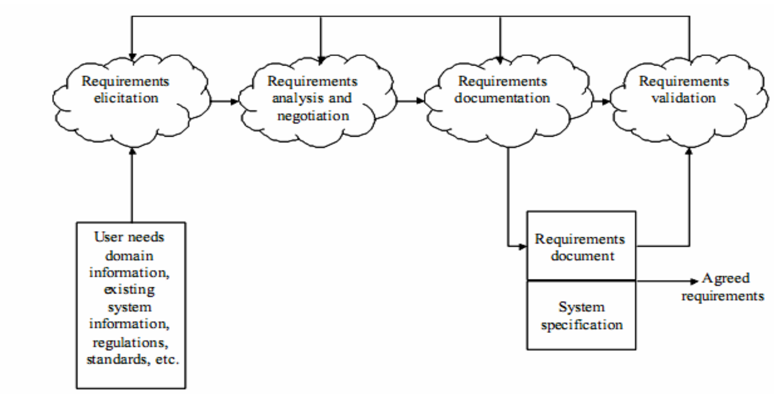


Fig. 1. RE process of Kotonya et Sommerville [Kotonya and Sommerville (1998)].

Macaulay provides a RE process purely linear (fig.2).

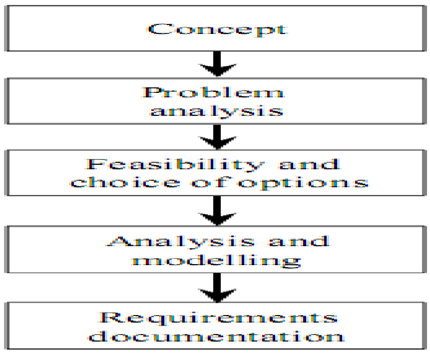


Fig. 2. RE process of Macaulay [Macaulay (1996)].

Loucopoulos and Karakostas propose a model of RE process iterative and cyclical (Fig.3). Their model shows the interactions between elicitation, specification, validation, user and problem domain.

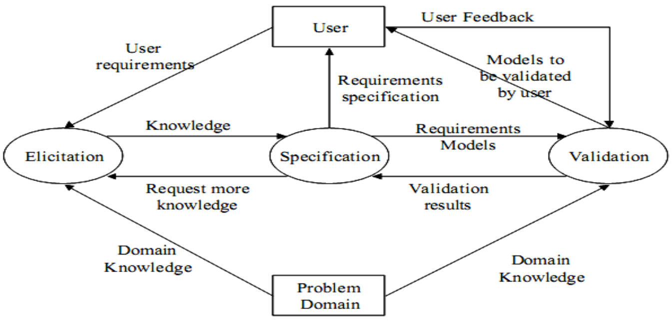


Fig.3. RE process of Loucopoulos and Karakostas [Loucopoulos and Karakostas (1995)].

The RE process proposed by Pohl (Fig.4) shows the four tasks of requirements engineering processes, their relationships as well as some potential players in the process, namely the elicitation, negotiation, specification / documentation and verification / validation of requirements. Typically, a requirement is first discovered. In a second step the various stakeholders negotiate about the requirement, agree on it or change it accordingly. The requirement is then in specification / documentation task integrated with the existing documentation and finally the validation / verification task checked if it meets the needs of user / customer or conflicts with other documented requirements.

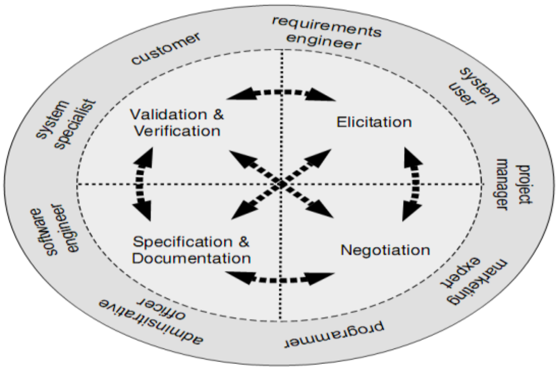


Fig.4. RE process of Pohl [Pohl (1994)].

Ian Sommerville determines that whatever the process used, some activities are fundamental to all RE processes namely elicitation, analysis, validation, negotiation, documentation and management (fig.5).

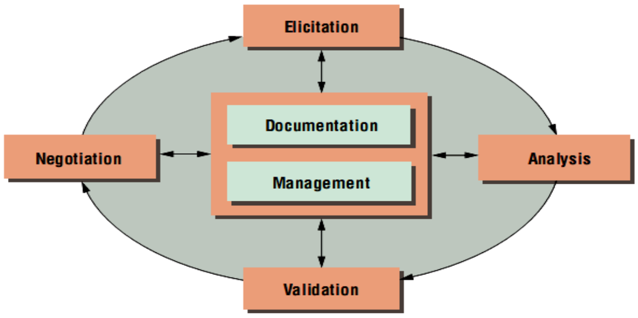


Fig.5. RE process of Ian Sommerville [Sommerville (2005)].

Wiegers presents the RE process as hierarchy (Fig.6).

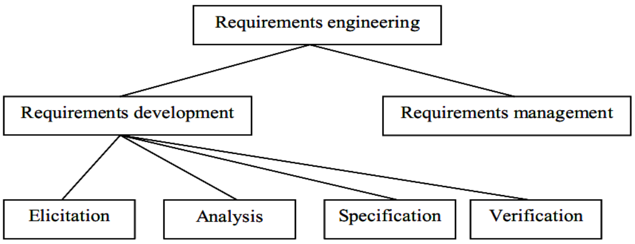


Fig.6. RE process of Wiegers [Wiegers (2000)].

Pandey and his colleagues have proposed an effective model of RE processes, which is shown in Fig.7. It mainly consists of four phases, namely: requirements development and elicitation, requirements documentation, requirements validation and verification, and requirements management and planning.

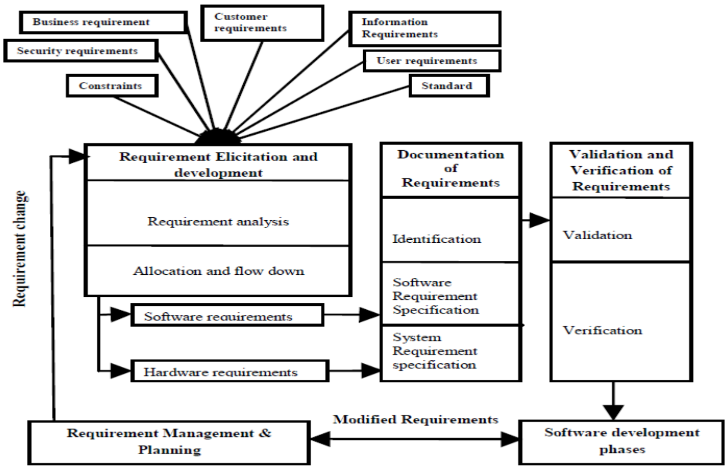


Fig.7. RE process of Pandey & al. [Pandey et al. (2010)].

2.2. Related Work

There are different techniques of conflict resolution. They can be divided by their degree of cooperation, field dependence, and automation. The degree of cooperation includes non-cooperative theories as well as cooperatives. The degree of domain dependence varies from specific knowledge for the application of the domain (depending on domain) to a common knowledge for the application of multiple domains (i.e. domain independent). The degree of automation is a relevant feature for the tool support for the different techniques of conflict resolution. These tools can be fully automated, partially automated general tool support or manual. Table1 (adapted from [Hoh and David (2004)]) shows conflict resolution skills grouped by their attributes above.

Table 1. Techniques of Conflict resolution

Techniques	**Cooperation**	**Domain dependency**	**Automation**
Zero sum game theory & non-zero sum non-cooperative game theory, bargaining theory	Non-Cooperative	Domain independent	Manual
Easterbrook's Synoptic	Cooperative	Domain dependent	Partially automated
Robertson's Oz, Sycara's PERSUADER	Cooperative	Domain dependent	Fully automated
Non-zero sum cooperative game theory, theory W, decision theory, Gilb, Rome Lab quality metrics reports, Pruitt's theory, Architecture Tradeoff Analysis Method	Cooperative	Domain independent	Manual
GIBIS, WinWin, Coordinator, Lotus Notes, Total Quality Management	Cooperative	Domain independent	General tool support
REMAP, Klein's tool, Goal-driven	Cooperative	Domain independent	Fully automated
Recently revised REMAP, SIBYL, NFRs, the MIT DICE Project, QARCC, S-COST	Cooperative	Domain independent	Partially automated

3. Proposed Approach

In this section, we present a new RE process and then the negotiation approach in this process in a cooperative context.

3.1. A new requirements engineering process

After studying a set of RE process, the process that we propose concerns all activities of the RE process because they are fundamental to it. More particularly, this process must represent the relationships between the various activities in it (Fig.8):

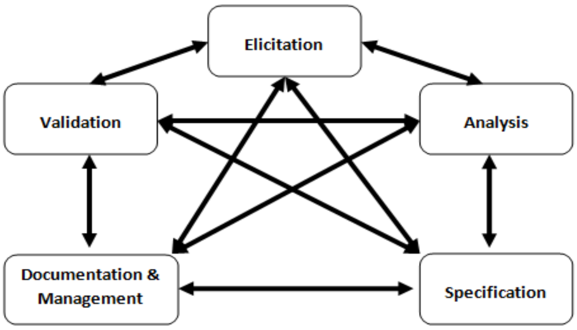


Fig.8. Proposed Requirements Engineering Process.

The selected RE process represents tasks and their relationships, i.e., the elicitation, analysis, specification, validation and management and documentation of requirements. Also we show all the potential actors in the process which constitutes the whole of the analysts of each company applied in the cooperation.

In our new RE process, the interactions between these different phases are carried out in both directions between them. Thus, each phase is connected to all phases of our process. Knowing that the order of execution of our process is as follows: elicitation, analysis, specification, validation and documentation, and management, these lasts are performed at long of this process. But after running the process hundreds or thousands times, we are forced to make backtracking either: the previous phase (analysis return to the elicitation), another phase (validation return to the analysis), in order to clarify, correct or remake again requirements. A process like this gives us a great advantage in our negotiation approach, since it runs in parallel with it. So we will not waste time to find solutions to conflict situations.

3.2. The proposed negotiation approach

Our negotiation approach consists of three main phases (Fig.9):

- Phase of conflicts detection;
- Phase of negotiation;
- Phase of evaluation and journalizing.

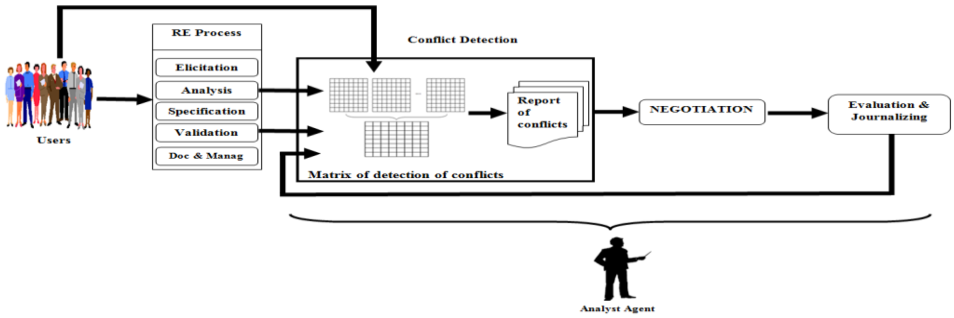


Fig.9. Negotiation Proposed Approach

3.2.1.Phase of conflicts detection

This phase consists of two steps (i) the matrix of conflict detection, (ii) the report of conflicts.

- *The matrix of conflict detection (MDC)*

The MDC is used mainly for the detection of conflicts between requirements. It aims to define the relationship between the different requirements of different users. In a very simplified view, MDC is a matrix of correlation values between the requirements. This matrix is created as follows:

(1) The rows and columns of the matrix MDC represent the requirements resulting from the steps of analysis and validation. After studying the various phases of RE, it was found that the negotiation takes place in phase: (i) Analysis (the purpose of requirements analysis is to clarify the requirements, remove inconsistencies and to ensure completeness and non-redundancy), (ii) validation (validation requirements is focused on the verification of the final requirements document to detect conflicts, omissions and deviations of standards [EIA (1999)]).

(2) Analysts are asked to complete the matrix with their collected correlations; these can be expressed in several ways, although a three-point scale ("a", "b" and "c") is used. The analyst agent sends for each analyst an empty matrix. Each analyst completes the matrix received as follows:

- ✓ It puts "a" for contradiction requirements;
- ✓ It puts "b" for fuzzy requirements (badly expressed) and;
- ✓ It puts "c" for requirements deviated from the norms.

(3) After filling all matrices, the analyst agent combines them into a global matrix MDC. Each cell of the MDC is represented by the same set of cells filled by the system analysts. Tables 2 and 3 show an example of filling an analyst matrix and an MDC involving four analysts (Ei: requirement):

Table 2. Analyst Matrix

	E1	E2	E3	E4	…
E1	b		a		
E2		c		a	
E3	a				
E4		a		b	
…					

Table 3. MDC matrix involving four analysts

	E1	E2	E3	E4	…
E1	bb..	….	aaaa	….	
E2	….	ccc.	….	aa..	
E3	aaaa	….	….	.aaa	
E4	….	aa..	.aaa	bcbb	
…					

- *The report of conflict*

After completing the MDC for each requirement, we calculate the index Rate of Conflict (RC):

$$\mathbf{RC = N/U} \tag{1}$$

Where **N** is the number of the analysts who announced that a requirement is presents a conflict. And **U** is the total number of the member agents.

Then, the agent analyst fills the conflict report. This report constitutes of two tables, one for the contradiction requirements and the other for the deviated requirements from the norms and fuzzy ones (poorly expressed).

- *Requirements that deserve to be negotiated*

For the types of requirements that deserve to be negotiated: (i) all requirements in contradiction; (ii) for fuzzy requirements or deviated with the norms we take those that have a conflict rate (RC) superior or equals to 0.5.

3.2.2. Phase of negotiation

In this phase, we will apply models of negotiation according to the type of conflict of requirements. For conflict in contradiction, we use the negotiation protocol heuristic, for conflict blurred or poorly expressed, we use the model of negotiation Contract Net Protocol (CNP) and for the conflicts deviated from the norms, we use the model of negotiation CNP with social constraints. These models of negotiation are not detailed in this paper because they are in experimental phase.

3.2.3. Phase of evaluation and journalizing

The purpose of this phase is to involve a process of traceability and evaluation of the various phases of this approach. This phase is involved in every action done and journalizes it for reuse in similar cases or when a change or modification of requirements, laws or recommendations. The causes of evolution may result from each phase of our approach. Almost, stakeholders are the sources of these changes. Traceability is made to follow the origins of each action. If a change occurs, we can locate the origin of the action. Also, we consider the evaluation to be able to control and examine the results of the negotiations carried out to improve the requirements and recommendations implied in the negotiation.

Description of agents

Like it was mentioned in the introduction, the concept agent adapts well with CIS. We show the type and the structure of the agents used in our approach. Each agent represents a company which has a competence in a particular field. This is defined in the profile of each agent, in order to help the other agents to choose the right partner during the negotiation.

The internal structure of an agent is a combination of internal components (modules), generally inaccessible to the other agents (fig.10). It defines its architecture which depends on the type of agent considered, since in our context of work the type of agent that we adopted is the cognitive agent. This representation is inspired of the one of [Briot and Demazeau (2001)]. The agent has the following mechanisms: knowledge, beliefs (represent its beliefs on the world, the other agents and itself), desires (corresponding to the goals and preferences of the agent), intentions (represent commitments of the agent and the action plans that will perform to satisfy its goals and desires), and mechanisms for planning, communication and decision-making. Moreover, since our work can be articulated on social constraints (SC) then we will add a mechanism that supports them.

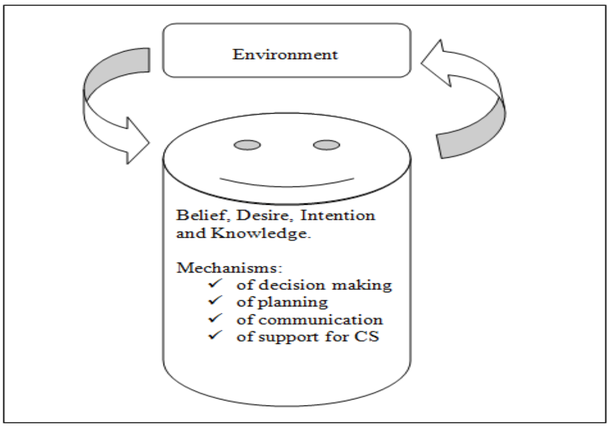


Fig.10. Internal structure of our agent.

The negotiation model that we propose uses two types of agents: agent Analyst and agent Member.

The agent Analyst: it is him which controls the system of negotiation. It has complex task to realize, it manages the other agents and the negotiation between them.

- It sends to the users the matrices to be filled, then it fills the matrix MDC;
- Then it generates the conflicts report by filling out the two tables by information which corresponds;
- Then, the agent analyst begins planning negotiation sessions between all agents member. So it uses the model appropriate to the type of conflict;
- Finally, it maintains a record of what is happening throughout the system.

The agent Member: represents any agent with which the agent analyst plans scenes of negotiation to try to eliminate the existing conflicts between requirements from the system to develop.

4. Some Aspects of Implementation

To objectively evaluate the proposed approach, we have implemented a system that allows the identification of CIS companies and these users, and filling the matrix MDC.

The system consists of two parts. The first focus on identifying companies and users of our CIS (Fig. 11 and 12), the second part is filling the matrix MDC (Fig.13). The purpose of these two parts: (i) have the necessary information for the agent profile of each company of our CIS (information from the company and the user), (ii) generate the conflicts report to know the types of conflict to be addressed in the negotiation phase.

Fig.11. Form of the company

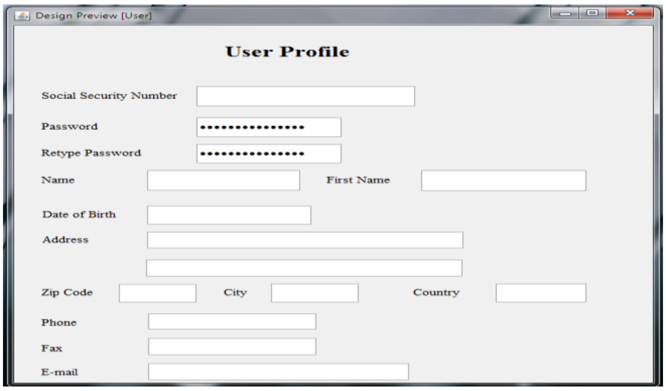


Fig.12. Form of the user

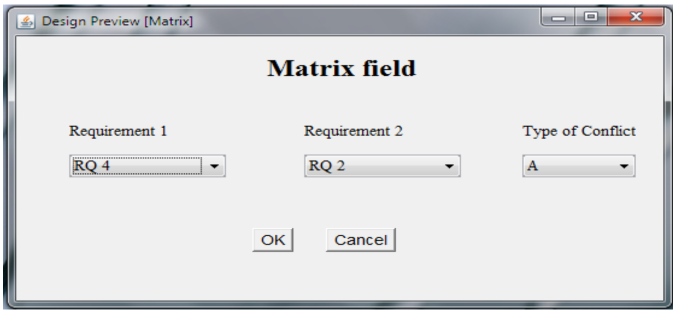


Fig.13. Filling of the matrix MDC

5. Discussion

According to these different models from of RE process that we mentioned above, we see that they have different structures, table 4 summarizes these differences

Table 4. Classification of RE process

Process of \ Characteristic	Linear	Linearly with iterations between activities	Iterative	Hierarchical	Containing all phases of RE process
Kotonya et Sommerville		+			
Macaulay	+				
Loucopoulos et Karakostas			+		
Pohl		+			
Ian Sommerville			+		
Wiegers				+	
Pandey & all			+		
Our procesus			+		**+**

The focus of our approach is cooperative, domain independent and semi-automatic, because this type of approach has:

- scalability advantages relative to approaches fully automatic;
- in addition it is more systematic of stakeholders negotiation than manual approaches or general tools support;
- and it is widely applicable to domains as approaches dependent on a specific domain.

Table 5. Techniques of Conflict resolution

Techniques	Cooperation	Domain dependency	Automation
Zero sum game theory & non-zero sum non-cooperative game theory, bargaining theory	Non-Cooperative	Domain independent	Manual
Easterbrook's Synoptic	Cooperative	Domain dependent	Partially automated
Robertson's Oz, Sycara's PERSUADER	Cooperative	Domain dependent	Fully automated
Non-zero sum cooperative game theory, theory W, decision theory, Gilb, Rome Lab quality metrics reports, Pruitt's theory, Architecture Tradeoff Analysis Method	Cooperative	Domain independent	Manual
GIBIS, WinWin, Coordinator, Lotus Notes, Total Quality Management	Cooperative	Domain independent	General tool support
REMAP, Klein's tool, Goal-driven	Cooperative	Domain independent	Fully automated
Recently revised REMAP, SIBYL, NFRs, the MIT DICE Project, QARCC, S-COST	Cooperative	Domain independent	Partially automated
Our Approach	**Cooperative**	**Domain independent**	**Partially automated**

6. Conclusions and Future Research

In this paper, we presented modestly a new RE process to especially value the relationships that exist between the activities of this process in a cooperative context. Also, we described an approach of negotiation in the proposed RE process. Negotiation provides a flexible means for finding solutions to conflict situations that may occur.

We focused mainly on this approach to show that the negotiation is transversal to the RE process and is not limited to a single phase or activity in that process as it was shown in the state of the art. We proposed for each type of conflict the model of negotiation which is appropriate to it. Consequently, conflicts are better supported and therefore we minimize the maintenance step in informatics projects.

We want to continue this work by implementing mechanisms for negotiating for the set of conflict situations. However, there is no universal model that can be applied to any situation of negotiation. This is due to the existence of a large number of parameters [Zarour (2004)]. These parameters depend on the techniques of negotiation used, the language which the agents use to negotiate, and the types of agents used.

References

Briot, J. P.; Demazeau, Y. (2001) : Principe et architecture des systèmes multi-agents. hermes.

Chaib-draa, B. ; Jarras, I. ; Moulin, B. (2001) : Système multiagents : principes généraux et application. In : Briot, J. P. et Demazeau, Y. Agent et systèmes multi agents Hermes.

Electronic Industries Alliance EIA. (1999): EIA-632: Processes for systems engineering. ANSI/EIA-632-1998.

Fiksel, J.; Dunkle M. (1991): Principles of requirement management automation. In Combined Proceedings and Workshops on the Reliability and Maintainability Computer-Aided Engineering in Concurrent Engineering, 1990 and 1991, pages 231-236, Leesburg, VA; Ellicott City, MD, USA, 9-11 Oct. 1990 and 30 Sept.-3 Oct. 1991.

Frost and Sullivan. (2006): Meetings Around the World : The Impact of Collaboration on Business Performance. Technical report, Frost and Sullivan, Verizon Business and Microsoft, http://www.frost.com.

Hoh, P. I.; David, O. (2004): Requirements Negotiation Using Multi-Criteria Preference Analysis. Journal of Universal Computer Science, vol. 10, no 4.

Jorge, G. F. (2007) : Annotation sémantique des specifications sémantiques de besoins par la méthode d'exploration contextuelle : Une contribution des méthodes linguistiques aux conceptions de logiciels. University of Paris VI SORBONNE, graduate school concepts and language.

Kotonya, G.; Sommerville, I. (1998): Requirements Engineering Processes and Techniques. John Wiley & Sons, UK.

Loucopoulos, P.; Karakostas, V. (1995): System Requirements Engineering. McGraw-hill Book Company Europe.

Lubars, M.; Potts, C.; Richer C. (1993): A review of the state of the practice in requirements modeling. Proc. IEEE Symp. Requirements Engineering. San Diego, CA.

Macaulay, L. A. (1996): Requirements Engineering. Springer-Verlag.

McGraw, K.; Harbison, K. (1997): User centred requirements, the scenario based engineering process. Lawrence Erlbaurn Associates Publishers.

Nuseibeh, B.; Easterbrook, S. (1999): The process of inconsistency management: A framework for understanding. IEEE CS Press pp. 364-368.

Nuseibeh, B.; Easterbrook, S. (2000): Requirements Engineering : A Roadmap. In Proceedings of International Conference on Software Engineering, ACM Press, Limerick, Ireland.

Pandey, D.; Suman, U.; Ramani, A. K. (2010): An Effective Requirement Engineering Process Model for Software Development and Requirements Management. International Conference on Advances in Recent Technologies in Communication and Computing.

Parametric Technnology Corporation. (2005): Conception d'un système de collaboration. Technical report, Parametric Technnology Corporation, http://www.ptc.com.

Pohl, K. (1994): The thrree dimensions of requirements engineering- A framework and its applications. Information systems, 19 (3), pp. 243-258.

Sommerville, I. (2005): Integrated requirements engineering: A tutorial. IEEE computer societt.

Wiegers, K. E. (2000): When Telepathy Won't Do: Requirements Engineering Key Practices. Originally published in Cutter IT Journal, available on: http://www.processimpact.com/articles/telepathy.html.

Wiegers, K. E. (2006): More About Software Requirements. Microsoft Press, USA.

Zarour, N. (2004) : Contribution à la modélisation de la coopération des systèmes d'informations distribués et hétérogènes: le système DAArACHE application aux entreprises de productions. PHD thesis in computer science, University of Constantine.

International Conference on Software Engineering and New Technologies

HYBRID FRAMEWORK FOR AUTOMATIC INTEGRATION OF ONTOLOGY

ARKAM MERIEM

Computer Department, Saad-Dahlab University, BP 270 Road to Soumaa,
Ouled-Aich, Blida, Algeria
arkam2406@yahoo.fr

OUKID-KHOUAS SALYHA

Computer Department, Saad-Dahlab University, BP 270 Road to Soumaa,
Ouled-Aich, Blida, Algeria
osalyha@yahoo.com

SOUAMI FERYEL

Computer Department, University of Sciences and Technology Houari Boumediene BP 32 El-Alia,
Bab Ezzouar, algiers, Algeria
souamifey@gmail.com

The heterogeneity of data sources and their diversity on the web are the main difficulties encountered by users. The data integration solutions take some research on approaches to mediation, data warehouses or semi-materialized approaches, but their main difficulty lies in the automatic interpretation of the meaning and semantics of heterogeneous data autonomous. The work presented in this paper focuses on the integration of data sources based on ontology with a hybrid or semi-materialized, this approach offers a compromise between response time and the time to update the data.

In this paper, we defined in first part, an algorithm for automatic integration of data sources based ontological (HybExtendOnto). This algorithm is used to create the data warehouse schema and the virtual (mediating) based on the following principle : All existing concept in a number greater than or equal to N number of data sources or being referenced by at least N sources will be included in the materialized, otherwise it will be integrated in the virtual. In second part, we have modeled our integration architecture based hybrid ontological, the results obtained so far are deemed satisfied.

Keywords: Data Integration, Mediator, datawarehouse, Hybrid Approach; Ontology Based DataBase, OntoDB.

1. Introduction

The need to examine different sources of information available in various autonomous information systems and makes it essential to use heterogeneous systems integration of data that allow the user a uniform view and transparent query information. These systems are very useful in the field of health because the activity of hospitals (medical and administrative) generates a large amount of data and information of various kinds more available on the Web, these data can be stored in a fully structured or not structured (eg text files) which takes us to the problems of redundancy and inconsistency of information from which the problem of integrating heterogeneous data sources and autonomous.

Several approaches for integrating heterogeneous data sources and independent have been proposed, namely multi-database systems [9], federated systems [16], mediators [20], data warehouses [19] the peer-to-peer [11] and also the semi-materialized approaches (hybrid) [14]. All these approaches aim to offer users a unique view of the different sources they want to question, however, without touching their autonomy.

We are interested in the hybrid approach or semi-materialized, this approach consists of a data warehouse and a mediator (hence its name), this approach was the use of experts in the field for the selection of data stored in the materialized and the data accessed directly from local sources which leaves the designer decides on the structure of the data warehouse and the mediator, we speak in this case a manual integration of the sources. In this case, a change in the structure of a local source or a new source of integration is too cumbersome because it usually involves the intervention of experts and an update of the overall system architecture, to solve this problem we propose an automatic integration of data sources according to a set of criteria, which makes the system more flexible and quicker to interact during the last days of local sources and adding new sources to long.

The choice of a hybrid architecture is based on the benefits it brings, one side it offers greater flexibility of data refreshing approach that fully materialized as long as it is part of the data that need to be always updated at the warehouse took less time for updates on the other hand it provides a better query optimization and a response rate faster than entirely virtual methods on the part of the data gathered in at the warehouse [14].

Many systems have been proposed integration remains their biggest challenge is the automatic interpretation of the meaning, the semantics of heterogeneous and autonomous data which represents a real challenge for this, several works on data integration based on ontologies have been reported ([2, 10, 14]), particularly in the biomedical field ([13, 18]), ontologies are used in this work to represent the global schema and / patterns or local system integration.

We are particularly interested in sources of data to Ontological basis for solving the problem of semantic interoperability among sources and the system, and between the system itself and the user, we propose an Automatic Integration of Data Base Ontology in a Hybrid Architecture. This approach will provide an introduction from a modular and fully automated platform and a hybrid automatic integration of local data sources within this architecture, on the other hand, a wider selection of relevant services to ask, and queries consistent (the results, from different sources, are systematically grouped to provide a comprehensive response to the request).

In the rest of the paper, we consider some previous work on hybrid integration approaches in Section 2, In Section 3, we present briefly the structure of ontology based databases as it was presented in the work of [8].

In section 4 we will discuss approaches to automatic integrations carried out by [12], in section 5 we give a formal-representation of the algorithms of integration data sources and in section 6 we detail our algorithm HybExtendOnto integration and we will discuss our proposed architecture and its different levels in terms of Section 7, and finally we finish with a conclusion and perspectives.

2. Related works

Several projects have been developed by adopting the approach semi-materialized, there is the project Squirrel [7] which uses a hybrid approach to develop a general and flexible framework mediator, but focuses only on data sources, relational or object-oriented. A framework for storage of web content has also been discussed in [15], it uses a hybrid approach in order to integrate data from the data warehouse with "necessary" information on the Internet. It uses ontologies to express domain knowledge related to web sources and the logical design of data warehouse. In addition, an ontology engine is deployed as an intermediate layer by defining mapping rules between web data and attributes of the data warehouse in ontologies using the structure of the data warehouse and requirements of the repair, also some data on the Web are selected for materialization, however, some requests can’t be answered using only materialized in the data warehouse.

In PICSET and HYLEME [14], a framework for integrating ontology was introduced on the basis of the approach entirely virtual. They build integrated "virtual" views according to the approaches of mapping between local ontologies and the global ontology. The sense here is that concept mapping one ontology corresponds to a search on other ontologies. Then, an application will be placed on the ontology "virtual" global and based on the mapping between the concepts of global ontology and local ontologies, the query is unfolded and the answers retrieved from the sources.

The proposal [10] extends the architecture of ontology integration using DL-Lite to express the global schema, and using the approach LAV mapping between databases "local" and the schema of the ontology overall. There is a remarkable difference between our work and theirs.

Alasoud [1] proposed a hybrid architecture to integrate heterogeneous data in order to reduce response times, some frequently requested data are materialized: some queries should be answered only from the materialized views, others have be answered directly from the sources, for the work of [3], a hybrid approach to improve performance of applications on systems integration of heterogeneous and autonomous data has been proposed, the recommended method to realize the domains of definition attributes Digital and listed in mediator. These annotations allow the mediator to select sources involved in the application and to avoid that the request be sent to all sources and improve the quality and response time.

Our architecture is based much more on that proposed in [1], except that it is based on the following principle: in the data warehouse of our architecture, only the existing ontological concepts that are referenced will be in data warehouse and data deriving from these concepts, the rest of the concepts will be integrated in the mediator, this integration is based on a fully automatic algorithm.

3. Description of Ontology Based Database (OBDB)

Displayed equations should be numbered consecutively, with the number set flush right and enclosed in parentheses. The equation numbers should be consecutive within the contribution

In our approach the data is structured like the fact that attention to databases. The databases are organized by their conceptual model, we focus more specifically to ontology-based databases, a concept developed within the LISI in the project OntoDB.

A database ontology based consists of a database that can store both:

- The usual content of a database
- The ontology that describes the significance of these data

Unlike conventional databases, consisting mainly of two parts (data and meta database). Databases based ontological formed of four parts, thus it takes its name architecture 4-quart:

- Meta Data Base: whole system tables describing the structure of data tables
- Content: The data or instances of the conceptual model
- Ontology: dictionary containing the meaning (semantics) of data
- Meta Schema: structure describing the model of the ontology used. It is for this ontology is the basis for meta-data.

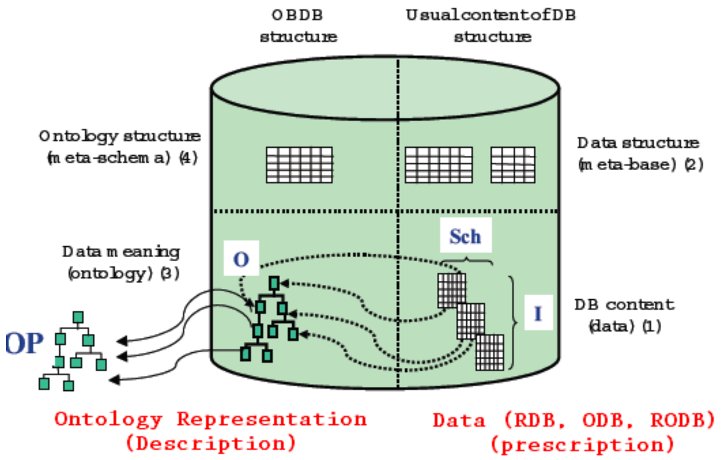


Fig. 1. Architecture of OBDB : OntoDB [8].

4. Approaches for automatic integration of OBDB

The integration of data sources based ontology has been the subject of the thesis Nugyen Xuan Dung [12] via the so-called warehouse materialized approach was adopted for the mediator approach, it has been addressed in memory of Amel Boussi [5]

We propose a merger of these two approaches to make the most of the advantages in both approaches; our approach is called Hybrid Approach for Integrating Ontology Based Databases.

In the work of [12] each data source has its local schema through its local ontology (source ontology based). The overall pattern is represented by the shared ontology (domain). Local ontologies are referenced as much as possible the shared ontology [4].Mappings between the two schemes can be achieved by integration scenarios.

We studied the following integration scenarios and through them, we adapted the process of fragmentation in our approach to the integration of ontologies a priori.

(1) FragOnto: Each local source is defined as a fragment (which may be horizontal or vertical mixed) of the shared ontology (a horizontal piece is the result of a selection operation applied to the shared ontology, a fragment resulting from a vertical operation projection applied to the shared ontology, a fragment is a mixed combination of horizontal and vertical fragmentation).

(2) ProjOnto: each source defines its own ontology (it does not instantiate any class of the shared ontology). Cons by the local ontology reference ontology shared by respecting the condition Smallest Class Reference Subsuming [4]. In this scenario, it would nevertheless incorporate instances of each source as instances of the shared ontology;

(3) ExtendOnto: each local ontology is defined as in the scenario ProjOnto, but you want to automatically enhance the shared ontology. Then all instances of data are integrated, without modification, within the integrated system. By following the proposed architecture and the integration algorithm we have modeled our architecture since the work of [12].

5. Formal representation for automatic integration of OBDB

An ontology O (local or global) is formally defined as a tuple: O: <C, P, Sub, Applic>, with:

- C: the set of classes used to describe concepts in a given field.
- P: the set of properties used to describe instances of all classes C.
- Sub: is the relation of subsumption (is-a and is-of-box), which in each class Ci of ontology, combines its direct subsumed classes.
- Applic: associates with each class of the ontology properties that are applicable for each instance of this class.

A Ontology Based Database (OBDA) is formally defined as a quadruplet: OBDB: <O, I, Pop, Sch>, with:

- O: Represents his ontology (O :<C, P, Sub, Applic>.
- I : is the set of data instances of the database. The semantics of these bodies is described by O.
- Pop: associates with each class instances that belong to him (directly or through classes subsumes it). So pop is the population of Ci.

- Sch : associates with each class Ci properties that are applicable to this class and are actually used to describe some or all instances of Pop (Ci).

Intergation Approach based on Ontology

Are OG :< C, P, Sub, Applic> global ontology (also called shared ontology) as domain ontology, and $S = \{S_1, ... , S_n\}$ be a set of OBDBs participating in the integration process. In this case we integrate the ontologies then the data.

The integration of ontologies is performed and expressed using subsumption relations between classes of different ontologies and import properties. We rely on the following assumptions:

(1) If each source has its own local ontology: $O_{Li} <C_i, P_i, Sub_i, applicC_i>$ in the global ontology referenced O_G. These references are stored in the source Si and Mi is an articulation between O_G and O_{Li}.

(2) Each class has its schema Sch_i C_i (C_i), and its instances $Pop_i(C_i)$. A source with a hinged If OG global ontology can be formulated as follows: Si: $<O_{Li}, I_i, Sch_i, Pop_i, M_i>$.

About classes, we also rely on the following assumptions:

(1) each class of each reference if his (or her) smaller classes subsuming in the global ontology.

(2) if each class of each of its important classes subsumed in the global ontology the properties it wishes to use, preserving their identifications

(3) If all classes and all the properties of S_i that are not identified with imported properties are specific to the source S_i.

6. Proposed Approach

In our architecture we keep the same formalism seen in [4, 8, 12].

The platform is divided into two parts, the virtual and materialized component but which are visible to the user by one and only one single interface.

We rely on an automatic integration of a priori ontology of all sources of data to Ontological Base by the following criteria: any shared concept, existing in a number greater than or equal to n number of data sources or being referenced by at least n sources will be included in the materialized, otherwise it will be integrated in the virtual part of the platform.

In the algorithm which proposed integration of ontology then data, we took as an example n = 2, however, the algorithm is easily adapted for any number fixed by a law of statistics or by the DBA.

6.1. *Proposed Integration Algorithm: HybExtendOnto*

In our approach, we adopted the algorithm ExtendOnto applied to the projection and fragmentation (Hybrid) so that hybrid integration is adopted by Intersection for the part and materialized by union Hybrid Integration least for part intersections virtual integrated system.

As in the algorithm of [12], we assume that each source S_i has its own ontology O_i and his specific classes. However the ontology O_i can reference whenever possible global ontology O_G as it may include fragments of the global ontology O_G. Sources can also have their own concepts and not found in the global ontology.
As ExtendOnto Algorithm, our system can integrate even extensions of each S_i. The Global Ontology will be extended by integrating local ontologies.
In our case we have two systems Materialized System Int_{DW} (ontology based data warehouse which look like a big OBDB only) and virtual system (ontology based mediator) Int_{MED}.
To do this we must find the final structure of integrated systems (Int_{DW} et Int_{MED}) : $\langle O_{int_{DW}}, Sch_{int_{DW}}, Pop_{int_{DW}} \rangle$ and $\langle O_{int_{MED}}, Sch_{int_{MED}} \rangle$
Initially we have :

$$O_{int_{DW}} = O_{int_{MED}} = O_G \quad (1)$$

$$Sch_{int_{DW}} Sch_{int_{MED}} = \varnothing \quad (2)$$

For each concept $C_G \in O_G$, we must we must check whether it belongs to at least one source S_i, if it is checked two cases arise:

1st case : $C_G \in O_G \;\&\; \exists S_i \neq S_j / C_G \in \left(S_i \cap S_j\right)$ then :

$$Sch_{int_{DW}} = Sch_{int_{DW}} \cup \left(\bigcup_{i=1..n} Sch_i\left(C_G\right) \right) \quad (3)$$

$Pop_{int_{DW}}$ represents the population of each concept of the integrated system Int_{DW} it is defined as follows:

$$Pop_{int_{DW}}(C) = i * Pop_i(C) \quad (4)$$

2nd case:
$C_G \in O_G, \forall S_i \neq S_j / C_G \in \left(S_i \cup S_j\right) \wedge C_G \notin \left(S_i \cap S_j\right)$ then:

$$Sch_{int_{MED}} = Sch_{int_{MED}} \cup Sch_i\left(C_G\right) \quad (5)$$

For each concept $C_p \in O_p$ not included in one source but is referenced by links [4], by at least one class $C_i \in S_j$ and $C_i \notin O_p$,
1st case : there are at least two classes C_i and C_j a belonging to two different but that reference both the same class C_G of Global Ontology O_G by articulation:

$C_G \in O_G \,\&\, C_i \in S_r \,\&\, C_j \in S_s \,/\, A_{C_i,C_G} < S_r, O_G, OntoSub_{i,G} > \wedge A_{C_jC_G} < S_s, O_G, OntoSub_{j,G} >$
then:
We must first expand the ontology of $Int_{DW}(O_{\mathrm{int}_{DW}})$ as follows:

- $C_{Int_{DW}} = C_{Int_{DW}} \cup C_i \cup C_j$
- $P_{Int_{DW}} = P_{Int_{DW}} \cup P_i \cup P_j$
- $Applic_{Int_{DW}}(C) = Applic_i(C)$
- $Sub_{Int_{DW}}(C) = Sub_i(C)$

For the population $Pop_{Int_{DW}}$ of each class C outside of $\mathrm{O_G}$, it is given by the following equation:

$$Pop_{\mathrm{int}_{DW}}(C) = Pop_i(C) \tag{6}$$

Finally the schema of each class of integrated system $\mathrm{Int_{DW}}$ is calculated as follows:

$$Sch_{\mathrm{int}DW} = Sch_i(C) \tag{7}$$

2nd case : there is a single class $\mathrm{C_i}$ among all classes of all data sources S_i referenced the class $\mathrm{C_G}$ of Global Ontology $\mathrm{O_G}$ by articulation:
$C_G \in O_G \,\&\, \exists! C_i \,/\, A_{C_i,C_G} < S_r, O_G, OntoSub_{i,G} >$ Then :
We must start to enrich the ontology of $Int_{MED}(O_{\mathrm{int}_{MED}})$ as the precedent case:

- $C_{Int_{MED}} = C_{Int_{MED}} \cup C_i$
- $P_{Int_{MED}} = P_{Int_{MED}} \cup P_i$
- $Applic_{Int_{MED}}(C) = Applic_i(C)$
- $Sub_{Int_{MED}}(C) = Sub_i(C)$

The schemes of each class of integrate Int_{MED} is given as follows:

$$Sch_{\mathrm{int}MED} = Sch_i(C) \tag{8}$$

6.2. *Framework of our Approach*

The approach taken in creating the ontology is the hybrid approach, using the method defined by [6] is made in three steps: building local ontologies (each database must be provided prior to its ontology or build) and build the common vocabulary and finally defining the mapping between the global and schematic diagrams premises.

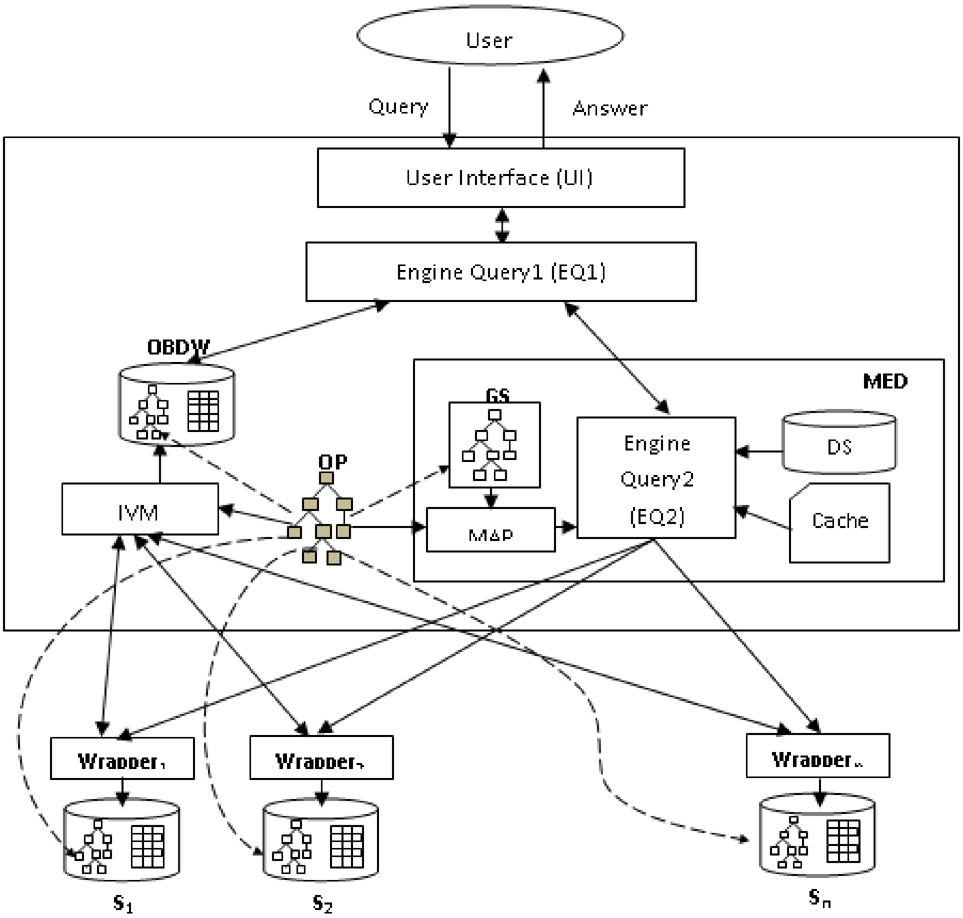


Fig. 2. Ontology Based Hybrid Framework of Integration of OBDB sources.

As shown in Fig. 2, our software architecture consists of:
A user interface for receiving user requests and issue responses
Two-engine queries (EQ1 and EQ2), EQ1 can filter the request sent by the user at the beginning is to say whether the request will be processed by the data warehouse (result fully recoverable since materialized views) or by the mediator (results obtained from local data sources), or by the two modules at a time that the query id (or subqueries) will be sent to two modules and the result will be rewritten in terms of EQ1 before issue, even for EQ2, its task is to decompose the queries into sub queries, locate sources of data relevancy, rewrite and rewrite the queries and then build the answer.

A mediator with the MED EQ2 Query Engine, a comprehensive scheme of sources (SG), a relational database for source description and possibly a DS cache

A global schema (GS): Represents a domain ontology (which references the shared ontology) Scheme provides a comprehensive overview on data sources or a portion of the data sources in addition, it provides a unique vocabulary for expressing user queries and content description of the sources which allows to introduce and extract the source data, the user poses queries in terms of its relations in the global schema.

The description of the sources (DS): Contains a set of abstract views on the sources.

A mapping module (MAP) to do the mapping (correspondence) between the global schema and local schemas based on the shared ontology, the mapping method used in our architecture is that Glav s' proves most interesting.

Adapters (wrappers) that translate the subqueries issued by the Ombudsman in the language specific data sources, the results will then be forwarded to the Ombudsman who deals with the present.

An Ontology Based Data Warehouse, it also contains the data actionable and relevant (most requested), the schema and local ontology describing the data warehouse, its contribution to our architecture is the gain response time.

An integrator of materialized views (IVM) for the integration of data embodied in the data warehouse (for updating and maintaining the data warehouse).

A shared ontology modeling the application domain and represents a common vocabulary for all local ontologies (including the ontology of data warehouse and the overall pattern of mediator), which facilitates the integration of sources data and solve the problem hétérogéiniété semantics.

Data Sources (S_i): sources of data are represented as well as databases based on ontology, that is to say, besides the data, we find the model design, ontology and metadata model of ontology, data sources are designed after the model OntoDB (see section 3)

7. Conclusion

In our paper, we first proposed an algorithm for integrating data sources based on ontology "HybExtendOnto" based on a shared ontology or known global ontology, each source is provided with its own local ontology which is an ontology of application-specific source, however, it refers to the best possible global ontology, our algorithm is inspired by the work of Mr. Xuan Dung Nugyen [12], then we have defined our hybrid architecture, however it remains to be validated by any experimental section, therefore, we are in phase to test our architecture we have implemented in simulation using the following tools: System Management and Oracle databases it was able to resume more bases medical data collected from several Algerian hospital, they were restructured format OntoDB, we implemented our algorithm using SQL and PL was used PHP scripting language for user interfaces.

We intend to subsequently validate our architecture for real tests on OBDB located at different sites.

References

[1] Alasoud, A.; Haarslev, V.; Shiri N. (2005). *A Hybrid Approach for Ontology Integration*, Proceedings of the 31st VLDB Conference, Trondheim, Norway.

[2] Arens, Y.; Hsu, C-N.; Knoblock, C-A, (1997). *Query processing inf the SIMS information mediator*, in HUHLS M-N et SIN§GH M-P, éditeurs : Reading in agents, Morgan Kaufmann, San Francisco CA, USA.

[3] Bakhtouchi, A.; Bellatreche, L; Balla, A. (7-8 december 2009). *Materializing Attributes Annotation: A hybrid approach for databases integration*. 1st doctoriale "STIC2009", Algeria-Msilla.

[4] Bellatrache. L. ; Pierra. G. ; Ngueyn Yuan. D. (2004). *An a Priori Approach for Automatic Integration of Heterogeneous and Autonomous Databases*, Proc. of Database and Expert Systems Applications - DEXA '04

[5] Boussi, A. (2007). *Intégration de sources de données à base ontologique dans un environnement p2p*, Mémoire pour l'obtention de diplôme de magister, Ecole Supérieur en Informatique (ex INI), Algeria.

[6] Buccella. A.; Cechich. A. (2003). *An Ontology Approach to Data Integration*, Journal of computer science and Technology - Citeseer.

[7] Cherat, L.; Ezziyyani, M.; Essaadi, M.; Bennouna, M. (2008). *Hybrid integration Approach for Heterogenous Information Sources : Applied to Medical Resources and Patients Medical Documents*, in New Technologies, Mobility and Security.

[8] Hondjack Dehainsala, H. (2004). *Explicitation de la sémantique dans les bases de données : Base de données à base ontologique et le modèle OntoDB*, PhD Thesis, Poitiers University, France.

[9] Litwin, W.; Abdellatif, A.; Zeroual, A.; Nicolas, B.; Vigier, Ph. (1989). *MSQL: A Multidatabase Language*.59-101.

[10] Mena, E.; Illarrmedi, A.; Kashyap, V.; Sheth; A-P. (2000). *OBSERVER : an approach for query processing in global information systems based on interoperation across pre-existing*.

[11] Nejal, W.; Wolf, B.; Qu, C.; Decker, S.; Sintek, M.; Naever, A.; Nilsson, M.; Palmer, M.; Risch, T. (2002). *Edutella : A P2P networking infrastructure based on rdf*, in Proceeding of the 11th International Conference on World Wide Web.

[12] Nguyen Yuan, D. (2006). *Intégration de Bases de Données Hétérogènes par articulation à prioi d'ontologie : Application aux catalogues de composants industriels*, PhD Thesis, Poitiers University, France.

[13] Perez-Rey, D.; Maojo, V.; Garcia-Remesal, M.; Alonso-Calvo, R.; Billhardt, H.; Martin-Sanchez, F.; Sousa, A. (2006). *ONTOFUSION : Ontology-based integration of genomic ans clinical databases*, Computers in Biology and Medecine.

[14] Rousset, M-C.; Reymaud, C. (2003). *PICSET and HYLEME : Two illustrative information integration agents, in Intelligent Information Agents – The AgentLink Perspective*, Springer.

[15] Sais, F. (2007). *Intégration sémantique de données guidée par une ontologie*, PhD thesis, Paris-Sud University, France.

[16] Sheth, A-P.; Larson, J-A. (1990). *Federated database systems for managing distributed, heterogeneous, and autonomous databases*. ACM Comput Surv.

[17] Shethamit, P.; Gala Sumit, K.; Navathe Shamkaut, B. (1993). *On automatic reasoning for schema integration*. International Journal of Intelligent and Cooperative Information Systems

[18] Stevens, R.; Baker, P-G.; Bechhofer, S.; NG, G.; Jacoby, A.; Faton, N-W.; Goble, C-A; Brass, A.; TAMBIS. (2000). *Transparent Access to Multiple Bioinformatics Information Sources. Bioinformatics*.

[19] Widom, J. (1995), *Research problems in data warehousing*, in Conference on Information and Knowledge Management.

[20] Wiederhold, G. (1992). *Mediators in the architecture of future information systems*.

International Conference on Software Engineering and New Technologies

GRAPH BASED APPROACH FOR ONTOLOGY MATCHING

TOUATI CHAHIRA

Laboratoire LITIO, Université d'Oran,
BP 1524, El-M'Naouer, 31000 Oran, Algérie
chahira40@yahoo.fr

BENAISSA MOUSSA

Laboratoire LITIO, Université d'Oran,
BP 1524, El-M'Naouer, 31000 Oran, Algérie .
moussabenaissa@yahoo.fr

LEBBAH YAHIA

Laboratoire LITIO, Université d'Oran,
BP 1524, El-M'Naouer, 31000 Oran, Algérie
ylebbah@yahoo.fr

Ontology matching is a key interoperability enabler for the Semantic Web since it takes the ontologies as input and determines as output correspondences between the semantically related entities of those ontologies. We present in this paper a graph based approach to tackle the ontology matching problem. The objective is to address the combinatorial aspects related to this issue. More precisely, our approach[a] consists in modeling the problem of extracting an alignment (matching) which satisfies cardinality constraints, as minimizing some cost on feasible flow problem defined on a bipartite graph. The found solution represents the best alignment which maximizes the global similarity between the entities of the two ontologies.

Keywords: Ontology alignment; ontology matching; graph based approach; feasible flow; minimum cost flow; global Similarity.

1. Introduction

Ontologies are at the present time in the middle of the work undertaken in the semantic Web. Aiming at establishing representations through which the machines can handle the semantics of information, the construction of ontologies requires at the same time a study of human knowledge and the definition of languages of representation, as well as the development of systems to handle them.

An ontology allows describing a domain by defining vocabulary and axioms that govern it. In this context, an ontology defines a set of concepts and their relationships with other concepts by specialization or through properties. Ontologies are often different because they were conceived with a view to achieve various goals and describe

[a] This work is supported by TASSILI research program 11MDU839 (France, Algeria)

sometimes numerous fields. Nevertheless, it is not rare to state the existence of common information between these ontologies [Ghomari(2009)]. For example, given two ontologies, the same concept can be indicated via several terms, or the case of two ontologies which express the same knowledge by using different specification (see the Fig. 1).

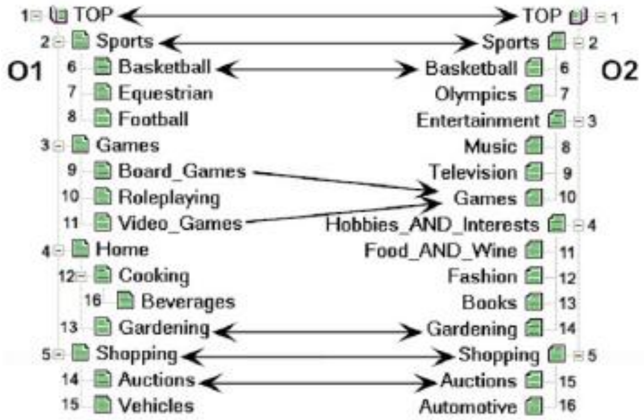


Fig. 1. matching between two ontologies.

Moreover users of ontologies do not use only their own ontologies, they must often integrate or adapt other ontologies to solve their problems. Unfortunately it is very difficult to use in a simultaneous way these ontologies for a new application. This problem of heterogeneity between knowledge expressed within each of them must be resolved. Therefore, semantic links between entities belonging to two ontologies different must be established, which is the purpose of the ontology alignment [Euzenat and Shvaiko(2007)].

Given two ontologies, the alignment produced a set of matches each linking two entities (eg concepts, instances, properties, terms, etc..) by a relation (equivalence, subsumption, incompatibility, etc..), which may include a degree of confidence. All correspondences, also called alignment, can be used to merge ontologies, migrate data between ontologies or translating queries formulated in terms of an ontology to another.

The idea of this paper consists in exploiting the graph theory and in particular the minimum cost flow algorithm to solve the problem of identifying an ontology alignment which satisfies the cardinality constraints and has a maximum global similarity. It is necessary to note on this level that the work presented here approaches the alignment extraction problem which has formal properties. On the other hand, it does not approach the aspect of similarities calculation which is supposed to be calculated in addition.

The paper is organized as follows: we begin by giving some related work and then in section 3 we introduce an illustrating example to present the overall approach. In section 4 we present some preliminary notions of ontology alignment necessary to understand the next sections. We present successively the notions of correspondence, alignment,

alignment properties, similarity measures and alignment extraction methods. Then, in section 5, we present our approach which consists of a flow model of the alignment extraction problem which satisfies the cardinality constraints and has a maximum global similarity. In this section we describe the minimum cost flow algorithm upon which our contribution is based. Then in section 6, we present our contribution which consists in the construction of network on which we apply the minimum cost algorithm to select the alignment with the necessary properties. Finally we present our experimental results and conclude the paper.

2. Related Work

There has been important background work that can be used for ontology alignment: in discrete mathematics for matching graphs and trees [Karp(1980)], [Papadimitriou and Steiglitz(1998)], in databases for reconciling and merging schemas [Rahm and Bernstein(2001)], in machine learning for clustering compound objects described in a restricted first order logic [Bisson(1992)].

In this section, we will look at related systems with a special focus on the topics of extraction methods:

Globally we can distinguish two approaches for handling the alignment extraction problem.

(1) Interactive approach: The user is involved in the alignment extraction process. One Way to implement this approach consists of displaying all entity pairs with their confidence measures and those judged the most relevant by the user are selected. This approach seems more relevant than the automatic one especially in traditional applications where large data sets are handled [Shvaiko and Euzenat(2008)]. In this case we can quote [Benaissa and Lebbah(2011)].

(2) Automatic approach: which our approach is based. Correspondences between entities are extracted automatically without the user intervention. Within this approach, various methods have been proposed in literature. These methods depend heavily on the properties of the target alignment. In this case we can quote two works:

   (i) The principle of this method [Meilicke and Stuckenschmidt(2007)] is based on the idea that it is possible to infer logical constraints by comparing subsumption relations between concepts of the ontologies to be matched. A standard algorithm to solve the problem of extracting correspondences is known as the Hungarian method [Kuhn(1955)]. In order to show how this method can be applied to this problem a few explanations have to be given. The Hungarian method expects a real-valued matrix as input and creates a one to one assignment, such that the sum of the chosen entries is minimal. To use the Hungarian method the input mapping M’ has to be transformed into a corresponding matrix H. Each concept of the source ontology corresponds to a row and each target concept corresponds to a column. Since the Hungarian method finds a minimal assignment an entry in the matrix has to be interpreted as distance between two concepts, where the distance between C_1 and C_2 is defined as 1−similarity. If there exists no such correspondence in M' the distance is set to ∞. In most matching situations it will not be possible to match all or even the majority concepts. Matching candidates will thus not be available. Therefore, the input matrix has to be extended by additional concepts that play the role of alternative matching candidates. We call these concepts phantom concepts. Thus, if n is the number of concepts in O_1 and m is the number of concepts in O_2, we add m rows to the input matrix corresponding to m phantom concepts and n columns corresponding to n phantom concepts. The value of the entries in these rows respectively columns is set to $1 + \varepsilon$ with $\varepsilon < 0$.

(ii) In this work [Cruz et *al.*(2009)], they provide an efficient solution to this problem by reducing it to the maximum weight matching in a bipartite graph and by adopting the Shortest Augmenting Path algorithm (SAP) [Karp(1980)]. They provide an alternative solution to this problem by reducing it to the maximum weight matching in the bipartite graph $G = (S \cup T, E)$, where S contains the source ontology concepts, T contains the target ontology concepts, and E contains an edge oriented from S to T for each correspondence with a similarity value higher than the threshold, weighted with the threshold value itself. They recall that a maximum weight matching M is a subset of the edges in E such that for each vertex in G at most one adjacent edge is contained in M and the sum of the weights (i.e., the similarity values) of the selected edges is maximized. Thanks to this transformation, they can adopt the Shortest Augmenting Path algorithm (SAP) to find the optimal solution in polynomial time.

Many diverse solutions of matching have been proposed so far, see [Shvaiko and Euzenat], [Doan and Halevy(2005)], [Noy(2004)], [Kalfoglou and Schorlemmer(2003)] and [Rahm and Bernstein(2001)]. But in our contribution we provide the best alignment between two ontologies which maximizes the global similarity and verifies the cardinality constraints.

3. Illustrative Example

Let O and O' be two ontologies. O contains the concepts $\{C_1,C_2,C_3,C_4\}$ and O' contains the concepts $\{C'_1, C'_2, C'_3, C'_4, C'_5, C'_6\}$. We assume that some techniques computing similarities between concepts of O and O' has produced the following similarity matrix S (Table1).We assume in this example that the cardinality constraints of the source ontology O is equal to 3 and the cardinality constraints of the target ontology O' equal to 2.

We require again that all similarity values must be greater or equal to a threshold s = 0.5. Filtering S over the above threshold gives the following matrix (Table2).

Table 1. Similarity matrix

O O'	C'_1 : 2	C'_2: 2	C'_3 : 2	C'_4 : 2	C'_5 : 2	C'_6 : 2
C_1 : 3	0.81	0.61	0.73	0.61	0.50	0.44
C_2 : 3	0,92	0.83	0.39	0.52	0.84	0.12
C_3 : 3	0.64	0.62	0.26	0.74	0.94	0.31
C_4 : 3	0.23	0.96	0.32	0.25	0.60	0.82

Table 2. Similarity matrix after filtering

O O'	C'_1 : 2	C'_2: 2	C'_3 : 2	C'_4 : 2	C'_5 : 2	C'_6 : 2
C_1 : 3	0.81	0.61	0.73	0.61	0.50	
C_2 : 3	0,92	0.83		0.52	0.84	
C_3 : 3	0.64	0.62		0.74	0.94	
C_4 : 3		0.96			0.60	0.82

The problem here is to extract automatically an alignment with the maximum global similarity between concepts of O and O' which satisfy the cardinality constraints. To handle efficiently this problem we propose an algorithm based on graph theory.

The algorithm proposed in our approach is the minimum cost flow algorithm. We give hereafter the flow graph for the above example Fig 2.

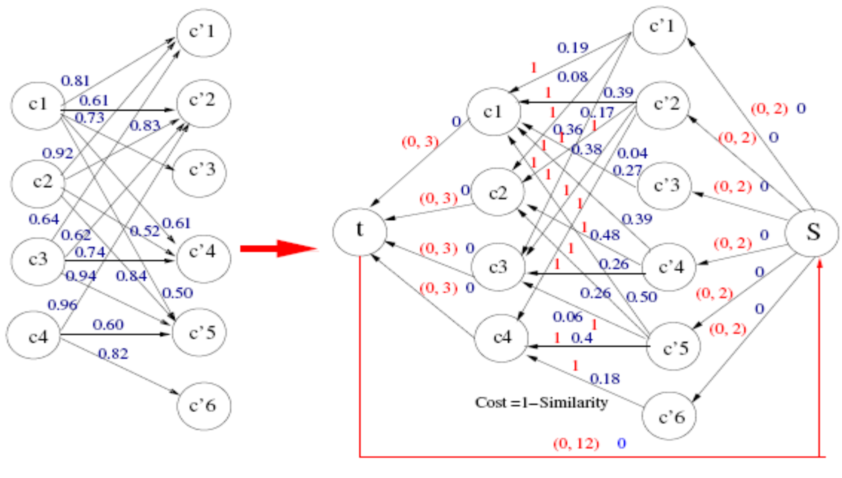


Fig. 2. Network flow of the example

We can find several solutions where the flow is max, for example {(C1,C'3),(C1,C'4), (C1,C'5), (C2,C'1), (C2,C'2), (C2,C'4),(C3,C'1), (C3,C'2), (C4,C'5), (C4,C'6)} with global similarity equal to 6.79. But after the application of the minimum cost flow algorithm, we obtain the following alignment: {(C1,C'1), (C1,C'3), (C1,C'4), (C2,C'1), (C2,C'2), (C2,C'5), (C3,C'4), (C3,C'5), (C4,C'2), (C4,C'6)} with global similarity equal to 8.2. Here is the contribution of this paper: providing the best alignment between two ontologies which maximizes the global similarity and verifies the cardinality constraints.

4. Ontology Alignment Problem

The process of alignment between ontologies aims to identify semantic correspondences between their entities. In this section we give some definitions on the key materials of this work [Euzenat and Shvaiko(2007)].

4.1. *Correspondence notion*

Let O and O' be two ontologies. A correspondence M between O and O' is a quintuple ⟨id, e, e', R, n⟩ where:

- id is a unique identifier for the correspondence M;
- e and e' are entities of O and O' respectively (e.g., concepts, roles or instances);
- R is a semantic relation holding between e and e' (for example, equivalence, more general, more specific, disjointness);
- n is a confidence measure (typically a value in [0,1]).

4.2. *Alignment notion*

The alignment can be defined as a set of correspondences. The alignment process has two ontologies O and O' as input and produces an alignment A between entities of O and O' as output (See Fig. 3). Other parameters can complete this definition, namely:

(1) A preliminary alignment A' to be completed or refined by the process.
(2) External resources r such as a thesaurus for example.
(3) Parameters p such as thresholds or weights for example.

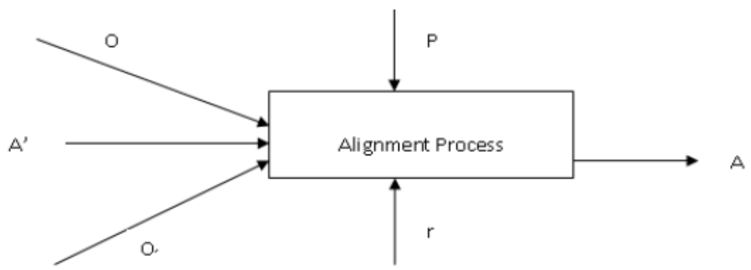


Fig. 3. Alignment process

4.3. *Alignment properties*

Total alignment: An alignment A is said to be total (or complete) from O to O' if and only if each entity of O has a corresponding entity in O'.

Injective alignment: An alignment A is said to be injective from O to O' if and only if for each correspondences C = <id_C, e_1, e_2, n, R〉 Є A and C' =〈id_C, e'_1, e'_2, n, R〉 Є A we must have: if $e_2 = e'_2$ then $e_1 = e'_1$.

4.4. *Similarity measures*

There are many different ways to compute the distance between entities of two ontologies (which can be as reduced as an equality predicate) and computing the best match between ontologies, i.e., the one that minimizes the total distance (or maximizes a similarity measure). Roughly speaking, they can be classified as (this complements the taxonomy provided in [Rahm and Bernstein(2001)] and only considers features found in actual systems):

- Terminological (T) comparing the labels of the entities; string based (TS) does the terminological matching through string structure dissimilarity (e.g., edit distance); terminological with lexicons (TL) does the terminological matching modulo the relationships found in a lexicon (i.e., considering synonyms as equivalent and hyponyms as subsumed);
- Internal structure comparison (I) comparing the internal structure of entities (e.g., the value range or cardinality of their attributes);
- External structure comparison (S) comparing the relations of the entities with other entities; taxonomical structure (ST) comparing the position of the entities within a taxonomy; external structure comparison with cycles (SC) an external structure comparison robust to cycles;
- Extensional comparison (E) comparing the known extension of entities, i.e. the set of other entities that are attached to them (in general instances of classes);
- Semantic comparison (M) comparing the interpretations (or more exactly the models) of the entities.

4.5. *The cardinality constraints and ontology alignment*

After the selection process, the result alignment must have the following properties:

(1) The global similarity should be maximal. We mean by global similarity, the sum of the values of the similarities of different correspondences that forms the alignment.

(2) The cardinality constraints must be verified. We distinguish in general the following cases:

  (i) Case 1: 1−1 constraints: each entity of the ontology source must correspond to one entity of the target ontology and each entity of the target ontology must correspond to one entity of the source ontology.

  (ii) Case 2: n − m constraints: each entity of the ontology source must correspond to at most m entities of the target ontology and each entity of the target ontology must correspond to at most n entities of the ontology source.

  (iii) Case 3: n − *, * − m,* −* constraints: in this case we use the symbol * to mean that we don't impose cardinality constraints.

We propose in this paper, a flow based model that can extract an alignment that satisfies the two conditions above. This model can be easily adapted to understand the case of total and injective alignment and consider more general cardinality constraints where an entity of a given ontology (source or target) must correspond to at most n entities and at least m entities of the other ontology.

In Fig. 4, we give some examples of the configurations of multiplicity between two ontologies.

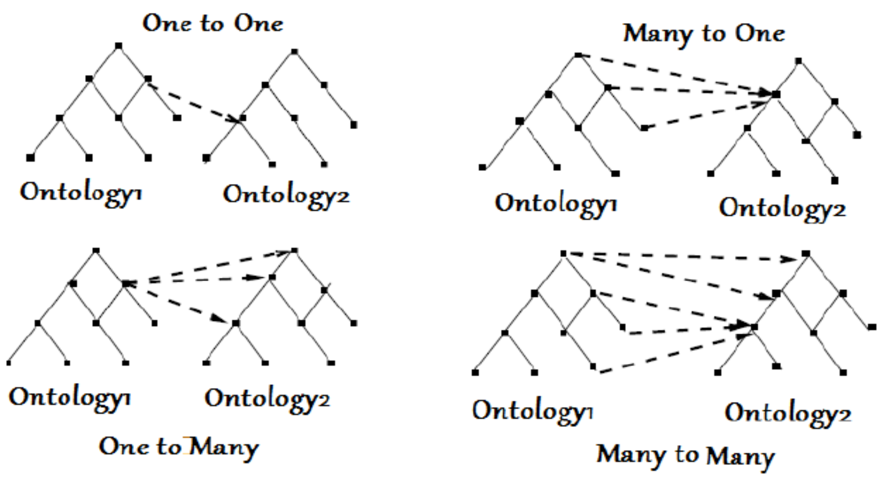


Fig. 4. Different kinds of alignments

4.6. *Alignment extraction methods*

We can distinguish two categories of methods for extracting an alignment.

- Methods based on local optimizations: The methods of this category extract the target alignment by iterating over correspondences belonging to the initial a alignment (typically the similarity matrix). At each step, similarity within each pair of entities is locally maximized [Euzenat and Valtchev(2004)].
- Methods based on global optimization: The methods of this category proceed by optimizing a global criterion rather than optimizing locally. Typically, we consider the global similarity between corresponding entities as an objective function to be maximized: $f = \sum_{C \in A} conf(C)$ where $conf(C) = conf\langle id_C, e_1, e_2, n, R\rangle = n$.

5. Preliminaries on Flows

In this work, we use the automatic approach and the method based on global optimization. Our objective is not to extract just an alignment but is to extract the best one which maximizes the objective function f.

Networks are especially convenient for modeling because of their simple nonmathematical structure that can be easily depicted with a graph. This simplicity also reaps benefits with regard to algorithmic efficiency.

In this section we describe the minimum cost flow algorithm used in our approach to solve the problem of extracting an alignment between two ontologies. The minimum-cost flow problem is a generalization of the maximum flow problem. It is one of the most fundamental network flow problems.

Suppose that we have a network G(V,E) with nodes V = 1, .., n, directed edges E = (i, j) Є V × V . Network G has two special nodes s and t called the source and the sink, respectively. For every directed edge (i, j) Є E, the cost of pushing one unit of flow from node i to node j is c(i, j), and the positive capacity is u(i, j). The minimum-cost flow problem is to find a maximum flow of minimum-cost from the source node s to the sink node t.

Different approaches have been proposed to solve the minimum-cost flow problem. Extensive discussion of this problem and its applications can be found in the book and paper of Ford and Fulkerson [Ford and Fulkerson(1962)], Edmonds and Karp [Edmonds and Karp(1972)].

Given a flow network G(V,E) with source s Є V and sink t Є V , where edge (i, j) Є E has capacity u(i, j), flow f(i, j) and cost c(i, j). The cost of sending this flow is f(i, j). c(i,j). You are required to send an amount of flow d from s to t.

The definition of the problem is to minimize the total cost of the flow: $\sum_{(i,j) \in E} c(i, j) \cdot f(i, j)$ with the constraints:

Capacity constraints:	$f(i, j) \leq u(i, j)$
Skew symmetry:	$f(i, j) = - f(j, i)$
Flow conservation:	$\sum_{w \in V} f(i,w) = 0$ for all $i \neq s, t$
Required flow:	$\sum_{w \in V} f(s,w) = \sum_{w \in V} f(w, t)$

Definition 1. *A feasible flow is a flow satisfying the following capacity and symmetry constraints for each arc (i, j) Є E:*

$l(i, j) \leq f(i, j) \leq u(i, j)$ where l(i, j) is the minimum capacity

$f(i, j) = -f(j, i)$

The objective is to found a feasible flow which has the minimum cost $\sum_{(i,j) \in E} f(i, j) \cdot c(i, j)$.

We show that we can bring the search for a feasible flow of minimum cost network R(G, l, u, c) to that of a maximum flow of minimum cost in a network R' provided with an input, an output, and whose edges are valued by a maximum capacity and cost. The construction of network R' is based on the sign of excess $e_l(x)$ vertices of G for the function l. We denote S+ (respectively S−) the set of vertices of G for the excess of l which is strictly positive (respectively negative) and α the quantity $\sum_{x \in S+} e_l(x) = \sum_{x \in S-} e_l(x)$. We assume α strictly positive [Beauquier et *al.*(2005)].

R' network vertices are the same of the G graph on which we add a source vertex s and a sink t. R' network edges are:

- edges (i, j) of graph G where each edge (i, j) has a null minimum capacity, a maximum capacity equal to u(i, j) − l(i, j) and a cost c(i, j);
- for each vertex x of S+ a new edge u_{sx}, from s to x, has a null minimum capacity, a maximum capacity $e_l(x)$ and a null cost;
- for each edge x of S− a new arc u_{xt}, from x to t, has a null minimum capacity, a maximum capacity $-e_l(x)$ and a null cost;
- return edge u_0 from t to s, has a null minimum capacity, a maximum capacity α and a null cost.

Proposition 1. *There exists a bijection between the compatible flow of the network R and the saturating flow of R' network. The image of a compatible flow of minimum cost network R is the maximum flow with the minimum cost of the R' network [Beauquier et al.(2005)].*

The algorithm runs in pseudo polynomial time. However, suppose the costs c(i, j) are integers, which are less than or equal to an integer C, Edmonds and Karp [Edmonds and Karp(1972)] have proven that the algorithm halts after at most $1 + (1/4)(n^3 - n)(n - 1)C$ flow augmentations, which can be equivalently rewritten as $O(n^4C)$. For more details we return the readers to [Liu(2003)].

The general schema of the algorithm is proposed in algorithm 1

Algorithm 1 Minimum Cost Flow (Ford and Fulkerson)

Initialise F and k à 0
Initialise the flow to 0
repeat
  search for a ameliorative chain π with minimum cost z, from s to t in the ecart graph
  let μ the corresponding chain of G
  if μ is found **then**
    Calculate δ, augmentation of possible flow on μ
    $\delta = Min(\delta, ReqF - F)$
    increase F and the flows of forward arcs of μ with δ units
    decrease the flows of backward arcs of μ with δ units
    $k = k + \delta.z$
  end if
until μ (not found) or $(F = ReqF)$

6. A Flow Model of the Alignment Extraction Problem

We present in this section our contribution. It is initially about the construction of a network which models the cardinality constraints with in particular a wise choice of the capacity constraints. Then the choice of the costs and the research of the maximum flow which has the minimum cost and ensures the optimality of the global similarity. We detail below network construction rules.

- Orienting each edge from each concept of ontology 2 to concept of ontology 1. For such an arc (u, v) : $l_{uv} = 0$ and $u_{uv} = 1$. The cost associated with the edge (u, v) is : c(u, v) = 1 − similarity(u, v).
- Adding a vertex s and an arc from s to each concept of ontology 2. For such an arc (s, a_i): the bounds of capacities are defined as follows: $l_{sai} = 0$, $u_{sai} = n$. The cost associated with each arc (s, a_i) is equal to 0. The value n represents the cardinality with source ontology O.
- Adding a vertex t and an arc from each concept of ontology 1 to t. For such an arc (x, t) : the bounds of capacities are defined as follows: $l_{xt} = 0$, $u_{xt} = m$. The cost associated with each arc (x, t) is equal to 0. The value m represents the cardinality with target ontology O'.
- Adding an arc (t, s) with $l_{ts} = 0$. The cost associated with the arc (t, s) is equal to 0. And $u_{ts} = m \times |$ the number of concepts of ontology 1 $|$.

After the construction of the network, we apply the minimum cost flow algorithm. This algorithm ensures that the flow obtained is compatible (it checks the capacity constraints, thus it checks also the cardinality constraints). On another side, it ensures that the global costs of the flow are minimum, therefore that the total similarity of alignment is maximum since the cost is equal to 1−sim, with sim the similarity, and that the maximum value of the similarity between two entities is equal to 1.

We notice that this network makes it possible to model all the cardinality constraints. An injective alignment, for example, can be obtained by fixing the maximum capacities at 1. The completeness property is assured if we set the maximum and minimum capacities of all the arcs (x, t) to 1. Moreover, this model makes it possible to take into account more general cardinality constraints. Indeed, it is possible to represent the following constraint in particular: "each entity of an ontology can be associated with at most n entities and at least m entities of other ontology".

7. Experimentation and Discussion

In the figures 5 and 6, we present the results of the minimum cost flow algorithm used in our approach (cardinality constraints considered in this case are (n−m)).

We detail in this section the results obtained. We limit to give the order of magnitude of the performance of our approach and some reports. Indeed, we did not find in the specialized literature other systems having tackled the question of the extraction of an alignment with multiple cardinality constraints and having provided detailed results being able to be used like support as comparison with our approach. Only work that we know is described in [Cruz et *al.*(2009)] but which does not provide detailed results only for squares matrices.

We noticed two main results from our experimentations:

(1) Rectangular similarity matrices: Our min cost flow algorithm handles efficiently the problem. For example with some matrix 100 × 1000 with cardinality constraints equal to 4 for ontology1 and 3 for ontology2, the min cost flow algorithm gives a solution after 5 seconds because it reacted directly on the matrix such as it is. But on square matrices and generally when the size of the matrix exceeds the 1500, the algorithm gives a solution after 10 minutes. In other words, our approach gives results better for ontologies where the difference between the number of concepts is important. Whereas that in the case of large ontologies and with weak variation the results are less powerful than in the preceding case.

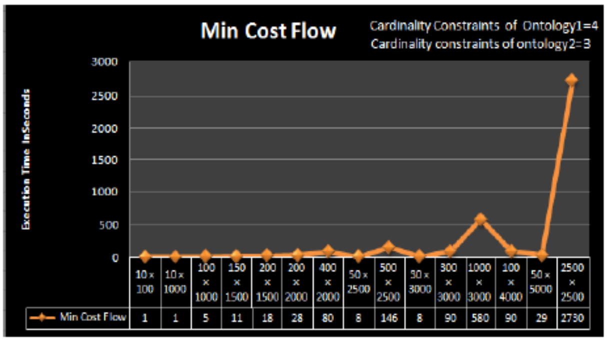


Fig. 5. Min Cost Flow algorithm representation

(1) Our algorithm depends on the cardinality constraints: If we change cardinality constraints we obtain the results represented in Fig. 6. With cardinality constraints equal to 20 for ontology1 and 10 for ontology2, the Min Cost Algorithm reacted more slowly than we use the algorithm with cardinality constraints equal to 4 for ontology1 and 3 for ontology2.

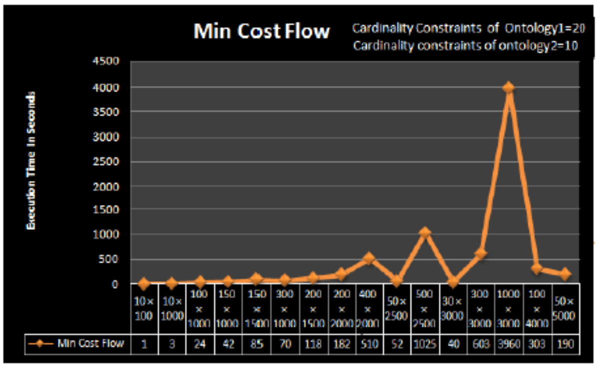


Fig. 6. Min Cost Flow algorithm representation

In Fig. 7, we compare the results of the Hungarian[b] algorithm [Kuhn(1955)] used in other alignment system with the results of the minimum cost algorithm used in our approach (the cardinality constraints considered in this cases are of type (1-1)).

The Hungarian algorithm works only on square matrices, and to adapt it and make it applicable to any type of matrix, we used the approach proposed in [Meilicke and Stuckenschmidt(2007)].

We noted two main results of our experiments:

(1) Rectangular similarity matrices: our min cost flow algorithm treats effectively the problem than the Hungarian algorithm. For example of a matrix 100 × 1000 the min cost flow algorithm finds a solution after 12 seconds because it reacts directly on the matrix. One of the reasons for this behavior is that the Hungarian algorithm transforms the matrix into a matrix 1100× 1100. It returned a solution after 38 seconds. So for large matrices the Hungarian algorithm is less efficient than the algorithm based on the flow.

(2) Square similarity matrices: The Hungarian algorithm is better. Since this algorithm is established in practice to find an optimal allocation of such matrices.

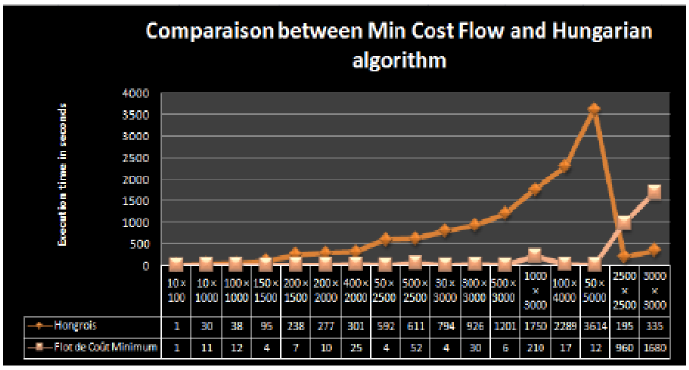


Fig. 7. Performance comparison between the Minimum Cost Flow and the Hungarian method on different input sizes

By exploiting complexities given of these two algorithms (i.e., Hungarian $O(n^3)$, and the minimum cost algorithm $O(n^4C)$), and since all the capacities are lower than 1, the complexity of the minimum cost algorithm can be written under $O(n^4)$. Therefore, it is provable that on the square matrices, the Hungarian algorithm is better. On the other hand it is not preferable in the contrary case i.e., on the rectangular matrices what is confirmed by our experimentation.

In Fig. 8, we compare the results of the algorithm used in [Cruz et *al.*(2009)] with the results of the minimum cost.

[b]The implementation is available at http://code.google.com/p/hungarianassignment/

The algorithm used in [Cruz et *al.*(2009)] is more effective than our algorithm, but it reacts only on the square matrices, which is seldom findable in the problem of alignment of ontologies, therefore we can say that our algorithm remains most usable in the general case (rectangular matrices).

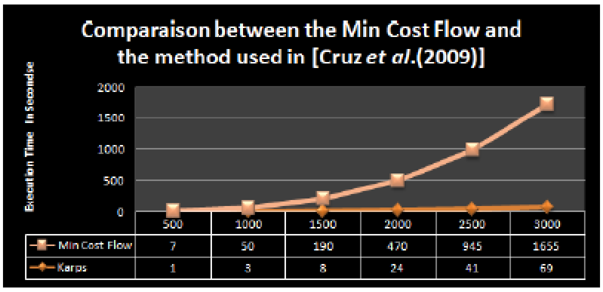


Fig. 8. Performance comparison between the Minimum Cost Flow and the algorithm used in [Cruz et al.(2009)] on different input sizes

Finally, we can conclude that the minimum cost algorithm has an unquestionable advantage compared to the Hungarian algorithm or the algorithm used in [Cruz et *al.*(2009)] in the context of the ontology matching. Indeed, ontologies generally correspond to rectangular matrices and it is very seldom to have in the reality of ontologies with the same number of concepts.

8. Conclusion

In this paper we showed that ontology alignment problem can benefit from the algorithmic techniques developed within flow theory. More precisely we modeled the problem of extracting an alignment which satisfies some cardinality constraints and the objective function defined as the global similarity between the ontologies entities as a graph network. In order to extract such an alignment we used the minimum cost flow algorithm. Then, we showed that the approach presented here handles efficiently the problem with rectangular similarity matrices.

References

Beauquier, D, et *al.*(2005): Eléments d'algorithmique.

Benaissa, M.; Lebbah, Y.(2011): Optimisation de l'extraction de l'alignement des ontologies avec la contrainte de différence. Revue des Nouvelles Technologies de l'Information (RNTI). N_ E.20.Extraction et Gestion des Connaissances (EGC 2011), Editions Hermann.

Bisson, G.(1992): Learning in FOL with similarity measure. In Proc. 10th AAAI, San-Jose (CA US), pp. 82-87.

Cruz, I, et *al.*(2009): Efficient selection of mappings and automatic quality-driven combination of matching methods. International Workshop on Ontology Matching (OM-2009), US.

Doan, A.; Halevy, A.(2005): Semantic integration research in the database community:A brief survey. Special Issue on Semantic Integration, pp. 26(1):83-94.

Edmonds, J.; Karp, R.(1972): Theoretical Improvements In Algorithmic Efficiency For Network Flow Problems. Journal of ACM, pp. 248-264.

Euzenat, J.; Shvaiko, P.(2007): Ontology matching. Springer.

Euzenat, J.; Valtchev, P.(2004): Similarity-based ontology alignment in OWL-Lite. The proceedings of the 15th European Conference on Artificial Intelligence (ECAI), pp. 333-337.

Ford, L.; Fulkerson, D.(1962): Flows In Networks. Press,Princeton.

Ghomari, L.(2009): Alignement d'ontologies de domaine : Etude Comparative Syntaxique versus Sémantique, Cas d'application COMA++ Vs. OWL CTXMatch. Ecole nationale Supèrieure d'Informatique (E.S.I).

Kalfoglou, y.; Schorlemmer, M.(2003): Ontology mapping: the state of the art. Engineering Review, pp. 18(1):1-31.

Karp, R.(1980): An Algorithm to Solve the m × n Assignment Problem in Expected Time O(mnlog n). Networks, pp. 143-152.

Kuhn, H.(1955): The Hungarian Method for the Assignment Problem. Naval Research Logistic Quarterly, pp. 83-97.

Liu, J.(2003) Algorithms for Minimum cost flows. Computer Science Department, The University of Western Ontario.

Meilicke, C.; Stuckenschmidt, H.(2007): Applying Logical Constraints to Ontology Matching. Proceedings of the 30th German Conference on Artificial Intelligence (KI-07), Lecture Notes in Artificial Intelligence, Springer, 4667, pp. 99-113.

Noy, N.(2004): Semantic integration: A survey of ontology-based approaches. SIGMOD Record, pp. 33(4):65-70.

Papadimitriou, C.; Steiglitz, K.(1998): Combinatorial Optimization : Algorithms and Complexity. Prentice-Hall.

Rahm, E.; Bernstein, P.(2001): A survey of approaches to automatic schema matching. VLDB Journal, pp. 334-350.

Shvaiko, P.; Euzenat, J.(2008): Ten challenges for ontology matching. The proceedings of the seventh International conference on Ontologies, Data Bases and application of Semantics (ODBASE).

Shvaiko, P.; Euzenat, J.(2005): A survey of schema-based matching approaches. Journal on Data Semantics, pp. 146-171.

ONTOLOGY – DRIVEN MVC: A VARIANT OF MVC ARCHITECTURAL STYLE

AMRITA CHATURVEDI

Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh, India
amrita@iitk.ac.in

PRABHAKAR T. V.

Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, Kanpur, Uttar Pradesh, India
tvp@iitk.ac.in

Despite separation of concern provided by the MVC architectural style, there still remain some inter – component dependencies which constitute the liabilities of the style. In this paper we highlight the drawbacks of the MVC architectural style and propose an ontology driven solution to address these problems. In the proposed solution the ontology is used as an active run time component. The ontology is generic and can be used for any application based on MVC style. We demonstrate the usage of this style through an example implementation based on both the traditional MVC style and the ontology – driven MVC style. We evaluated the impact of ontology on the quality attributes of the MVC style. We observe that the proposed solution apart from addressing the drawbacks of MVC style also enhances maintainability and reusability of an information system based on the proposed variant style.

Keywords: MVC architectural style; observer pattern; ontology; software quality attributes.

1. Introduction

MVC (Model-View-Controller) style is an architectural pattern first conceived by Trygve Mikkjel Heyerdahl Reenskaug in 1979 for Graphical User Interface (GUI) design. The main goal was “to support the user's mental model of the relevant information space and to enable the user to inspect and edit this information” [Reenskaug (2003)]. MVC provides clear separation of concerns by decomposing an application into three parts: the model, the view and the controller. The model contains application data and the logic for accessing and manipulating that data. It notifies the view when a state change occurs in the model. The view is responsible for rendering the state of the model. The controller is responsible for intercepting and translating user inputs into actions to be performed by the model or view.

A pattern solves a particular problem but its application has various side effects, which may raise new problems [Buschmann *et al.* (1996)]. Some of these can be solved by other patterns. MVC separates data and core functionality from the user interface, thus providing adaptable user interfaces. But this design decision makes the view dependent on the model state. To synchronize the view and model, a change propagation mechanism needs to be established so that the dependent views are updated automatically whenever the state of the model changes. The observer pattern [Gamma *et al.* (1996)] has been largely used to provide this mechanism where the model embodies the role of the subject while the dependent views play the role of observers. However, several drawbacks of observer pattern have also been recognized [Gamma *et al.* (1996); Fowler (2006a)]. Besides that, there are several other dependencies between the M, V and C components in the MVC style, which need to be handled, or else they constitute the liabilities of the style. All such drawbacks of the MVC style are discussed in section 3.

Ontology is one of the most widely used semantic technologies and is formally defined as “an explicit specification of a shared conceptualization” [Gruber (1993)]. It can be used to model domain knowledge in the form of a set of concepts and relationship between them and can be expressed in ontology languages like OWL, OIL, DAML etc.

We have used a generic ontology as a run time component in MVC architectural style to overcome its drawbacks. All the inter component dependencies in the MVC style are encoded in the ontology. So that the components (model, view and controller), instead of directly accessing each other, access the ontology to draw the information about the component they wish to access and then accordingly invoke the methods of the component based on the information provided by the ontology. This approach not only overcomes the drawbacks associated with the MVC style but also enhances maintainability and reusability of an information system based on the style.

The paper is structured as follows: section 2 covers the related research on variants of MVC style and how ontologies are used in this style. Section 3 highlights the problems and dependencies lurking in the MVC style. Section 4 describes the ontology driven MVC style – the proposed solution. Section 5 summarizes the conclusion and ultimately the references are enlisted.

2. Related Research

2.1. *Tactical variants of MVC style*

Flow Synchronization [Fowler (2006b)] uses explicit updating of screens sharing domain data while Observer Synchronization [Fowler (2006b)] uses implicit updating (event invocation) of screens. Presentation Model (PM) or Application Model [Fowler (2006b)] and Model – View – ViewModel (MVVM) [Gossman (2005)] extract the state and behavior of the View out into a Model class with the difference that ViewModel also provides a specialization of the Model that the View can use for data-binding. Both PM and MVVM enhance testability while MVVM further decouples Model and View. To enhance testability, the View is made thinner in Supervising Controller [Fowler (2006b)] and Passive View [Fowler (2006b)]. Unlike Supervising Controller, in Passive View, Controller is responsible for updating the Views so no synchronization work required to be done by the View. Model – View – Presenter (MVP) [Potel (1996)] and Model – View – Adapter (MVA) [Simler (2009)] further decouple the Model and View by placing Presenter and Adapter components respectively in between the Model and View. Unlike MVA, MVP also separates the UI from the presentation logic, thereby enhancing reusability and testability. Presentation – Abstraction – Control (PAC) [Buschmann *et al.* (1996)] defines a structure for interactive software systems in the form of a hierarchy of cooperating agents. Every agent is responsible for a specific aspect of the application's functionality and consists of three components: Presentation, Abstraction, and Control. This subdivision separates the human-computer interaction aspects of the agent from its functional core and its communication with other agents. The Hierarchical – Model – View – Controller [Cai *et al.* (2000)] pattern decomposes the client tier into a hierarchy of parent-child MVC layers. HMVC is similar to PAC with the difference that model and view are not totally decoupled in HMVC. openMVC [Barrett and Delany (2004)] is a framework based on the MVC pattern which allows the Style Information, Layout and Validation Constraints components to be updated with no code change or recompile requirement. This variant reduces redundancy and facilitates faster maintenance.

2.2. *MVC variants with additional component*

Jorg Stocklein *et al.* [Stocklein (2009)] extend the MVC design pattern by an additional dimension "Environment" to capture elements and constraint from the real world. Although the real world sensor information can be handled as Controller events in the MVC model, but this can lead to complex and obscure models. This extension (MVCE) helps in iterative design and development of next generation user interfaces where each component can be refined independently. MVCD (Model-View Controller-Data) style [Zhang (2006)] just abstracts the data source to form an independent data layer. The data layer transforms data received from the model layer into data in the same storage format with the appointed database and stores it to the database. Therefore, the MVCD model can be independent of database. The MVCD model inherits various kinds of advantages of MVC model and has better expansion and maintainability. Chen and Han [Chen and Han (2008)] present an improved MVC style where the controller draws out the parameters from the user request objects, and then makes an instantiation or obtains a corresponding Servlet in memory according to the parameters to respond the request. The corresponding Servlet is called Action in the structure. The servlet inherits from abstract Action class and overrides the run method to respond to the particular request. This improvement enhances maintainability of the style.

2.3. *Use of ontology in MVC*

Bolognini *et al.* [Bolognini *et al.* (2009)] employ ontology based MVC style to implement a sophisticated, user-friendly, generic, customizable metadata editor. The proposed metadata editor uses different OWL ontologies to create custom forms for the input of instance values for an ontological metadata schema. The separation of the ontologies makes it possible to create metadata editors for any schema, and to customize the forms themselves to provide the best user experience in filling in the required metadata values. The main use of ontologies in [Bolognini *et al.* (2009)] is to generate user interfaces (metadata editor), which are generic and Model (metadata schema) independent. Gómez-Pérez and López-Cima [Gómez-Pérez and López-Cima (2004)] present a rapid web application development framework to build semantic portals based on MVC style. In the proposed variant, the Model is composed of domain and cross-domain ontologies and the Controller has ontology based navigation model for controlling the user’s navigations. The main difference between this work and ours is that the described navigation ontology is domain specific and cannot be reused contrary to our MVC ontology which is domain independent. Also, the ontologies described are only helpful in navigating the user from one view to another or generating a composed view. Unlike this work, our variant also deals with model related user actions, model – view – controller dependencies and view notification mechanism. Paulheim [Paulheim (2009)] presents a framework that uses ontologies for building user interfaces from independent, loosely coupled modules (plugins). The information describing a plugin is distributed across three ontologies namely the application ontology, domain ontology and plugin ontology. Plugins communicate with each other through event exchange mechanism. It provides an extension of MVC style where each plugin consists of at least four parts viz. Model, View, Controller and Plugin Ontology, which defines the components and interactions of the plugin. One reasoning component is also there responsible for processing the different ontologies and triggering the appropriate events. User action events generated by a plugin's user interface are read by the reasoner component. The reasoner queries the ontologies for components defining interactions triggered by this type of event, and notifies the respective plugins. It proposes a centrally mediated communication pattern between different components (plugins) which are themselves based on MVC architectural style. It basically employs ontologies for removing inter-component dependencies where each component is based on MVC style. Unlike this, our paper deals with intra component dependencies in an MVC style and attempts to remove its drawbacks by proposing ontology driven solution.

3. MVC Architectural Style – Problems and Dependencies

3.1. *Problems with the MVC style*

The MVC architectural style was designed to balance two forces:

(1) It should be easy to modify the user interface

(2) The functional core of the software should not be affected by user interface modification

In an attempt to balance these forces, the core functionality, input and output were separated and encapsulated in three components namely model, controller and view respectively. But this design gives rise to another problem by making the view dependent on the model state. Whenever the model changes, all its dependents must be updated accordingly. This change propagation mechanism is largely implemented by employing the observer pattern where the model embodies the role of the subject while the views play the role of observers.

However, there are several drawbacks associated with observer pattern (adapted from [Gamma *et al.* (1995)] and [Fowler (2006a)]):

- **Implicit coupling between the view and the model:** Although the main intent of observer pattern is to minimize coupling but still there remains an implicit coupling between the model and view. The model has all the observing views registered with it, which need to be notified when a change occurs in the model.

- **Unexpected updates:** A small operation on model may lead to cascade of updates to observers and their dependents. Also since the dependency criteria aren't well defined or maintained, it may lead to spurious updates which may be hard to track down.
- **The push and pull model of updates:** In the push model, the model sends detailed information about the change to the views whether they want it or not. In the pull model, the model sends minimal notification and views ask for details explicitly thereafter. The push model increases coupling because the model must have some information about the views to send detailed information about the change. The pull model may be inefficient because the views must ascertain what has changed without the help of the model.
- **Dangling references:** When a view (observer) registers with the model, the model actually carries a reference to the view. If the view (observer) doesn't remove itself from the model when the screen is closed, it gives rise to zombie or dangling references and a memory leak.
- **Implicit event invocation:** When there is chain of event invocations, it becomes hard to follow and debug by looking at the code.

3.2. *The model, view and controller dependencies*

Although MVC style separates business logic from presentation, yet there still remain some implicit dependencies which need to be handled:

(1) Views depend on each other and/or a parent child relationship exists between the views. There might be system requirement to update the parent views before their children or some particular views might have higher update priority. An example of view and sub – view relationship is shown in Figure 1. It shows a partial view hierarchy of General Demographic Survey Information System. In Figure 1, the boxes denote the views and the arrows denote the sub – view relationship.

(2) The controller is responsible for converting the user action into system response. To do so, it directly calls the model and/or view methods to generate an appropriate response upon receiving the user input. This makes the controller dependent on the model and view interfaces, as a result a change in model and/or view interface tends to break the controller code.

(3) The model needs to have some knowledge of its observing views in order to notify them of changes. This implicitly couples the model and view.

4. Ontology – Driven MVC Architectural Style – The Proposed Solution

All the problems and dependencies discussed in the previous section can be solved by introducing ontology as a runtime component in the MVC style (and hence the name Ontology – Driven MVC[*]). This section describes the generic Model – View – Controller

[*] The 'Ontology-Driven' phrase has been taken from [Guarino (1998)] where the author has clearly described the difference between 'Ontology-Aware Information System' and 'Ontology-Driven Information System'

Ontology and the static, dynamic and implementation aspects of the proposed style along with its consequences.

4.1. *The Model – View – Controller ontology*

In order to model the MVC domain, we have created a generic ontology, which we call *Model – View – Controller (MVC) Ontology*. The MVC ontology comprises of MVC domain concepts like Model, WindowView, Controller, UserAction, UpdatePriority and source code concepts like Class, Constructor, Method, Query, VariableType. The MVC ontology maps the MVC domain concepts with the source code concepts and hence encodes the entire dynamic flow of the MVC style. Figure 2 shows a partial view of the MVC ontology concepts and Object Properties connecting two concepts. For any given user action (UserAction), the ontology returns the response method (Method) along with its class (Class) name. When a particular model (Model) information changes, the ontology returns the views (WindowView) displaying the information that has changed along with the updating methods (Method), classes (Class) and update priorities of the corresponding views.

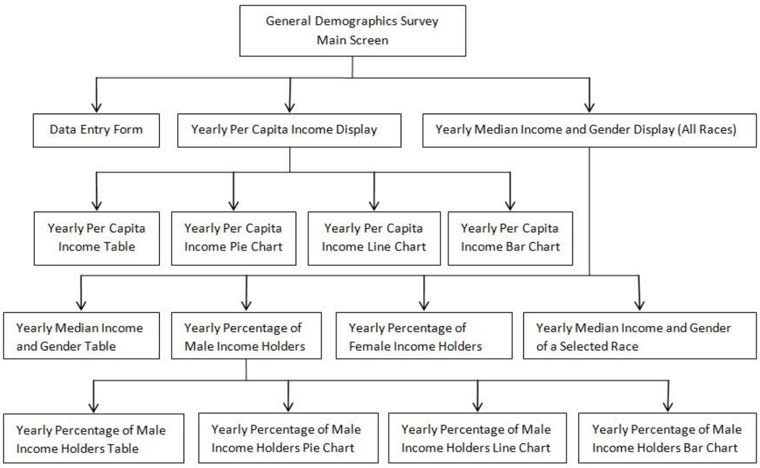


Fig. 1. Partial Hierarchy Showing Views and Sub-views of General Demographics Survey Information System.

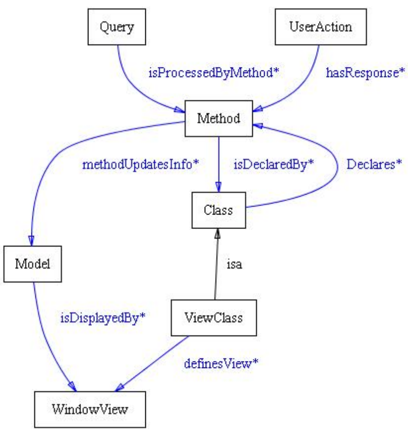


Fig. 2. Partial View of MVC Ontology Generated by the Ontoviz Plugin of Protege3.4.

Several MVC frameworks like Struts [McClanahan (2000)] and Spring [Johnson (2003)] use XML based routing configuration file to map user action to system response. But the MVC ontology in our proposed solution is employed not merely for configuration or metadata purposes but for modelling the MVC programming domain which provides several benefits over the XML based configuration file. The object properties along with their characteristics (Transitive, symmetric property) and domain – range restrictions define the MVC space concept relationships. The cardinality restrictions on the properties aid the developer in verifying the consistency of the resulting instantiated ontology. Different atomic views ‘compose’ (transitive object property) the compound and complex views which can be refreshed if they display the updated information rather than refreshing the entire view. Moreover, if the response to a particular user action requires that a view be displayed and if there is an error displaying the view, then any other view rendering similar information can be displayed. Moreover, the ontology also enables the application to perform priority based updating of the views rendering the model information that has changed during the runtime. Figure 3 shows a partial view of the MVC ontology concepts along with is-a relationship and instances. The portion of the image above the red dotted line shows the intension and that below shows the extension part of the MVC ontology. In the following text MVC ontology is simply referred to as ‘ontology’.

4.2. *Structure*

Figure 4 shows the difference between the structure of the traditional MVC style and the ontology – driven MVC style. In Figure 4, the numbers along the edges denote the sequence of communication. In the traditional MVC style (Figure 4 (a)):

(1) The controller receives the user action/input from the view.
(2) The controller calls the model and/or view methods to generate the desired system response.
(3) The model notifies the view of change in model state and also passes the changed state.
(4) The view may call model method to query the model state.

The proposed ontology – driven MVC style consists of four components viz. Model, View, Controller and Ontology with each one of them having their own APIs. The entire communication dependencies between the components are encoded in the ontology, which is domain independent. The sequence of communication between the components occurs as follows (Figure 4 (b)):

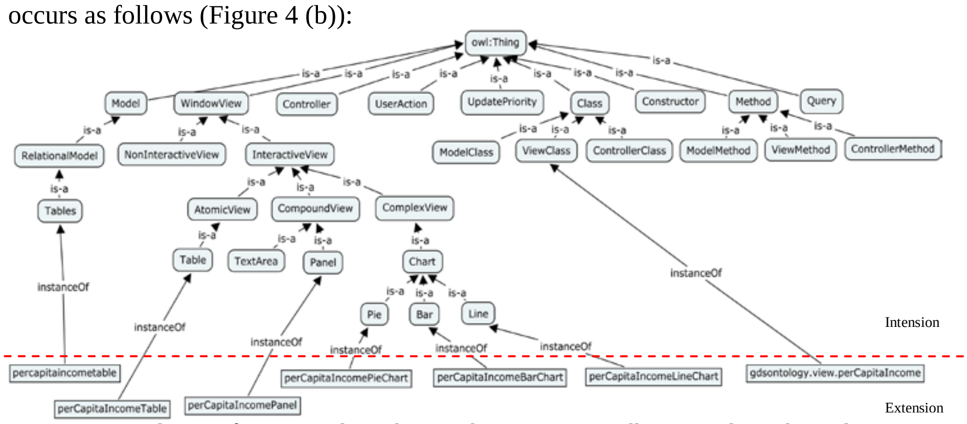


Fig. 3. Partial View of MVC Ontology Showing the Concepts as well as is-a Relationship and Instances

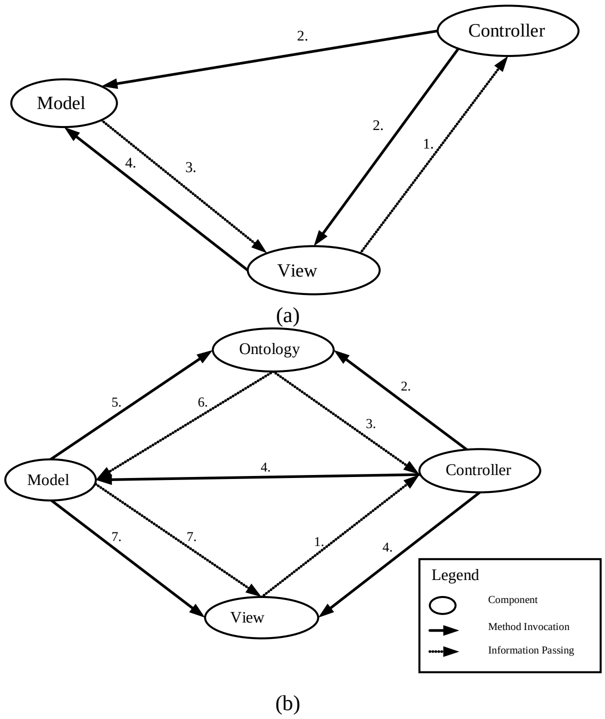


Fig. 4. The Structure of (a) Traditional MVC Style (b) Ontology – Driven MVC Architectural Style.

(1) The controller receives the user action/input from the view.
(2) The controller consults the ontology to find the system response for that particular input.
(3) The ontology sends the method information which should be invoked to achieve the response.
(4) The controller invokes the appropriate methods of the model or view to achieve the desired output.
(5) If in this process the model state is changed, then the model's NotifyAll method consults the ontology to find out the views rendering the information that has changed.
(6) The ontology sends the view information of those views which need to be updated.
(7) The NotifyAll method then, accordingly passes the changed information and invokes the updating methods of the views returned by the ontology in the order of their update priorities (as supplied by the ontology component).

Thus the ontology plays the pivotal role in the entire working of the style from receiving input to delivering output.

4.3. *Dynamics*

We have taken three common scenarios to depict the dynamic behavior of the proposed architectural style.

Scenario I

This shows how the controller gets the data from the model desired by the user (Figure 5).

- The controller handles the request of the user to get model data.
- The controller retrieves the query from the ontology which should be executed to get model data.
- The controller retrieves the model method from the ontology which should be executed to get the data.
- The controller retrieves the model class defining the method from the ontology.
- The controller then, invokes the model method whose detail was derived from the ontology.
- The model method after completing execution, returns the data to the controller.
- The controller then, returns from its getModelData method.

Scenario II

This shows how the controller updates the model data (Figure 6).

- The controller handles the user request to set model data.
- The controller retrieves from ontology, the model property updated by the view through which the request is sent.
- The controller retrieves the model method from ontology which updates the model property.
- The controller retrieves the model class defining the method from the ontology.
- The controller then, invokes the model method whose detail was retrieved from ontology.
- The model method, after updating the model property, calls the NotifyAll method of the model which is responsible for notifying all the views rendering the model property that has changed/updated.
- The NotifyAll model method retrieves from the ontology the views rendering the model property which is updated.
- The NotifyAll model method calls it's sub-routine to arrange the views according to their update priority.
- For each view, the updating method is retrieved from the ontology and is invoked by the NotifyAll model method.

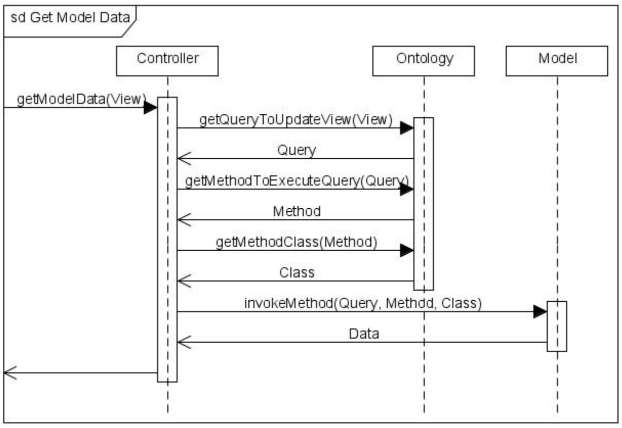


Fig. 5. Sequence Diagram Showing Interaction between the Components to Get Desired Data from the Model.

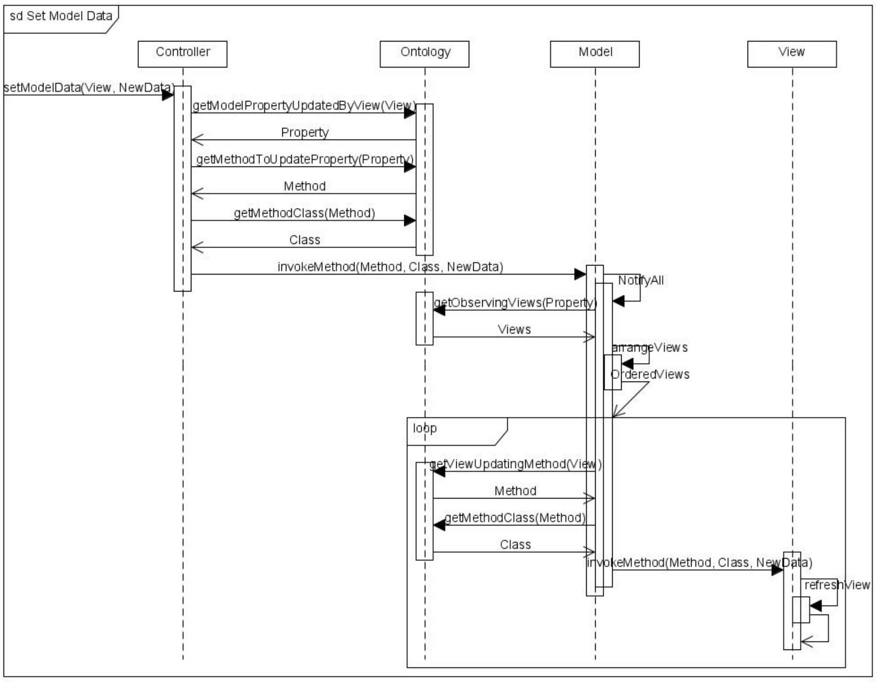


Fig. 6. Sequence Diagram Showing Interaction between the Components to Update the Model

Scenario III

This shows how the controller displays views desired by the user (Figure 7).

- The controller handles the user request to display a view.
- The controller retrieves the response method, method type and method class from the ontology.
- The controller then, invokes the view method whose detail was retrieved from the ontology.

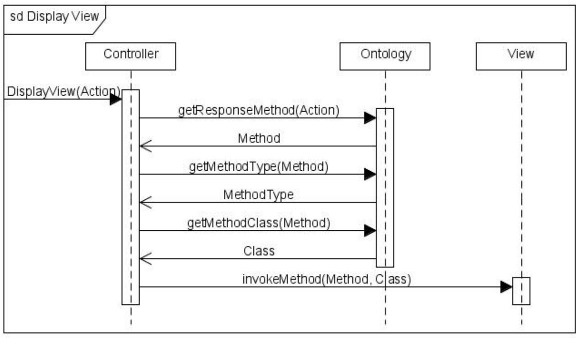


Fig. 7. Sequence Diagram Showing Interaction between the Components to Display the View Desired by the User.

4.4. *Implementation aspects*

The proposed architectural style gives rise to a new programming technique where the main principle is to encode the inter component dependencies in ontology. The following steps should be followed to develop an application based on the proposed architectural style:

(1) Design and develop the model: Encapsulate the data and core functionality of the application in the model component. Define functions for accessing and manipulating the data and expose the desired functions of the model to the controller.

(2) Design and develop the views: Design the appearance of main views as well as their children and dependent views. Dedicate a view class to initialize all the views rendering model data. Implement the draw and update methods to display and update the views respectively. The update method is called by the model method performing notification on model state change.

(3) Instantiate the MVC ontology based on the application: The MVC ontology used for the proposed architectural style is generic and can be used for any application based on MVC architectural style. The ontology needs to be instantiated based on the different views, models and their methods used. A partial view of concepts and their relationships in the MVC ontology is shown in Figure 2 and instances of some concepts are shown in Figure 3.

(4) Design and implement a generic and reusable controller or extend and/or reuse the controller: The controller used in the proposed architectural style is completely generic (does not depend on the application) and is reusable. The controller can be implemented in any desired programming language following the principles laid down in the proposed solution. The resulting controller is reusable and has methods defined like setModelData, getModelData, displayView which are ready to use. Moreover, the controller methods do not depend either on the model or view interfaces. They rather derive their desired information from the ontology and hence enable the controller to be reusable with any set of models or views.

(5) Reuse the NotifyAll method of the model: The NotifyAll model method initiates the change propagation mechanism. When the model state changes, the NotifyAll method retrieves the dependent views from the MVC ontology and invokes their update methods. It derives the required information from the ontology which makes it application independent and reusable.

We have identified the software quality attributes affected by the proposed solution through a General Demographics Survey (GDS) information system based on the proposed architectural style. The information system is coded in java language and jena2.5.7 [Jena] is used as the ontology API. Our MVC ontology is in OWL – DL [W3C OWL Working Group (2004)] and is developed using the protégé 3.4 ontology editor [protégé ontology editor]. We also developed the same information system based on the traditional MVC style and then analyzed and compared the quality attributes exhibited by both the styles. Our findings are illustrated in the next subsection.

4.5. *Consequences*

To analyze the quality attributes affected by the proposed solution, we use the Dromey's software quality model [Dromey (1995)] to map the characteristics exhibited by the information system based on the traditional MVC style as well as on the proposed style to high level quality attributes. Among the several software quality models (like ISO/IEC standard [ISO/IEC (2010)], IEEE standard [IEEE Computer Society (R2009)]), we have chosen the Dromey's software quality model because it is the only model which focuses on software code or implementation. The model advocates that software does not manifest quality attributes. It, rather exhibits certain characteristics which can be mapped to high level quality attributes. Since we have software code at hand, we can easily apply the Dromey's model to obtain the high level quality attributes exhibited by softwares based on two different architectural styles.

4.*5.1. Benefits*

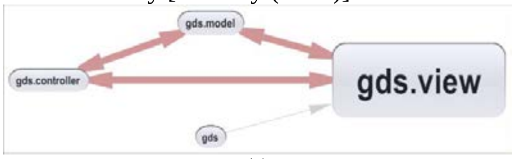
Since the main identified drawbacks with the MVC style deal with inter – modular issues and the proposed solution targets to solve the identified problems so among the four basic quality carrying properties identified by Dromey [Dromey (1995)], the properties affected by the proposed solution are modularity properties (high level, inter – modular design issues). Among the six modularity properties identified, we find that the proposed solution enhances three of them which are parameterized, loosely coupled and abstract.

(1) Parameterized: "A module is said to be parameterized if it contains as parameters all and only the necessary and sufficient inputs and outputs to characterize a particular well-defined function/procedure" [Dromey (1995)]. Unlike the traditional MVC style, our proposed solution offers a completely parameterized controller and NotifyAll part of the model. In the controller, the parameters define the user action and the system response while in the NotifyAll the parameters define the entire notification mechanism when some model property changes along with the changed model property. The high level quality attributes affected by this property are maintainability and reusability [Dromey (1995)].

(2) Loosely Coupled: We use the JArchitect tool [JArchitect] to analyze the java code of our information system based on the MVC style and on the Ontology – Driven MVC style separately. Figures 8 (a) and 8 (b) show the dependency cycle graph between the components of the GDS Information System based on the traditional MVC style and on the Ontology – Driven MVC style respectively. Single arrow edge from A to B indicates that A is using B and double arrow edge indicates that A and B are mutually dependent. The thicknesses of edges are proportional to the strength of coupling in terms of the number of members involved. Figure 8 (a) shows that there exists a strong coupling between the gds.model, gds.view and gds.controller components and they are mutually dependent. Figure 8 (b) shows only a single arrow from gdsontology.view to gdsontology.controller which shows that view accesses the generic controller methods to hand over the user action. Apart from this, no other dependency exists between the model, view and controller components. Figures 9 (a) and 9 (b) show the dependency matrix between the components of the GDS Information System based on the traditional MVC style and on the Ontology – Driven MVC style respectively. A black matrix cell denotes that the package in row and the package in column are both directly using each other. The number 2 in black matrix cell denotes that A is using B and B is using A so A

and B are involved in a dependency cycle of length 2. A blue matrix cell denotes that the package in column is using the package in row and a green matrix cell denotes that the package in column is used by the package in row. There are 6 black matrix cells in Figure 9 (a) which means that all the three model, view and controller components are engaged in a dependency cycle of length 2. Whereas, there are no black matrix cells in Figure 9 (b) so there is no mutual dependence between the model, view and controller components. The high level quality attributes affected by this property are maintainability, reusability and reliability [Dromey (1995)].

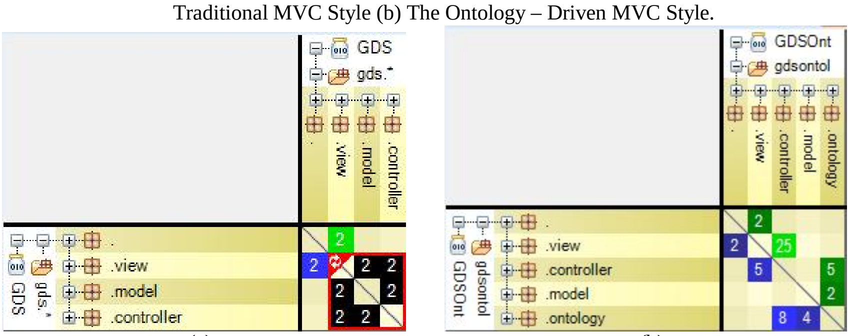
(3) Abstract: "An object/module is sufficiently abstract if there is no obvious, useful higher level concept that encompasses the structural form" [Dromey (1995)]. The controller and NotifyAll part of the model are sufficiently abstract as very high level generic methods are used for handling input, delivering output and notifying the observers. The high level quality attributes affected by this property are maintainability and reusability [Dromey (1995)].



(a)

(b)

Fig. 8. Dependency Cycle Graph between the Components of the GDS Information System based on (a) The Traditional MVC Style (b) The Ontology – Driven MVC Style.



(a) (b)

Fig. 9. Dependency Matrix between the Components of the GDS Information System based on (a) The Traditional MVC Style (b) The Ontology – Driven MVC Style.

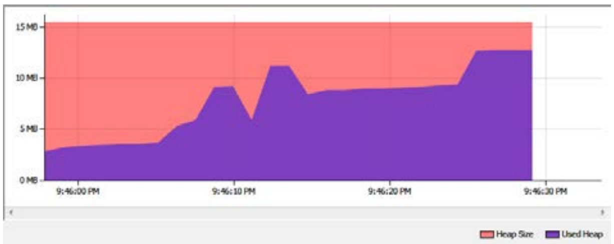
Apart from the added enhancement of the above mentioned quality attributes, the proposed solution also overcomes the drawbacks associated with the MVC style (discussed in Section 3). This summary is presented in Table 1.

Table 1. Table showing how the proposed solution overcomes the drawbacks of MVC style

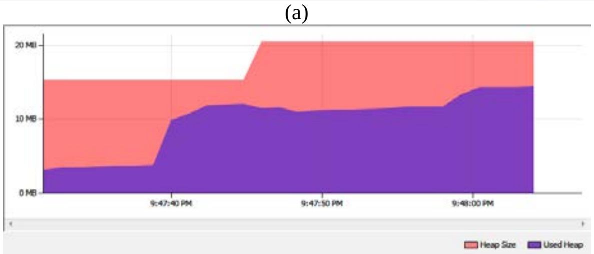
MVC style Drawbacks	Proposed Solution
Implicit coupling between the view and the model	No coupling exists between view and model as shown in Figures 8 (b) and 9 (b)
Unexpected updates	Since the dependency criteria of observers are well defined and maintained in the ontology, there is no chance of spurious updates.
The push and pull model of updates	Since the entire notification information is provided by the ontology, there is no need of push or pull model of updates
Dangling references	Since the observers don't register themselves with the model, there is no risk of dangling references.
Implicit event invocation	Since there is no coupling between the model, view and controller (Figure 8 (b) and 9 (b)), there is no chance of complicated event invocations.

4.5.2. Liabilities: Although the proposed solution enhances certain quality attributes, but it also causes an indirection in the flow of control and uses reflection technique and hence negatively impacts performance. As shown in Figure 10, memory consumption is more in case of ontology – driven MVC style because of the processing of the ontology involved.

Although, to develop an information system based on the proposed style, the developer has to instantiate the ontology once but the effort is recovered over multiple implementations in the maintenance phase.



(a)



(b)

Fig. 10. Runtime Memory Heap Graph of GDS Information System based on (a) The Traditional MVC Style (b) The Ontology – Driven MVC Style.

5. Conclusion

In this paper, we have highlighted the problems of the MVC style and have proposed an ontology driven solution to address the identified problems thus defining a new variant of MVC architectural style: Ontology – driven MVC architectural style. In the proposed solution the ontology is used as an active run time component. The ontology is generic and can be used for any application based on MVC style. All the inter component dependencies in MVC style are encoded in the MVC ontology so that the components instead of directly accessing each other, draw the communication details from the ontology and subsequently invoke appropriate methods. We have demonstrated the utility of the proposed style through an example implementation based on the proposed style. We found that the proposed solution addresses all the drawbacks of MVC style and also enhances software quality attributes like maintainability and reusability. The implementation aspects reveal that the proposed solution gives rise to a new programming technique where the ontology drives the entire communication sequence, breaking the communication dependencies between the components. This enhances the maintainability and reusability of the components but negatively impacts the overall performance because of the indirection in the flow of control and use of reflection technique. The ontology – driven MVC style can be used in the following contexts:

(1) When the model and/or the view are expected to change in near future: For example, the model of the GDS information system may have to be changed by recording the educational attainment of the population in addition to their mean income. In this scenario apart from the model changes, the view also requires minor changes. There is another scenario where the view requires major changes but the model remains unchanged. This is the case when additional presentations of the same data have to be made like adding 3D pie charts, animated charts, histograms etc. There is another scenario where both the model and view need to be changed. For example the GDS information system is to be converted into an election information system for recording and displaying the election data. In this worst-case scenario, both the view and model are implemented and connected with the reusable controller and notification mechanism via the MVC ontology.

(2) When the view is rich: This is the case when there are several widgets on the same screen displaying different parts of the model and handling several user actions. For example in a medical information system, one text area may be displaying symptoms of a disease and another text area displays the recommended medicines for the same disease. When the user clicks a medicine, a third text area displays the effects and side effects of the medicine. Such rich widgets and user actions can be better developed and maintained by being declared in the MVC ontology rather than defined imperatively in scattered code.

(3) When runtime changes in the view can be made: This scenario arises when the user is allowed to change the widgets at runtime. The response to all user actions is encoded in the MVC ontology and hence this can be flexibly done.

(4) When the user interface (UI) designer wants to generate and fix various different types of user interfaces: Since the system response to various user actions is encoded in the MVC ontology, the designer can flexibly change the structure of UI and test run the various UI designs

It should be possible to reduce the overhead imposed by the ontology by efficient implementations of this component using main-memory based RDF or RDBMS components.

Acknowledgments

The authors would like to thank Indian Institute of Technology, Kanpur, Uttar Pradesh, India for providing sufficient infrastructure and resources which aided in the successful completion of this work. We would also like to thank Dr. Philippe Dugerdil, Professor of Software Engineering, Department of Information Systems, Geneva School of Business Administration, University of Applied Sciences, Geneva, Switzerland for his valuable comments and feedback.

References

Barrett R.; Delany S. J. (2004): openMVC: A non-proprietary component-based framework for web applications. WWW 13th International World Wide Web Conference, pp. 464-465.

Bolognini V.; Iorio A. D.; Duca S.; Musetti A.; Peroni S.; Vitali F. (2009): White B.: Exploiting ontologies to deploy user-friendly and customized metadata editors. In: Proceedings of the IADIS Internet/WWW conference. Rome, Italy.

Buschmann F.; Meunier R.; Rohnert H.; Sommerlad P.; Stal M. (1996). *Pattern-Oriented Software Architecture A System of Patterns*, volume 1. Wiley.

Cai J.; Kapila R.; Pal G. (2000): HMVC: The layered pattern for developing strong client tiers, http://www.javaworld.com/javaworld/jw-07-2000/jw-0721-hmvc.html?page=2, Accessed on 21 April 2012.

Chen C.; Han H. (2008): The research on modern distance education system based on improved mvc pattern. In: International Conference on Computer Science and Software Engineering, **5**, pp. 462 – 465.

Dromey R. G. (1995): A model for software product quality. IEEE Transactions on Software Engineering, **2**, pp. 146-163.

Fowler M. (2006a): Organizing presentation logic: observer gotchas, http://martinfowler.com/eaaDev/OrganizingPresentations.html#observer-gotchas, Accessed on 20 March 2012.

Fowler M. (2006b): GUI architectures, http://martinfowler.com/eaaDev/uiArchs.html, Accessed on 20 March 2012.

Gamma E.; Helm R.; Johnson R.; Vlissides J. (1995). *Design Patterns Elements of Reusable Object-Oriented Software*, Addison-Wesley.

Gómez-Pérez A.; López-Cima A. (2004): Advanced design of semantic web portal: ODESeW2. In: Poster Proceedings of 3rd International Semantic Web Conference (ISWC2004), Hiroshima, Japan.

Gossman J. (2005): Introduction to Model/View/ViewModel pattern for building WPF apps, http://blogs.msdn.com/b/johngossman/archive/2005/10/08/478683.aspx. Accessed on 21 April 2012.

Gruber T. (1993): A translation approach to portable ontology specification. In: Knowledge Acquisition, **5**(2), pp. 199-220.

Guarino N. (1998): Formal Ontology and Information Systems. In: Proceedings of Formal Ontology in Information Systems (FOIS '98), IOS Press, Trento, Italy.

IEEE Computer Society (R2009). *IEEE Standard for a Software Quality Metrics Methodology*, IEEE Std 1061-1998.

ISO/IEC (2010).*Systems and software engineering —Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality Models*, ISO/IEC FDIS 25010 International Standard, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=35733, Accessed on 15 August 2012.

JArchitect: A tool to evaluate java code base, http://www.javadepend.com/. Accessed on 10 April 2012.

Jena: A Java Framework for building semantic web applications, http://sourceforge.net/projects/jena/files/Jena/Jena%202.5.7/, Accessed on 5 March, 2012.

Johnson R. (2003): Spring framework, MVC based web application framework, http://www.springsource.org/, Accessed on 25 April 2012.

McClanahan C. (2000): Apache struts, MVC based web application framework, http://struts.apache.org/2.x/index.html, Accessed on 25 April 2012.

Paulheim H. (2009): Ontology-based modularization of user interfaces. In: Proceedings of the 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems, pp 23-28.

Potel M. (1996): MVP: Model-View-Presenter The Taligent programming model for c++ and java, http://www.wildcrest.com/Potel/Portfolio/mvp.pdf, Accessed on 21 April 2012.

Protégé ontology editor, http://protege.stanford.edu/, Accessed on 5 March 2012.

Reenskaug T. (2003): The Model-View-Controller (MVC) its past and present, http://heim.ifi.uio.no/~trygver/2003/javazone-jaoo/MVC_pattern.pdf, Accessed on 2 March 2012.

Simler K. (2009): Model – View – Adapter, http://blog.palantir.com/2009/04/20/model-view-adapter/, Accessed on 21 April 2012.

Stocklein J.; Geiger C.; Paelke V.; Pogscheba P. (2009): Poster: MVCE - a design pattern to guide the development of next generation user interfaces. In: IEEE Symposium on 3D User Interfaces, pp. 153 – 154.

W3C OWL Working Group (2004): Web ontology language overview, http://www.w3.org/TR/owl-features/, Accessed on 5 March 2012.

Zhang L.; Ma F.; Zhong C; Zhang L.; Wang Y. (2006): A MVCD model and its application in university project management. In: The Sixth World Congress on Intelligent Control and Automation (WCICA), **2**, pp. 7008 – 7012.

International Conference on Software Engineering and New Technologies

AN ONTOLOGY-BASED APPROACH TO MODELING MULTILINGUAL E-LEARNING RESSOURCES

HADHEMI ACHOUR

Tunis High School of Management, University of Tunis
Tunis, Tunisia
Hadhemi_Achour@yahoo.fr

MAROUA ZOUARI

Tunis High School of Management, University of Tunis
Tunis, Tunisia
zouari.maroua@gmail.com

In this paper, we propose a semantic representation model of online multilingual educational resources, that aims to improve their search, access and reuse. This model is based on a set of ontologies, whose concepts are used for indexing learning resources that are available in a system of e-Learning. The ontological concepts and relations that we propose, provide a means of guided navigation in these resources. We also show how these ontologies can be helpful in performing multilingual search of specific learning resources and in facilitating their share and reuse for building new educational resources.

Keywords: e-Learning; ontology; multilingual educational resource; educational resource indexation; e-Learning standards; educational resource search.

1. Introduction

The significant use of ICT has substantially changed practices in the world of education. Learning, teaching and training processes are increasingly relying on the web, which has led to the development of a considerable number of various educational web applications, such as e-Learning platforms (LMS: Learning Management Systems), learning portals, learning object warehouses, blogs, wikis, virtual communities and educational social networks, …

The increasing use of e-Learning systems is accompanied by a growing development and setting on line, of various electronic educational resources such as online courses, course materials, presentation slides, FAQs, exercises ...

Considering the growing number and the diversity of educational resources available within e-Learning websites, as well as their heterogeneity, it becomes increasingly difficult to manage, classify and access these resources [Chaput *et al.* (2004)], or to find the relevant resources that meet user queries. This situation raises a need to increase or enhance intelligence in online learning environments [Psyché *et al.* (2003)].

Currently, Semantic Web technologies and mainly, ontologies are becoming increasingly prized solutions in research works on e-Learning environments [Lando.

(2006)] [Psyché *et al.* (2003)]. Our paper comes within the general framework of these works, and is particularly interested in using ontologies for improving access, retrieval and reuse of online multilingual learning resources. We indeed propose in this paper, a semantic ontology-based representation model of diverse educational resources that may be available in a learning portal. The proposed ontologies, which will be used for indexing learning resources, provide a means of guided navigation within these resources. They also allow performing multilingual retrieval of specific learning resources, which can be, not only entire courses, but also course fragments (or elements). The idea of our model is in fact, to enable a user interested in Object Oriented Programming (OOP) for example, to ask queries such as:

- I am looking for an *OOP course outline*, written in *Arabic*.
- I am looking for an *English course chapter* dealing with the *inheritance* notion.
- I am looking for a *French definition* of an *abstract class* in OOP.

…

The next few sections of this paper are organized as follows: Section 2 presents the problem of indexing and retrieving on line educational resources. In this section, we highlight the limitations of current indexing standards in the field of e-learning (such as *LOM*, *SCORM*, ...), and show the need of semantic indexing for improving learning resources retrieval. Section 3 defines the concept of ontology and presents a brief state of the art of main ontology uses and contributions in learning resource indexing. Section 4 presents the proposed semantic representation model of educational resources of a multilingual learning portal, which is based on 3 ontologies: *learning domain ontology*, *learning theme domain ontology*, *learning resource ontology*. Finally, section 5 concludes this paper by discussing the contributions and limitations of our model. We also propose next steps to follow, for the improvement and the operational implementation of our work.

2. Indexation of Learning Resources

2.1. *Learning objects and the need for indexing*

In e-Learning, the basic form of an educational resource is the "learning object" (LO) [Iles *et al.* (2008)]. Indeed, the main objective of creating learning objects is to build course fragments or small parts, that can be used and reused in different learning contexts [Ghebghoub *et al.* (2009)]. This is how we can build a course by assembling various elements, such as fragments of text, images, sounds, simulations, ... which will form chapters, then courses or modules and then comprehensive trainings [Fory (2004)]. Having therefore, these small units of resources, can lead to a faster and more efficient creation of new educational resources [Wiley (2002)].

The precise definition of a learning object has raised many debates and several definitions exist in the literature. The Institute of Electrical and Electronics Engineers (*IEEE*) defines a learning object as “any entity, digital or non-digital, that may be used for learning, education or training"[a]. In our work, we retain Wiley's definition, which defines a learning object as "any digital resource that can be reused to support learning"[Wiley (2002)]. Still according to Wiley, "this definition includes anything that can be delivered across the network on demand, be it large or small. Smaller reusable digital resources include digital images or photos, small bits of text, animations, ... Larger reusable digital resources include entire web pages that combine text, images and other media or applications to deliver complete experiences, such as a complete instructional event"[Wiley (2002)].

In order to reuse these learning objects, it is necessary to index them, to facilitate their access, search, share and reuse. This indexing operation consists in describing each learning object, by using metadata that identify it and represent it, and which help find it later. Indeed, a non-indexed learning object can not be found and, thus will be unusable [Ghebghoub *et al.* (2009)].

2.2. *Standard metadata schemas in e-Learning*

The indexation of online educational resources, actually uses descriptive metadata, which are usually structured in various categories of properties that characterize an educational resource. Many works in the field of education have indeed dealt with the description of educational resources, as well as its standardization. These works led to the development of a number of metadata conceptual schemas in the field of norms and standards for e-Learning and ICT, and of which we can mention as examples:

- **DC** (*Dublin Core*)[b] : is a general scheme of metadata describing online resources, using 15 basic elements on content, intellectual property and technical aspects (such as the date, type, username ...). It is considered a general standard for bibliographic description[c].
- **LOM** (*Learning Object Metadata*)[d] : is a more sophisticated schema proposed in 2002 by the IEEE consortium. It defines a metadata structure using a set of 80 elements distributed in 9 groups of properties (general, life cycle, meta-metadata, educational, technical, rights, relation, annotation, classification).
- **SCORM** (*Sharable Content Object Reference Model*)[e] : allows structuring the educational resource contents, as well as their interactions with the operating environment, and is therefore proposed for distance learning platforms. Structured resources according to SCORM are composed by SCOs (Sharable Content Objects), which are autonomous and reusable components. These components are identified using educational and informative metadata (such as title, author, creation date, level of difficulty, learning objectives, ...). SCORM also describes the information needed for communication between content (resource) e-Learning platform (how SCOs fit together with each other, and what resources they use).

[a] Learning Technology Standards Committee 2002, p. 45
[b] http ://dublincore.org/
[c] http://wiki.univ-paris5.fr/wiki/Indexation
[d] http ://ltsc.ieee.org/wg12/files/LOM 1484 12 1 v1 Final Draft.pdf
[e] http://www.scorm.fr/

- **IMS-LD** (*Instructional Management System Learning Design*)[f]: is a norm which aims to bring some elements of pedagogy in an online learning system via a modeling language of learning processes. [Hernandez *et al.* (2008)] explain that in this model, a unit of learning ("part") is structured as a set of "actions", which are composed by "partitions" involving "activities" to "roles" (teacher, learner, ...)

Indexation based on standard metadata schemas are thus, a solution for describing learning resources. It remains however not enough efficient for searching, accessing and reusing educational resources, and this is mainly due to the lack of content semantic representation provided by these standards. [Hernandez *et al.* (2008)] show that current e-Learning standards offer a partial solution to the learning object re-usability, because they do not describe the semantic content of metadata, and do not represent the various existing relations between learning objects. [Ghebghoub *et al.* (2009)] raises the problem of semantic ambiguity of some of the standard schemas elements, and their subjectivity. Such is especially the case of the pedagogical category of LOM standard, which makes difficult the use of its descriptors. [Ghebghoub *et al.* (2009)] actually summarize the causes of difficulties in the use of LOM metadata, to the complexity of its structure schema, the large number of its elements and misinterpretation of users for certain elements. In addition, [Toledano *et al.* (2005)] show that, when filling the metadata fields for describing a given resource, we may use different vocabularies to express the same concept, and this may cause problems when searching these resources.

This is why, to overcome these shortcomings, a growing number of recent works are using ontologies for indexing learning resources. Ontologies allow among other things, to set a common vocabulary that avoids ambiguities and allow better communication between resource users. Ontologies are also able to enhance searching capabilities, through an inference mechanism which is able to replace searched terms by other semantically related concepts in the ontology [Toledano *et al.* (2005)]. We present some of these works in the next section after defining the concept of ontology.

3. Ontology-Based Indexation of Learning Resources

3.1. *Ontologies*

We point out that the term "ontology" originally comes from the metaphysics branch of philosophy, which deals with the nature of being, that is of general properties of what exists[g]. It was in the 80s that it began to be used in the field of artificial intelligence as a form of knowledge representation, and is now frequently exploited in computer and information science, it is particularly considered as one of the key technologies of the Semantic Web [Wilson (2004)].

[f] http://www.imsglobal.org/learningdesign/
[g] Wikipedia (http://en.wikipedia.org/wiki/Ontology)

One of the most cited definitions of an ontology, is that of Gruber: "An ontology is an explicit specification of a conceptualization" [Gruber (1993)]. This definition was slightly modified in [Borst (1997)] as follows: "an ontology is defined as a formal specification of a shared conceptualization". In fact, ontologies provide a representation of a field by specifying its concepts and the existing relationships between these concepts. These relationships between entities can be hierarchical ones, or can be of any other semantic relationships. Ontologies, thus provide a common and shared understanding of a domain, which can be communicated and exploited by humans as well as machines.

Currently, Semantic Web technologies and mainly, ontologies are becoming increasingly prized solutions in research works on e-Learning environments [Lando (2006)] [Psyché *et al.* (2003)], and are especially more and more used for indexing learning resources.

3.2. *Ontology-based indexation of learning resources*

Ontology-based indexation has developed significantly in recent years, especially in the field of e-Learning. We cite in what follows, some works that have used different kinds of ontologies in order to describe educational resources:

[Duitama *et al.* (2003)] propose to manage adaptative learning systems using 3 models: a domain model which is an ontology that represents concepts from the knowledge domain, a user model that considers knowledge, goals and preferences of learners, and an educational component model that allows interaction among components.

[Fontaine *et al.* (2006)] (Memorae project) propose to manage the resources, information and knowledge of a training, using a training organizational memory based on ontologies. Resource indexing uses ontology concepts, and learners access the knowledge through ontology exploration. In MEMORAe two domain ontologies are considered: the application domain ontology, which specifies the concepts to teach and relations between them, and the training domain ontology which specifies all useful concepts about training in general (user types, resource types, material formats, ...).

[Ghebghoub *et al.* (2009)] shows the limits of LOM metadata schema and proposes an indexing solution of learning resources, based on an ontology of LOM (LOMonto), which more clearly identifies the elements of the schema description.

[Hernadez *et al.* (2008)] proposes a representation model of learning objects, which is based on ontologies. This model considers several aspects for describing learning objects: LOM and SCORM description, content thematic description (domain of theme ontology), learning theories description (educational theory ontology), pedagogical scenarios description.

[Bianchi *et al.* (2009)] (AquaRing project) aims to facilitate and improve access to digital content concerning the aquatic environment and its resources. In this work, 2 types of ontologies are used to annotate and retrieve learning content: ontologies that focus specifically on the aquatic domain (referring to what a content is about), and an educational ontology that has been designed and used to enrich the description of the learning (referring to educational purposes of contents). This latter ontology is based on 5 classes: context, objective, resource, resource feature, user.

All of these presented works are enriching learning resource descriptions by including different semantic aspects, that can help accessing, retrieving and reusing learning objects. We can notice that most of them are using at least one ontology which is the theme to teach domain ontology. This ontology usually contains the specific concepts of the domain (concepts to teach or to learn) and relations between these concepts. The advantage of this ontology is to provide values (that are ontology concepts) of metadata that represent the semantic of the resource content. This facilitates the indexing operation by reducing the metadata vocabularies to the ontology concepts, and enhances the possibilities of learning resources access: accessing resources would be possible in 2 manners:

(1) by navigation through the ontology concepts
(2) by searching a specific resource with specific ontology concepts.

However, we noticed that relations defined between concepts of theme domain ontologies proposed in the previous works, are often semantic relations that are specific to the considered domain. We think that it is useful to include another type of more general semantic relations such as translation relation, synonymy relation, … Translation would be useful in performing multilingual search on resources (as an example, asking for an Arabic resource with a French concept). Synonymy will be helpful for a free search using a concept that is different but equivalent to the one used in the describing resource metadata.

We also noticed that previous work did not provide detailed description of the content type of an educational resource. There is indeed a semantic aspect that indicates the nature of the resource (definition, syllabus, example, exercise, exam, ...). This aspect would be useful to modelize with an ontology of educational resource content, which allows a user to find a particular type of elemental resource and possibly use it to build other new resources.

These are the main ideas, we will propose in our semantic representation model in the following section.

4. An Ontology-Based Model for Accessing and Retrieving Learning Resources

In this work, we are particularly interested in multilingual searching, accessing and reusing various types of educational resources, that may be available through a learning portal. By multilingual search, we mean the ability to find a resource in one language, while the searched concepts are expressed within the user query, in a different language. On the other hand, we want to find, share and reuse different types of learning resources, not only entire courses, but also course fragments (or elements). Our system would thus enable a user interested in Object Oriented Programming (OOP) for example, to ask queries such as:

- I am looking for an *OOP course outline*, written in *Arabic*.
- I am looking for an *English course chapter* dealing with the *inheritance* notion.
- I am looking for a *French definition* of an *abstract class* in OOP.

…

Search and access to learning objects which are course fragments or components encourages reusing these objects to develop new monolingual or multilingual resources.

The semantic representation model of online educational resources, that we propose is based on both e-Learning standards (and particularly on LOM schema), as well as a set of ontologies that we present in the following sub-sections:

4.1. *Learning domain ontology*

This ontology describes the field of learning or teaching. Regarding the field of computer science, we use an existing ontology that corresponds to The ACM Computer Classification system[h]. The 2012 ACM Computing Classification System has been developed as a poly-hierarchical ontology that can be utilized in semantic web applications. It replaces the traditional 1998 version of the ACM Computing Classification System (CCS). Figure 1 is an extract of this ontology:

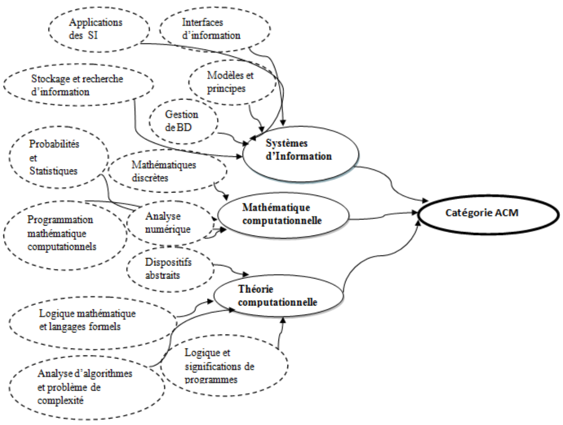


Fig. 1. Learning Domain Ontology

4.2. *Learning theme domain ontology*

We propose to define this ontology, that aims to represent the semantic content of a course (theme to de studied). This ontology contains mainly, the concepts of the course as well as the existing relations between these concepts. These ontology concepts correspond to the course concepts (that are the notions to be presented to learners). As for the relations, we propose a set of general relationships between concepts regardless of the course nature. These relationships can be applied to any course concepts and are not

[h] http://www.acm.org/about/class/2012

specific to a particular theme. On the other hand this ontology constitutes an onto-terminological resource, including terms related to the domain concepts, and expressing several semantic relationships between terms such as synonymy and translation. This makes this ontology very useful for searching multilingual educational resources. The semantic relations between concepts of the course that we propose are in fact of two types:

(1) Semantic relations which are more related to pedagogy and express the way in which course concepts can be studied (how and in what order).
(2) Semantic relations that are more related to terminology and express relations between terms (synonymy, abbreviation, translation). That is why we consider this ontology as an onto-terminological resource.

Considered relations in the theme domain ontology are: « *is a sub-concept of* », « *includes* », « *has as prerequisite* », « *is translated in language L by* », « *is abbreviated by* ». Figure 2 shows this ontology, considering 3 resources languages: English, Arabic and French:

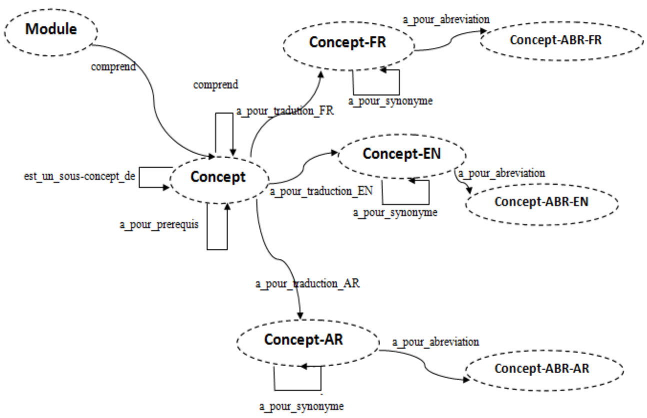


Fig. 2. Learning Theme Domain Ontology

Note that, when browsing this ontology, we can explore the different concepts of the theme or subject to teach or learn, in the different available languages, and thus see for each concept its equivalent in another language.

4.3. *Learning resource ontology*

The third component of our model is a classification model of the content nature of educational resources. This classification is presented in the form of an ontology that gives a resource typology and relations between different types of resources. Learning resource ontology is given in figure 3. It should be noted that:

- in this representation, learning resources may include several course components, which can be nested into each other. A resource may take a particular form (video, audio, image, text, mixed). We also consider different types of content resources (introduction, chapter, conclusion, exercise, definition, discussion, references, lesson plan, a series of exercises, ...).

- our representation model takes into consideration the LOM e-Learning standard, since it describes educational resources using a LOM application profile, which includes LOM description elements, but also adds other elements, such as the learners and teachers opinions (about this resource. This element is given with a score between 1 and 5, that consists in a sort of evaluation, that will be useful in the search of learning resources and can be one of classification criteria of the found resources in response to a user query.

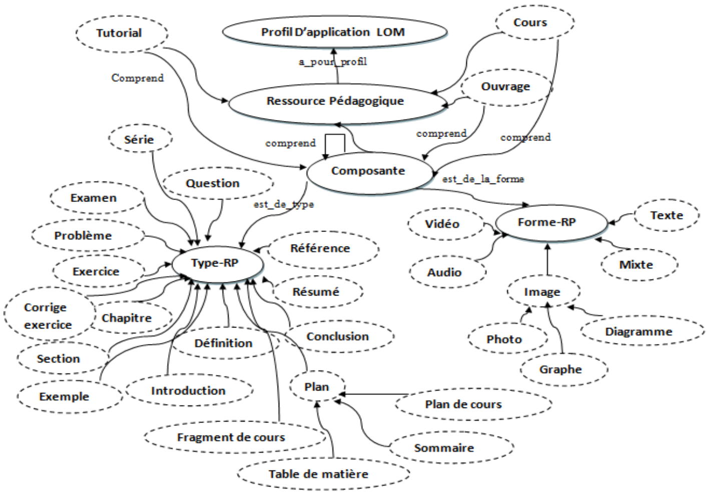


Fig. 3. Learning Domain Ontology

5. Conclusion

In this paper, we proposed a semantic representation model of online multilingual educational resources, that aims to improve their search, access and reuse. This model is based on 3 ontologies: *learning domain ontology*, *learning theme domain ontology*, *learning resource ontology*. Ontology concepts are used for indexing learning resources. The ontological concepts and relations provide a means of guided multilingual navigation in these resources, and are helpful in performing multilingual search of specific learning resources and in facilitating their share and reuse.

The next step of this work is to implement this model by creating real ontologies and developing a multilingual learning portal, where educational resources are indexed using the proposed ontologies. This work has already started using the Protégé Editor [i] for creating ontologies and J2EE for developing the learning portal.

[i] http://protege.stanford.edu/

References

Abel M.-H., *et al.* (2003a) : Gestion des ressources pédagogiques d'une e-formation. Document numérique 2003/1-2, Volume 7, p. 111-128.

Abel M-H., *et al.* (2005b) : Ontologies pour le web sémantique et le eLearning. Actes Plate-forme AFIA'2005 « Web Sémantique pour le e-Learning » , 31 Mai – 3Juin 2005, p. 44-96.

Bianchi S., *et al.* (2009). Use of Ontologies to Annotate and Retrieve Educational Content: the AquaRing Approach. Je-LKS Journal of e-Learning and Knowledge Society — Vol. 5, n. 1, february 2009 (pp. 211 - 220).

Borst W. N. (1997): Construction of Engineering Ontologies. Center for Telematica and Information Technology, University of Tweenty, Enschede, NL.

Bouzeghoub A., *et al.* (2005) : Un modèle de description sémantique de ressources pédagogiques basé sur une ontologie de domaine. *Revue STICEF, Volume 12*, 2005, ISSN : 1764-7223, mis en ligne le 15/03/2006, http://sticef.org. (pp. 1-17).

Chaput, B., *et al.* (2004): Une expérience de construction d'ontologie d'application pour indexer les ressources d'une formation en statistique. In *Actes électroniques* des 36es Journées de Statistique, SFDS, Montpellier, 24-28 mai 2004.

Duitama J.-F., *et al.* (2003): A Framework for the the Generation Of Adaptive Courses based on Semantic Metadata. Proc. MAWIS Workshop, with OOIS'03 Conference, Geneva, september 2003.

Florescu V.; Aleca O. E. (2009): The issue of semantic Modeling of The Learning Organizational Memory for e-Learning. Annales Universitatis Apulensis Series Oeconomica, 11(1), 2009.

Fontaine D.; Benayache A.; Abel M-H. (2006) : L'accès aux ressources d'une formation guidé par des ontologies. International Journal of Technologies in Higher Education, 3(3). 2006.

Fory L. (2004) : Les caractéristiques d'une ressource pédagogique et les besoins d'indexation qui en résultent - La vision du côté TICE (technologies de l'information et de la communication pour l'enseignement). L'indexation des ressources pédagogiques, 2004, enssib, Villeurbanne. Available online, on October 2012 at http://www.enssib.fr/bibliotheque-numerique/document-1234.

Ghebghoub O.; Abel M.H.; Moulin C. (2009): LOMonto : Une ontologie pour l'indexation d'objets pédagogiques. Atelier Construction d'Ontologies GBPOnto. Available on October 2012 at http://www-limbio.smbh.univ-paris13.fr/GBPOnto.2009/web4-ouafiaetall.pdf.

Gruber, T. (1993): A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, Volume 5, Issue 2, 1993.

Hernandez, N., *et al.* (2008): Indexation multi-facettes des ressources pédagogiques pour faciliter leur ré-utilisation. Institut de Recherche en Informatique de Toulouse, available online, on October 2012 at ftp://ftp.irit.fr/IRIT/SIG/2008_RNTI_HMRRS.pdf.

Iles N.; Chikh A.; Chouiti S. M. (2008) : Architecture d'un entrepôt pédagogique et sa modélisation sémantique. International Conference on Web and Information Technologies ICWIT'08.

Lando P. (2006) : Conception et Développement d'Applications Informatiques utilisant des Ontologies: application aux EIAH. 1ères Rencontres jeunes chercheurs en EIAH, RJC-EIAH'2006.

Megder El.; Cherkaoui C.; Mammass D. (2008). Towards a design Methodology for an Organizational Memory based on E-Learning and ontologies. ICGST-AIML Journal, ISSN: 1687-4846, vol. 8, Issue III. Dec. 2008.

Psyché V.; Mendes O.; Bourdeau J. (2003) : Apport de l'ingénierie ontologique aux environnements de formation à distance. Sciences et Technologies de l´Information et de la Communication pour l´Éducation et la Formation, Volume 10, 2003.

Sander E.; Meunier J-M. ; Bosc-Miné C. (2004) : Approche ontologique et navigation dans un EIAH. Le cas de l'enseignement des statistiques. Sciences et Technologies de l´Information et de la Communication pour l´Éducation et la Formation, Volume 11, 2004.

Toledano B.; Baron M.; Giroire H. (2005) : Un outil pour la qualification de ressources, à base d'ontologies. In Dieng-Kuntz, Rose ; Grandbastien, Monique ; Hérin, Danièle (eds), *Plate-Forme AFIA, Web sémantique pour le e-learning*, pp. 39--44.

Wilson R. (2004): The Role of ontologies in Teaching and Learning. *Online Version*. Retrieved 10 14, 2012, from the World Wide Web: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.7066&rep=rep1&type=pdf.

Wiley, D. A. (2002): Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In D. A. Wiley (Ed.), *The Instructional Use of Learning Objects: Online Version*. Retrieved 10 14, 2012, from the World Wide Web: http://reusability.org/read/chapters/wiley.doc .

International Conference on Software Engineering and New Technologies

MODELING KNOWLEDGE MANAGEMENT SYSTEMS FOR COMPONENT-BASED SOFTWARE ENGINEERING

Mohammed Amine Mostefai

ESI (Ecole Supérieure d'Informatique),Algiers, Algeria
m_mostefai@esi.dz

Mohamed Ahmed-Nacer

USTHB (Université Sciences et Technologie Houari Boumediène), Algiers, Algeria
anacer@cerist.dz

Component-Based Software Engineering (CBSE) is a knowledge-intensive discipline where all activities imply the utilization and the transfer of knowledge between collaborators.

A better utilization, transfer and application of this knowledge are indispensable factors to face the challenges of CBSE. In this perspective, Knowledge Management (KM) aims at fulfilling these factors.

Knowledge Management Systems (KMS) are the technological infrastructure making KM concrete. Similar to software engineering techniques, knowledge engineering techniques are used to model the KMS

In this paper, we present our approach to model KMS for CBSE organizations in an organization-independent context. Our approach is based on CommonKADS, one of the most adopted knowledge modeling techniques or the design of KMS.

Keywords: Component-Based Software Engineering, Knowledge Management, Knowledge Management Systems, Knowledge Modeling, CommonKADS.

1. Introduction

Component-based software engineering (CBSE) is a discipline that consists in the development of systems as assemblies of parts (software components), the development of parts as reusable entities, and the maintenance and upgrading of systems by customizing and replacing such parts [Crnkovic (2003)]. A software component is a unit of composition with contractually specified interfaces and explicit context dependencies only. A software component can be deployed independently and is subject to composition by third parties [Szyperski (2002)].

CBSE activities involve many people working in different phases and activities. The knowledge used and produced in these activities is various, growing and hardly traceable: organizations have problems keeping track of what this knowledge is, where it is, and who owns it [Rus and Lindvall(2002)]. In this perspective, Knowledge Management (KM) provides mechanisms to create, share, transfer and apply knowledge.

KM relies on a technological infrastructure called Knowledge Management Systems (KMS) [Alavi and Leidner(1999)]. The design of KMS uses a set of methods similar to

Software Engineering Techniques. These methods are called Knowledge Engineering Techniques.

Our work addresses modeling KMS for the CBSE domain by using a popular Knowledge Engineering technique: CommonKADS.

This paper is structured as follow: the Section II tackles the motivations of implementing KM in CBSE and quotes related work, Section III introduces knowledge engineering techniques and justifies the choice of CommonKADS and describes it, Section IV details the steps we took to model the KMS and finally, the last section is a conclusion and future work.

2. Motivations and Related Work

The main propose of CBSE is to reduce time-to-market delays by using (or reusing) components. However, there are many risks in CBSE that can be summarized in the following points [Crnkovic (2003), Voas(1998)]: 1) an increased time and effort required for the development of components, 2) the consequences of bad component selection, 3) the dependence to the component supplier and 4) the costs of the maintenance of components and sensitivity to their changes.

These risks can be avoided or at least reduced if the knowledge existing in the brain of collaborators, in the corporate memory or outside the organization is effectively captured and used. The motivations of using KM in CBSE are justified by: 1) Capturing and using process knowledge, 2) Capturing and using product knowledge, 3) Identifying knowledge holders, 4) Better risk analysis by consulting internal or external knowledge sources for previous experiences about a component or a component provider, 5) Avoiding making the same errors, 5) reduce the sensitivity to change by using an automated knowledge-based approach for the impact of changing a component and 6) promoting collaboration among individuals.

The implementation of KM does not intend to be a replacement of the current software processes and tools in the organization. It is a support tool that targets making knowledge more available and effectively shared while executing these processes.

One of the most famous achievements in implementing KM in software engineering (SE) is the experience factory brought by the Basili et al. in [Basili, Caldiera et al.(1994)]. However, the experience factory is based on a proprietary methodology called QIM (Quality Improvement Paradigm) and addresses software engineering in general and not especially CBSE. In another hand, while our work focuses only on modeling, [Basili et al.(1994)] reaches the implementation level for the experience factory.

In [Rus and Lindvall(2002)], the authors emphasize on the motivations of implementing KM in SE but don't propose a model for implementation. In [Kneuper(2001)], the author proposes an approach for supporting software processes using KM. Again, the solution is not based on a standard methodology and there is no particular focus on CBSE.

3. Designing the Knowledge Management System

The implementation of KM in the organization is achieved by a special class of information systems called knowledge management systems (KMS). KMS are the tools for managing knowledge, helping organizations in problem-solving activities and facilitating decisions making [Speel, Schreiber et al.(2001)].

KMS inherit from Knowledge-Based systems (KBS). KBS are designed using knowledge engineering techniques which are similar to software engineering techniques but with an emphasis on knowledge rather than data processing [Studer et al.(1998)]. Knowledge engineering techniques are based on knowledge modeling which takes a special form of requirement engineering.

A knowledge model provides an implementation-independent specification of knowledge in an application domain [Speel, Schreiber et al.(2001)].

The development process of a KBS is similar to the development of any other software system; phases such as requirements elicitation, system analysis, system design, system development and implementation are common activities.

The stages in KBS development are: business modeling, conceptual modeling, knowledge acquisition, knowledge system design and KBS implementation.

Modeling contributes to the understanding of the source of knowledge, the inputs and outputs, the flow of knowledge and the identification of other variables such as the impact that management action has on the organizational knowledge [Davenport and Prusak(2000)].

Knowledge modeling is used to provide a conceptual view of the KMS. The constructed models make the KMS more explicit and faster to develop.

There are many techniques of knowledge modeling such as CommonKADS [Schreiber and Akkermans(2000)], Protégé 2000 [Grosso et al.(1999)] or MIKE [Angele et al.(1996)].

We have chosen CommonKADS because of the pertinence of the models it includes and their large functional coverage. In addition to that, CommonKADS is one of the most used methodologies for modeling knowledge. Moreover, the models of CommonKADS permit to give a very detailed view of all the aspects of the KMS by using the UML notation, largely used in Software Engineering community.

Protégé 2000 has a very intuitive user interface and provides a very strong way to model knowledge but it’s not really a knowledge engineering methodology. It focuses rather on creating and editing ontologies.

MIKE is suitable to a KMS in a development state but does not support standards such as UML.

The CommonKADS has evolved from the KADS to the formulation of [Schreiber and Akkermans(2000)]. The CommonKADS methodology provides an integrated framework for knowledge management, which also encapsulates knowledge engineering activities.

CommonKADS supports structured knowledge engineering techniques, provides tools for corporate knowledge management and includes methods that perform a detailed analysis of knowledge intensive tasks and processes.

CommonKADS incorporates an object-oriented development process and uses UML diagrams such as class diagrams, use-case diagrams, activity diagrams and state diagrams. CommonKADS also has its own graphical notations for task decomposition, inference structures and domain schema generation.

Common KADS addresses the complexity of the system by using different models [Schreiber and Akkermans(2000), Motta(2001)].

The organizational model provides a view of the organization where the KMS will be deployed. It also addresses the impact of the KMS on this organization.

The task model refers to the characteristics of the business processes such as: the inputs and outputs, the preconditions, performance and quality, the function of the agents that will carry out the processing, the structural coupling of those agents, the flow of knowledge between the agents, their overall control, the knowledge and competences of the agents and the resources available to deliver the business process.

While the task model specifies what needs to be done, the agent model specifies who does it. The task model describes "the characteristics of agents, their competencies, authority to act, and constraints in this respect". The term 'agent' in CommonKADS has a generic connotation: an agent can be a human being, a robot or a software program.

The knowledge model is used to describe the application-related knowledge used to perform tasks and the role of the knowledge in problem-solving activities. The knowledge model of CommonKADS has three categories of knowledge: task knowledge that describes the order of execution for the reasoning (inference) steps, inference knowledge that describes the reasoning step (inference) performed using the domain knowledge and the domain knowledge that describes the application concepts.

The communication model describes the inter-agent communication needed when performing the tasks. The design model specifies the target software and hardware platform, the various software modules included in the target system, their functional and technical specifications and the mapping between these modules and the conceptual components identified during the analysis phase.

In CommonKADS, designing the KMS entails the construction of the models listed above. The models are structured in three levels, the context level, the concept level and the artefact level. Each level is composed of one or many models and each model needs the fulfilling of the corresponding worksheets.

The models of the context level are the input for the concept level models which are, in their turn, the input to the final model (the design model).

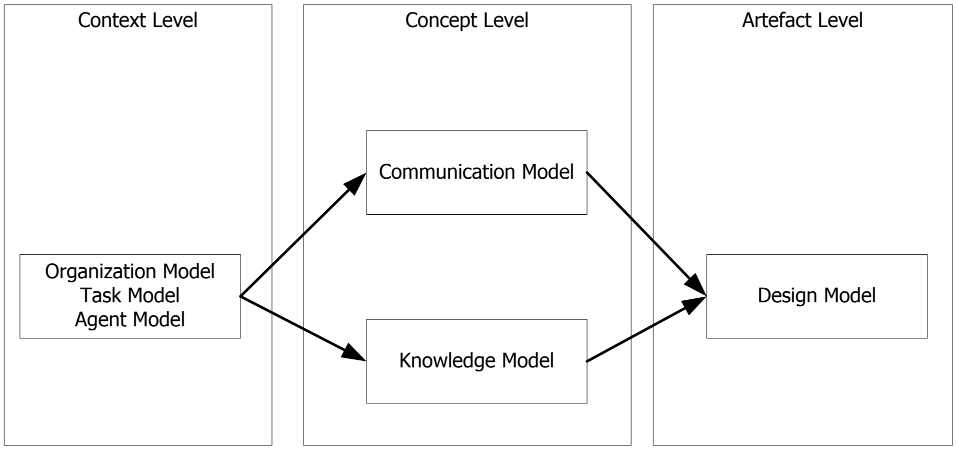


Fig 1. The CommonKADS models

The context level includes the organization model, the task model and the agent model. The organization model (OM) focuses on the organization structure, the processes, the staff and resources [AIFB(2001)]. The OM is composed of five worksheets OM-1 (problems, opportunities and solutions), OM-2(structures and processes), OM-3(detailed tasks), OM-4(knowledge assets) and OM-5(feasibility). The task model focuses on the processes executed in the organization and is composed of the worksheets TM-1 (Tasks details) and TM-2 (Tasks knowledge). The agent model concerns the staff profiles in the organization and their missions. This model includes one worksheet AM-1(Agent description). An additional worksheet OTA-1 (impacts and improvement decisions) summarizes this level.

The concept level includes the communication model (Worksheet CM-1 and CM-2) and the knowledge model. The artefact level involves the design model which is the technological infrastructure of the KMS.

4. Modeling Knowledge Management Systems for CBSE

4.1. *Modeling Considerations*

The models of CommonKADS are designed with the vision of being independent from organization-specific aspects such as tools, methods or culture.

In the implementation phase, each organization has to adapt the models to its own context by extending the model and / or including organization-specific considerations in the implementation phase.

In this phase, we apply the CommonKADS to construct the models level by level and consequently, in three phases. In the first phase, we build the context level models: the organization model, the task model and the agent model. In the second phase, we build the knowledge model and the communication model. Finally, the models built in the previous phases will be the input for the design model.

4.2. *Phase 1- The Context Level*

4.2.1. OM-1, Problems and Opportunities

This worksheet focuses on the essential problems (or opportunities) concerning CBSE in an organization, attempts to contextualize these problems and provides some propositions of solutions. The main targeted problems are the motivations of implementing KM in SE discussed in section II. The problems are divided in two categories. The first category concerns general software engineering problems summarized in the motivations provided by [Rus and Lindvall(2002)] such as capturing processes knowledge, making employees more efficient and indirectly, enhancing quality and reducing time to market. The second category concerns CBSE-specific problems, outlined in [Crnkovic (2003)] or [Takeshita(1997)] such as the impacts of bad component selection, insufficient analysis / design, architecture mismatch or the costs of development and maintenance.

All the problems related to both categories have formal and informal methodologies. For example, all (formal and formal) component evaluation methodologies cited in [Goulão and Abreu(2004)] are intensively based on knowledge. An efficient management of this knowledge will be an excellent support to these methodologies.

Knowledge management does not intend to be the direct solution to all the problems. It is a supporting solution for 1) the existing software processes, 2) meeting the current challenges in SE and CBSE and 3) improving the usage, transfer and application of software team’s individual and collective knowledge.

4.2.2. OM-2, Structure and processes

In this worksheet we focus on the organization structures, people involved, the corresponding processes, the utilized knowledge and resources and the related culture.

The main CBSE structures can be projected on CBSE activities: each activity will be a logical structure. The CBSE activities, by inheritance to standard SE, can be structured in two categories: product-oriented activities and process-oriented activities [Rombach and Verlage(1995)].

The process-oriented activities include improvement and modeling of processes. The product-oriented activities are subdivided in two categories: managerial activities (product management, project management, project data management...) and technical activities [Crnkovic et al.(2006)] (requirements, design, implementation, integration, component selection, component adaptation, component integration, test, release, maintenance...).

Each activity determines involved people, the used resources and the manipulated knowledge. For example, Test phase involves testers, quality assurance tools, automated test tools, test-oriented knowledge ...etc. Implementation involves coders, integrated development environments, compilers, etc. Component selection involves component selectors, developers, project manager, component database, internet, component-oriented knowledge, etc.

Each process will be described using UML notation. For example, the activity diagram (fig 2) provided in the example illustrates a typical component testing activity involving two user profiles (tester, developer) and various knowledge assets (requirements, test plans ...).

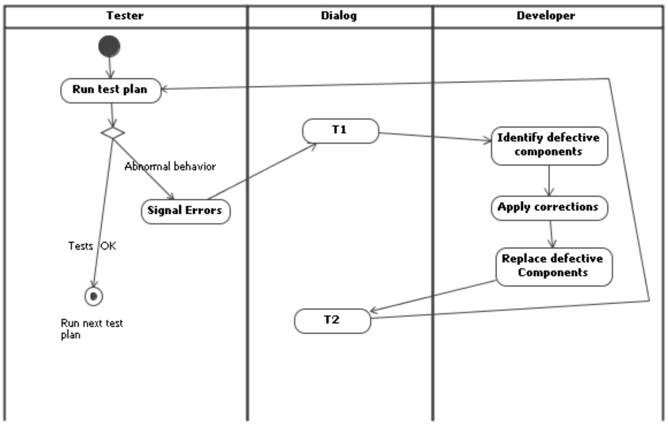


Fig2. Example of an activity diagram illustrating the component testing process

4.2.3. OM-3, Process Breakdown

In this phase, the processes listed in OM-2 are detailed using the OM-3 worksheet provided by CommonKADS. In this phase, every single process is assigned a unique identifier and a name. For each task, we indicate 1) who performs it, 2) where it is performed, 3) the knowledge it requires, 4) if it is knowledge-intensive and 5) its significance.

The table below (Table I), gives our example of process breakdown. The processes we took in the examples are a subset of CBSE activities based on the general process model for CBSE provided by [Dogru and Tanik(2003)] and [Crnkovic et al.(2006)].

ORGANIZATIONAL MODEL		OM-3, PROCESS BREAKDOWN				
TASK NO	TASK	PERFORMED BY	WHERE?	KNOWELEDGE ASSET	INTENSIVE?	SIGNIFICANCE
1	SOFTWARE SYSTEM SPECIFICATION	P. MANAGER ARCHITECT	MANAGEMENT ARCHITECTURE	Specification documents, individual experience	YES	5
2	DECOMPOSITION	P.MANAGER ARCHITECT	MANAGEMENT ARCHITECTURE	Individual experience, previous projects packages, components interfaces	YES	5
3	COMPONENT SPECIFICATI	P.MANAGER ARCHITECT	MANAGEMENT	Individual experience, templates	YES	5

	ON		ARCHITECT URE			
4	COMPONENT SEARCH	P.MANAGER ARCHITECT DEVELOPER	MANAGEM ENT ARCHITECT URE DEV WORKSHOP S	Component database, internet, individual experience	YES	5
5	COMPONENT MODIFICATI ON	DEVELOPER	DEV WORKSHOP S	Component documents and manuals, individual experience	YES	5
6	COMPONENT CREATION	DEVELOPER	DEV WORKSHOP S	Developer knowledge, manuals, trainings, documents, Source codes	YES	5
7	INTEGRATIO N	DEVELOPER	DEV WORKSHOP S	Developer knowledge, reports, components database, internet	YES	5
8	TEST	TESTER	TEST DEP	Tester knowledge, test templates, trainings, documents	YES	5

Table 1 - EXAMPLE OF CBSE PROCESSES BREAKDOWN

All the tasks listed in (table 1) are knowledge-intensive and make usage of two types of knowledge: tacit knowledge (individual and informal knowledge stored in individual's brain: individual knowledge and experience) and explicit knowledge (explicitly formalized and generally stored in an electronic format: documents, databases, manuals, internet...) [Nonaka and Konno(1998)].

For example, because of administrative and technical reasons, the component search can be done by managers, architects and developers. This task is achieved in the corresponding (logical) structures and uses both kinds of knowledge.

4.2.4. OM-4, Knowledge Assets

In this step, for each knowledge asset listed in the previous step (OM-3), we specify the agent possessing the asset, the tasks where the asset is used in, it is in the right form, in the right place, in the right time and in the right quality [Schreiber and Akkermans(2000)].

For example, the generic knowledge asset "individual experience" is used in all tasks. The knowledge asset "test templates" is used in the tasks associated to test and is possessed by the user with "tester" profile. The questions "right place?", "right time?", "right form?" and "right quality" depends on the organization and the culture of the environment.

Because of its importance and sensitivity, the challenge is to transform the tacit knowledge (individual experience) to a more exploitable form (externalization process [Nonaka and Konno(1998)]). However, in most of times, it is not easy to do. The designers of the KMS and the KM project managers have to give importance to the socialization process too (direct and informal knowledge sharing among individuals).

4.2.5. OM-5, Feasibility decisions

The OM-1 to OM-4 steps had for role to specify the context of CBSE (organization structure, agents, processes, knowledge). The OM-5 establishes judgment on feasibility of implementing the KMS in the context of CBSE. To study the feasibility of the KMS project, we study the factors which can slow down the project or cancel it.

The factors that can eventually hamper KM implementation are structured in three axes: 1) financial factors, 2) technical factors and 3) social factors.

The most significant financial factor is that KM implementation will involve more people and will require more time and consequently more money. The project managers are concentrated on finishing current projects on time rather than focusing on long-term additional projects.

However, even if KM implementation would be costly in the short term, it addresses risks that can be fatal to the organization such as 1) knowledge loss, 2) knowledge lack, 3) repeating mistakes or 4) unavailability of people owning key knowledge [Rus and Lindvall(2002)].

The technical factors are that 1) organizations have already a set of installed tools and PSEE (Process-centered Software Engineering Environments) and the integration of the KMS should be light and transparent, and 2) the complexity of representing and using tacit knowledge.

Fortunately, most software engineering tools and environments are built on mechanisms allowing an easy integration such as APIs or web services [Wicks(2006)]. Moreover, the SE domain has the particularity that people are accustomed to information systems and IT tools and that a very significant part of data is already stored in an electronic form [Schneider(2001)].

The knowledge model, discussed in the next sections, will address the knowledge representation problem.

Finally, social factors such as the non-development of share culture, reluctance to share and / or receive knowledge, fear of sharing negative experiences [Rus and Lindvall(2002)] could make such a project a failure.

To face these issues, implementing KM does not rely only on a technological infrastructure; it is accompanied by a suitable methodology and strategy to ensure success factors. These factors include motivational practices, knowledge champions, corporate awareness campaigns, anonymous participation, ...etc. [Rus and Lindvall(2002), Davenport et al.(1997)].

4.2.6. TM-1, Tasks descriptions

In this step, each task (obtained in OM-3) is analyzed from two points of view: managerial and informational. The managerial point of view details the goal of the tasks, the added-values and measures used to qualify the quality and the performance of the task [Schreiber and Akkermans(2000)].

The informational point of view discusses the functional view of the task (schematized by an activity diagram), the structure of manipulated objects, their relationships (schematized by a class diagram) and the description of the temporal order of and control over the subtasks (schematized by a state diagram) [AIFB(2001)].

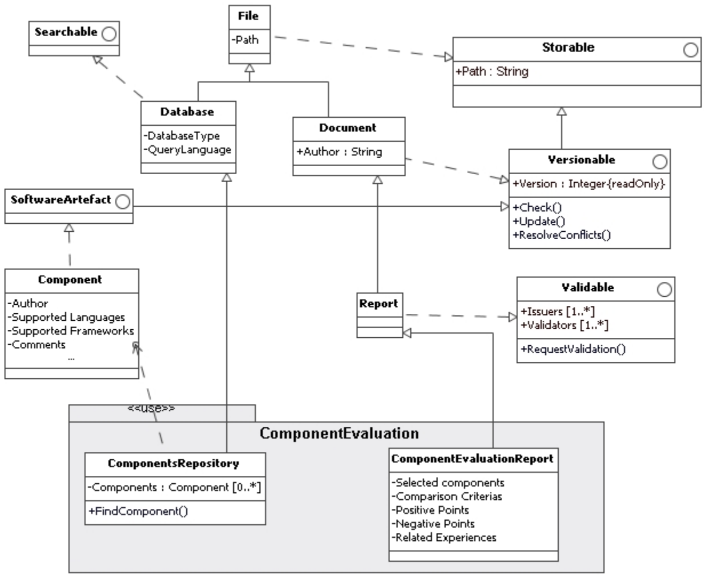


Fig. 3. Component Selection objects modeled by a UML class diagram

For example, figure 3) models a subset of component selection activity objects. The two main objects are the evaluation report and the component repository.

The evaluation report inherits from report and indirectly from document and file. Because the evaluation report is a report, it is a “Validable” object and because it is a document, it implements the “Storable” and “Versionable” interfaces. In another hand, “ComponentRepository” is a database containing a set of components versions. While the static aspect of the task is described using class diagrams, dynamic aspects are model using activity diagrams or state diagrams.

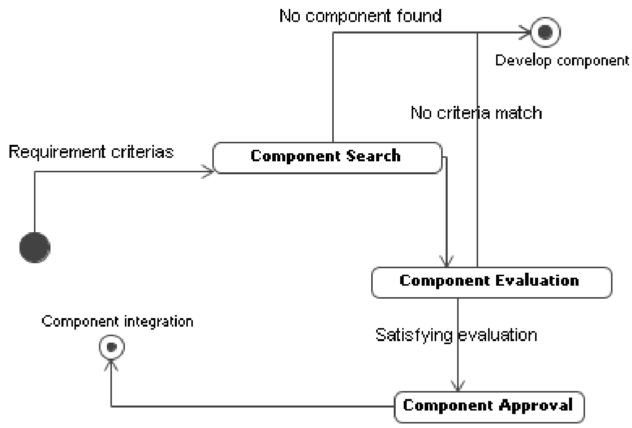


Fig. 4. Component Selection subtasks modeled using UML states diagram.

For the example above (fig 4), the subtasks of the selection induce three states: component search, component evaluation and component approval. In the first state, the component repository modeled previously is used. In the other states, the component evaluation report is used.

4.2.7. TM-2, Task Analysis

This step is a refinement of the OM-4 step, where we specify for each knowledge asset of OM-4 the following criteria: the knowledge nature (formal, rigorous, empirical, heuristic, action-based, incomplete...), the knowledge form (mind, paper, electronic ...) and the knowledge availability (limitation in time, limitation in space, limitation in access...,etc.). In this step, we have to specify if each criterion entails bottlenecks or have to be improved.

The analysis in this step cannot be generic and is toughly organization-dependent. For example, a test plan can be more formal in an organization than another depending on the adopted culture and methods.

4.2.8. AM-1, Agent model

In this step we focus on the various roles and competencies in a general CBSE environment. Using the worksheet, we indicate for each agent: his name, his organization (basing on OM-2), the tasks he is involved in (TM-1), the agents he communicates with, knowledge he possesses (OM-4), any additional competencies and his responsibilities and constraints.

In addition to AM-2 worksheets, UML use case diagrams can be used to schematize agent actions. For example, the table 2 illustrates the AM-2 worksheet for the "component selector" role.

Agent criteria	Value
NAME	COMPONENT SELECTOR
ORGANIZATION	COMPONENT SELECTION COMPONENT EVALUATION DEVELOPMENT
INVOLVED IN	COMPONENT SELECTION TEST
COMMUNICATES WITH	PROJECT MANAGER DEVELOPER TESTER
KNOWLEDGE	INDIVIDUAL EXPERIENCE COMPONENT EVALUATION REPORTS
OTHER COMPETENCIES	(ORGANIZATION-DEPENDANT)
RESPONSIBILITIES AND CONSTRAINTS	- FIND AND EVALUATE COMPONENTS - THE EVALUATION QUALITY IS VITAL TO THE SUCCESS OF THE PROJECT

TABLE 2 - AM-1 WORKSHEET FOR THE ROLE "COMPONENT SELECTOR"

4.2.9. OTA-1, Impacts and Improvements

This is the final step of the context phase. It is about the impacts of the implementation of the KMS and the changes that will have to be adopted for that.

In a modeling point of view, the availability, the efficiency of transfer channels and the quality of knowledge are vital factors to improve the various CBSE tasks and activities.

If correctly implemented and accompanied with a suitable strategy, the KMS will be widely benefic and will be a part of the solution of all knowledge-intensive tasks problems. In an implementation point of view, each organization has to adapt and / or ameliorate the model to fit its culture and integrate with its tools. In that sense, OTA-1 is widely, organization-dependent.

4.3. *The Concept Level*

4.3.1. The KNOWLEDGE model

a) Building the Knowledge Model

The knowledge model of an application provides a specification of the data and knowledge structures required for the application [Speel et al.(2001)].

The knowledge is structured in three categories: 1) domain knowledge including static concepts and domain types, 2) task knowledge which is a functional decomposition of the activities and is goal-oriented and 3) inference knowledge that includes basic inference steps (using domain concepts) to undertake an action.

b) Domain and Task Knowledge

The domain knowledge captures static structures and types concerning CBSE. Ontologies are used to specify this knowledge. Ontology is defined as an explicit specification of the terms and their meaning in the domain of interest [Speel et al.(2001)]. The role of ontologies is to provide a clear and consensual knowledge structure, reduce the ambiguity due to terminological conflicts and enhance knowledge share and reuse.

Ontologies specify knowledge using classes (concepts) and slots (properties). While ontologies handle the formal aspect of concepts as classes, knowledge bases hold instances of these concepts as class instances.

There has been many attempts to define ontologies related to SE [Ruiz and Hilera(2006)], but in order to reach the consensus, the most promising achievements are based on the project SWEBOK [IEEE(2004)]. The project SWEBOK aims at providing a standardized set of all SE engineering definitions and concepts.

Most large-coverage SE ontologies, such [Mendes and Abran(2005)] or [Sicilia et al.(2005)] are based on the SWEBOK body of knowledge. For our work, we will use the ontology (and ontology development methodology) of [Mendes and Abran(2005)] because it provides a high level of details (more than 4000 concepts) and a clear methodology [Mendes and Abran(2004)].

According to the SWEBOK project, the SE domain has been divided in ten categories representing ten knowledge areas in the ontology: requirements, design, construction, testing, maintenance, configuration management, software engineering management, software engineering processes, tools and methods and finally, software quality [IEEE(2004)].

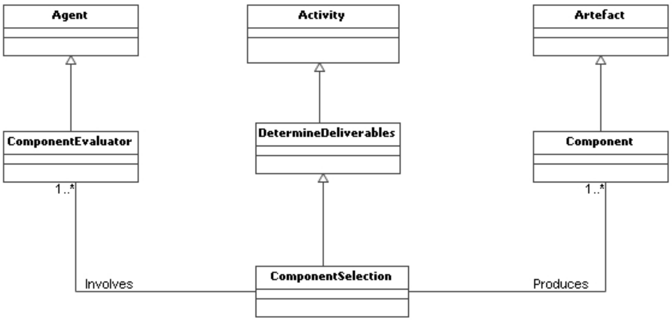


Fig. 5. The ComponentSelection related concepts derived from SWEBOK concepts

While each knowledge area possesses its own concepts and relations, some “shared concepts” are used in more than one knowledge area. These shared concepts inherit essentially from the concepts Artefact, Activity or Agent [Abran et al.(2006)].

For example (figure 5), a software component is a shared concept descending from the concept “Artefact”. Component selection activity inherits from the “determine deliverables” activity which belongs to “software engineering management” knowledge area [IEEE(2004)].

c) Inference knowledge

While domain and tasks knowledge address CBSE specific concepts, inference knowledge is about basic reasoning tasks explicitly or implicitly used in every-day activities.

Inference knowledge is represented using functions and roles. A function is a basic reasoning step included in one or many tasks. A knowledge role is knowledge used as input or output for functions [Schreiber and Akkermans(2000)].

The inferences are the lower level of tasks hierarchy. Each CBSE task implies one or generally a set of inferences. In another hand, a single inference can be used by more than one task.

For the modeling of the KMS, even if the activities enumerated in [IEEE(2004)] are in a high level (tasks), the modeling of inference will be achieved by decomposing these tasks and extracting basic reasoning activities as inferences.

Here again, inferences will be categorized in two groups: shared inferences used in more than a knowledge area and specific inferences exclusive to only one knowledge area.

For example, the “IntegrateComponent” inference (figure6) has for input the component to integrate, the module to integrate into, the integration manual and an environment for integration (IDE).

This basic inference is used in the “Coding” activity (integrate a component), individual component test (integrate the component in test projects) or in the maintenance process (replace a component by another more satisfactory component or another version of the same component).

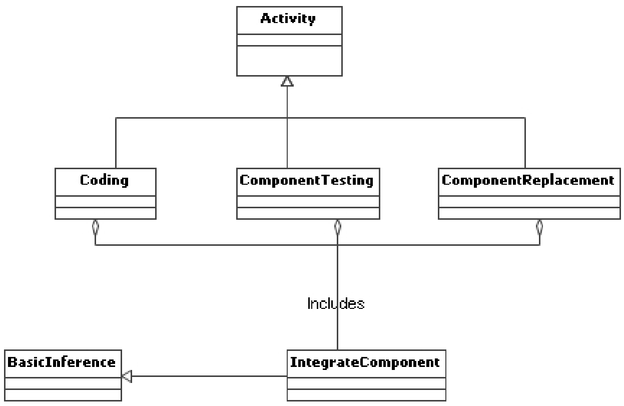


Fig. 6. The inference “IntegrateComponent” used in more than one task

4.3.2. The Communication Model

The communication model provides descriptions about communications between the different agents. It is the model that specifies knowledge transfer between agents.

In the communication model, communications between agents are composed of communication units called “transactions”. In a lower level, each transaction is composed of a set of messages.

Consequently, the process of building the communication model goes in three layers: 1) the communication plan which governs the full dialog between the agents, 2) the individual transactions that link two (leaf) tasks carried out by two different agents and 3) the information exchange specification that details the internal message structure of a transaction [Akkermans et al.(1998)].

In the process of constructing the communication model for CBSE, we list all the tasks (Task model) that involve more than one agent and detail the transactions engendered by these tasks.

The CommonKADS methodology provides two worksheets CM-1 and CM-2 allowing the construction of this model [Schreiber and Akkermans(2000), Akkermans et al.(1998)].

The CM-1 worksheet is applied to each transaction in the deducted communication plans. It specifies the transaction name, the knowledge elements included, the agents involved and the owner communication plan. The CM-2 details the messages structure of each transaction.

Messages are classified according to their types and their purpose: communication init messages (START, FINISH), acceptance messages (ACCEPT, REJECT), information messages (INFO, WARNING, ERROR) and request messages (REQUEST, PROPOSE). Each message indicates the agent who sent it, the reference, the contents, eventual attachment and the receivers.

4.4. *The Artefact Level : The Design Model*

The design model provides a description of the software modules that will compose the KMS, describing the used data structures and preparing the implementation. The models generated by the analysis phase are the input for the design model. However, to remain context-independent, we only provide the software architecture of the KMS. The implementation details will be individually processed by the organizations according to their platforms and tools.

4.4.1. The KMS functions

In addition to the standard function provided in the built model, the KMS should ensure the following additional functions [Staniszkis(2003)]: 1) Contents repository: storage of knowledge instances, 2) Knowledge dissemination: making knowledge accessible to KMS users, 3) Content integration: Integrating contents provided by external sources (documents, files, databases, etc.), 4) Collaboration: communication mechanisms among KMS users (mails, chats, wikis, forums, etc.), 5) Security: protecting knowledge from loss disallowed access

4.4.2. The KMS functional model

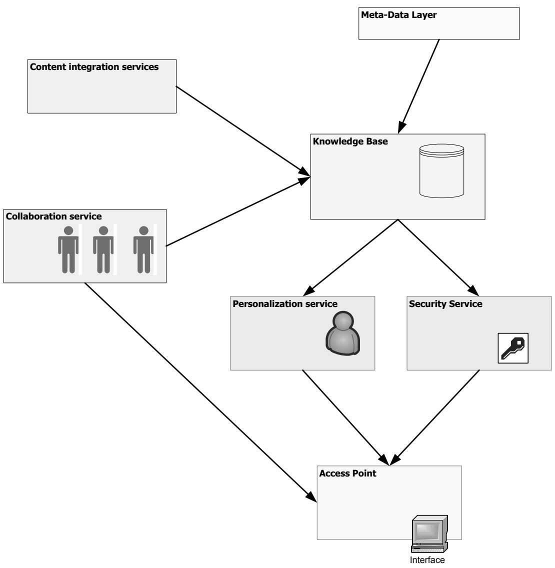


Fig. 8. The KMS functional model

The KMS functional model (see figure 8) provides the architectural foundation for implementing the KMS in the organization.

According to the requirements expressed in [Staniszkis(2003)] and the models created in the analysis phase, the modules composing the functional model are:

a) The meta-data layer

The meta-data layer includes all the knowledge definitions. It contains the classes and relationships expressed in the knowledge model. To suit the organization requirements, the classes and relationships are structured as an ontology extending the ontology provided in the knowledge model.

b) The Knowledge Base

While the meta-data layer stores the definitions of classes, the knowledge base (KB) will store all the instances of these classes called knowledge instances. The knowledge base allows the persistence of objects and provides mechanisms to query the KB. For example, the KB module can be implemented using relational databases or object databases.

c) Content Integration Services

Ensures the automatic collection of knowledge from corporate or external knowledge sources such as files, source codes, databases, document repositories, internet, external KMS,.., etc.

d) Personalization Service

Adapts (form and content) the knowledge elements to the profile of the current user. For examples, filters information judged “too technical” for the “manager” profile.

e) Security Service

Authenticates collaborators and processes and defines their access level and action perimeter.

f) Collaboration service

Provides a set of tools allowing the collaborators to be synchronously or asynchronously linked (for example forums, chats, mailings …). The collaboration service implements all the transactions listed in the communication model. In addition to that, this service provides a gateway to the content integration service for capturing knowledge created or transferred in communications (chat captures, resolved forum queries, etc.).

g) The Access Point

The access point allows entering knowledge queries, creating or updating knowledge instances and evaluating existing knowledge instances.

5. Conclusion and Future Work

Our work was about modeling Knowledge Management Systems for Component Based Software Engineering.

Among Knowledge Engineering Techniques, we have chosen CommonKADS because of the models it includes target all the aspects of the KMS from knowledge assets to functional and communication behavior. When applying CommonKADS methodology, we have considered as a constraint the independence from organization-specific context such as tools, methods or culture and the inclusion of integration of the KMS with the organization tools.

In an organization-independent context, we constructed the CommonKADS models: the task model, the agent model and the communication model.

For the knowledge model, instead of developing an ontology from scratch, we adopted a consensual ontology constructed on the basis of the SWEBOK project.

We finally brought a design model serving as basis for the implementation of KMS in the context of the organization.

Organization-specific context will be integrated by extending the KMS models and also in the implementation phase.

However, this paper focuses only on the modeling aspect. There are some practical and technological issues that have to be addressed to list guidelines for the implementation of the KMS.

References

Abran A.; Cuadrado J. J, García-Barriocanal 01E., Mendes O., Sánchez-Alonso S., and Sicilia M. A., "Engineering the Ontology for the SWEBOK: Issues and Techniques " in Ontologies in Software Engineering and Software Technology: Springer-Verlag, 2006.

AIFB, "The CommonKADS Methodology," Institut AIFB Universität Karlsruhe 2001.

Akkermans H.; Gustavsson R., and Ygge F., "An integrated structured analysis approach to intelligent agent communication," in IFIP World Computer Congress, Information Technologies and Knowledge Systems Vienna/Budapest: Austrian Computer Society/IFIP, 1998, pp. 305-318.

Alavi M. and.Leidner D. E, "Knowledge management systems: issues, challenges, and benefits," Communications of the AIS, vol. 1, 1999.

Angele J., Fensel D., and Studer R., "Domain and task modelling in MIKE," In A. Sutcliffe et al. (Eds.), Domain Knowledge for Interactive System Design, Chapman & Hall, 1996.

Basili V. R; Caldiera G., and Rombach D., "The Experience Factory," Encyclopedia of Software Engineering, vol. 2, 1994.

Crnkovic I., "Component-based software engineering: New challenges in software development," CIT. Journal of computing and information technology, pp. 151-161, 2003.

Crnkovic I.; Chaudron M., and Larsson S., "Component-Based Development Process and Component Lifecycle," in Software Engineering Advances, 2006.

Davenport T. H., De Long D.W., and Beers M.C., "Building Successful Knowledge Management Projects," Center of Business Innovation 1997.

Davenport T. H. and Prusak L., Working Knowledge: How organizations Manage What They Know, 2000.

Dogru A. H. and Tanik M.M., "CommonKADS Context Models," IEEE Software, vol. 20, pp. 34-41, 2003.

Grosso W. E.; Eriksson H.; Fergerson R. W.; Gennari J. H., Tu S. W. and Musen M. A., "Knowledge modeling at the millennium (The design and evolution of Protégé-2000)," Proceedings of the Twelfth Workshop on Knowledge Acquisition, Modeling, and Management, Banff, Canada, 1999.

Goulão M. and Abreu F. B., "Software Components Evaluation: an Overview," 5a CAPSI, 2004.

IEEE, Guide to the Software Engineering Body of Knowledge 2004 version: IEEE Computer Society, 2004.

Kneuper R., "Supporting Software Processes Using Knowledge Management," Handbook of Software Engineering and Knowledge Engineering, vol. 2, 2001.

Mendes O. and Abran A., "Software Engineering Ontology: A Development Methodology," Metrics News, vol. 9, pp. 68-76, 2004.

Mendes O. and Abran A., "Issues in the development of an ontology for an emerging engineering discipline," in First Workshop on Ontology, Conceptualizations and Epistemology for Software and Systems Engineering (ONTOSE) Alcalá de Henares, Spain, 2005.

Motta E., "The knowledge modelling paradigm in knowledge engineering," Chang, SK (ed.), Handbook of Software Engineering and Knowledge Engineering, vol. 1, 2001.

Nonaka I. and Konno N., "The concept of "Ba": building a foundation for knowledge creation," California Management Review, vol. 40, pp. 40-54, 1998.

Rombach H. D. and Verlage M., "Directions in Software Process Research," Advances in Computers, vol. 41, 1995.

Ruiz F. and Hilera J. R., "Using Ontologies in Software Engineering and Technology," in Ontologies in Software Engineering and Software Technology: Springer-Verlag, 2006.

Rus I. and Lindvall M., "Knowledge management in software engineering," IEEE Software, vol. 19, pp. 26-38, 2002.

Schneider K., "Experience Magnets - Attracting Experiences, Not Just Storing Them," in Product Focused Software Process Improvement, PrOFES'01 Kaiserslautern, Germany, 2001.

Schreiber G. T. and Akkermans H., "Knowledge Engineering and Management: The CommonKADS Methodology,," in MIT Press, Massachusetts, 2000.

Sicilia M.Á; Cuadrado J.J, García E.; Rodríguez D., and J. R. Hilera, "The Evaluation of ontological representations of the SWEBOK as a revision tool," in International Computer Software and Application Conference (COMPSAC) Edinburgh, UK, 2005.

Speel P. H.; Schreiber A. T.; van Joolingen W.; van Heijst G., and Beijer G. J., "Conceptual Models for Knowledge-Based Systems,," Encyclopedia of Computer Science and Technology, Marcel Dekker Inc, New York, 2001.

Staniszkis W., "Feature Requirements of a Knowledge Management System," Rodan Systems S.A. 2003.

Studer R.; Benjamins V. R. and Fensel D., "Knowledge engineering: principles and methods," Data & Knowledge Engineering, vol. 25, pp. 161 - 197, 1998.

Szyperski C., "Component Software. Beyond Object-Oriented Programming. Second Edition," in Addison-Wesley, 2002, p. 41.

Takeshita T., "Metrics and risks of CBS," international symposium on software tools and technologies, IEEE, 1997.

Voas J., "COTS software: the economical choice?," Software, IEEE, vol. 15, pp. 16-19, 1998.

Wicks M. N., "Tool integration in software engineering: an annotated bibliography. Technical Report.," Heriot-Watt University 2006.

International Conference on Software Engineering and New Technologies

DESCRIPTION OF CONNECTIONIST SYSTEMS BY THE MDA APPROACH*

Sohaib HAMIOUD

Computer Science Department, Badji Mokhtar University,
Annaba, Algeria
sohaib21000@hotmail.com

Djamel MESLATI

LRI Laboratory, Badji Mokhtar University,
Annaba, Algeria
meslati_djamel@yahoo.com

Fadila ATIL

LRI Laboratory, Badji Mokhtar University,
Annaba, Algeria
atil_fadila@yahoo.fr

MDA (Model Driven Architecture) aims to provide a conceptual, technological and methodological framework, in which models are central to the activities of software engineering. The classic use of MDA is to describe systems as models and their development as model transformations. Using MDA to describe the functioning itself of a system is not common or easy to perform. In this paper, we suggest the modeling of connectionist systems by models and their functioning by model transformations. A model is used to model a system in a given state among the different states that exist during its functioning. The interest is to generate from a particular model, a new model just by applying a model transformation. The generated model describes the system in the next state.

Keywords: Connectionist systems; model driven architecture; model transformation.

1. Introduction

MDE (Model-Driven Engineering) advocates the use of models during software development, to capture the structure of different concerns of a software system, at different levels of abstraction, whether business or technical concerns [3]. The model represents a key concept of MDE and its importance can be compared to that of the concept of object in object-oriented programming [1].

The MDE relies mainly on the initiative MDA (Model-Driven Architecture) published in 2000 by OMG (Object Management Group) [10]. This approach has caused drastic changes in the software development cycle. It is to separate the functional specifications of a system from details of its implementation on a given platform. For this, MDA defines an architecture specification structured in several types of models. Each type corresponds to a particular concern in the system.

MDA suggests mapping of abstract model associated with the functioning of a software system to platforms, and thus to different implementation technologies [3]. The result of this mapping is a platform-specific model that belongs to a more concrete level of abstraction. From the latter, the executable code of the system can be generated.

In this article, and contrary to the conventional use of MDA, the description by models does not affect the development process but the system's functionning itself. Systems involved in our study are connectionist systems. The connectionist approach is the most natural approach to mimic the massive parallelism inherent in the processing performed by the human brain [11]. It is based mainly on Artificial Neural Networks (ANNs).

An ANN is a structure composed of a set of artificial neurons connected to each other via connections. It is characterized by an ability to learn and allows the solving of many computational problems not easily solvable by conventional computers (e.g. pattern recognition, character recognition, etc.).

In our approach, and based on the MDA approach, the functioning of connectionist systems is described as a succession of states in which these systems can be found. Perform the functioning means that the modeled system must pass through all the states until the end of the functioning that corresponds to the final state of the system. This is where the model transformation takes place; a model is used to model a system in a given state among the different states that exist during its functioning. To move from one state to another, just apply a model transformation. It represents the processing operated on the input model to create an output model that describes the system in the next state.

This paper is organized as follows: Section 2 describes the connectionist approaches and their principles. Section 3 is devoted to the MDA approach and model transformations. Section 4 presents the connectionist system on which we perform the modeling task (a system that calculates the XOR function). Here we describe this system by models and model transformations, and conclude in section 5.

2. Connectionist Approaches

Connectionist approaches (neural networks) are a strong current for the modeling of complex phenomena. A neural network is a structure composed of entities that can perform calculations and interacting with each other. Indeed, its operation is inspired from that of biological neural cells, and is therefore different from analytical methods of calculation that is used commonly.

There are several neural networks of different types used in various fields. However, in this paper we focus on the MLP (MultiLayer Perceptron) (section 2.3) which we model (case study: the XOR function).

2.1. *From biological neuron to artificial neuron*

A biological neuron is a nerve cell whose function is to transmit an electrical signal under certain conditions. The body of a neuron (soma) is connected, on the one hand, to a set of dendritics (neuron input) and, on the other hand, to an axon, stretched portion of the cell, which represents its output. The neuron is usually connected to other nearby neurons via synapses (axon/dendritic junction between two neurons) [4].

To mimic the biological neuron, the neuropsychiatrist McCulloch and the logician Pitts have used a basic mathematical model (i.e. a formal neuron) [7]. A formal neuron is a basic information-processing unit that forms the basis for designing artificial neural networks. We can identify three essential elements that characterize the neuron model [5]:

- A set of synapses (or connecting links), each of which is characterized by a weight w_{kj} where k is the recepient neuron of a signal x_i, and j refers to the input end of the synapse (i.e. input neuron). The synaptic weight of an artificial neuron may lie in range that includes positive as well as negative values.
- An *adder* for summing the input signals, weighted by the respective synapses of the neuron.
- An *activation function* (or transfer function) for limiting the amplitude of the output of the neuron. Typically, the normalized amplitude range of the output of a neuron is written as the unit interval [0,1] or alternatively [-1,1].

2.2. *Artificial neural networks*

An ANN (Artificial Neural Network) is composed of many artificial neurons that are massively interconnected and operating in parallel, each of which performs a simple treatment independently of the others, but all led to the emergence of global properties of interest, this means that network capabilities exceed those of its composing elements [4].

The fact that ANNs are bio-inspired systems based on the human brain function makes these networks a powerful calculation tool that is characterized by very interesting properties (e.g. nonlinearity, adaptivity, fault tolerance, parallelism, generalization) [5].

2.3. *Multilayer perceptron (MLP)*

Multilayer perceptron is one of the most used neural networks, especially for some difficult problems such as pattern approximation, recognition, classification and prediction [12]. It usually consists of an input layer of source neurons, one or more hidden layers of computation neurons, and an output layer of computation neurons. A multilayer perceptron has three fundamental characteristics [5]:

- The model of each neuron in the network includes a nonlinear activation function (excluding input neurons). A commonly used function that satisfies nonlinearity is a sigmoïdal function. This is an important feature that allows MLPs to solve complex problems.
- The network contains one or more layers of hidden neurons allowing it to learn complex tasks by extracting progressively more meaningful features from the input patterns.
- The network exhibits a high degrees of connectivity, determined by the synapses of the network. A change in the connectivity of the network requires a change in the population of synaptic connexions or their weights.

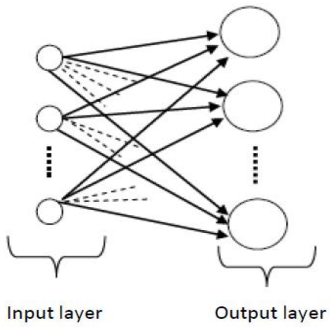


Fig. 1. Perceptron.

The combination of these characteristic gives the multilayer perceptron its computing power.

A particular and simpler form of MLP is the single-layer perceptron (Fig. 1) that correctly classifies patterns linearly separable (i.e patterns that lie on opposite sides of a hyperplane). When a perceptron is trained using a learning procedure called the *perceptron learning rule*, which is intended to adjust the synaptic weights in the right direction when a misclassification occurs (i.e. a supervised learning where there is a teacher which corrects the network whenever it makes a mistake), the rule converges, for a finite number of iterations, to a correct separating hyperplane defined by Eq. (1) (see Fig. 2).

$$\sum_{j=1}^{m} w_j\, x_j + b = 0 . \quad (1)$$

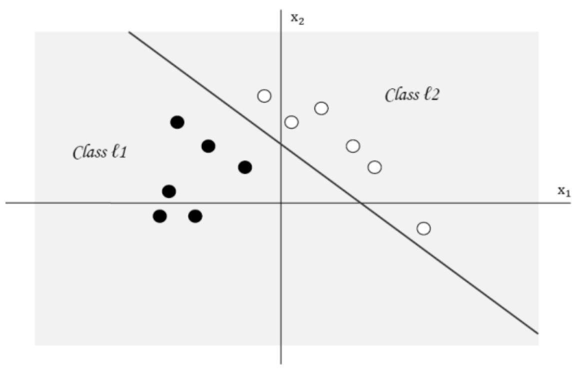


Fig. 2. A separating hyperplane for a two-dimensional pattern classification.

The inability of a single layer perceptron to solve interesting problems (nonlinear problems in general) is demonstrated by Minsky et Papert [9]. So Many pattern classification problems require a nonlinear partition of the input patterns which these perceptrons can not solve. An example often used to illustrate this limitation is the Exclusive OR (XOR) problem shown in Fig. 3. To resolve this problem, just add a single

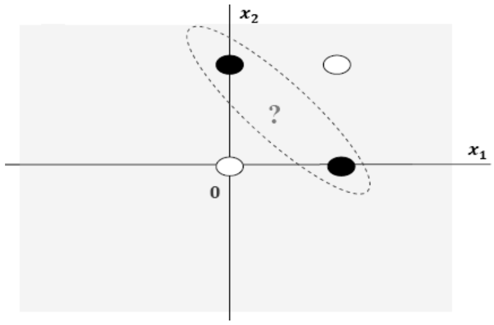


Fig. 3. XOR problem.

hidden layer with two neurons. Each neuron (except input neurons) uses a threshold function for its activation function [5].

2.4. *Bio-inspired systems*

Bio-inspired systems can be characterized according to several criteria. Systems based on connectionist approaches (neural networks) are bio-inspired systems that can be characterized by an epigenetic process [8]. Any system based on an epigenetic process is characterized by an easily alterable structure and the ability to learn and to adapt.

The architectural structure and environment where the system operates are two fundamental parts involved in connectionist systems just like all bio-inspired systems. The architectural structure of a connectionist system describes the system's components and the way they are interconnected.

The environment generates stimuli that will be processed by the system. The system can be adapted for a particular purpose. Fig. 4 illustrates these parts and their participation in the functionning of connectionist systems.

The environment is represented by the stimuli unit. It defines the data set recieved by the system. In the context of artificial neural networks, this set can be used either to train the network or to test it or to obtain an output.

The architectural structure unit contains the system topology and defines its components. In the context of artificial neural networks, the topology is defined in terms of three concepts: layers, nodes (neurons) and connections (synapses) between nodes.

Parameters unit contains a set of parameters used by the system. They may be learning parameters (e.g. learning rate, maximum number of iterations, etc..).

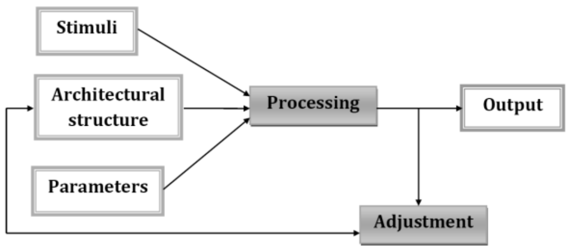


Fig. 4. Units participating in the connectionist system's functionning.

With the availability of the stimuli unit and parameters unit, the connectionist system defined by the architectural unit will perform a set of processing operations on the input data set and produces an output. The latter will be used, if it is wrong, to adjust the system (Learning Phase) to correct its behavior. The adjustment primarily affects the synaptic weights or the system architecture. Thus, there are two essential steps: a processing step and an adjustment step. These two steps follow one another, and, for a training phase, there may be several iterations of this sequence.

3. Model-Driven Engineering

Model engineering, and particularly the MDA (Model Driven Architecture) initiative, is an approach for describing software systems through models and their development as model transformations [3].

The intent of MDA is to separate business logic from its implementation. The benefits of this approach which considers the model as a primary artifact and a reference point throughout the development process, have attracted considerable attention during the last decade.

A model of a system is a description or specification of the system and its environment. MDA is to describe systems using models that can be of different types: platform-independent models that do not have any technology-specific implementation information (e.g. J2EE, .Net, PHP, etc.) and platform-specific models that have technology-specific implementation information. Models are defined in three different levels of abstraction (Fig. 5). The MDA enables transforming models from one level of abstraction to another (i.e. model transformation).

Model transformation lies at the center of the MDA approach. OMG defines model transformation as the process of converting a model into another model of the same system based on transformation rules and metamodel correspondence [10]. A more general definition considers model transformation as the automatic generation of one or multiple target models from one or multiple source models, according to a transformation description [2]. Fig. 6 illustrates the components of a model transformation.

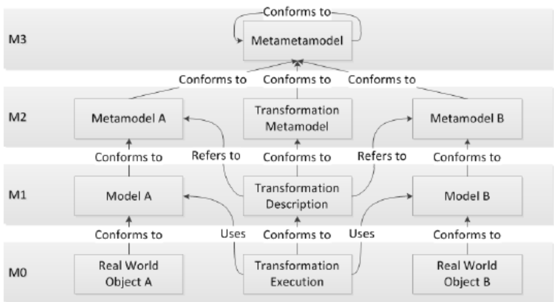


Fig. 5. The four levels architecture.

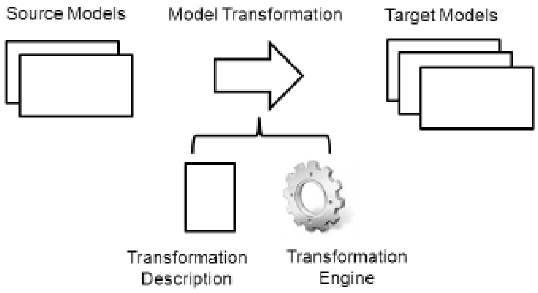


Fig. 6. The components of a model transformation.

A model transformation description is written in a model transformation language which provides a vocabulary and grammar with well-defined semantics for performing model transformations. The model transformation language is built on a model transformation approach (e.g. mandatory, operational, functional, declarative or relational). Most model transformation languages are rule-based. In other words, the transformation description is a set of model transformation rules (i.e. a set of entities describing how fragments of the source model can be transformed into frangments of the target model).

4. Description of Connectionist Systems by the MDA Approach

This section presents our modeling approach which aims to utilize MDA techniques to describe the functionning of a connectionist system.

4.1. *XOR problem*

To illustrate the functionning of a connectionnist system described as models and model transformations, we rely on the famous pattern classification system that solves the XOR problem (section 2.3). The latter is a MLP with one hidden layer.

One of the main characteristics of the connectionnist systems is their ability to learn. A training phase is required before a given system becomes able to perform a particular task. There are several learning rules but the most common and widely used algorithm for multilayer feedforward neural networks is the back-propagation algorithm. In the context of our work, the learning phase is not taken into account, and the system is considered valid (i.e. it behaves correctly). We focus on the functionning of the system, that is to say, data propagation from one layer to another (processing step shown in Fig. 4).

Fig. 7 depicts the network topology and the bias of each node represented as an additional input. *b1*=-3/2 and *b2*=-1/2 are respectively the bias of node 3 and node 4. *b3=-1/2* is the bias of node 5. It should be noted that the bias shown as an additional input (external parameter) is obtained by multiplying the input value by the connection weight value of this input. For exemple, *b1* is obtained by multiplying +1 by -3/2.

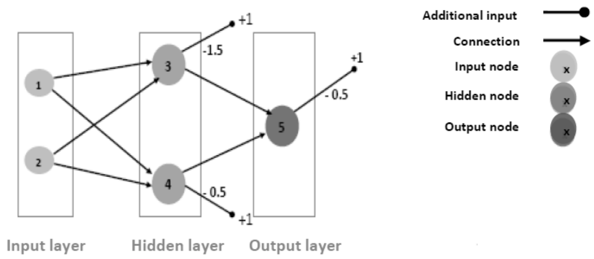


Fig. 7. Connectionist system that computes the XOR function.

4.2. *Artefacts modeling*

We denote by an artifact a data structure in the composition of the system. In the context of artificial neural networks, two essential elements form a network: artificial neurons (nodes) and the links between them. Fig. 8 shows a metamodel that allows the development of different types of artificial neural networks. It contains two meta-classes, *ANeuron* (*Artificial Neuron*) and *Link*. The meta-class *ANeuron* represents artificial neurons. It has a set of attributes:

- The attribute *name* indicates the unique name of the neuron.
- The attribute *bias*.
- The attribute *activation* contains the activation value of the neuron (the output of the neuron).
- The attribute *activated* indicates if the neurone is activated or not.
- The attribute *tfunction* (for *transfer function*) indicates the type of the transfert function employed by the neuron. The attribute type is enumeration (linear, threshold, sigmoïde, etc.).

The meta-class ANeuron thus has two references:

- The reference *inL* (for *input Link*) allows you to browse through the links (meta-class *Link*). This reference defines all input links of the meta-class *ANeuron*.
- The reference *outL* (for output *Link*) allows you to browse through the links (meta-class *Link*) coming out of the artificial neuron (meta-class *ANeuron*). This reference lists all the links coming out of the meta-class *ANeuron*.

The meta-class *Link* represents the links that connect the artificial neurons. It has two attributes. The attribute *id* that indicates the unique identifier of the link and the attribute *weight* for the weight of the link.

The meta-class *Link* thus has two references:

- The reference *src* (for *source*) allows you to browse through the start node (meta-class *ANeuron*) of the link.
- The reference *trg* (for *target*) allows you to browse through the target node (class *ANeuron*) of the link.

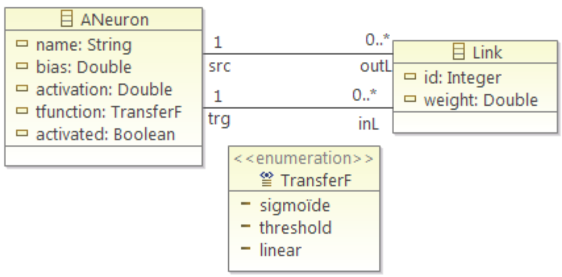


Fig. 8. Artificial neural network metamodel.

The metamodel of Fig. 8 is very general and does not explain the properties that characterize some artificial neural networks. Thus, the definition of a metamodel specific to a particular type of artificial neural network is necessary. We will develop a type-specific metamodel used to compute the XOR function. We speak then of MLP (Fig. 9).

This metamodel consists of four meta-classes. The meta-class *System* represents the connectionist system and has an attribute *name* to indicate the system name. The meta-class *Layer* represents the concept of layer. The meta-class *Node* defines the artificial neurons that form the layer. The meta-class *Link* which is totally similar to that of the metamodel of Fig. 8.

The meta-class *System* includes four references of composition: *InLayer* (for *Input Layer*), *HLayer* (for *Hidden Layer*), *OutLayer* (for *Output Layer*) and *link*.

- *InLayer* refers to the input layer of the system where the type of this reference is the metaclass *Layer*. At the opposite of this reference, one finds the corresponding reference *SysInlayer* in the meta-class container *Layer*. The system contains only one input layer.
- *HLayer* refers to the hidden layers of the system. The reference type is then a meta-class *Layer*. At the opposite of this reference, there is the reference *SysHlayer* in the meta-class container *Layer*. As indicated in the metamodel, the system may contain one or more hidden layers.
- *OutLayer* refers to the output layer of the system where the type of this reference is a metaclass *Layer*. In the referenced meta-class using *OutLayer* (meta-class *Layer*), there is a corresponding reference *SysOutlayer*. The system has only one output layer.
- *link* refers to the links contained in the system. This reference is of type *Link*. *system* is the corresponding reference in the meta-class *System*.

The meta-class *Layer* includes four references: *SysInlayer* (for *System Inputlayer*), *SysHlayer* (for *System Hiddenlayer*), *SysOutlayer* (for *System Outputlayer*) *and node*.

- *SysInlayer refers to the meta-class container System.*
- *SysHlayer* refers to the meta-class container *System*. It indicates that the hidden layer belongs to the referenced system.
- *SysOutlayer* refers to the meta-class container *System.*
- *node* is a composition-type reference that allows navigation via the nodes that compose the layer. For each reference *node*, there is a corresponding reference *layer* in the meta-class *Node* referenced by *node*.

The meta-class *Node* contains three references: *inL*, *outL* et *layer*. The meta-class *Link* comprises three references: *src*, *trg* and *system.*

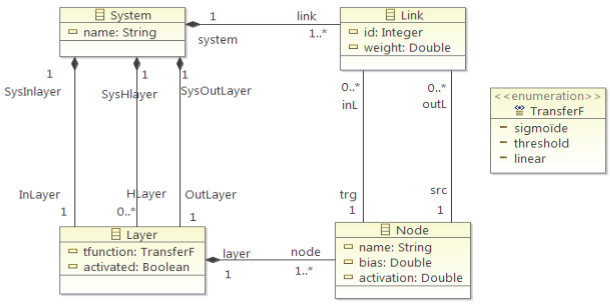


Fig. 9. Perceptron-based connectionist systems Metamodel (*SystemMM*).

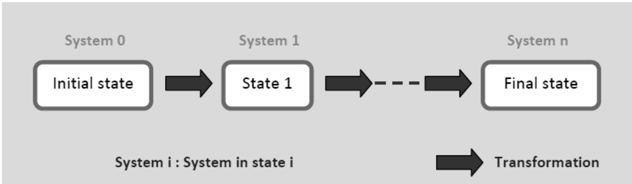


Fig. 10. System evolution during functioning.

Several differences can be noticed between the metamodel of Fig. 8 and that of the Fig. 9. In addition to adding the concept of the layer, some attributes are moved from the metaclass *ANeuron* (equivalent to the meta-class *Node*) to the meta-class *Layer*. Both *tfunction* (transfer function) and *activated* (attribute that indicates whether the node is activated or not) are replaced because of the characteristics of MLP. The transfer function adopted by neurons composing a layer is the same. In other words, the attribute *tfunction* of a layer describes the type of the transfer function used by all nodes that form this layer. Furthermore, parallelism is a fundamental characteristic of MLPs. The transmission of data from one layer to another in the perceptron generates simultaneous activation of all nodes of the receiving layer.

4.3. *Transformation modeling*

We denote by transformations actions performed on the artifacts. In the context of the modeled connectionist system (XOR function), operate these transformations on the artifacts modeled in the previous section should contribute to achieving the output of the XOR function (the value of 0 or 1).

4.3.1. Principle of transformation

Similar to a software development process, the functioning of a connectionist system can be seen as a succession of steps. Each step corresponds to a particular state of the system. Therfore, we no longer speak of steps, but rather of states in which the system can be found during functioning. By the state of the system, we mean the set of values that explain its evolution during functioning. The initial state of the system corresponds to the beginning of functioning while the final state corresponds to the end of functioning. The transformation is to change the system from one state to another until it reaches the final state. Fig. 10 summarizes this principle.

To describe the system "XOR function" as models and its functioning as model transformations, we must define all states in which it can be found during functioning and the nature of transformations that allow passage from one state to another. Any MLP-based system consists of propagating data (stimuli) from one layer to another until they reach the output layer. Some computation will be performed on this data throughout the propagation which causes its modification. This in turn may change the state of the system. We can deduce that the transformation is indeed a data propagation from one layer to another. Fig. 11 shows the application of two successive transformations on the system model "XOR function" which is in its initial state.

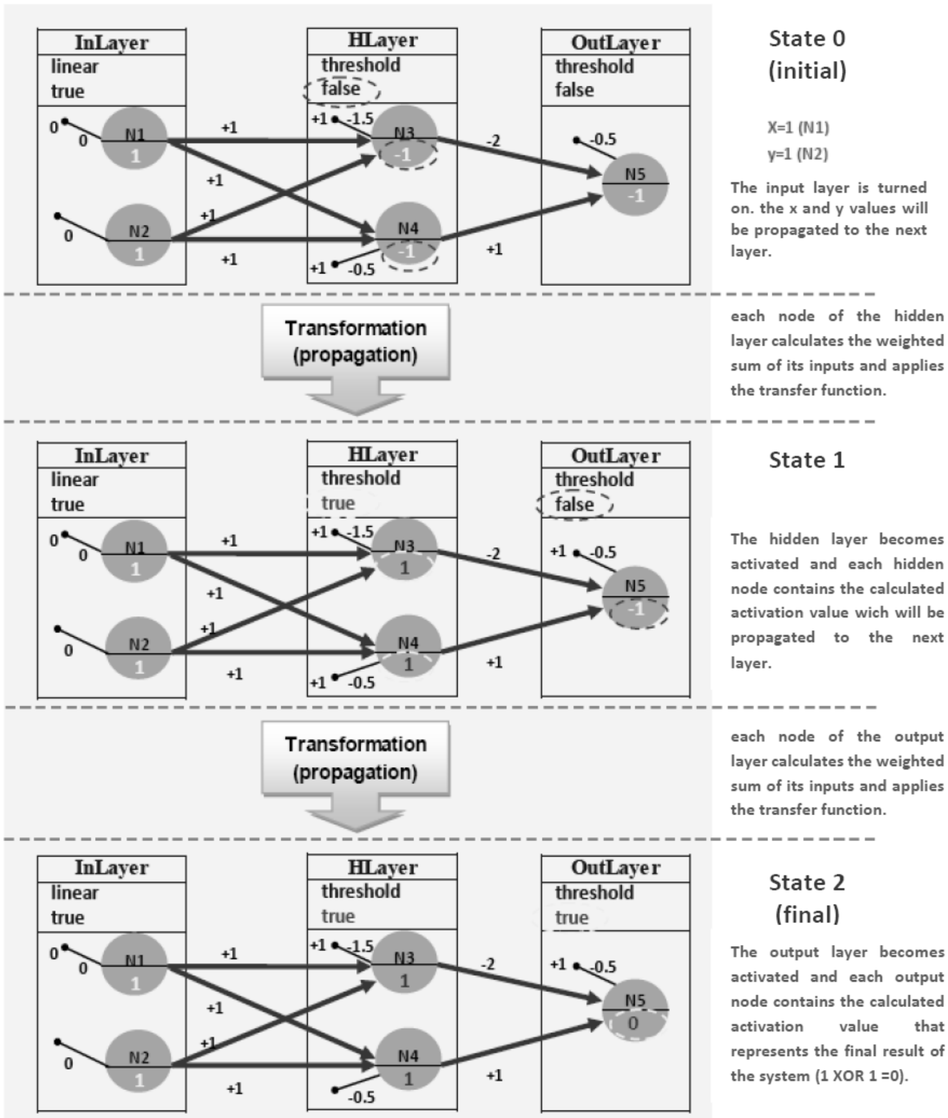


Fig. 11. Functioning of a connectionist system as model transformations.

We can list the following actions:

- Determination of system status.
- Propagation of data from previous layer to next layer if the system has not yet reached the final state, otherwise no modification will be made.
- For each propagation, the nodes that constitute the layer situated after the last activated layer in the system perform the weighted sum of their inputs and apply the transfer function of this layer. After data propagation is done, the layer becomes activated.

- The weighted sum of the inputs is calculated in two stages. The first is to multiply the activation value of each node in the activated layer by the weight attached to the respective link that connects it to the node performing this calculation. The second is to sum the values obtained from the previous stage and add to this sum the bias.

Listing 1. ATL transformation model.

```
module Propagation;
create OUTSystem : SystemMM refining INSystem : SystemMM;
helper context SystemMM!Layer def:isTheInLayer() :Boolean =
if not self.SysInlayer-> oclIsUndefined()
then true else false endif;
helper context SystemMM!Layer def:isTheFirstLayer() :Boolean =
if self.node.first().inL.first().src.layer.isTheInLayer()
then true else false endif;
helper context SystemMM!Node def:previousLayerIsActivated() :Boolean =
if self.inL.first().src.layer.activated then true else  false endif;
helper context SystemMM!Layer def:activatedUpdater():Boolean=
if  self.isTheInLayer()  then true
else if self.isTheFirstHLayer()  then true
      else if self.node.first().previousLayerIsActivated() then true
            else false
            endif
       endif
endif;
helper context SystemMM!Node def:inSignal(): Real =
self.inL->iterate (l; sum : Real = 0 |
(l.weight*l.src.activation)+ sum)+self.bias;
helper context SystemMM!Node def: tFunction() :Real =
if  self.layer.isTheInLayer()
then self.activation
else if not self.previousLayerIsActivated() then self.activation
      else if self.layer.tfunction=#sigmoid  then 1/(1+((0-1)*self.inSignal).exp())
            else if self.layer.tfunction=#threshold
                  then if self.inSignal>=0 then 1
                          else 0
                           endif
            else self.inSignal
            endif
      endif
    endif
endif;

rule Layer2Layer{
from
s : SystemMM!Layer
to
t :SystemMM!Layer (
            activated <- s.activatedUpdater
 )
 }
rule Node2Node {
from
s : SystemMM!Node
to
t :SystemMM!Node (
            activation <-s.tFunction()
 ) }
```

4.3.2. Transformation model

To describe the transformation using a model, we utilize the transformation language ATL (ATLAS Transformation Language) [6]. The choice of this language is motivated by several reasons:

- ATL is inspired by the OMG QVT requirements.
- ATL is an hybrid transformation language, declarative and imperative, that uses OCL expressions to create queries, manipulate and navigate through the elements of a source model.
- ATL is a widely used language, in both the academic and industrial domain.
- ATL is integrated in the Eclipse development environment and can handle models based on EMF.

The model transformation description, written in ATL, which expresses the transformation described in the previous section, is shown in Listing 1. We call this model "*Propagation.atl*". The source model is the model "*INSystem*" which describes the system "*XOR function*" in its initial state. It conforms to the metamodel "*SystemMM*" which defines perceptron-based connectionist systems. We call the target model "*OUTSystem*". The latter conforms to "*SystemMM*". The transformation is then horizontal and endogenous.

4.3.3. Capabilities and limitations

The proposed models and the transformation description *Propagation.atl* are characterized by a set of capabilities and are suffering from limitations:

- The ability to transform a model representing a perceptron-based system (whether multilayer or single-layer) from one state to another. The change of state is performed by propagating data from one layer to another.
- It is possible to involve as an input to the transformation a set of models each representing a different system, and for each system, the transformation can propagate data from one layer to another.
- The transformation requires that the source model be syntactically correct (conforms to metamodel *SystemMM*) so that it can produce a correct target model. For example, if the model does not represent a system with the structure of a perceptron (e.g. the presence of a loop or the absence of input layer) transformation does not work.
- The transformation requires that the source model be semantically correct (represents a perceptron-based connectionist system in a consistent state) so that it can produce a correct target model. For example, for a model representing a system that respects the structure of perceptrons but the values of its attributes are inconsistent (e.g. A system that contains two hidden layers; the first is not activated while the second is activated), the transformation does not work or produces a model representing a system in an inconsistent state.
- If the source model describes a system in its final state, the transformation is performed and produces the same model without any modification. In other words, it

does not indicate whether the system has reached its final state or not. It is for the user to determine.

- The metamodel SystemMM does not specify links direction. This means, for example, that a model representing a system based on a neural network that exhibits a layered structure and is disconnected (i.e. contains for example an isolated layer), conforms to this metamodel. However, the transformation accepts only perceptron-based systems.

5. Conclusion

In this paper, we focused on the MDA (Model Driven Architecture) and new concepts that it brought and allows us to describe connectionist approaches using models and model transformations. Contrary to the conventional use of MDA, in the context of our work, the description by models does not concern the development process but the functioning of a system based on the connectionist approach.

Based on the idea adopted in the MDA process, we considered the functioning of systems as a succession of states in which they can be found. Perform the functioning means that the modeled system must pass through all the states until the end of functioning which corresponds to the final state of the system. This is where model transformation takes place; A model is used to model a system in a given state among the different states that exist during its functioning. To pass from one state to another, just apply a model transformation. It represents the operations performed on the input model to create an output model that describes the system in the next state.

In the context of bio-inspired systems that integrate an epigenetic process, we can apply the same idea used to describe the functioning to model the training phase of the system (transformations model the operations performed to calculate errors committed by the system, and the learning rule adopted for the learning process). Thus, all artificial systems that adopt a bio-inspired process of evolution [8] (ontogeny, phylogeny or epigenesis) can be represented by models and model transformations.

References

[1] Bézvin J. (2004): In Search of a Basic Principle for Model Driven Engineering, The European Journal for the Informatics Professional, vol. 5, **2**, pp. 21-24.

[2] Biehl, M. (2010). Literature Study on Model Transformations. Technical report, Royal Institute of Technology, ISRN/KTH/MMK/R-10/07-SE, Stockholm, Sweden.

[3] Blanc, X. (2005). *MDA en Action: Ingénierie Logicielle Guidée par les Modèles*. Ed. Eyrolles.

[4] Blayo, F.; Verleysen, M. (1996). Les réseaux de neurones artificiels. Presses Universitaires de France, Que sais-je ? 3042.

[5] Haykin S. (1999). *Neural networks: A comprehensive foundation*, 2nd edn. Prentice-Hall, Englewood Cliffs.

[6] Jouault, F.; Kurtev, I. (2006): Transforming Models with ATL. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, **3844**, pp. 128–138.

[7] Mcculloht, W. S.; Pitt, W. (1943): A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, **5**(115–133).

[8] Meslati, D. (2006). Mage: Une approche ontogénétique de l'évolution dans les systèmes logiciels critiques et embarqués. Phd thesis, University of Annaba.

[9] Minsky, M. (1975). *A framework for Representing Knowledge*. The Psychology of Computer Vision, Wiston (Ed.), MacGraw-Hill.

[10] Omg (2010). Model-Driven Architecture, the official homepage, www.omg.org/mda

[11] Virole, B. (2001). Réseaux de neurones et Psychométrie. Editions du Centre de Psychologie Appliquée.

[12] Weiss, S. M.; Kulikowski, C. A. (1991). *Computer systems that learn*. Morgan Kaufmann Publishers Inc, San Mateo, California.

A SUPPORT TOOL FOR REUSE IN SOFTWARE ARCHITECTURE DESIGN

Sofiane BATATA

ESI (Ecole nationale Supérieure d'Informatique),
Algiers , Algeria
s_batata@esi.dz

Mohamed Amine MOSTEFAI

ESI (Ecole nationale Supérieure d'Informatique),
Algiers , Algeria
m_mostefai@esi.dz

Mohamed AHMED-NACER

USTHB (Univ. Sciences Technologie Houari Boumediene),
Algiers , Algeria
anacer@mail.cerist.dz

The design of component-based software architectures is performed regardless of the existing components in the warehouses, which does not promote the reuse. In this paper we present a support tool to assist the architect in the design phase of component-based software architecture, the tool implements an approach exploiting external descriptions of components, architectural elements and quality attributes. A new characterization model based on facets is proposed for components, it supports quality attributes of the ISO 9126 standard enhanced with metrics of measurement proposed to assist the architect in the evaluation of the quality of the architecture being designed.

Keywords: Software components characterization; Software components selection; Component-based software architecture; Software quality; ISO-9126.

1. Introduction

Companies in software development move increasingly towards large-scale reuse. Indeed, the cost of developing new software from scratch is very significant. Moreover, it is commonly admitted that the reuse of existing software enhances the reliability of the new system [Gill (2006)].

The most important benefit provided by the component-based development is the reuse of existing components in building new systems; this allows companies to save money in terms of time, effort and quality in software development [Wang and Krishnan (2006)]. The component-based development can be described according to two principles: « reuse but do not reinvent (the wheel); assemble prebuilt components rather than coding line by line.» [Wang and Qian (2005)].

It is difficult to find the components that meet exactly the requirements. The selection is performed usually in an ad-hoc manner [Dong *et al.* (2005)]. The identification and selection phase has a direct influence on the remaining phases of the process. Consequently, a wrong selection can have a major impact on project costs and outcomes, which may lead to considerable losses. So we consider that the success or failure of a project depends largely on a correct assessment of software components during the selection phase [Sjachyn and Beus-Dukic (2006)].

The design of component-based software architectures is performed regardless of the existing components in the warehouses, which does not promote the reuse. In this paper, we present a support tool to assist the architect in the design phase of component-based software architecture.

The paper is organized as follows: Section 2 presents the most serious barriers to the reuse of software components that meet the architects to design a component-based architecture. In Section 3 we describe the different steps of our approach implemented through the support tool for the reuse of components. In Section 4 we present the results of our first experiments. Finally, section 5 concludes the paper.

2. Selection in the design of components-based software architecture

Reuse of software components in the design of software architectures is a challenge so far, and few works in the domain of software architectures exploit the selection of existing components in the design phase. The reuse of a component or configuration of components to meet needs is not a simple task and many factors are behind:

- First, Component-based development and software architectures are two areas that have evolved independently and that address different problems at different level of abstraction [Crnkovic (2002)].
- Most of the investigated research systems offer to the user, as a result of his research, a list of Atomic components, but little work focused on the selection of composite components (configuration of components) and do not provide the means to exploit them [Mancebo and Andrews (2005)].
- Architectures can be analyzed through the use of quality attribute models like McCall or BOHEM. Some quality attributes have well established analysis models, such as performance. Other attributes have less mature models, such as variability, testability and security [Kazman *et al.* (2006)]. However, the ISO-9126 standard currently tends to become a reference model for software quality. Even if measures metrics of architecture quality are essentially the only way to evaluate architecture in an automatic way, they cannot replace the assessment of experts [Raiha (2008)].

3. A support tool for reuse in architecture design

The aim of our work is to assist architecture designer by providing a support tool in the design phase of components-based software architectures, to allow a better selection of reusable components or configurations of components.

This support tool will enable the architect to specify his functional needs in natural language, moreover he will be able to introduce the architectural specifications of the expected components, and quality requirements expected from these components.

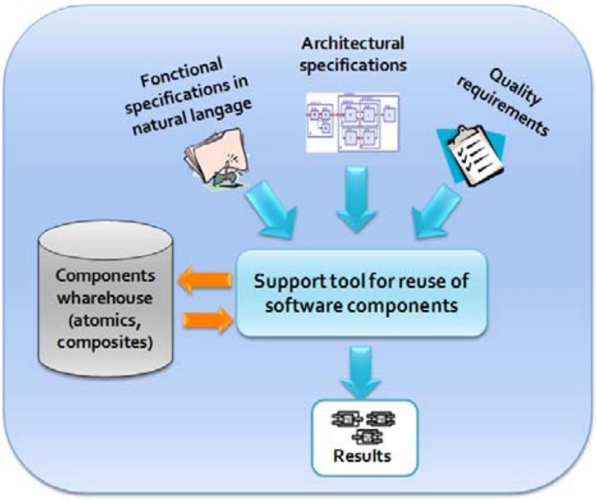


Fig. 1. Overall view of our support tool

Then, our tool will perform a search in the warehouse of the software components, exploiting both:

- The documentation provided with the components.
- Configurations (compositions) of existing components.
- The architectural properties of existing components.
- The quality characteristics of the components.

To achieve our goal, the first step is to specify the characterization model for the components in our warehouse.

3.1. *Characterization model for the software components*

Software components characterization consists in determining the properties that define a software component. The goal behind a new characterization is to improve the search and use of components by providing a better cataloging [Gill (2006)].

Many works have been done in this area, among others, the works of [Ben Sassi *et al.* (2003), Dong *et al.* (1999), Yacoub *et al.* (1999)], in which authors introduce models for characterizing components giving different weights for each attribute (criterion) selected in the model.

Based on the works cited above and the work of [Mezeghrane (2010)] which proposed a facets-based characterization approach, including the study of some markets of components like “componentsource” and “sourceforge”; we retained the facets which we consider essential in our solution. Table 1 shows the different facets adopted in our search system.

Table 1. Facets-based characterization model

View	Facets
General	• Component's name • Intention • Cost / Licence • Author • Date of creation • Version • Size • Encapsulation • Contact • Domain
Technology	• Standard • Development language • Operating system
Specificity	• Comportemental aspect • Required interfaces • Adaptation type • Structural aspect • Provided interfaces
Related components	• Similar Components • Collaborators

3.2. *Search system architecture*

The search process is implemented in three phases; our support tool returns after the execution of each phase a score for each component existing in the warehouse, each component will have three scores:

(1) The first score is returned by the search based on the external description; this score represents the functional capacity of the component to meet the needs of the user.

(2) The second score represents the ability of the component or the configuration to meet the architectural specifications of the user. This is achieved by taking into account the architectural features (interfaces, connectors, ports, etc ...) of a software component (or configuration). This score represents the architectural quality of the expected component (or configuration).

(3) The last score will allow to rank the selected components taking into account other quality attributes such as reliability, portability, performance.

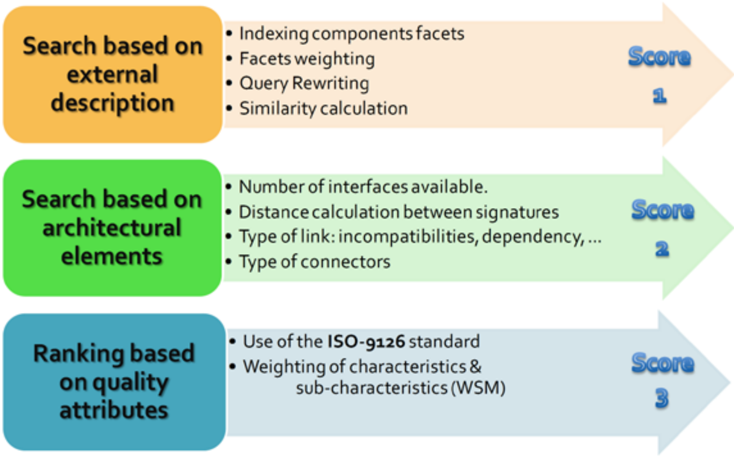


Fig. 2. Overview of the search system phases.

3.2.1. Search based on external description

This first phase is to search for components based on their external descriptions. Figure 3 provides an overview on its different stages.

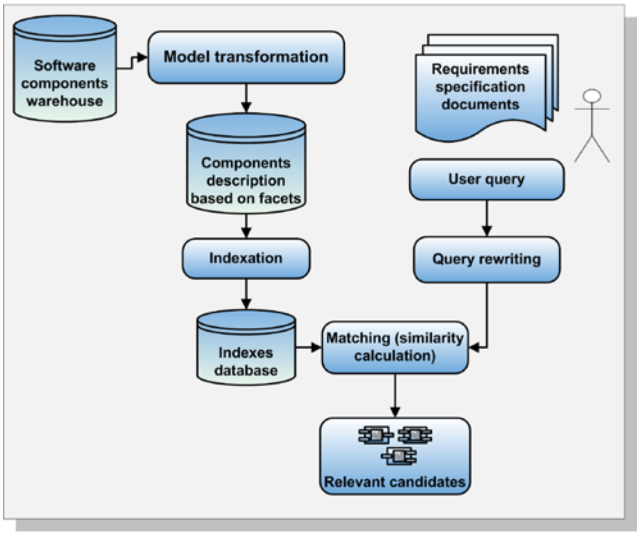


Fig. 3. Phase 1: Research Process based on external description

For this phase, we exploited the models and techniques used in the data-mining to search information in the descriptions of components, starting with the indexing process.

(1) The indexing process

For each component described in the components descriptions database, we take its facets one at a time and each facet is processed according to the following steps:

- Extracting words from different facets by removing separators.
- Elimination of stop words from the list of words resulting from the previous step.
- Extracting the lemma (grammatical root) from each word obtained from the second step. For that, we used Porter's algorithm.
- Measuring the importance of each lemma compared to others in the same facet. This step consists in calculating the frequency of appearance (term frequency) *tf* of each word in the facet, and after finishing with all facets of the component, we calculate the *idf* (Inverted Document Frequency) of each word. There are several form of *idf* in the literature and the formula we have adopted is :

$$idf = \log_2\left(|D| / df\right) \qquad (1)$$

Where $|D|$ is the number of existing components and *df* is the number of components where the term appeared in, at least, one of its facets. The terms of all facets of all components form the dictionary terms.

(2) The query Rewriting.

We used the vector model to search for information in the descriptions of components. The vector model is based on the representation of data and queries in a multidimensional vector space. The search is carried out by performing a similarity calculation between these vectors.

The query rewriting consists in adapting the query for the calculation of similarity with the index of components. For each query term, the system calculates the appearance frequency *tf*. As a result, we obtain a vector containing the query terms and their *tf* then we take the terms of the query vector, one at a time, and if the term exists in the dictionary of terms, we multiply the *tf* by the associated *idf*.

This step gives us a vector containing the final representation of the user query. At this point, there is no need to detail every step of the query rewriting process, since we find again - with some differences - the same steps of the indexing process.

(3) Similarity calculation

After running the indexing process we obtain a vector with three elements:

- The first contains the components.
- The second is the dictionary of terms.
- The third element is a vector that contains, for each component, the weight of its terms. These weights are calculated by the formula *tf * idf*.

The principle of the search mechanism is to find the component vectors that are closest to the query vector. For this, a similarity calculation is performed by calculating the cosine between the vector representing the user query and the vectors containing the weights of terms.

At the end of this phase, each component will have a score. This score represents, in our case, the functional capacity of the component to meet the expected needs.

3.2.2. Search based on architectural elements

In this phase, we evaluate the candidate components considering their contribution in the design of the architecture of the current system. At the end of this phase, each atomic or composite component will have a score that represents its quality from an architectural point of view. Two cases occur in this phase:

- If the component is atomic, the system will evaluate its ability to meet architectural specifications expressed by the user by calculating the convergence between the signatures of the component interfaces and the user's query.
- If the component is a composite, and to be able to classify the composite components with respect to their architectural features, we identified a set of metrics that will be used in assessing the quality of candidate components from an architectural point of view.

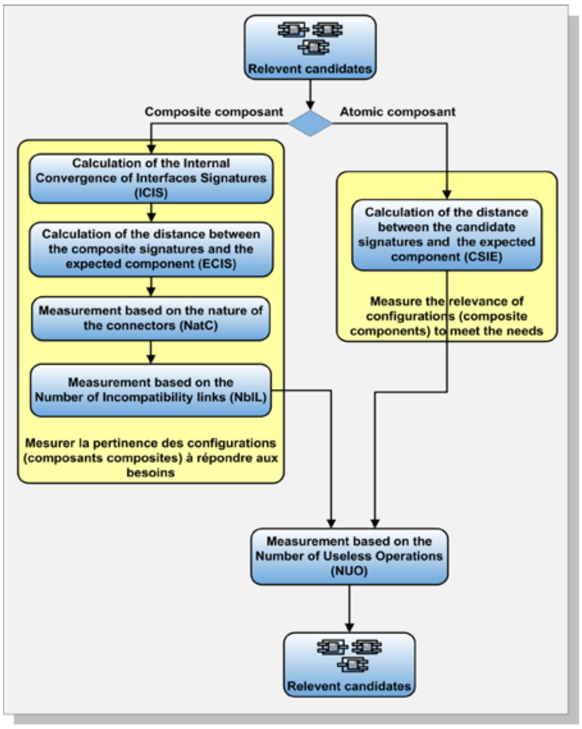


Fig. 4. Phase 2: Research Process based on architectural elements.

We proposed a set of metrics to measure the importance of a component or configuration from an architectural point of view.

- *ICIS: Internal Convergence of Interfaces Signatures*: Consist in calculating within a same configuration of components (composite component) the distance between signatures of the interfaces involved in this composition. Our formula of convergence between two signatures is based on the work of [Bonnel and Le Guernic (2006)]. It is calculated from the distance between the entities involved in these two signatures. An entity may be a primitive or structured type, a class, an interface, or a pointer. The distance between two entities is obtained by calculating the sum of the costs associated with the different existing paths between these two entities. Each path has an associated cost, and the distance used is determined by the minimum of the costs of all paths found between these two entities.

- *ECIS: External Convergence of Interfaces Signatures: represents the distance* between the signatures of the expected component and different signatures contained in the user query. The same formula for calculating the ICIS was used to calculate the external convergence.
- *CNat: Type of connector:* Our system gives a higher weight to the component with more atomic connectors than composites connectors. To each candidate component is assigned a weight. The number of connectors (atomic and composite) in the candidate is calculated and a weight is assigned to the candidate using the following formula :

$$weignt_{connector} = \frac{1}{(10 \times (nbc\_atomic + 1))} + \frac{9}{(10 \times (nbc\_composite + 1))}$$

The weight of component decreases at each increase in the number of connector.

- *NbIL: Number of Incompatibility links:* This is the number of internal components that are not compatible in a candidate composition. its weight decrease when the number of incompatibility links increases, we use the following formula:

$$Weight_{incomp} = \frac{1}{(1 + nb\_link_{incomp})}$$

- *NUO: Number of Useless Operations:* Assuming that the additional operations not useful have a negative influence on the costs in terms of money, response time and time to market. This metric gives more importance to the component that has the smallest number of operations not useful. Thus, the weight of the component decreases if the number of operations of interfaces increases unnecessary. The following formula is applied:

$$Weight_{operation} = \frac{1}{nb\_operation}$$

3.2.3. Ranking based on quality attributes

The third phase of our system is to rank the candidates components retained in the first two phases using other metrics of quality measures, essentially, the metrics that cannot be automatically calculated and those that require human intervention, such as: the quality of the documentation, the quality of the support provided by the manufacturer, security,…etc.

To obtain the final score for the quality at the end of this phase, all the quality characteristics of our software components were weighted using the WSM technique for decision making.

4. System validation

Usually, the validation tests of a retrieval system such as that we have proposed are done empirically using benchmarks. However, to our best knowledge, there is no Benchmark in form of warehouse software components.

Therefore, to test the kernel of our system, we have assembled a collection of components, taken from different components marketplaces such as: ComponentSource[a], sourceforge[b] ...etc., then transformed manually according to our facets-based description model.

A prototype of this support tool has been developed, through its interface, the user can update the components in the warehouse, and he can specify the search criteria and the interfaces signatures of the expected component. In addition, the user can give, within his query, a greater weight for a quality attribute versus another.

The first experiments that we performed indicated a very high recall and recorded a high response time. But with the lack of benchmark, these two values may not be sufficiently representative.

5. Conclusion and perspectives

In this paper, we presented a support tool in order to promote reuse in the design phase of software architectures. The tool uses a new approach of characterization based on facets, and the selection is done by combining the approach based on external descriptions and interface specifications of the expected components.

Our future works, first time, will be directed towards the use of ontologies to characterize existing components automatically. Later, it would be interesting to get the architectural specifications from architectures described with an ADL instead of specifying them manually through the user interface.

References

Ben Sassi, S., Labed Jilani, L. ; Ben Ghezala, H. H. (2003) COTS Characterization Model in a COTS-Based Development Environment. *10th Asia-Pacific Software Engineering Conference (APSEC'03)* pp. 352-352). Chiang Mai, Thailand.

Bonnel, N. ; Le Guernic, G. (2006) Système de recherche de méthodes Java basé sur leur signature. Lorient, France.

Crnkovic, I. (2002). *Building Reliable Component-Based Software Systems* Norwood: Artech House, Inc.

Dong, J., Alencar, P. S. ; Cowan, D. D. (1999) A Component Specification Template for COTS-Based Software Development. *International Workshop on Ensuring Successful COTS Development, in conjunction with ICSE-21* Los Angeles, USA.

Dong, J., Yang, S., Chung, L., Alencar, P. ; Cowan, D. (2005) A COTS architectural component specification stencil for selection and reasoning. pp. 1-4).

Gill, N. S. (2006). Importance of software component characterization for better software reusability. *SIGSOFT Softw. Eng. Notes* 31: 1-3.

Kazman, R., Bass, L. ; Klein, M. (2006). The essential components of software architecture design and analysis. *Journal of Systems and Software* 79: 1207-1216.

[a] Component source. Available: http://www.componentsource.com
[b] Source forge. Available: http://www.sourceforge.net

Mancebo, E. d. ; Andrews, A. (2005) A strategy for selecting multiple components. pp. 1505-1510).

Mezeghrane, W. (2010) La réutilisation des composants logiciels : Approche de caractérisation. In: Faculté Electronique et Informatique. USTHB, Algiers, Algeria.

Raiha, O. (2008) Applying Genetic Algorithms in Software Architecture Design. University of Tampere, Finland.

Sjachyn, M. ; Beus-Dukic, L. (2006). Semantic Component Selection - SemaCS. *iccbss* 0: 83-89.

Wang, A. J. A. ; Qian, K. (2005). *Component-Oriented Programming* New Jersey: Wiley-Interscience.

Wang, L. ; Krishnan, P. (2006) A Framework for Checking Behavioral Compatibility for Component Selection. pp. 49-60).

Yacoub, S., Ammar, H. ; Mili, A. (1999) Characterizing a Software Component. *In Proceedings of the 2nd Workshop on Component-Based Software Engineering, in conjunction with ICSE'99* Los Angeles, CA,USA.

International Conference on Software Engineering and New Technologies

A MDE PARAMETERIZED TRANSFORMATION TO IMPROVE USABILITY OF ADAPTIVE USER INTERFACES

Wided BOUCHELLIGUA

CES, ENIS, Soukra Road km 3.5, B.P: w 3038,
Sfax, Tunisia
wided_bouchelligua@yahoo.fr

Takwa HAMOUDA

CES, ENIS, Soukra Road km 3.5, B.P: w 3038,
Sfax, Tunisia
takwa.hamouda@gmail.com

Adel MAHFOUDHI

CES, ENIS, Soukra Road km 3.5, B.P: w 3038,
Sfax, Tunisia
adel.mahfoudhi@fss.rnu.tn

Mourad ABED

LAMIH, UVHC, F-59313,
Valenciennes, France
mourad.abed@univ-valenciennes.fr

The automatic generation of the adaptive user interfaces deteriorates the ergonomic quality of the generated user interfaces. In order to improve the ergonomic quality and the usability of adaptive user interfaces, we propose an ergonomic meta-model. The proposed meta-model is introduced as a parameter model of a parameterized transformation in the framework of the model driven engineering approach, for the transformation of an abstract user interface into a concrete user interface.

Keywords: adaptive user interface; usability; ergonomic meta-model; model driven engineering

1. Introduction

In Human Computer Interaction (HCI), the plasticity property denotes the capacity of an User Interface (UI) to be adapted to its context of use in respect of its usability [23]. By context of use, we mean the triplet (user, platform, environment). We are often interested in adjusting the UI to its context of use [18, 15]. The usability, also called ergonomic quality, was the focus of HCI. It was treated by standards such as ISO 9241-110 [11], ISO/IEC 9126 [10] and ISO/IEC 25010 [12]. In QUIM [20], although Seffah includes the majority of the usability works, few studies have been conducted to prove the usability and ergonomic quality of the interface during the process of adaptation.

It is in this context that our research work lies to integrate usability in the process of adaptive UI generation. To do so, we use an ergonomic evaluation. The referential of Bastien and Scapin [1] is considered as a good indicator of the UI quality. Our approach is based on these ergonomic criteria to propose an ergonomic meta-model. For the UI generation, our approach is based on a new paradigm that is emerging today in software development, which is the Model Driven Engineering (MDE). MDE goes beyond the framework of Model Driven Architecture (MDA) [16], which can be summarized in the elaboration of the Platform Independent Models (PIM) and in their transformation into Platform Specific Models (PSM) [2], to cover the methodological aspects. Thus, our ergonomic meta-model is introduced as a parameter model in a model driven approach for the generation of adaptive or plastic UI.

The remainder of this paper is structured as follows. Section 2 presents a state of the art on the approaches that focused on the preservation of the usability of adaptive user interface. Then section 3 defines the principles of the parameterized transformation of MDE. Next section 4 describes the proposed approach in terms of ergonomic meta-model and the method which is used to insert the ergonomic criteria in transformation modules presented in our previous work [3]. Section 5 provides a case study illustrating the approach. Finally, section 6 draws the conclusion and provides perspectives to future research.

2. State of the Art

Generally speaking, we are interested in adapting the UI to their context of use while preserving usability. The supports of adaptation may be redistribution and/or remodeling of the interface:

- Redistribution changes the state distribution system across all the interactive platforms involved. It functions by the migration of any part of interactive system.
- Remodeling is a local adaptation to the platform which is a design of the interface without changing the distribution. This reorganization may involve the presentation of information, functionalities, or the functional core (changing concepts and services of the applicable domain).

Based on these two supports, a reference framework that covers both the design and implementation of plastic UI is presented in [5]. It allows the production of UI whose areas of plasticity are known.

Several methods exist for the evaluation of user interfaces, which identify usability problems of interactive systems. As an example, Correani et al. [6] propose an inspection-based tool for improving web site usability. He defines and implements a number of design criteria for vision-impaired users. In the same direction, Leporini et al. [14] provide a MAGENTA tool for supporting inspection-based evaluation of accessibility and usability guidelines. In addition, building on a method of assessing compliance with the recommendations, Vigo et al. [25] propose an application evaluation that considers specific device features in the evaluation process. A design environment GUIDE2ux is suggested by Meskens et al. [17] to identify usability problems automatically and facilitate to designers the verification of their designs on the target device easily.

However, although several evaluation methods exist, most of them are targeted for the evaluation of final products. But today, with the expansion of model-based approaches for development user interfaces, research is oriented to integrate the evaluation at the level of modeling steps.

Among these research works we can cite that of Frey [9] offering QUIMERA, a quality meta-model. QUIMERA is composed of criteria that can be decomposed into subcriteria. The meta-model provides different recommendations specified for each criterion. QUIMERA covers the evaluation methods that are specified by metrics and/or practices. However, this meta-model is not yet implemented in order to use at design time and runtime.

The work of Sottet [22] is among the pioneering work that dealt with the connection of the MDE-based concepts to the UI, in the purpose of generating a plastic UI. Actually, Sottet [22] proposed not only meta-models and model transformations to derive plastic UI, but also a meta-model that allows the characterization of the changes in models with ergonomic criteria. In [22], Sottet proposed an adaptation controlled by an intelligent system in order to automate the generation of plastic UI while respecting certain ergonomic criteria. This intelligent system allows the choice of the appropriate transformation to a given context, respecting such ergonomic properties. For example, if the user makes many mistakes, we should choose the UI that limits his false manipulations; i.e., to throw the rules of model transformations classified as "protection against error". In this case, it is necessary to create an N transformation to an N context and the intelligent system chooses the transformation that best suits a given situation. Building on this idea, we suggest a process of generating plastic UI respecting a set of ergonomic criteria. Such a process is based on the transformation modules and plays the same role as an intelligent system.

3. Principles of the Parameterized Transformation of MDE

Our objective is to handle the adaptation of the UI to the context of use (platform, environment, and user) and to improve usability of adaptive UI. To do so, our research work will build on the parameterized transformations defined by [24]. The cited work describes a parameterized transformation within the framework of the model driven engineering for a contextual development. The authors propose a parameterized transformation focusing on PIM to PIM transformations (Fig. 1).

By the use of this transformation technique, the contextual parameter identified into the model will be contextualized with the parameterizable element which represents context information [26]. Such correspondences are guaranteed by the transformation parameter setting, whose basic principle is to take into consideration the properties of the context during the specification of transformation rules (right of Fig.1). Quoting [19], "a parameter specifies how arguments are passed into or out of an invocation of a behavioral feature like an operation. The type and multiplicity of a parameter restrict what values can be passed, how many, and whether the values are ordered".

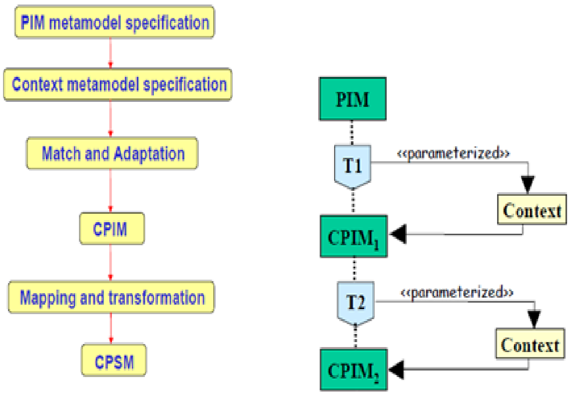


Fig. 1. MDE Parameterized transformation [24].

Indeed, Frankel [8] indicates the importance of the parameterization in the operations within the models by associating the tagged values with PIM and PSM. Tagging model elements allows an easy filtering of some specific elements.

The use of the parameterized transformations is envisaged with the aim of either improving new features (values, properties, operations) or changing the behaviour of an application. For that purpose, the designer has to specify the parameters intended to be inserted during the phase of transformation. In his work, [24] proposes that these parameters are the context information, thus after the transformation, the application will join the context information specified into the parameters as illustrated in Fig.1

A PIM model can be developed without considering the contextual information: the name of the user, his profile, the platform type, and the location etc. can be added as parameters that will be used during the phase of transformation. The same PIM model can be transformed and refined several times by adding, or deleting each time the information relative to the context, thus obtaining different CPIM (Contextual PIM). In fact, to the same PIM we can attribute various CPIM, just by modifying the contextual information. A CPIM in turn, can generate a CPSM (Contextual PSM) by resorting to the traditional techniques of transformation. CPSM specifies operation system requirements, programming languages, middleware architectures and networking.

Building on the concept of transformation parameterized by the context as defined within the framework of MDE. Our previous research works have focused on the generation of multi-platform user interfaces [4]. We have proposed a complete approach for generating a UI adapted to the full context use [3].

In this paper, the proposed approach assures the usability of the UI during the adaptation process. The generation process consists of three transformation modules starting with an Abstract User Interface and generating a Concrete User Interface by inserting the user, platform and environment model, respectively. The ergonomic model serves as a parameter in the three transformation modules.

4. Proposed Approach

Faced with the multitude of the existing recommendations, Bastien and Scapin have conducted, since 1997, the synthesis of about 900 recommendations in the field of computing ergonomics at the large sense [1]. Their research work has led to a list of 18 criteria divided into eight dimensions. The set of these criteria can help the evaluator to estimate the ergonomic quality of the UI in terms of usability.

The ergonomic evaluation can be carried out at different stages of the development cycle of the UI and is usually performed in the final generation of the interface. But aiming at an early detection of the major problems of usability of an interface, we will incorporate the ergonomic assessment in the different transformation modules of the process of generating the plastic or adaptive UI that has already been conducted in our previous work [3].

4.1. *Parameterized Transformation By The Context Of Use*

In [3], we have proposed a methodological framework for the generation of plastic UI. The proposed approach in [3] is shown in Fig.2. The abstraction levels of the Cameleon framework [5] incorporated in our approach are: Abstract User Interface (AUI) and Concrete User Interface (CUI).

In our approach, the Abstract User Interface allows the transformation of the specification in the modelling of the abstract components of the interface. Indeed, the AUI meta-model which describes the hierarchy of the abstract components “UIComponent” corresponding to the logical groups of interactions “UISpace”. The modelling of the abstract interface of an application is then made by one or several “UIGroup” which model containers forming coherent graphic elements (a window in a Windows environment, for example). Each “UIGroup” consists of one or several “UIUnitSuit” and/or “UIUnit”. A “UIUnit” gathers a set of interaction elements which cannot be separated from a logical business standpoint of the application (a treatment form for example). It can include one or several “UISubUnit”. The advantage of this modelling is to allow the creation of the application by assembling the existing elements, resulting in a strong reusability.

The Concrete User Interface is deduced from the Abstract User Interface to describe the interface in terms of graphic containers, interactors and navigation objects. The CUI meta-model consists of one or several windows represented in the meta-model by the “UIWindow” class (graphical modality) and by the “UIVocalForm” class (vocal modality). Besides, the “UIPanel” class (respectively “UIVocalGroup”) allows the modelling of the possible hierarchies of containers. The interactors presented by the “UIField” class (respectively “UIVocalComponent”) of the concrete interface are classified according to their types in three groups: “UIFieldMultimedia”, “UIFieldData” and “UIFieldControl”. In our CUI meta-model, a functional service is connected to any type of container and to all the constituents belonging to it.

The service of personalization can ensure several types of personalization: a linguistic personalization “useoflanguage” dependent on the language of the user, a guide personalization “useoftooltip” according to the skills of the user (computing and business), a presentation personalization of the interface (background, font, color) according to the preferences of the user and to the environment in which the application is executed, and so forth. Default values are given to the class attributes “PersonalizationService” which are used when none of the rules of transformation would be valid. The use of the functional services at the level of the concrete user interface has the advantage of being able to apply the impact of several properties of the context from the very phase of modelling. The AUI meta-model and the CUI meta-model are presented in [3].

The passage to the concrete level has for objective the generation of a plastic interface adapted to the planned context. Our approach facilitates the adaptation of the UI to the user, because the latter is in the center of all the problem of the UI and everything revolves around him. The first transformation (T1 in Fig.2) allows the generation of the first concrete user interface (CUI1 in Fig.2) adapted to the preferences of the user having received the information suitable to him and echoing them on this intermediate interface.

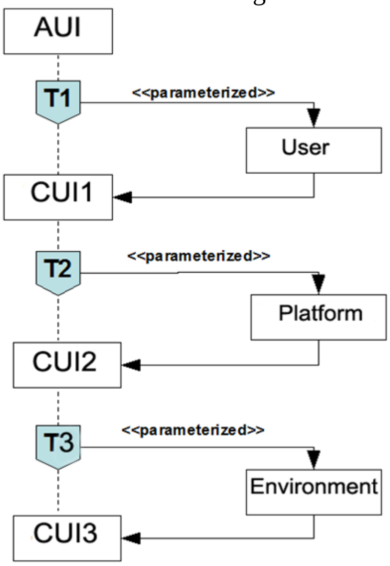


Fig.2. Parameterized transformation by the context of use.

On the other hand, we are interested in the injection of the characteristics of the platform used to assure the plasticity towards this context. Indeed, we opted for the choice of this injection order of the characteristics for multiple reasons. Indeed, it is around the user that revolves everything and it is his characteristics that are going to impose the choice of the platform. Besides, it is the user who decides about the device on which he even wishes to post the information. In fact, this variation is going to require the appearance and the disappearance of the other devices of interaction. Furthermore, it is according to his preferences that the modality: graphic, hearing or even olfactive is going to be chosen. Then, in case of change at the level of one of the contextual dimensions, an adaptation is launched to protect the usability. Certainly, the specific properties and the capacity characteristics of the target device have to satisfy the needs of the user. This second transformation (T2 in Fig.2) adapts the first CUI1 to the characteristics of the platform which is going to host the application, from which the second CUI (CUI2 in Fig.2) results.

So, having fixed and adapted the characteristics of the target platform to his own motivations and intentions, the user has now nothing but to choose the environment which is going to host the application. In fact, this environmental variant has to be in agreement with the characteristics of the user and the target platform. It is the profile of the user, defined as being a first entity for the process of adaptation, as well as these accompanied intentions, naturally, symptomatic of the platform that are going to determine the environmental aspects. The latter are going to be implemented during the process of adaptation to succeed in the generation of a plastic UI while taking into account three speeds of the context. Hence, in the third place, we are going to inject the environmental properties in the third transformation (T3 in Fig.2) to have the interface (CUI3 in Fig.2).

Therefore, the generation of the concrete user interface is made on three phases. To do so, we proposed, in [3], three meta-models (the user, platform and environment meta-models), so as to implement the parameterized transformation principle to illustrate the process of adaptation.

The user model has to contain information allowing the characterization of the user. Our meta-model [3] presents four categories:

- Information staff (the name and the first name of the user, the age, the kind);
- Knowledge (The expertise level of the user in computer science, the expertise level regarding task or manipulated concept);
- Preference (The modality of interaction (graphic, vocal, olfactive, tactile, etc.), police, the character size, colour and the sound volume);
- Capacity (physical (sensory and engine) and cognitive capacities).

Our platform meta-model [3] consists of:

- Calculation resources: These resources do not only include the material aspect, such as the memory or processor but also the software aspect as the supported operating system;
- Interaction resources: We identify two classes of interaction devices (the input devices and the output devices). Certain devices inherit both classes and are thus input/output devices, such as the touch screen.

Our environment meta-model [3] consists of:

- The ambient environment that surrounds the interactive system. But with the invasion of the ubiquitous computer science, the ambient conditions are changeable from one moment to another. The ambient environment is formed by (climatic environment, luminous environment and sonorous environment);

- The Temporal environment;
- The social environment;
- The spatial environment (geographical location of the interactive system).

The ergonomic evaluation can be carried out at different stages of the development cycle of the UI and is usually performed in the final generation of the interface. But aiming at an early detection of the major problems of usability of an interface, we will incorporate the ergonomic assessment in the different transformation modules of the process of generating the adaptable UI. The three transformations (T1, T2 and T3) are parameterized by ergonomics criteria involved in proving the usability of the generated CUI as shown in Fig.3.

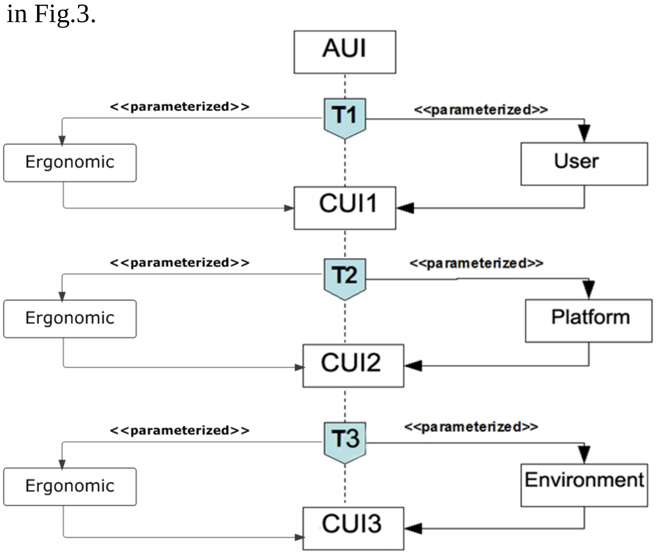


Fig. 3. Parameterized transformation by the context of use and ergonomic model.

4.2. *Ergonomic Meta-model*

The choice of the ergonomic criteria to be applied is done in the processing modules automatically, using an intelligent system that is implicitly implemented. That is to say, if the criterion “protection against error” is one of the criteria in the model of ergonomic criteria, then during the performance of transformation modules, along with the transformation rules common to all UI, the appropriate rules for this criterion are automatically enforced.

In the process of building the UI, Sottet has proposed a meta-model that allows the characterization of the model transformation by ergonomic criteria [22]. Building on this idea and the ergonomic criteria of Bastien and Scapin [1], we propose our own meta-model of ergonomics which seeks to explain the ergonomic evaluation. An ergonomic model conform to this meta-model is taken as a parameter for improve usability of UI during the adaptation process. Fig.4 shows the proposed meta-model.

In this meta-model, we attempt to present the elementary criteria that are really integrated in the transformation model. But, for legibility reasons of the model, we have not discarded the other criteria, which are eight. Some of the criteria are divided into sub-criteria, 18 of which are counted elementary. So, our meta-model is composed of three classes.

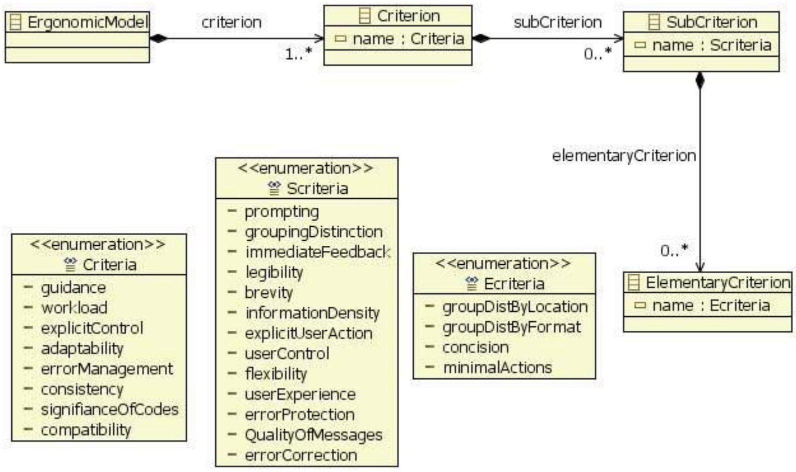


Fig. 4. Ergonomic Meta-model.

The first class is the class “Criterion”, in which we have specified a unique attribute that is “name” of the type “Criterion”. We have defined an enumeration called “Criterion”, whose values are the eight basic criteria.

We have also defined a second class called “SubCriterion” which consists of one attribute “name” of the type “Scriteria”, whose enumeration values “Scriteria” are the sub-criteria of the basic criteria.

Because certain sub-criteria are in turn subdivided into other criteria, we have added a third class “ElementaryCriterion”. Just like the other two classes, this class contains one attribute “name” of the type “Ecriteria”, whose enumeration values are the sub-criteria of the sub-criteria.

The notion of priority between the criteria does not appear in this meta-model, but is introduced implicitly in the processing modules. Indeed, the addition of criteria to the processing modules is done so that some priority is respected in function of the user preferences outlined in the user model, as well as the characteristics of the platform and environment presented in the platform model and the environment model. For example, to a large screen, if the “minimal actions” criteria are taken into account, then the “informational density” sub-criteria will be ignored.

Several constraints are added to the meta-model of ergonomics. As an example, the code below is intended to ensure that the hierarchy of criteria is respected. For instance, if you want to insert the “guidance” criterion in the ergonomics model, only “prompting”, “grouping/distinction of items”, “immediate feedback” and “legibility” can be put under this criterion.

```
invariant crt:
self.criterion->forAll(c : Criterion |
  //Criteria with subcriteria
(c.name = Criteria::guidance implies
c.subCriterion->forAll(sc : SubCriterion |
sc.name = Scriteria::prompting
or sc.name = Scriteria::immediateFeedback
or sc.name = Scriteria::legibility
or (sc.name = Scriteria::groupingDistinction
implies sc.elementaryCriterion->
forAll(ec : ElementaryCriterion |
ec.name = Ecriteria::groupDistByLocation
or ec.name = Ecriteria::groupDistByFormat))))
and (c.name = Criteria::workload implies ...
  //criteria have not sub criteria
and (c.name = Criteria::consistency implies
c.subCriterion->isEmpty()) and ... ;
```

4.3. *Injection Of Ergonomic Criteria In Transformation Modules*

Our work lies within the reference of Bastien and Scapin [1] and adopts the perspective [7] in the distribution of the criteria of ergonomics in the process of generating UI. We are interested only in the transformations applied to generate a CUI from an AUI. The task-abstract user interface transformation is not the subject of this paper.

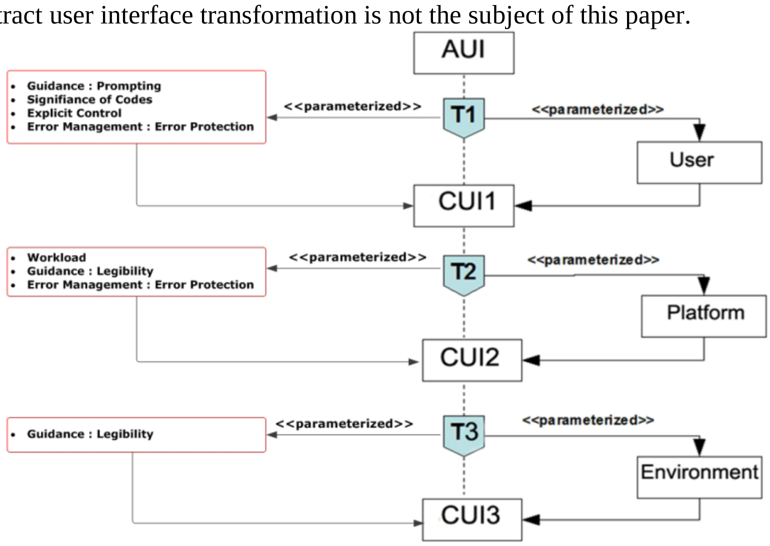


Fig. 5. Injection of ergonomic criteria in transformation modules.

The transformation AUI to CUI is a presentation choice for workspaces and their contents as well as the sequencing between spaces. These transformations are governed by several criteria.

The first transformation module consists in concretizing abstract containers while at the same time taking into consideration the characteristics of the user. This concretization is motivated by the “incitation”, “explicit actions”, “user control” and “protection against errors” criteria. We decided to insert the “incitation” criterion at this level because the concretization of the containers allows the specification of the container type and therefore, it offers the possibility of associating a label and/or another additional indication with the container of text fields type. Moreover, because the type of the container is known, then we can associate buttons with concrete containers and therefore the “explicit actions” and “user Control” criteria can be inserted at this level. Besides, the text fields can be replaced by drop-down lists to minimize the risk of seizing erroneous values. At this stage, it is the “protection against errors” criterion that is implemented.

Concerning the second transformation, we have tried to find a solution to have the possibility of specifying the platform and injecting the corresponding ergonomic criteria in order to create the adequate interface independently. Therefore, the second transformation module is controlled by several criteria among which “minimal actions”, “informational density”, “legibility” and “protection against errors” can be mentioned. Thus, the injection of the ergonomic criteria depends on the choice of the platform. That is why we can favor one ergonomic criterion to another with regard to the platform characteristics, namely the screen size. So, we have introduced the notion of priority between the criteria in an implicit manner.

The third transformation module takes the environment meta-model as a parameter. Certainly, this meta-model includes all the facets of the environmental context susceptible to react directly or indirectly to the interactive system and to this level of transformation; the only ergonomic criterion on which we decided to inject this module is the “legibility” criterion.

4.3.1. *First Transformation Module Parameterized By User Model And Ergonomic Model*

The injection of ergonomic criteria in the first transformation module parameterized by user model can improve four ergonomic criteria:

Incitation:

It is to associate a label with each text field to guide the user. Besides, at the level of the label, further clues about the entry format of data can be added. For an input field having enumeration type (unit of measurement or some other symbol), we add a label next to the input field to properly guide the user.

In this first transformation module a component of the type “specify" in AUI is transformed into a “FieldIn” (Textfield) in the first concrete user interface generated. The code below is an excerpt from the kermeta [13] code which outlines the method explaining the incitation criteria.

```
    //prompting criterion
var nat : Nature init uic.nature
var tp : AnnotationType
init lnk.uicomponentannot.type
if (nat == Nature.Specify) then
 //TextField creation
 createFieldIn(uiw,uic,lnk)
 if (prompting == true) then
  if (tp == AnnotationType.EEnumerator) then
    //Creation of Label with symbol
    createStaticField(uiw,uic,lnk,symbol)
  end
 end
end
```

Explicit actions:

The system should be an explicit action of validation by the user (eg. : Entry, Validation, OK) following an entry of data (FieldIn, DropDownList, RadioButton, CheckBox). The following code lines present the insertion of this criterion following an entry of data of the type of text field.

```
operation createFieldIn(uiw :UIWindow,
uic : CollapsedUIUnit, lnk : Link)
is do
var fi : UIFieldEdit
init UIFieldEdit.new
fi.name := lnk.uicomponentannot.data
uiw.uifieldW.add(fi)
stdio.writeln("creation of FieldIn" + fi.name)
   //explicit actions
 if (ExplicitUserAction == true) then
   var bt : UIButton init UIButton.new
   bt.name := "OK"
   uiw.uifieldW.add(bt)
   stdio.writeln("creation of UIButton" + bt.name)
end
end
```

User control:

Allowing the user to interrupt an action or processing in progress at any time using the button “cancel”.

```
  //user control
```

```
var btCancel : UIButton init UIButton.new
btCancel.name := "Cancel"
uip.uifieldP.add(btCancel)
stdio.writeln("creation of UIButton" + btCancel.name)
```

Protection against error:

The protection of the user against error can be translated at this level, by the fact of creating a list happening instead of the text field. In fact, the user has only to choose the appropriate values and he is protected against entering the incorrect values.

```
if (ErrorProtection == false) then
  createFieldIn(uiw,uic,lnk)
else
  if (lnk.uicomponentannot.enumNB > 3)then
   createDropDownList(uiw,uic,lnk)
  else
   createRadioButton(uiw,uic,lnk)
  end
end
```

4.3.2. *Second Transformation Module Parameterized By Platform Model And Ergonomic Model*

The already-realized work in [21] presents a development process of a plastic user interface that is applicable to the context of use. The latter taking into account the context of use is considered especially on the platform. an UI for PC is produced first, and following a series of iterations starting from the, an UI of the PC for an iPhone is produced. That is to say, to create an interface for the iPhone, it is necessary at the first place to create an interface for PC. The distribution of ergonomic criteria in [21] is done in iteration and certain criteria taken into consideration for a given platform can be non-applicable for another platform. In our work we have tried to find a way to be able to specify the platform and create the appropriate interface independently (the relationship between UI for different platforms is horizontal).

The injection of ergonomic criteria relies then on the choice of the platform. So we can focus on one ergonomic criterion instead of another depending on the characteristics of the platform, namely the size of the screen, which is how we introduced the implicitly-made notion of priority between the criteria.

For a large-sized screen: the ergonomic criterion “minimal actions” can be proved.

Minimal actions:

This is to reduce the path length of interaction and limiting, particularly, the actions of navigation, which do not contribute to the achievement of the business task. Of course, to limit the navigation, we replace the windows with panels while respecting the relationship between windows regardless of being sequential or simultaneous.

If the relationship between windows is sequential, then the target window becomes panel in the source window, without forgetting the deletion of target window.

```
if screenSize < 100 then
  //treatment
else //Minimal Action
   //sequential relationship
 if (String.clone(newrs.type).equals("Sequential")) then
   stdio.writeln("window source: " + uiwsrc.name)
   stdio.writeln("window target: " + uiw1.name)
   createPanel(result,uiw1,uiwsrc)
   stdio.writeln("Creation of panel in window " +uiwsr.name
+"in the place of window " + uiw1.name)
   result.uiwindow.remove(uiw1)
 end
end
// operation createPanel
operation  createPanel(outputmodel  :  ConcreteUserInterface,
uiw1 :UIWindow,uiwsr : UIWindow) is do
  //create new Panel
  var uip: UIPanel init UIPanel.new
  uip.name := uiw1.name
  getCUIPanel(uiw1).each{p|createUIPanelP(p,uip)}
  getCUIFieldW(uiw1).each{f|createFieldPanel(uip,f)}
  uiwsr.uipanel.add(uip)
end
```

If the relation between windows is simultaneous, then both windows become two panels in source window if it exists or in a new window.

```
operation createPanels(outputmodel :
ConcreteUserInterface,uiw1 : UIWindow,
uiwsr : UIWindow,nwin : UIWindow ) is do
  //first panel
 var uip1: UIPanel init UIPanel.new
     uip1.name := uiw1.name
     getCUIPanel(uiw1).each{p|createUIPanelP(p,uip1) }
     getCUIFieldW(uiw1).each{f|createFieldPanel(uip1,f)}
      nwin.uipanel.add(uip1)
  //second panel
 var uip2: UIPanel init UIPanel.new
    uip2.name := uiwsr.name
    getCUIPanel(uiwsr).each{p|createUIPanelP(p,uip2)}
    getCUIFieldW(uiwsr).each{f|createFieldPanel(uip2,f)}
    nwin.uipanel.add(uip2)
 end
```

On the other hand, in order to satisfy the sub-criterion “minimal actions”, each panel composed only of buttons is deleted and replaced by destinations panels related to these buttons.

```
//panels
var panels : OrderedSet<UIPanel>
panels := getCUIPanel(uiw1)
var panel : UIPanel init panels.each{p|
stdio.writeln(p.name)
var fields : OrderedSet<UIField>
fields := getCUIFieldP(p)
var test : Boolean init true
var Nb : Integer init 0
fields.each{f|
if (f.getMetaClass() != UIRadioButton)
then test:= false
else
Nb:=Nb+1
stdio.writeln(Nb.toString)
end}
stdio.writeln(test.toString)
if test == true then
var rs : OrderedSet<UIRelationShip>
  //UIRelationShip
    rs := getRelationShip(inputModel)
var rscp : UIRelationShip
init rs.detect{u|u.source == p.name}
var ps : OrderedSet<UIPanel>
ps := getCUIPanel(uiw1)
var panelt : UIPanel init ps.detect{pt|
rscp.target == pt.name}
from var j : Integer init 0
until j == Nb
loop
var fs : OrderedSet<UIField>
fs := getCUIFieldP(panelt)
fs.each{f|stdio.writeln(f.name)}
createPanelW(panelt,uiw1)
j := j + 1
end
  //remove the panels
uiw1.uipanel.remove(p)
uiw1.uipanel.remove(panelt)
end}
```

For a small-sized screen:

The height of a small-sized screen is not sufficient to display all information and the user must scroll to watch the entire window. This informational density negatively influences the performance of the user who can easily fall into error if he does not see the rest of the window.

The insertion of the two criteria of “informational density” and “protection against errors” at this level of transformation is crucial to solving this problem.

Informational density:

To reduce the informational density, we replace the panels with windows. The window should also be divided into multiple windows appropriate to the size of the screen.

```
if scs < 100 then
getCUIPanel(uiw1).each{cuip|
  // RelationShip Treatment
var uirs : OrderedSet<UIRelationShip>
uirs := getRelationShip(inputModel)
var rscp : UIRelationShip init uirs.
detect{u|u.source == cuip.name}
var newrs : UIRelationShip init
UIRelationShip.new
newrs := rscp
createWindow(result,cuip,uiw1,srv)
stdio.writeln("Creation of window in the place of panel " +
cuip.name) }
```

Legibility:

The fact of replacing the panels by the window increases the legibility and clarity of the interface.

Protection against error:

The navigation between windows is realized by the buttons “next” and “previous”.

As a consequence, the quality of the interface in terms of informational density and protection against error increases.

4.3.3. *Third Transformation Module Parameterized By Environment Model And Ergonomic Model*

The injection of ergonomic criteria in the third transformation module parameterized by environment model can mainly improve the legibility criteria.

Legibility:

The performance is enhanced when the presentation of information on the screen reflects the characteristics of the environment. Good legibility facilitates reading the information presented. For example, if we are at night, we use an appropriate interface. Hence the criterion of clarity depends sometimes on environmental characteristics, which explains the insertion of this criterion at the level of this transformation module.

5. Illustrating Example

In this paper, we have proposed an ergonomic meta-model that is introduced as a parameter in the generation approach of plastic UI [3]. We have illustrated our contribution in modeling and adaptation with a case study that had already presented by [22].

The application is concerned with a Home Heating Control System (HHCS) that enables the user to control the temperature of the room for a given context of use. The Fig.6 shows the tree-based description of abstract user interface of HHCS.

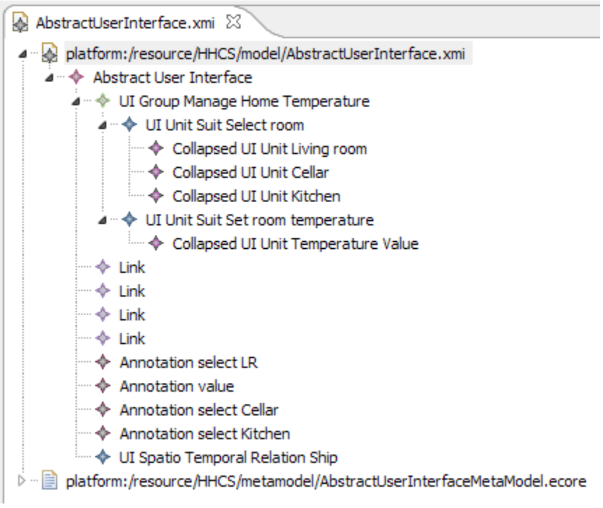


Fig. 6. The tree-based description of abstract user interface of HHCS.

In this model, the interface is composed of a “UIGroup” called “Manage Home Temperature” that presents the global task. This “UIGroup” contains two tasks represented by two “UIUnitSuit” “Select room” and “Set room temperature”. The first “UIUnitSuit” contains “CollapsedUIUnit” that presents the name of the room (Living Room, Cellar, Kitchen). These three values are enumerated in “Annotation” linked to this “CollapsedUIUnit”. The second “UIUnitSuit” “Set room temperature” also contains a “CollapsedUIUnit” that presents this time the temperature we want to associate with the room with a “symbol” Celcius that we want to display to guide the user.

The Fig.7 presents a visualization of the XML abstract user interface by AbstractUserInterface editor.

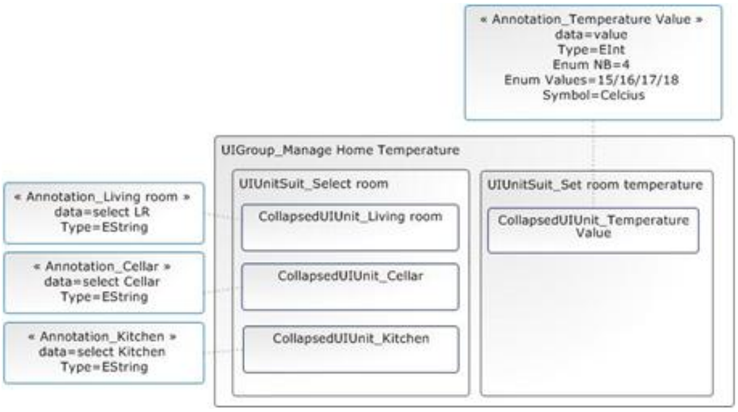


Fig. 7. Abstract user interface of HHCS.

5.1. *Parameterized Transformation By The Context Of Use*

The AUI is being transformed by injecting the characteristics of the user, platform and environment, respectively. The resulting concrete user interface is presented by the tree-based description (Left of Fig.8). Right of Fig.8 produces the CUI visualized with our ConcreteUserInterface editor developed thanks to the Graphical Modeling Framework (GMF) tool of Eclipse.

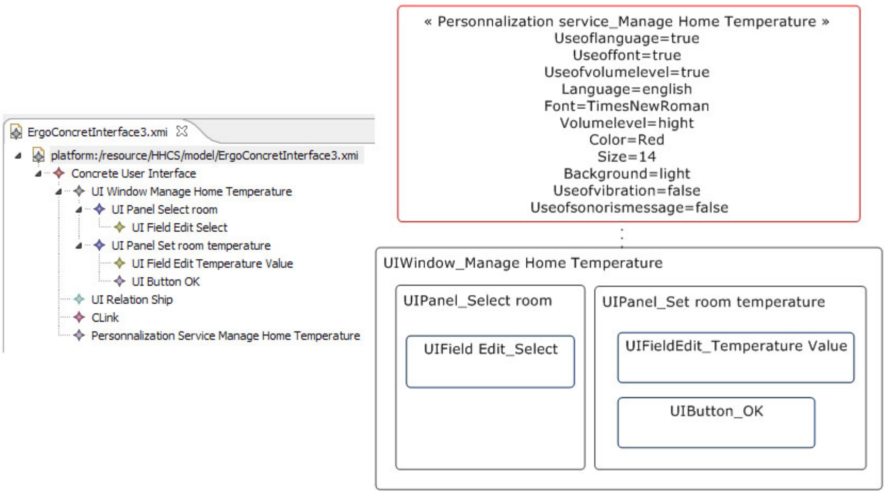


Fig. 8. (Left) The tree-based description of concrete user interface of HHCS. (Right) Concrete User Interface specific to the context model.

The resulting CUI consisting of a “UIWindow” “Manage Home Temperature” which contains two panels of the type “UIPanel”. Concerning the first panel “Set room temperature”, it contains a text field for entering the name of the room. As for the second panel “Set room temperature”, it contains a text field for entering the temperature without displaying the “symbol” Celsius. Therefore, this concrete user interface does not meet the specific criteria in “incitation”, “protection against errors”, “explicit actions”, “user control”. So we add the ergonomic model as a parameter of the transformation.

5.2. *Parameterized Transformation By The Context Of Use And Ergonomic Model*

For our case study and to better explain the impact of the ergonomic model of the transformation, we choose only a few ergonomic criteria for injecting into the process of generating the user interface. The criteria are “guidance”, “workload”, “explicit control” and “error management”. The ergonomic model that we added as a parameter is presented by the following xml code.

```
<?xml version="1.0" encoding="ASCII"?>
<Ergonomic:ErgonomicModel xmi:version="2.0" xmlns:xmi=
"http://www.omg.org/XMI"                          xmlns:xsi=
"http://www.w3.org/2001/XMLSchema-instance"
xmlns:ErgonomicMetaModel="ErgonomicMetaModel"
xsi:schemaLocation="ErgonomicMetaModel../metamodel/Ergonomi
cMetaModel.ecore">
  <criterion name="guidance">
    <subCriterion name="prompting"/>
    <subCriterion name="legibility"/>
  </criterion>
  <criterion name="workload">
   <subCriterion name="brevity">
    <elementaryCriterion name="concision"/>
    <elementaryCriterion name="minimalActions"/>
    </subCriterion>
   <subCriterion name="informationDensity"/>
  </criterion>
  <criterion name="explicitControl">
    <subCriterion name="explicitUserAction"/>
    <subCriterion name="userControl"/>
  </criterion>
  <criterion name="errorManagement">
    <subCriterion name="errorProtection"/>
  </criterion>
</ErgonomicMetaModel:ErgonomicModel>
```

5.2.1. *Step1: First Transformation Module Parameterized By User Model And Ergonomic Model*

Following the addition of our ergonomic model, the concrete user interface resulting from the first transformation contains three radio buttons that correspond to the names of the rooms instead of a text field for entering names. The second panel “Set room temperature” contains a list box instead of the text field for entering temperature. Hence the criterion “protection against errors” is satisfied. We have posted the symbol “Celsius” as a label with the drop-down list in order to meet the “incitation” criteria. The injection of the two sub-criteria “explicit actions” and “user control” resulting in the two buttons “Ok” and “Cancel”. The protection against error is translated by the fact of replacing the text field either by the radio buttons or a list according to the number of elements. The top of Fig.9 illustrates the tree-based description of HHCS concrete user interface adapted to the user and ergonomic models. The bottom of Fig.9 is the graphical presentation of our ConcreteUserInterface editor.

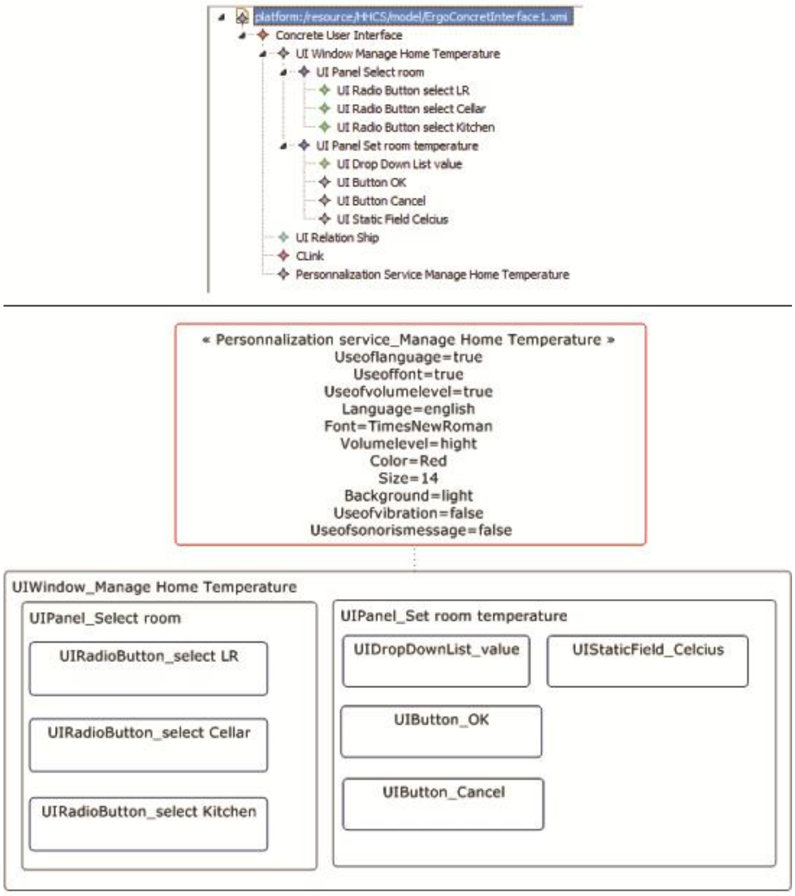


Fig. 9. (Top) The tree-based description of HHCS concrete user interface adapted to the user and ergonomic models. (Bottom) Concrete User Interface specific to the user and ergonomic models.

5.2.2. *Step2: Second Transformation Module Parameterized By Platform Model And Ergonomic Model*

Now the concrete user interface generated by the first transformation will play the role of the source model for this second transformation. Indeed, this transformation will take the platform and quality models as parameters. Top of Fig.10 describes the tree-based description of concrete user interface of HHCS generated after the transformation. The graphical presentation of the CUI is showed bottom of Fig.10.

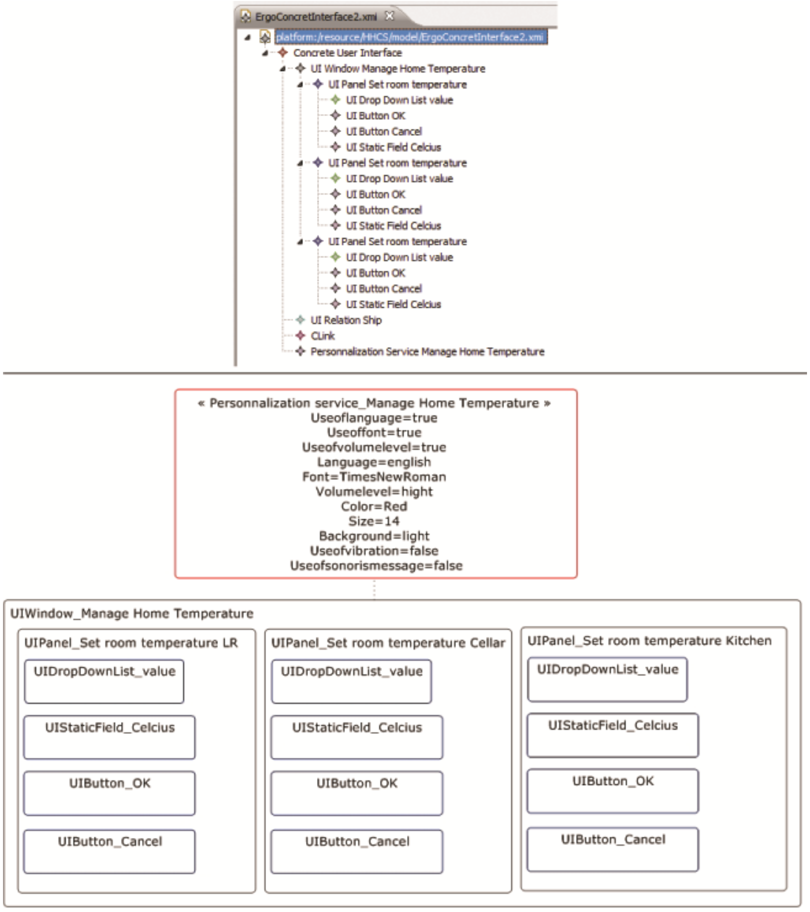


Fig. 10. (Top) The tree-based description of HHCS concrete user interface adapted to the user, platform and ergonomic models. (Bottom) Concrete User Interface specific to the the user, platform and ergonomic models.

In this second concrete user interface, the “minimal Actions” sub-criteria are taken into account. During the second transformation module, the panel “Select room” which is made only by three radio buttons is deleted and replaced by three panels “Set room temperature” with a panel for each button. And before proceeding to replace our two panels into three, our program demonstrates that the platform characteristics (screen size, for example) can achieve this transformation and therefore allow the insertion of this criteria.

5.2.3. *Step3: Third Transformation Module Parameterized By Environment Model And Ergonomic Model*

This last transformation allows the generation of the final concrete user interface presented by Fig.11. The dark letters on a light background are easier to read than the other way around and therefore taking into consideration the sub-criteria “legibility” is translated by the addition of the attribute Color which presents the color of the letters having the value “Blue” and the background attribute that the program gives a value of “light”.

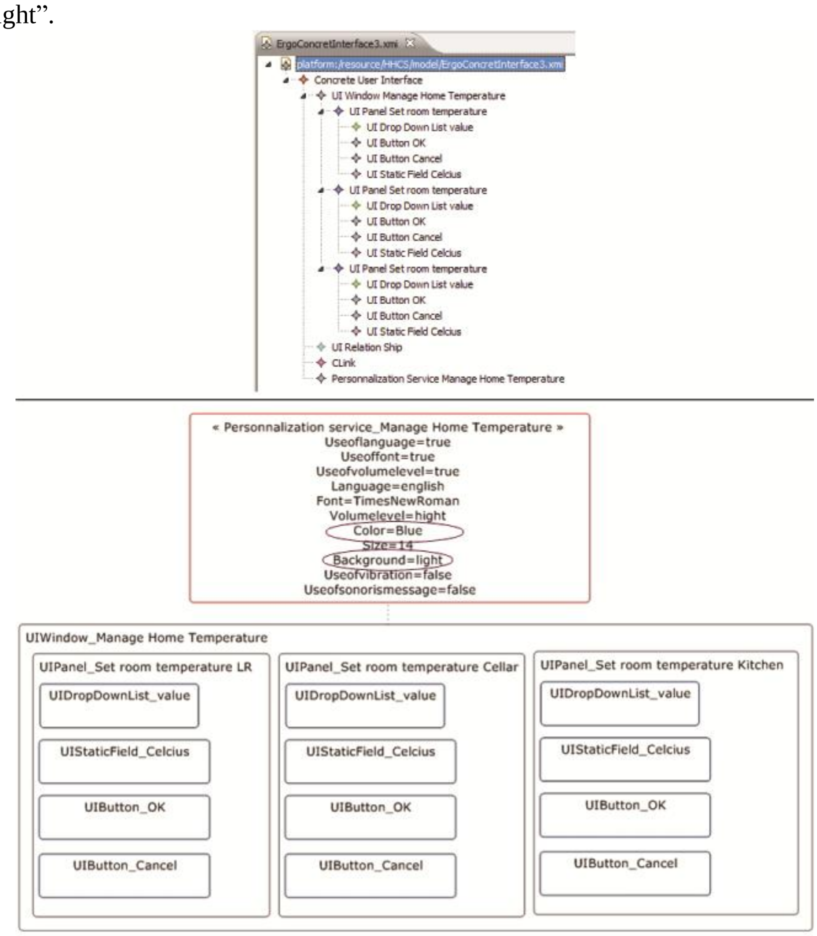


Fig. 11. The tree-based description of HHCS concrete user interface adapted to the context of use and ergonomic models.

6. Conclusions

Plasticity claims in its definition the preservation of usability. The process automation of generating the plastic UI, while respecting the ergonomic quality is not a simple matter to achieve. In this paper, the emphasis is on the preservation of the plastic UI usability. We propose an ergonomic meta-model that serves as a parameter for the process of generating a UI adaptable to the context of use. The generation process consists of three transformation modules starting from an abstract user interface and generating a concrete user interface by inserting the user, platform and environment model, respectively. The ergonomic model serves as a parameter in the three transformation modules.

The continuation of our work will naturally lead to the study of the possibility of merging the three transformation modules into a single model. The latter has a source model which is the abstract interface, a target model which is the concrete interface and four parameter models that are the user, platform, environment and ergonomic. The problem that arises is the causal relationship between the different thirds of the context of use as well as the priority between the different ergonomic criteria of the ergonomic model.

References

[1] Bastien, J. M. C.; Scapin, D. L. (1993): Ergonomic Criteria for the Evaluation of Human-Computer Interfaces}, version 2.1. Technical report N°156, INRIA.

[2] Bézivin, J.; Blay, M.; Bouzeghoub, M.; Estublier J.; Favre, J. M. (2005): Action spécifique CNRS sur l'Ingénierie Dirigée par les Modèles. Rapport de synthèse.

[3] Bouchelligua, W.; Mahfoudhi, A.; Abed, M. (2012): A Model Driven Engineering Approach Toward User Interfaces Adaptation. International Journal of Adaptive, Resilient, and Autonomic Systems (IJARAS), Vol **3**, pp 65-86.

[4] Bouchelligua, W.; Mezhoudi, N.; Mahfoudhi, A.; Daassi, O.; Abed, M. (2010): An MDE Parameterized Transformation for Adaptive User Interfaces. G.a. Tsihrintzis, E. Damiani, M. Virvou, R.j. Howlett, L.c. Jain (Ed.), Intelligent Interactive Multimedia Systems and Services, Springer-Verlag, Berlin Heidelberg, ISBN 978-3-642-14618-3, pp 275-286.

[5] Calvary, G.; Coutaz, J.; Thevenin, D.; Limbourg, Q.; Bouillon L.; Vanderdonckt, J. (2003): A unifying reference framework for multi-target user interfaces. In Journal of Interacting With Computer, Elsevier Science B.V, Vol **15/3**, pp 289-308.

[6] Correani, F.; Leporini B.; Paterno, F. (2006): Automatic inspection-based support for obtaining usable web sites for vision-impaired users. Universal Access in the Information Society, **5(1)**:82-95.

[7] Daassi, O. (2007): Les COMETs : une nouvelle génération d'Interacteurs pour la Plasticité des Interfaces Homme-Machine. PhD thesis.

[8] Frankel, D. (2005): Reflection of the state of MDA. MDA Journal, Meghan-Kiffer Press, pp 1-9.

[9] Frey, A. G.; Ceret, E.; Dupuy-Chessa, S.; Calvary, G. (2011): QUIMERA: a quality metamodel to improve design rationale. Proc. of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems. Pisa, Italy.

[10] ISO (2006): ISO 9241-110: Ergonomics of human-system interaction - Part 110: Dialogue principles. ISO.

[11] ISO/IEC (2001): ISO/IEC 9126-1: Software engineering. Product quality - Part 1: Quality model. ISO.

[12] ISO/TEC (2009): ISO/IEC CD 25010.3: Systems and software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Software product quality and system quality in use models. ISO.
[13] Kermeta (2012): Kernel Meta-modeling Framework. http://www.kermeta.org/
[14] Leporini, B.; Paterno, F.; Scorcia, A. (2006): Flexible tool support for accessibility evaluation. Interacting with Computers, **18(5)**: 689-890.
[15] Limbourg, Q.; Vanderdonckt, J. (2004): UsiXML: A User Interface Description Language Supporting Multiple Levels of Independence. Engineering Advanced Web Applications, Rinton Press, Paramus, M. Matera and S. Comai edition, pp 325-338.
[16] MDA (2012): Model Driven Architecture. http://www.omg.org/mda
[17] Meskens, J.; Loskyll, M.; Seibler, M.; Luyten, K.; Coninx, K.; Meixner, G. (2011): GUIDE2ux: A GUI Design Environment for Enhancing the User eXperience. Proc. of the 3rd ACM SIGCHI Symposium on Engineering Interactive Computing Systems. Pisa, Italy.
[18] Mori, G.; Paterno, F; Santoro, C. (2003): Tool Support for Designing Nomadic Applications. Proceedings of the International Conference on Intelligent User Interfaces, pp 141-148.
[19] OMG (2003): UML 2.0 Superstructure. OMG document ptc/03-08-02.
[20] Seffah, A.; Donyaee, M.; Kline, R.; Padda, H. (2006): Usability measurement and metrics: A consolidated model. Software Quality JournalSoftware Quality Journal, **14(2)**:159-178.
[21] Serna, A.; Calvary, G.; Scapin, D. (2010): How Assessing Plasticity Design Choices Can Improve UI Quality: A Case Study. In Proceedings of the 2nd ACM SIGCHI symposium on Engineering interactive computing systems (EICS'10).
[22] Sottet, J. S.; Calvary, G.; Coutaz, J.; Favre, J. M. (2007): A Model-Driven Engineering Approach for the Usability of Plastic User Interfaces. In the proceedings of Engineering Interactive Systems 2007 joining Three Working Conferences : IFIP WG2.7/13.4 10th Conference on Engineering Human Computer Interaction, IFIP WG 13.2 1st Conference on Human Centred Software Engineering, DSVIS - 14th Conference on Design Specification and Verification of Interactive Systems, University of Salamanca, Spain.
[23] Thevenin, D.; Coutaz, J. (1999): Plasticity of User Interfaces: Framework and Research Agenda. In Proceedings of INTERACT'99, Edinburgh, Scotland, A. Sasse & C.Johnson Eds., IFIP IOS Press Publ., pp. 110-117.
[24] Vale, S.; Hammoudi, S. (2008): Context-aware Model Driven Development by Parameterized Transformation. In MDISIS'2008, Proceedings of Model Driven Interoperability for Sustainable Information Systems.
[25] Vigo, M.; Aizpuruaa, A.; Arruea, M.; Abscala, J. (2009): Automatic device-tailored evaluation of mobile web guidelines. The New Review of Hypermedia and Multimedia, **15(3)**:223-244.

International Conference on Software Engineering and New Technologies

USING ADAPTIVE AGENTS FOR DOS ATTACKS DETECTION*

M.Cheikh

Computer Sciences Department,
University of August 20, 1955,
Skikda / LIRE Laboratory
Algeria
d.mohamedcheikh@gmail.com

S.Hacini

LIRE Laboratory, Computer Sciences Department,
Mentouri University,
Constantine, Algeria
salimahacini@gmail.com

Z.Boufaida

LIRE Laboratory, Computer Sciences Department,
Mentouri University,
Constantine, Algeria
zboufaida@gmail.com

A denial-of-service attack (DoS attack) is an attempt to prevent a service from functioning efficiently or at all, temporarily or indefinitely. This attack is typically launched against application servers or web servers and can be done by overloading the network with useless information. The intrusion detection systems are powerful tools for the detection of attempted DOS (Denial Of Service) attacks. However, they suffer from a number of problems such as high rate of false positives and negatives. In this paper, we present a new adaptive mechanism for DOS intrusion detection based on the use of agents. This self-learning mechanism ensures detection of DOS attacks with a reduction of false alarms.

Keywords: Behavioral approach, adaptive intrusion detection, DOS, network security, Intrusion Detection System, Multi-Agent Systems.

1. Introduction

Nowadays, networks and computer systems have become essential for the efficiency of most companies. However, the interconnection of these various systems and networks to the Web has made them vulnerable. Indeed, sensitive information of these companies may be disclosed, altered, or destroyed. Consequently, the efficiency of their systems can be hindered and their brand image shaken. Once these business networks have emerged as potential targets of attacks, their security has become an unavoidable issue. Besides the establishment of firewalls and authentication systems, it is necessary to improve this security policy, using IDS (Intrusion Detection System). IDSs are tools of powerful monitoring and control employed for the audit of information systems and the detection of potential intrusions [J, Andersonl (1980)]. , [D.E, Denning (1987)].

The behavioral approach ,which is used in this paper for the detection of DOS attacks, can detect unknown intrusions and therefore does not require prior knowledge about intrusions [A, Sundaram (1996)].,[K, Boudaoud (2000)]. In this approach, a normal pattern of activity is defined to build a profile of user activity. Any deviation from the established norm is considered as abnormal.

The behavioral approach uses multiple detection techniques [K, Boudaoud (2000)]., [K, Ilgun (1992)]., [A, Mounji (1997)]. ,[S, Kumar (1995)]. The most common is to generate a model of normal behavior through a statistical model. Despite the observed development of these behavioral techniques, the problem of high false alarms rate, especially for DOS attacks, remains posed.

The proposed approach is based on the adaptation of the normal profile of a subject to detect the DOS attack and minimize the number of false alarms (positive and negative). Agents are used in our work to ensure three main tasks:

- The learning and the detection.
- The enrichment of the normal profile as and when new legitimate activities are discovered.
- The filtering used to discover illegitimate activities in the normal profile. Note that the learning phase is not assumed safe and that filtering is used, in our approach, against the circumvention of IDS.

This paper presents an adaptive model for DOS attacks agent-based detection. This adaptation induces an update of the normal profile. Section 2 of this article describes the proposed approach. Section 3 quotes some works dealing with agents-based IDSs. Finally, Section 4 concludes the paper and suggests some perspectives to improve this work.

2. Proposed Detection Approach

Normal behavior of a subject cannot be fully represented, as it can change over time. That's why we designed a mechanism based on agents to adapt this behavior as and when it evolves.

Detection based on the behavioral approach is generally performed in the learning phase and detection phase [K, Boudaoud (2000)]. The proposed approach provides adaptation of the normal profile (on which detection is based) created during the learning phase in two specific cases:

1. The presence of activities from a legitimate user. These activities are not known a priori by the system. They do not exist in the initial knowledge base (already created during the learning phase). The latter will be enriched in every operation of adaptation by new legitimate activities, thereby reducing the rate of false positives.
2. The presence of activities from an illegitimate user (DOS attack), following an attempt to bypass IDS. This minimizes the rate of false negatives.

So the normal profile must adapt positively (by adding legitimate activities) or negatively (by the removal of unlawful activities). Therefore, it can detect DOS and minimize the number of false alarms. The adaptation of the knowledge base leads to a reduction in the rate of false positives and negatives in the detection.

2.1. *Proposed architecture*

The proposed detection system is applied to network traffic. The various parameters observed during the detection thus involve packets routed in the network. A set of nine agents is used to distribute the DOS detection task. The observed parameters were chosen based on the work of [A, Mitrokotsa and C,Douligeris (2005)]. They concern a traffic network and include:

(1) Duration

(2) Source bytes

(3) Destination bytes

(4) Count

(5) Same service rate

(6) Connections with SYN errors

(7) Connections-Same service-SYN errors

(8) Destination-Host-SYN error rate

(9) Destination-Host-Same-Service error rate

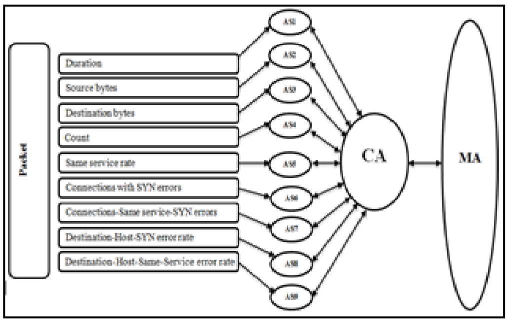


Fig. 1. Overall architecture

It is characterized by three types of agents:

- Agents Sensors (ASi, i = 1, ..., n). These agents are involved in the capture of parameter values of DOS attacks, acquired from the same packet.
- The Control Agent (CA) handles the collection of agents sensors' decisions and decide on the application of adaptation.
- The Manager Agent (MA) is responsible for the overall decision making on the trigger or not of an alert, based on the decisions of ASs.

The proposed detection approach is supported by a set of assumptions:

- n: represents the number of ASs.
- (KB1, KB2, , KB9) are knowledge bases associated respectively with (AS1, AS2, ..., AS9).
- (V1, V2, , V9) represent values obtained on the same packet for the parameters (Pr1, Pr2, ..., Pr9) respectively.
- (dA1, dA2... dA9) denote the decisions of (AS1, AS2... AS9) for respectively values (V1, V2... V9).
- We denote by 1: normal packet and by 0: a confirmed DOS Attack.

2.2. *The intrusion detection DOS*

The proposed mechanism is divided into two phases: the learning phase and detection phase. The first phase is learning and the creation of KBs during a definite period of time. At the end of this phase, the system is capable of detecting DOS while allowing the system of possible adjustments to KBs.

2.2.1. *The learning phase*

It uses a statistical approach commonly used for the generation of a model of normal behavior of a subject. The task of generating the model of normal behavior (learning phase) is partitioned between ASi (i=1,2..,9). Each ASi is taking steps to model a statistical Pri on the normal profile of a packet. This modeling leads to the construction of the KBi. During this phase, each AS, builds according to the observed parameter:

1. A list of normal values in the case of a parameter not quantifiable, as number of legitimate destination port: 80, 443, etc...
2. An interval in the case of a measurable parameter, such as the length in bytes of the data source destination worm (Source Bytes) in one packet: eg [100, 250]

After the learning phase the system is ready for the detection of DOS.

2.2.2. *The detection phase*

The detection phase brings together the various decisions of ASs to obtain a global decision that leads to the triggering or not of an alarm. Before calculating the overall decision, the decisions of ASs (equal to 0 or 1) are sent first to the CA which in turn transmits them to the MA. It performs some calculations to make an overall decision. These calculations are as follows:
After retrieving the decisions (dA1, dA2,… dA9) sent by the ASs to CA, the MA calculates the average of these decisions AD. We consider a decision range of values [0, 1] divided as follows:
[0, 0.4[(0 to 0.4), [0.4, 0.6] (0.4 to 0.6) and] 0.6, 1] (0.6 to 1) (This subdivision is just given as an example).
Where (0.4 to 0.6) represents a decision threshold to cover extreme cases (ambiguity).
MA determines to which range of values belongs AD. There are three possibilities:
AD is between 0 and 0.4 $\Rightarrow$ DOS attack, or AD between 0.6 and 1 $\Rightarrow$ no attack, and AD between 0.4 and 0.6 $\Rightarrow$ Ambiguity.

In the third case, the number of agents who decide that a packet is normal is very close or equal to the number of agents deciding that this same packet is a DOS attack. In this case, the decision of the MA is based on FC (the Factor of Confidence) associated to each AS.

Two new factors are introduced Ap and An. They represent, respectively, the number of positive adaptation operations (adding in the KB) and the number of negative adaptation operations (the deletion in the KB) performed by each AS (initially An = 0 and Ap = 0).

With each adaptation of AS*i*, its Ap*i* or An*i* are incremented respectively if the adaptation is positive or negative. Each FC*i* (factor of confidence) is computed by the CA using the following formula “Eq. (1)”:

$$FC_i = \frac{Ap_i + 2 * An_i}{\sum_{j=1}^{n}(Ap_j + 2 * An_j)} \quad (1)$$

These FC*i* (*i*=1,2,..9) give the advantage to the decision of agents that made few updates to their KBs since it is assumed that the ASs that keep their initial KBs (with little or no change) are more stable. Thus, they are better than the ASs that made many updates of their KBs.

2.2.3. *The adaptation of the knowledge base*

Two reasons induce the use of the adaptation. The first reason is that, in most cases, the learning phase is carried out in an unreliable environment and the second reason is that the duration of the learning phase is too small in comparison with the detection system life cycle.
The adaptation of the ASs' knowledge bases is carried out during the detection phase. This adaptation is performed under the following two assumptions:

1. If the number of ASs deciding that a packet is normal is greater than or equal to SP (SP represents a Majority Threshold) and strictly less than *n*, an adaptation request is sent by the CA to all ASs which decide that this same packet is a DOS attack".
2. If the number of ASs deciding that a packet is a DOS attack is greater than or equal to SN (SN represents a majority threshold) and strictly less than *n*, an adaptation request is sent by the CA to all ASs which decide that this same packet is normal.

After the reception of the adaptation requests, each AS performs an update of its KB. This adaptation aims the KB to be enriched by new legitimate values of the normal profile or to be filtered and hence to remove some illegitimates values from KBs.

To ensure that KB created during the learning phase remains valid, ASs employ a technique to validate the addition and removal of the normal profile values of the packets. To do so, with each AS*i* is associated two lists of Adaptation (LP*i* (Positive List): Enhancements to the KB, LN*i* (Negative List): for delete in KBs). Both lists are initially empty. Then, whenever a Request for Adaptation (positive or negative) acquired a new value V is sent by the CA.

They are two cases:

a) Positive adaptation: the AS*i* checks whether the number of occurrences of V in the list LP*i* is greater than or equal to an integer RP*i* associated with AS*i*, representing a factor validation of a Pr*i*. The new value V must be added to KB*i* and removed from the LP*i*. By cons, if this number is less than RP*i*, AS*i* adds V to the LP*i*.
b) Negative adaptation: AS*i* checks whether the number of occurrences of V in the list LN*i* is greater than or equal to an integer RN*i* associated with ASi, representing a factor validation of a Pr*i*. The new value V should be deleted on KB*i* and removed from LN*i*. By cons, if this number is less than RN*i*, AS*i* adds V to the LN*i*.

Moreover, if an AS has performed an operation of adaptation, it is imperative to send a message for CA to confirm adaptation. The adaptation is achieved by:

1. A change in a range of decision such that the interval length data in a packet. Given the decision interval [Vi, Vs] (Vi: lower value, Vs: upper value), and V a new value to add or remove:
   a) For addition :
      IF (V ≥ Vs) THEN Vs ←V.
      IF (V < Vi) THEN Vi ←V.
   b) For delete :
      IF V ≥ ((Vi+Vs)/2) THEN Vs← (V-1) ELSE Vi←(V+1).
2. Addition of new values to a list of decisions such as the list of legitimate destination IP addresses, or removal of illegitimate values in case of negative adaptation provided that the list (KB) is not empty.

2.3. *Operating principle*

It is essential to assign first, to each detection parameter an AS. After this initialization, the learning phase is triggered. During this phase, the KB is generated for each AS. At the end of this phase, the detection system is put into service.

At each occurrence of a packet, each ASi (according KBi and its associated parameter) acquires a value V of the associated parameter of the packet observed after the acquisition of parameter values by ASs.

Each of them makes a partial decision observed on the packet (see Figure. 2).

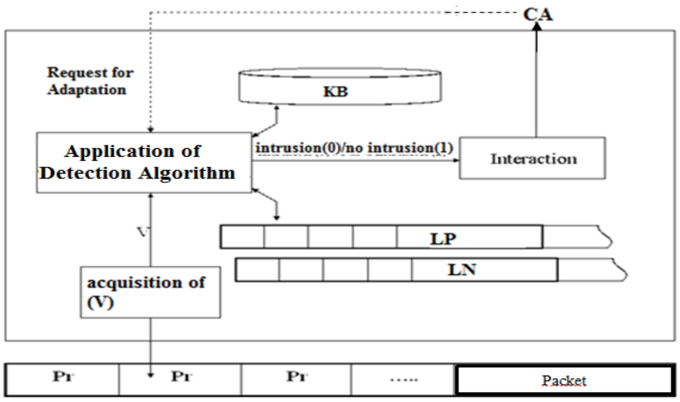


Fig. 2. Architecture of AS

The second step of our detection mechanism is performed by the CA (see Fig. 3).

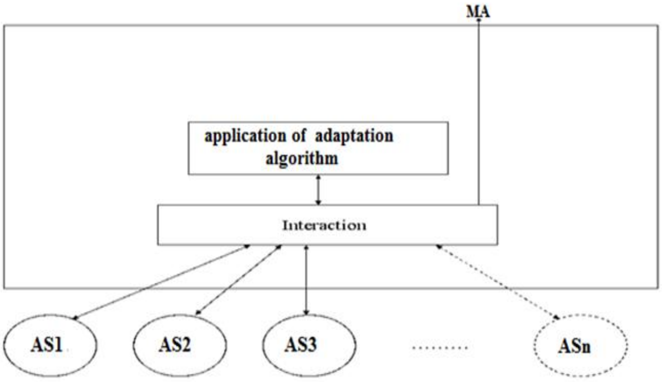


Fig. 3. Architecture of AC

It should be noted that during the learning phase, the decisions relating to the (normal) packet are made by the administrator of security, whereas after the training phase, the detection system makes decision.

In parallel with the operation of adaptation, the MA decides whether the packet analyzed is normal or not (see Figure.4).

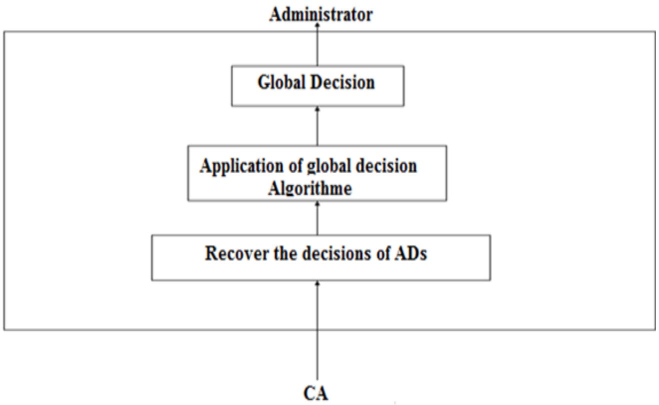


Fig. 4. Architecture of MA

3. The Related Work

Several studies have investigated the use of agents in intrusion detection systems. We cite as an example, the project 'The computer immunology "[S, Forrest et al (1997)]. [C. Warrender *et al* (1999)]., JAM (Java Agent for Meta-learning) [S. J, Stolfo et al (1997)]. , and AAFID (Autonomous Agents for Intrusion Detection) [J, Balasubramaniyan *et al* (1998)]. Each project addresses the problem of intrusion detection in different ways.

The project 'The computer immunology "explores the design of an intrusion detection system on the basis of the idea of biological immune systems. He helped develop a sense of autonomy of agent security for computer programs by observing the normal set of system calls executed by them. The JAM project uses distributed Java agents and the concept of "data mining" to learn models of fraud and intrusive behavior.

The project AFFID, meanwhile, has developed an intrusion detection system using agents for prototyping and cross-platform compatibility. The project also analyzed the agent-based approach to intrusion detection. The environment AFFID is agent-based intrusion detection, capturing and correlating data to perform intrusion detection.

More recently, Kannadiga and Zulkernine [P, Kannadiga and M, Zulkernine (2005)]. presented a new intrusion detection system called DIDMA (Distributed Intrusion Detection using Mobile Agents). It uses a set of mobile agents that can move from one node to another in a network, and perform the task of aggregation and correlation of data that has a relationship with the intrusion. These data are from a different set of static software entities called agents. Chong-jun and Shi-fu [W. J. W, Chong-jun and W. J. C, Shi-fu, (2006)]. also presented a model for BDI agents used, and achieved a detection system based on multi-agent cooperation which effectively improves the traditional distributed intrusion detection.

Vongpradhip and Plaimart [S, Vongpradhip and W, Plaimart(2007)]. proposed a new architecture capable of handling the attack over the network using agent technology with the design of the network topology.

Boughaci *et al* [D, Boughaci *et al* (2006)]. have established a mechanism for distributed intrusion detection based on autonomous agents and mobile. These agents are used mainly to cover information from the audit logs, to define the behavior of users, analyze data, and send alert messages in case of intrusion detection.

Chung Ming-Or [O. Chung-Ming (2012)]. proposed an IDS inspired by the theory of risk of the immune system.

There are relatively few studies concerned with the detection-based adaptive agents. For example, the work of Carb'o *et al* [J, Carb´o *et al* (2003)]. proposes an agent system based on the prediction of intrusions. There are also works that treated the detection of DOS / DDOS [J, Mirkovic and P, Reiherl (2005)]. , [J, Mirkovic *et al*(2006)].

All the presented works offer improvements in the detection and demonstrate the effectiveness of the use of agents in intrusion detection. Most of these works are based on a special mechanism that organizes and directs the agents to accomplish the detection task. However, the high false alarm rate remains a challenge for systems based specifically on the use of the behavioral approach.

The intrusion detection mechanism proposed in this paper, minimizes the positive and negative false alarm rates and enhances detection by use of a self-adaptive learning.

4. Conclusion

The mechanism of intrusion detection DOS presented in this paper is based on the use of a set of agents. These agents allow the detection of DOS attack, but also the adaptation of positive and negative profile of normal behavior. This allows both minimizing the rate of false alarms and improving detection of real DOS attacks.

The proposed detection system is mainly applied to the DOS attack, but it can also be used to detect other network attacks. To improve the detection rate, we consider assigning each parameter a weight representing its importance. We consider, in fact, that the importance of certain parameters is such that their faults should trigger a warning.

References

A. Sundaram., (1996), An Introduction to Intrusion Detection, Technical Report, Purdue University.

A. Mounji, (1997), Rule-Based Distributed Intrusion Detection, Thèse de l'université de Namur, Institut d'Informatique, Belgique.

A. Mitrokotsa, C.Douligeris, (2005), Detecting Denial of Service Attacks Using Emergent Self-Organizing Maps, IEEE International Symposium on Signal Processing and Information Technology.

C. Warrender, S. Forrest, and B. Pealmutter, (1999), Detecting intrusions using system calls : Alternative data models, In Proceedings of the 1999 IEEE Symposium on Security and privacy, pages 133-145. IEEE Press.

D.E. Denning, ,(1987), An intrusion-detection model, IEEE Transactions on software engeneering, SE-13: 222– 232.

D. Boughaci, H. drias, A. Bendib, Y. Bouznit and B. Benhamou, (2006), A Distributed Intrusion Detection Framework based on Autonomous and Mobile Agents, International Conference on Dependability of Computer Systems (DEPCOS-RELCOMEX'06) IEEE.

J. Anderson, (1980), Computer security threat monitoring and surveillance.

J. Balasubramaniyan, J. O. Gracia-Fernandez, D. Isacoff, E. H. Spafford, and D. Zamboni, (1998) , An architecture for intrusion detection using autonomous agents, Technical Report COAST TR 98-05, Purdue University Departement of computer Sciences.

J. Carb´o, A. Orfila and A. Ribagorda, (2003), Adaptive agents applied to intrusion detection, international Central and Eastern European conference on multi-agent systems No3, pp. 445-453.

J. Mirkovic and P. Reiher, , July (2005), D-WARD: a source-end defense against flooding Denial-of-Service attacks, IEEE Transactions Dependable and Secure Computing, vol. 2,no.3, pp. 216-232.

J. Mirkovic, M. Robinson, P. Reiher, and G.. Oikonomou, (2006), A framework for collaborative DDoS defense, Proc. of the 22nd Annual Computer Security Applications Conference, pp. 33-42.

K. Boudaoud, (2000), Détection d'intrusions : Une nouvelle approche par systèmes multi-agents, thèse de doctorat de l'école Polytechnique Fédérale de Lausanne.

K. Ilgun, (1992), USTAT: A Real-time Intrusion Detection System for UNIX, Master thesis, University of California.

O. Chung-Ming, (2012), Host-based intrusion detection systems adapted from agent-based artificial immune systems, Neurocomputing Volume 88, Pages 78− 86.

P. Kannadiga and M. Zulkernine, (2005), DIDMA: A Distributed Intrusion Detection System Using Mobile Agents, First ACIS International Workshop on Self-Assembling Wireless Networks (SNPD/SAWN'05) IEEE.

S. Kumar, (1995), Classification and Detection of Computer Intrusions, Technical Report Coast-TR-95-08, Coast Laboratory, Purdue University.

S. Forrest, S. A. hofmeyr and A. somayaji, (1997), Computer immunology, Communications of the ACM, 40(10) :88-96 ; Octobre.

S. J. Stolfo, A. L. Prodromidis, S. Tselepis, W. Lee, D. Fan, and P. k. Chan, August (1997) , JAM : Java Agents for meta-learning over distributed databases, In Proceedings of the 3rd International Conference on knowledge discovery and Data Mining, Pages 74-81, Newport Beach, CA, USA.

S. Vongpradhip and W. Plaimart, (2007) , Survival Architecture for Distributed Intrusion Detection System (dIDS) using Mobile Agent, Sixth IEEE International Symposium on Network Computing and Applications (NCA 2007).

W. J. W. Chong-jun and W. J. C. Shi-fu, (2006), Dynamic Hierarchical Distributed Intrusion Detection System Based On Multi-Agent System, 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW'06) .

International Conference on Software Engineering and New Technologies

DISASTER EMERGENCY SYSTEM

APPLICATION CASE STUDY: FLOOD DISASTER

MOSTAFA EZZIYYANI, SOUMAYA EL MAMOUNE, LOUBNA CHERRAT

Information Technologies System Laboratory, University Abdelmalek Essaadi, FST Tangier, Morocco
ezziyyani@gmail.com, Soumaya.elmamoune@gmail.com, cherrat81@yahoo.fr

MOHAMED ESSADI, OMAR KASSARA, IMANE SEFIOUI, MAROUANE MABROUK, AHMED ZIANI

Information Technologies System Laboratory, University Abdelmalek Essaadi, FS of Tetuan, Morocco
essaidi@gmail.com, okassara@gmail.com,imane.sefrioi@gmail.com, elmabrouk@gmail.com

Human security means protecting fundamental freedoms that are the essence of life. It also means protecting the individual against serious threats or generalized. It's necessary to build on the strengths and aspirations of each individual, but it also means creating systems that give people environmental elements essential to their survival, their lives and their dignity. In this context, we present in this paper a new concept to create an adaptive system to save lives and help the individual in all kinds of danger, the system called Disaster Emergency system "DES".

Keywords: Disaster; Flood; SMA; Sensor; Optimization, user profile.

INTRODUCTION

The term "Disaster" or "catastrophe" in French means the damaging effects of a brutal phenomenon, lasting or intense, natural or human origin. The consequences of a disaster are not measurable fracture organized continuity and comfort acquired. We distinguish particular situations including loss of life and widespread destruction. Contrary to accidents, disasters have resulted in new thinking on how to implement to avoid or to mitigate the disastrous effects. These reflections, which led in particular to the formulation of the principle of precaution, prevention and intervention, can lead to the creation of new standards or new legal requirements but to establish a prevention plan, you need reliable information and something that is difficult to obtain. The growing development of new information technologies, especially computer and telecommunications has all simplified concept that may be difficult or impossible. These technologies facilitate the extraction and communication of reliable information which will play the lead role in the prevention plan or rescue.

1. Objectif of the System

Seeing of the great danger posed by disasters and exploiting technological developments, this article presents the system "Disaster Emergency System" which manages any type of disaster or accident to meet the needs of those facing a disaster and aims to provide a rescue solution with the lowest cost possible.

The main objective of this system is the implementation of a new mechanism that serves to render services in a fast and efficient physical or moral danger and help them to cope with less damage or guide their choices according to their profiles and using a solution which is based on the integration of sensors with a network simulation methods via multi-agent systems and constraint programming.

Our interests are the trust and the quality of our support that this guarantees an efficient operation depending on the services offered by the system that occurs spontaneously cited as the best rescue system that gives individuals the essential elements of survival and several partial objectives were set as:

- Reduce the probability of accidents: After studying the tree of information and regarding the existence of an accident, the system can prevent the likelihood of the accident, evaluate the probability that he prevents the accident and notify the user.
- Inform the public about preventive behavior to be disaster or emergency: If the system receives information concerning a disaster, it shall inform any person may be exposed by this disaster.
- Reduce the severity, if they occurred despite the precautions taken: the severity of a disaster can be judged by the amount of damage and the latter to limit the severity our system improves the efficiency and quality of aid by implementation rules as call emergency rescue, fire …

2. D.E.S Services

To meet the requirements mentioned above and a complete system that handles disasters, there has been a process of working chained to provide multiple services into three categories:

2.1. *Before the Disaster:*

This is the first category where we try to predict the existence of a disaster in a region by studying the factors and causes that can lead to disaster. When a disaster occurs, it's necessary to analyze these factors (root cause) to prevent a similar accident does not happen again (capitalization of experience). All this to inform the user and any person subject to risk all measures to avoid a situation or an accident and make people understand what is the risk and why it is necessary to change (or control) the behavior according to the proposals offered by the system.

2.2. *During the Disaster*

During this step, the good deed done by the user can cost more in his life and for the fear and astonishment of the person subject to the catastrophe, he can't make good decisions. That is why our system offers to our users the solutions that can be adopted in a safe and reliable change as the path to follow, report emergencies by taking into account user profiles.

2.3. *After Disaster*

In this category, our system provides a means of real-time decision for emergencies and disasters in a time interval strictly minimum. This system is based on the collection of information, coordination with emergency units and the allocation of available resources to save the victims according to the severity of their situations. To simulate, we use the wireless sensor network to control the spread of dangers and collect data detected as intelligent agents are used to represent the virtual world in a different types of actors that interact to intervene in a situation.

3. Case Study: Flood

In this case study, we wanted to raise the complexity of the factors involved in a disaster to be able to grasp the sequence of components involved. It seemed important to investigate how to protect a company against risks and it requires a very good understanding of this complexity. In addition, this study should lead users to accurately handle the concepts of hazard, vulnerability, risk and disaster.

Fig1: Levels of areas risk

For the first category before the disaster, we try to provide a disaster and for each case study of the tree of cause and collet is the only information that leads us to reality. In our case that is flooding, the main causes of this disaster are due to:

- Rain and snow precipitation: Part of rainwater or snow is retained by the soil, absorbed by vegetation or evaporated floods occur when soil and vegetation can’t absorb any water and runoff cause an elevation of the bed of the stream. Most often, it doesn’t overflow, but sometimes the water runs in quantities that can’t be transported in the beds of rivers, or retained in natural ponds and artificial reservoirs behind dams. The river overflows and then produced a flood.
- Temperature: We also include the influence of temperature changes: the thaw causes an increase of the amount of water in the soil and rivers.
- Infrastructure: the nature of the terrain, the height of the region to study existence of rivers, dams ... are elements to be considered for a clear decision.

After identifying the critical factors influencing on having a disaster such as flood, we establish the tree causes or effects to notify users. Taking this case, it's assumed that the main factors of flooding rain precipitation, infrastructure, nature of earth ... based on these elements, we construct the following tree:

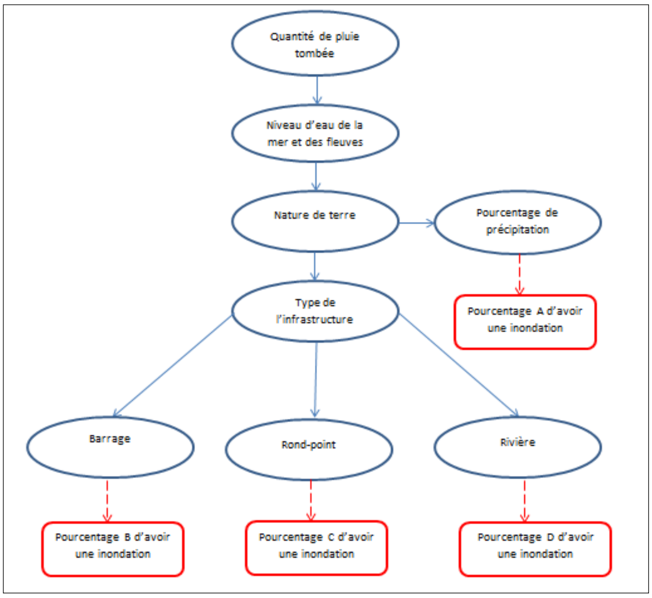


Fig2: Example of a tree cause

According to the values of photometer, causes of a disaster, our system can predict the percentage of having a disaster and inform users who may be exposed to this disaster so that they are careful. The information received will be a message, a ringtone or alert according to the configuration of the system by the user.

The disaster system offers several services according to user situation:

- Person affected by the disaster: if the user is hit by flooding, it may notify the fire department, a person emergency ambulance ... and all these issues will be configured by the user before using the system, given that the call automatically after user authorization. Thus, if the user has the possibility to move, he can see his own card to know the most secure way to get away with less damage. The system can also provide backup solutions most suitable for the victim.
- Person exposed by the disaster: if the user is near a flood by a distance, that is also configured, the system sends an alert to the user so they can make a decision or seek the path most secure and least cost, notify a person or not to move. The right decision will be made using the system map. In this case, the system

receives information from the network of wireless sensors that detect the level of water in the place flood.

- After the flood and after each disaster, there is more material damage and human, this point addressed in this class: how can we intervene to save people? According to several reports entered by users and collected automatically via sensors, the system can establish the action plan according to user states starting with the living that are in critical cases or dying.

4. Adapted Solution

To meet the needs identified and implement a robust system and develop the functionalities planned, we made reference to the following technologies:

4.1. *Multi-Agent -System*

MAS is a system composed of a set of agents situated in some environment and interacting according to some relations. An agent is an entity characterized in that it is, at least partially independently. It can be a process, a robot, a human being, etc. The main purpose of our project in SMA is to form an interesting type of modeling companies and have a wide application, up to the humanities.

4.2. *Wireless Sensor Network*

Our system uses dynamic data such as temperature, distance, acceleration, tilt, displacement, pressure, humidity, pH ... These variables are called variable instrumentation to be picked up by a sensor. The latter provides an output data of a specific physical quantity. The input value is called the measurand (the physical quantity that we want to know) and it causes a response in the sensor. To capture a series of values in different regions, we use a wireless sensor network: is a set of devices, wirelessly connected and are able to obtain information from the environment. They have no wired infrastructure or centralized administration. Its main features are as follows:

• They have limited processing power

• They have very low power consumption

• They are low cost

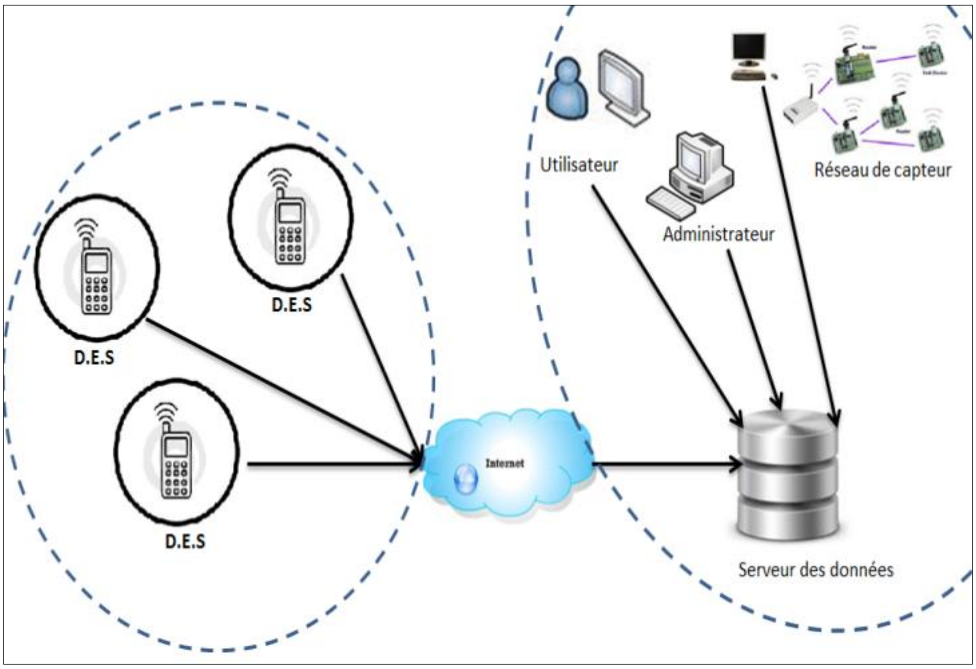


Fig3: D.E.S functional Schema

As shown in this diagram, our system will monitor the client-server architecture: the server has data that are indispensable information for the system and the client enjoys the services offered by the system are based on the data stored on the server and accessible by the Internet.

To mount the usefulness of this architecture, we can give the example of the configuration of a card: Each user can describe a geographic area by positioning several objects that have different properties, which is why each user will have its own database which will be adjustable to give greater freedom to manage the additional information needed or to be stored in the server. Also, there is a sensor network that collects real-time information and stores so that they are accessible and usable at all times.

For the proper management of the server, there are two types of users:

Administrator: a person who has all the rights to manage data grapple for general information stored, delete unnecessary information...

Customer: the person who uses our system and don’t have all the rights, but he has the rights that are related to their profile.

4.3. *Meta data*

The data provide essential information for the system so that each user can describe a geographic area by positioning several objects that have different properties, which is why each user must have its own database which will be configurable. To meet the needs of all users, we use a meta-database.

Meta-database is a technique appropriate to allow users the freedom to add more fields and tables as they want without ever having to fear either their volume or ease of interrogation. And all this without changing the architecture of the database will be ontology for the global schema definition on which different users query the database. It's the management of the ontology is a meta-database that manages other database will be stored in a data server accessible via the Internet.

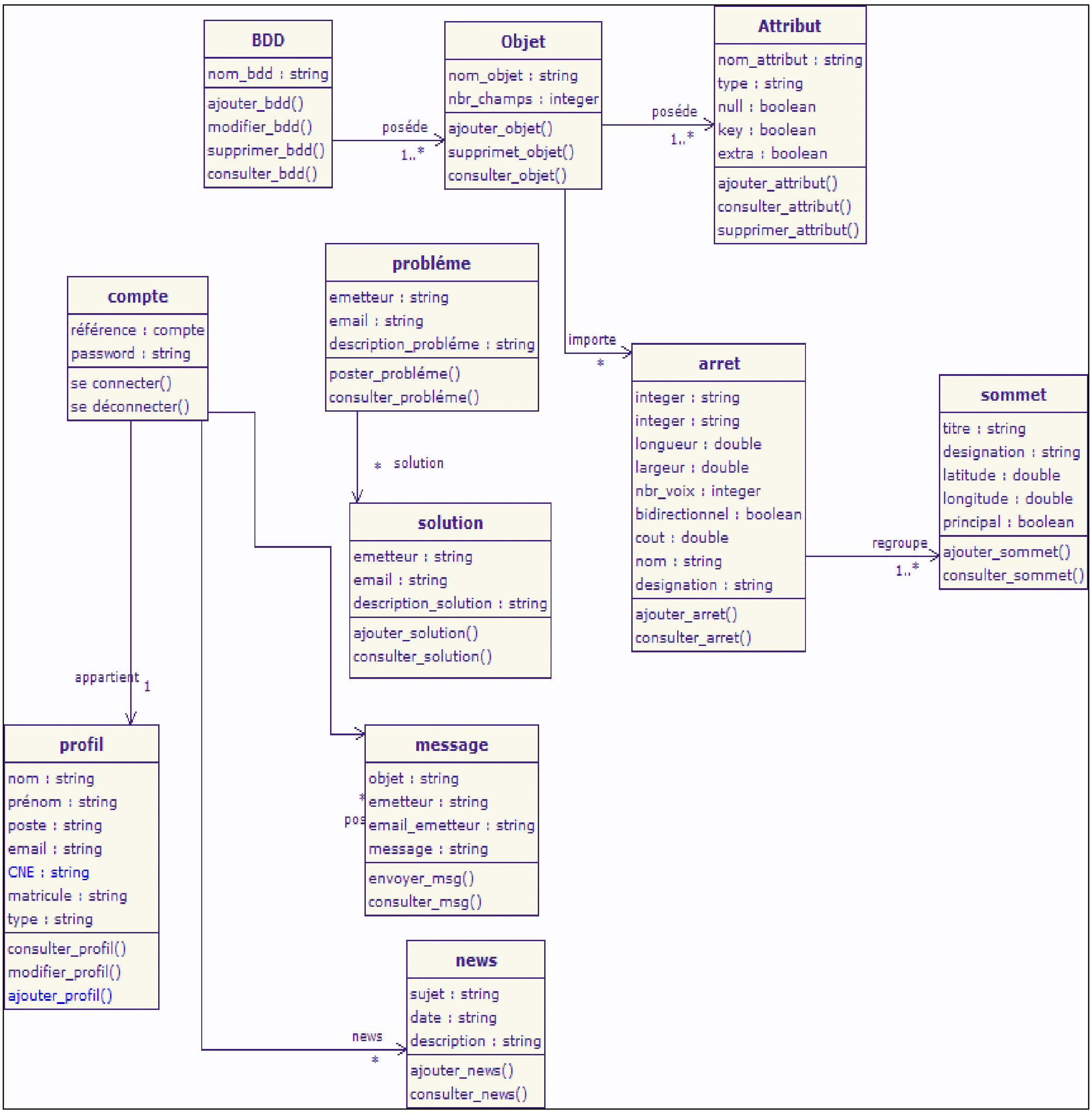


Fig4: Class diagram of meta-database

5. Layer architecture of the D.E.S system.

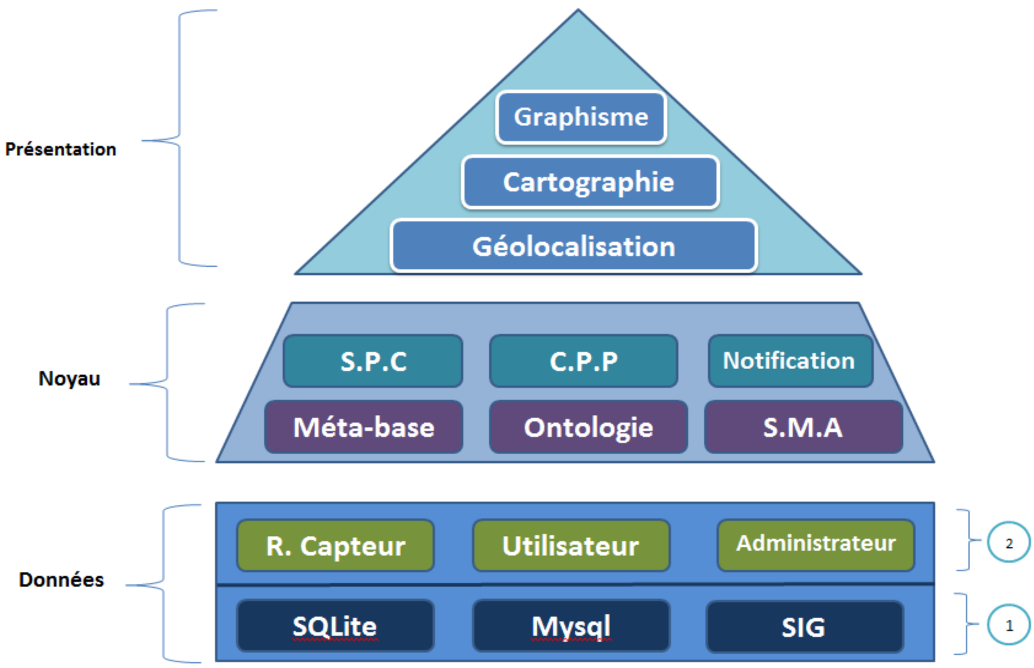


Fig5: Architecture of D.E.S

Our system consists of three layers:

Data: This is where the base layer manages the data that are essential in our project information data can be classified into two types data static data (1) and dynamic data (2). Static data are stored as different SGBD SQLite and MySQL and information sources such as GIS.

The dynamic data are changeable data and interact with the system as the sensors which transmit to each of the different time values depending on the situation.

Core: This is the layer where the system performs its functions based on the data layer. In this layer, we find the performance of different algorithms of our system as the shortest path (CPP), the position of objects on the map (SPC), user notification, management of the first layer...

Presentation: This is the final presentation of solutions to the user on the screen of his mobile and it is used for geolocation and mapping to create maps and graphics for the message in good shape.

The development of component technology for years played a crucial role in the development of computer science and to develop an application that fits with the latest trends and new news, we must seek the latest updates and the latest devices on the market. To achieve this system and meet the needs of our users, we have used new technologies to make it a complete system of these technologies includes:

6. User profil Ontology

The personalization of information is a major challenge. The relevance of the information provided its intelligibility and its adaptation to use and user preferences are key factors of success or rejection of these information systems. To do this, one of our goals is to focus on modeling user preferences in information retrieval, thus defining architecture for the user profile taking into account the size of center user interest based on historical question, then integrating this profile in a process of finding information to customize the results returned by the system DES. In this context the use of ontologies profiles for:

- facilitate access to multiple sources of information,
- exchange information in order to develop and participate in a collaborative process of solving distributed
- Distribute and balance workload.

We propose the class diagram focus of implementing the relationship between domains and users

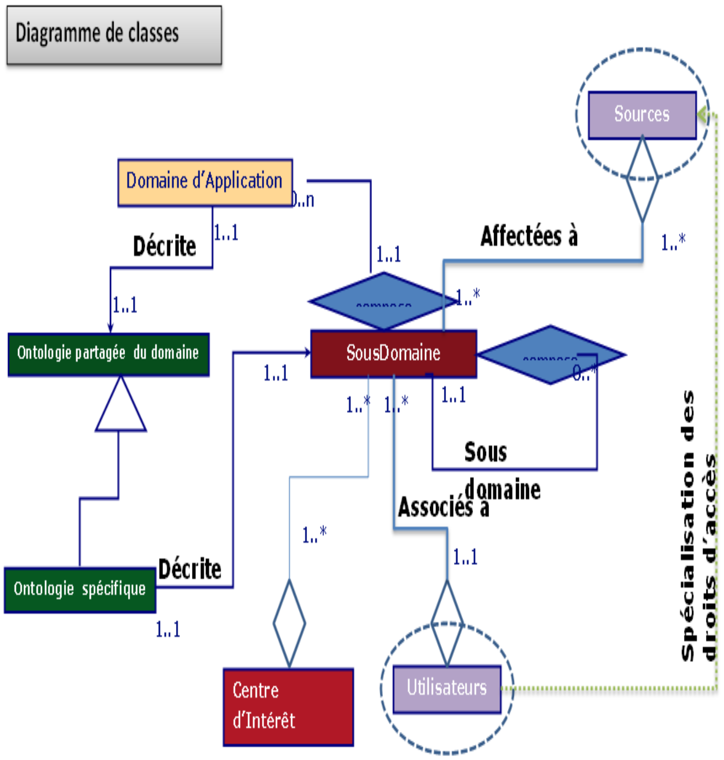


Fig6: Class diagram of interest center.

6.1. *Hierarchical organization of interest center*

The center of public interest is common to all users, but can be customized dynamically to the user when he interrogate the system, forming interest groups adapted to different users of the system. Indeed, the focus on a public or shared specific areas (research or not), provided by experts, which will be used by the system for the construction of interests of users classified by categories. They contain knowledge and information shared between groups of users of the same mode of operation of the system concerned.

Knowledge of each focus is copied into the user workspace when the user selects it. The system can then complete this focus by protected or private specific knowledge. Generally, interests provided by the system are not reduced because it is a prototype, but we plan to create a collection of interests-like directories categories of types of users to limit the interests in number.

Against our application context, we are particularly interested users of DES system for disaster management in real time. We identified three groups of these users for which our approach can improve the effectiveness of their work online:

- Administrative staff who want to compose static and newspapers for decision making.
- Agents who wish to intervene to rescue the situation by their victims and the type of disaster.
- Victims who want optimum support in real time during a disaster.

The main common point of these users is unaware of the exact type of disaster and damage when using the system. It is therefore necessary to adapt the system so not only specific to each user, but also as and navigation of the same user. To do these use ontologies cooperatives reusing techniques from personalization, customization and adaptive questioning to improve and optimize resources and improve the interrogation process systems.

For example, we suppose a security guard needs to save through a victim's vehicle. In the initial study, the system is to recommend ways compatible and not damaged by the disaster. In uncertainty, this question requires the use of the ontology of profiles for the same group as the basis for this guide. Therefore, the system can recommend paths according to a set of constraints and knowledge of the ontology of profiles (such as bridges without crossing or avoid flooded roads). In addition, the ontology can help agents to search for information about the travel path, it can search for suitable locations for emergency if necessary. Dimensions of interest center

The focus is to bring together in a knowledge base for personalized search results with the profile. This knowledge is used by the monitoring tool for communicating news sources on a topic or in a specific area.

The focus is therefore a continuous updating of knowledge a user or group of users vis-à-vis the data sources integrated in the system. There are therefore two types of selective knowledge: one is the provision of reference meets the profile of a single user:

the diffusion is then strictly "individual" is called "custom", the other is the provision of reference corresponding to the profile of interest from a homogeneous group of users. It is a "group profile" which usually results in a proposal "standard profiles". The users have the different objectives, different interests; therefore different perceptions of the concept of relevance, the same user may have different interests at different times. Based on the assumption that a user has different interests and it can switch from one focus to another during successive search sessions.

These dimensions are integrated in perspective in a knowledge base represented by an ontology exploited in a process of matching queries and querying historical sources and interaction with the DES system. We propose a method for this purpose cyclic proceeds in two steps. The first is to extract from the history of interactions a focus candidate qualified 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 hypothesis respecting the diversity of possible interests. This reflects the learning phase of interests which induces changes in the profile.

In order to assess the degree of change in contexts related to interests of the user, the context is described by an adaptive ontology. In the end, we describe the focus of four dimensions. The first is the history of its interactions with the system ON. The second recurring feature needs quality information, it is inferred and evolves from the first dimension. The third set of security operations and access to data sources. In the end, the fourth grouped on knowledge and data semantics and functionality offered by the mediation systems.

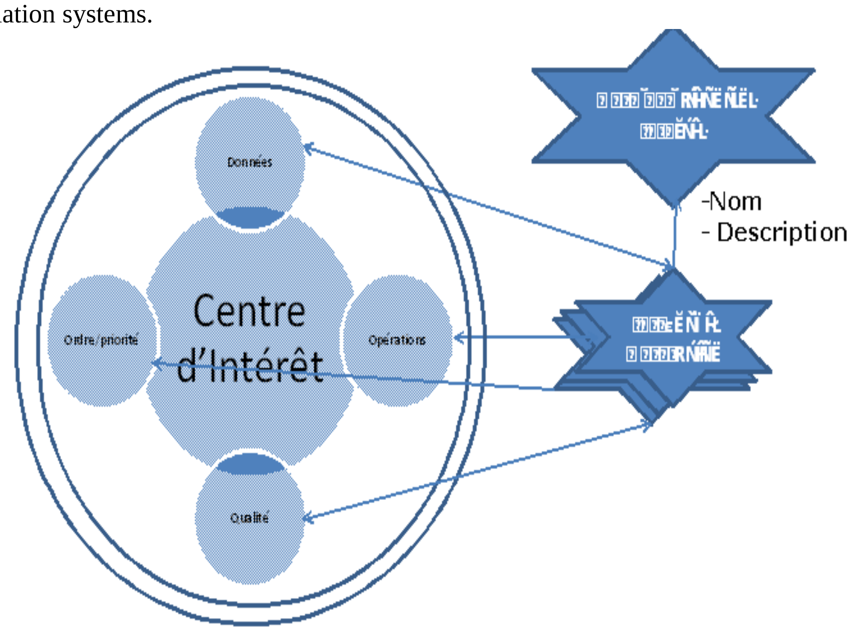


Fig 7: Focus

6.2. *Personal Data*

Personal data to better manage users. For example, they generalize the preferences of customers on the basis of demographic criteria (women are more vulnerable to disasters). We can consider this dimension as an organizer and in this case it should contain data on personal and professional contacts. In this dimension we find the data to identify a user (Citizen). In our case these data are grouped into two main sub-categories: static data and dynamic data or invariant. For the first category distinguishes between identity data: containing information identifying the person as his name or CIN. Data and demographic information on the personal and professional user number as address, contact etc. For the second category data are likely to evolve over time and space and ask for automatic updating via the profile manager. In many approaches these data do not play a role in the process of finding information, but serve as a bargaining chip against personalization services.

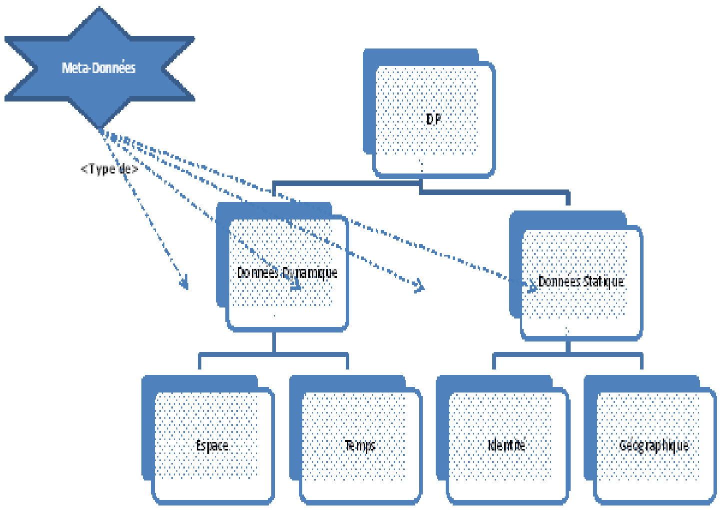


Fig8: Private data

6.3. *Data Quality*

One of the key factors is the quality of the personalization of the information provided. This dimension provides the possibility to express preferences as extrinsic origin of the information, accuracy, response time allowed, the accuracy of the data or the order of preference. In this dimension we give the possibility to users to describe their preferences on the quality of the query execution and quality data delivered. In the quality of the profile process the user expresses its requirements for the operation of the customization process. Timeliness of data includes the time since the creation of the information or the time of last update. For their part the precision and accuracy require the advice of an external expert and contain information on the quality of the data. There may be a user having specific preferences on the presence or absence of the information. To the side of subjective preferences into account, we introduce a hierarchical order of criteria.

6.4. *Historical inquiry and interaction*

This dimension is integrated in perspective in a knowledge base represented by an ontology exploited in a process of matching queries and querying the history of DES system. We propose a method for this purpose cyclic proceeds in two steps. The first is to extract from the history of interactions a focus candidate qualified 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 hypothesis respecting the diversity of possible interests. This reflects the learning phase of interests which induces changes in the profile.

6.5. *Data Security*

Since the user's profile contains data strictly personal, a major problem that arises is that of information security. The user can specify his requirements of the security level, but the aspect of customization must be supported by the profile manager that must comply with corporate and legislative safety requirements. In our approach we identify three aspects of security: data that the application and the identity of the client.

In the category of data security profile, the user will have the possibility to express requirements of the security level of data query results and those contained in the profile during an exchange with the outside. This level of security depends on the techniques used. In our context of DES system the user can request that the exchanges are on a secure connection and the data is encrypted. It may be that sometimes the predicates of a query are meaningful private and strictly personal and in this case we need to protect the execution of the application and hide the fact that an application or part of it has been launched. An example of a query that requires scripts that remains hidden is the case where an agent wants to get medical rescue sensitive data on a particular victim. In this case it will query all sources included in geographically remote system in real time and get the results that must remain hidden with a restriction that the right of access. In the security of the identity of the user, explains the conditions for access to the profile. It is in some cases to hide the identity of the client, to explain which applications have access to what data and for what purposes these data will be used. It defines how these rights will be awarded. Indeed the DES system requests information from the user profile. In this case it is assumed that the user has previously defined a role in providing access to their profile information Part keywords centers of interest. Therefore, the user does not need to be explicitly notified of the request for information and the exchange of information can be done automatically. To simplify the organization of access to data sources integrated by the DES system, we propose a clustering technique based on the principle of clustering and classification of users into categories defined by the domain ontology. Indeed, one definition of Groups of Users (GU) for Under Common Domain (SDC) with the possible merge groups if a sub domain to a single user group. Depending on the nature of GU, it has a set of rules and restrictions related to access profiles, sources and results of query execution (semantic cache). These restrictions are shared between the three categories, those data sources and queries.

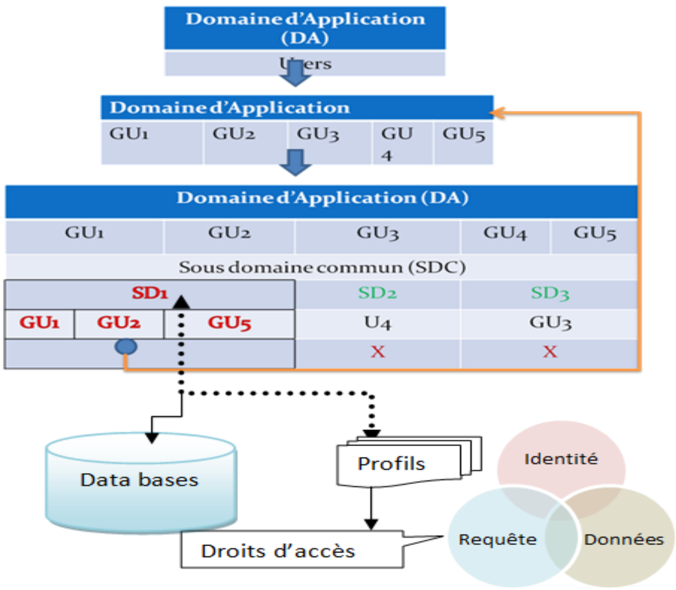


Fig9: restriction sbetweencategories

Research in the field of graph theory led to the creation of several algorithms to compute the 'CPC'. The majority of these algorithms carry almost the same principle of calculation, except that the major point of difference lies in the ability of individuals to treat a dense graph. More concretely, some algorithms calculate the 'CPC' between two vertices only known both in advance (one-to-one or one-to-all), which is reflected in their calculation speed, others allow to know the 'CPC' between input vertex and all other vertices remaining (one-to-all-in-one or all), there are even those who are widespread and can thus calculate the 'CPC' between any input vertex and any vertex of destination (all-to-all or in-all). While the latter type of algorithm has better processing capacity much denser graphs, but this goes at the expense of speed.

An algorithm to compute the shortest path is distinguished by the speed, but it should be noted that this speed depends primarily on the ability of an algorithm to deal with a large graph. This gives rise to a compromise-speed processing capability. This point has been a main axis in our research because it operates the system as the change of the solutions proposed by the acquired data for very high speed at informing users or in response or rescue case of danger. Among the algorithms, we chose the algorithm and Floyd algorithm we:

- G = (N, A) Graph evaluated positively with N and A are respectively the set of vertices and the set of edges.
- dist (i, j): Weight of edge (i, j).
- i0: Summit input.
- d (i): the shortest distance found at the top of i0 i

```
//initialisation
d(i0) = 0
Pour tout i∈N ;i ≠ i0 faire
d(i) = c(i0,i)
si (d(i) <∞) alors p(i) = 1
sinon p(i) = NULL
fin_si
fin_pour
//Itérations
S = {i0}
Tant que (S ≠ N) faire
Choisir i ∈ (N-{S}) tel que d(i)= minj ∈ (N-{S}) (dj)
S = S U {i}
Pour tout j ∈ Γ+(i) ∩ (N-{S}) faire
Si d(j) > d(i) + dist(i,j) alors
d(j) = d(i) + dist(i,j)
p(j) = i
fin_si
fin_pour
fin_tant que
```

Fig10: Algorithm of shortest distance

7. Geo-location and sensor deployment

Geolocation is a method for positioning an object (a person, a place, a car, etc.) on a plan or map using geographic coordinates. It is done by using a terminal capable of being located through a satellite positioning system and a GPS receiver for example, or by other techniques, and publishes real-time geographic coordinates. Recorded positions can be stored in the terminal and be retrieved later, or be transmitted in real time to a software platform geolocation.

At the base of this operation, there are several geolocation techniques such as geolocation geocoder, satellite, WiFi, but according to studies by the team, the decision was the use of GPS.

7.1. *Sensors mapping*

Mapping refers to the construction and study of maps and geological. The main principle of the mapping is the representation of data on a reduced space representing a generally real. The objective of the creation of this map is to make an effective and simplified representation of real data. For our system, we will take this principle to identify or locate an object that can be user, a place, an animal ... on the surface of the planet which is why we use a different system called benchmark map.

7.2. *Sensors connection and communication*

We present in this section the use of the sensor in the DES system after selecting a geographic area as an example to study.

To explore this area, we must seek the real dimension to take into account the distances to see the feasibility of components to use. It can be identified by surrounding area and calculating the distances from the actual geographic coordinates. Deviser and conquer is a practical principle which allows to decompose and identify the major studies of small studies manageable. The same principle was used to study well and place the sensors in this area.

Fig11: Calculation of distances in the study area

Before presenting the zones made, we present the architecture of sensor chosen to monitor the area.

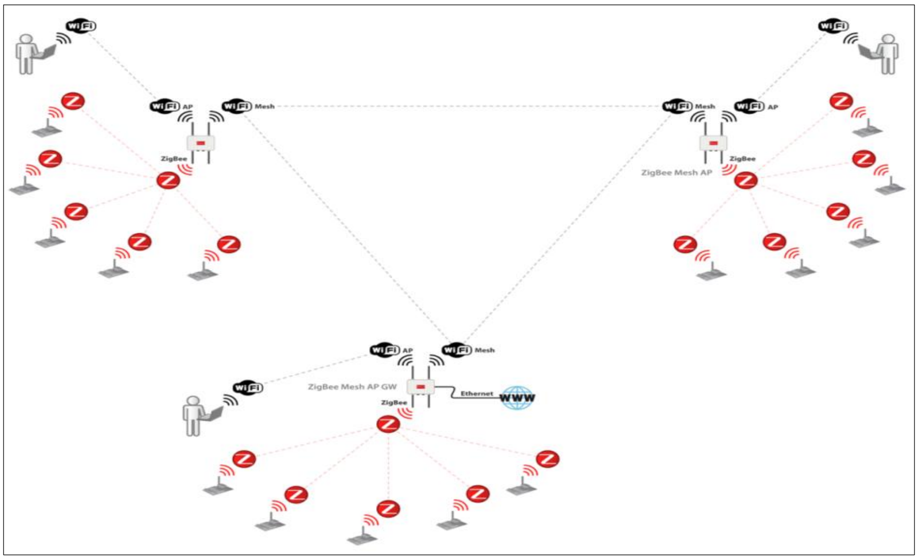


Fig12: Sensor network architecture proposed for the study area.

We will converse the area into three parts and place the selected sensors Meshlium waspmote and respecting the architecture presented before.

Fig13: Sensor positioning on the field

Note: there isn't a standard rule for positioning the sensors but we just take into account the precision and cost study. If we look more precisely, should be placed over sensor that is why the cost is higher.

8. Implementation Technology

8.1. *Android*

The end result of our system is a mobile application and that the application fits with the latest trends and new news we decided to develop our system by android; Android SDK is provided by Google, which is based on the Java platform. This is complete: It includes several libraries, debugger, profiler, emulator and sample code to start and more information.

8.2. *Web service*

In computing, the term data processing refers to a series of processes that extract information or to produce knowledge from raw data. These processes, once programmed, are often automated using computers. If the final results are intended for humans, their presentation is often essential to appreciate the value. This assessment, however, is variable among individuals.

This information processing can then take the data fusion of information retrieval or processing of representation. For example, the merger may be to combine multiple data sources to compile the information into a safer and extraction can be processed to synthesize data.

The data is very important in any project also in our system because they are essential for the functioning of our system using the web service for exchanging data between the server and our system. The web service is a computer program that allows communication and data exchange between heterogeneous applications and systems in distributed environments. It is therefore a set of functionality exposed on the Internet or on an intranet, by and for applications or machines without human intervention, and synchronously.

Conclusion

We present in this article the DES system that manages all the type of disaster by presenting the following modules: the objective of the system, the services offered by DES before or during or after a disaster, block diagram and architecture layer the system and the solutions adopted and the technology used as the sensor network, android and web services.

References

[1] AvgoustinosFilippoupolitis, X. Q. (2007, March 14). An adaptive system for movement decision support in building evacuation. , 80(17), 3673–3677.

[2] GokceGorbil, AvgoustinosFilippoupolitis, &ErolGelenbe (2007, November15). Intelligent Navigation Systems for Building Evacuation. , 8(10),3487–3492.

[3] AvgoustinosFilippoupolitis, X. Q. (2007, March 14). An adaptive systemfor movement decision support in building evacuation. , 80(17), 3673–3677.

[4] GokceGorbil, AvgoustinosFilippoupolitis, &ErolGelenbe (2007, November15). Intelligent Navigation Systems for Building Evacuation. , 8(10),3487–3492.

[5] AvgoustinosFilippoupolitis, X. Q. (2007, March 14). An adaptive systemfor movement decision support in building evacuation. , 80(17), 3673–3677.

[6] GokceGorbil, AvgoustinosFilippoupolitis, &ErolGelenbe (2007, November15). Intelligent Navigation Systems for Building Evacuation. , 8(10), 3487–3492.

[1] Bang-Jensen, J.; Gutin, G. (2000). Digraphs: Theory, Algorithms and Applications. Springer.

[7] D. Maquin, ‘Elements of graph theory and linear programmation’, Éléments de théorie des graphes et programmation linéaire, Institut Nationale Polytechnique de Lorraine, École Nationale Supérieure d’Électricité et de Mécanique, 2008.

[8] P.E.Hart, A Formal Basis for the Heuristic Determination of Minimum Cost Paths, IEEE Transactions on Systems Science and Cybernetics SSC4, vol. 4, no 2, p. 100–107, 1968.

[9] Balakrishnan, V. K. (1997-02-01). Graph Theory (1st ed.). McGraw-Hill. ISBN 0-07-005489-4.

[10] Berge, Claude (1958) (in French).’Graph theory and its applications’, Théorie des graphes et ses applications. Dunod, Paris: Collection Universitaire de Mathématiques, II. pp. viii+277. Translation: . Dover, New York: Wiley. 2001 [1962].

[11] Biggs, Norman (1993). Algebraic Graph Theory (2nd ed.). Cambridge University Press.ISBN 0-521-45897-8.

[12] Bollobás, Béla (2002-08-12). Modern Graph Theory (1st ed.). Springer.ISBN 0-387-98488-7.

[13] Diestel, Reinhard (2005). Graph Theory (3rd ed.). Berlin, New York: Springer-Verlag. ISBN 978-3-540-26183-4.

[14] Peters, L. D., Saidin, H., 2000, IT and the Mass Customization of Services: the Challenge of Implementation, International Journal of Information Management, 20(2), 103-119

Fuzzy Tracking Control Based on H_∞ Performance and Virtual Desired Model for Nonlinear Systems

Djamel Ounnas, Salah Chenikher.LABGET,
University of Tebessa,
Department of Electrical Engineering,
Tebessa, Algeria

No Institute Given

Abstract. This paper deals with the synthesis of fuzzy tracking controller via virtual reference model for a nonlinear systems with disturbance. First, a Takagi-Sugeno fuzzy model represents the nonlinear system with distrubance. Second, we combine an integral control schemes, and feedback fuzzy control whith H_∞ performance, to develop a robust feedback tracking controller. Third, the concept of virtual reference is used to to simplify the design of model reference. In light of this concept, the design procedure is split into two steps: i) Determine the virtual desired variables from the generalized kinematics; and ii) Determine the control law. The fuzzy controller is designed to track a reference signal and guaranteed a minimum disturbance attenuation level for the closed-loop system. The gains of fuzzy controller are obtained by solving a set Linear matrix inequalities (LMIs) techniques. Finally, we present an applocation on Permanent Magnet synchronous Machine. Satisfactory results validate the claims.

nonlinear systems, fuzzy tracking control, linear matrix inequality problem (LMIP), model reference, Takagi and Sugeno (TS) fuzzy model.

1 Introduction

The Takagi-Sugeno (T-S) fuzzy model has become popular because of its efficiency in controlling nonlinear systems. The T-S fuzzy model has been proved to be a good representation for a class of nonlinear systems. The main property of T-S fuzzy model is to describe the local dynamics by linear models. The overall model of the nonlinear system is obtained by fuzzy blending of these linear models through nonlinear fuzzy membership functions, The basic idea for the approach is to decompose the model of a nonlinear system into a set of linear subsystems with associated nonlinear weighting functions. Then by using well-formulated linear control theory, the control objective is achieved [1].

In recent years, control using Takagi-Sugeno (TS) fuzzy model has attracted wide attention. This is due to the systematic approach to analysis and design of controllers where multiple objectives can be considered in a uni?ed manner. The controller design using parallel distributed compensation (PDC) and stability analysis are carried out using Lyapunov direct method where the control problem is then formulated into solving linear matrix inequalities (LMIs). Then using powerful numerical LMIs toolboxes [2] - [9].

The tracking control design based on the Takagi-Sugeno fuzzy model has been treated by several researchers, for instance, [10] - [14]. The most important issue for fuzzy tracking control systems is that the output of the nonlinear system tracks a reference signal. In [10], the tracking problem of nonlinear systems is solved using a synthesis of the fuzzy control theory and the linear multivariable control theory. Simulation results show that the proposed tracking control system can make the output of the system to asymptotically track the reference signal. The robust fuzzy tracking controller based on internal model principle is introduced to track a reference signal in [11].

The problem of tracking control, using the TS fuzzy model based approach, has been limited to nonlinear model following, where all states are assumed known [15], or output feedback tracking for a linear reference model, where the reference input is considered as disturbance and is attenuated using a robust criterion [16], [17]. For these situations, it often results in bounded tracking errors between the system states and reference states. We also note that the model following approach is not easily used to deal with some control tasks where only some outputs need to be controlled and other internal state behaviors are hard to be described. Hence, it is hard to de?ne the reference model in this case [14]. To make a determination of reference model easily, the new concepts of virtual desired variables is introduced by [14]; However, the study [14] has not considered the disturbance.

In light of the above, we are motivated to propose a fuzzy tracking control for nonlinear systems with disturbances. First, a TS fuzzy model represents nonlinear system. Next, we combine an integral control schemes [18], [19], and feedback-based fuzzy control to develop a robust feedback tracking controller. Third, we formulate the control objectives into suf?cient conditions by stability analysis using Lyapunovs method. By solving a set of LMIs, we obtain controller gains with guarantee H_∞ control performance [20], [21]. Forth, based on virtual reference model [14], we design a set of virtual desired variables. Finally, we illustrate the effectiveness of the controller design on an numerical example. The satisfactory results is consistent with the claims.

This paper is organized as follows. In Section 2, we formulate the control problem. In Section 3, the tracking control with guarantee H_∞ is designed. Section 4, we explain the virtual desired variables and generalized kinematics design. In Section 5, we persent an numerical example of a Permanent Magnet Synchronous Machine. The last section gives a conclusion on the main works developed in this paper.

2 Problem Formulation

Consider a general nonlinear dynamic equation :

$$\dot{x}(t) = f(x(t)) + g(x(t))u(t) + \upsilon(t). \tag{1}$$

where $x(t)$ is the state vector, $u(t)$ is the control input vector,$\upsilon(t)$ is the varying disturbances vector, $f(x)$ and $g(x)$ are nonlinear functions with appropriate dimensions. The nonlinear system (1) can be expressed by the TS fuzzy system [1]: If z_1 is F_{1i} and ... and If z_g is F_{gi} Then

$$\dot{x}(t) = A_i x(t) + B_i u(t) + D_i w(t), \qquad i = 1...r.$$

where $z_1...z_g$ are the premise variables which would consist of the states of the system, $F_{ji}(j = 1, 2, ...g)$ are the fuzzy sets, r is the number of fuzzy rules, A_i, B_i and D_i are system matrices with appropriate dimensions.The global fuzzy model is inferred as follows:

$$\dot{x}(t) = \sum_{i=1}^{r} h_i(z(t))(A_i x(t) + B_i u(t) + D_i w(t)). \tag{2}$$

where

$$\lambda_i(z(t)) = \prod_{j=1}^{r} F_{ij}((z(t))).$$

$$h_i(z(t)) = \frac{\lambda_i(z(t))}{\sum_{i=1}^{r} \lambda_i((z(t))}.$$

for all $t > 0$, $h_i(z(t)) \geq 0$ and $\sum_{i=1}^{r} h_i(z(t)) = 1$.
For tracking feedback control, the control objective is required to satisfy:

$$x(t) - x_d(t) \to 0 \quad as \quad t \to \infty$$

where $x_d(t)$ denotes the desired variables.
Let $\tilde{x}(t) = x(t) - x_d(t)$ denote the tracking error. The time derivative of $\tilde{x}(t)$ yields:

$$\dot{\tilde{x}}(t) = \dot{x}(t) - \dot{x}_d(t).$$

$$\dot{\tilde{x}}(t) = \sum_{i=1}^{r} h_i(z(t))(A_i[x(t) - x_d(t)]) + \sum_{i=1}^{r} h_i(z(t))D_i w(t)$$
$$+ \sum_{i=1}^{r} h_i(z(t))B_i u(t) + \sum_{i=1}^{r} h_i(z(t))A_i x_d(t) - \dot{x}_d(t). \tag{3}$$

In (3), we assume the latter as follows:

$$\sum_{i=1}^{r} h_i(z(t))B_i\tau(t) = \sum_{i=1}^{r} h_i(z(t))B_i u(t)$$

$$+\sum_{i=1}^{r} h_i(z(t))A_i x_d(t) - \dot{x}_d(t). \tag{4}$$

Using (4), then the tracking error system (3) is rewritten below:

$$\dot{\tilde{x}}(t) = \sum_{i=1}^{r} h_i(z(t))(A_i\tilde{x} + B_i\tau(t) + D_i w(t)). \tag{5}$$

A new fuzzy controller $\tau(t)$ is designed to deal with the tracking control systems as:
If z_1 is F_{1i} and ... and If z_g is F_{gi} Then

$$\tau(t) = -K_i\tilde{x}(t), \qquad i = 1...r.$$

where $z_1...z_g$ are the premise variables which would consist of the states of the system, $F_{ji}(j = 1, 2, ...g)$ are the fuzzy sets, r is the number of fuzzy rules, A_i, B_i and D_i are system matrices with appropriate dimensions. The inferred output of the PDC controller is determined by the summation:

$$\tau(t) = -\sum_{i=1}^{r} h_i(z(t))K_i\tilde{x}(t), \qquad i = 1...r. \tag{6}$$

In order to reject slow varying disturbances (according to the system), an integral action ([18], [19]) is added (see Fig. 1).

$$\tau(t) = -\sum_{i=1}^{r} h_i(z(t))K_i\tilde{x} - \sum_{i=1}^{r} h_i(z(t))F_i\tilde{x}_I. \tag{7}$$

Where

$$\tilde{x}(t)_I = \int_0^{t_f} \tilde{x}(t)dt.$$

from (7), we can write:

$$\tau(t) = -\sum_{i=1}^{r} h_i(z(t)) \begin{bmatrix} K_i & F_i \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{x}_I \end{bmatrix}.$$

The new fuzzy controller can be written as:

$$\tau(t) = -\sum_{i=1}^{r} h_i(z(t))\bar{K}_i\bar{X}(t). \tag{8}$$

The TS model with an integral action can be written in the augmented form:

$$\dot{\bar{X}} = \sum_{i=1}^{r} h_i(z(t))(\bar{A}_i\bar{X}(t) + \bar{B}_i\tau(t) + \bar{D}_i w(t)). \tag{9}$$

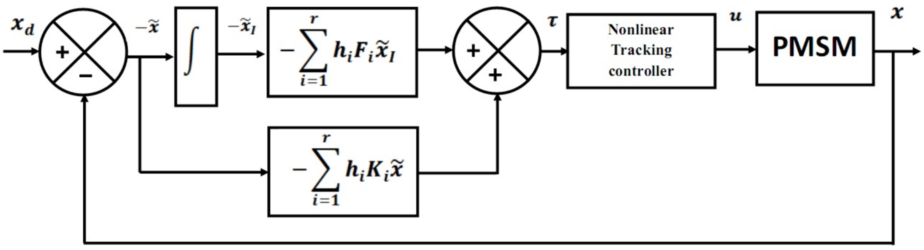


Fig. 1. Control scheme using an integral part

$$\bar{A}_i = \begin{bmatrix} A_i & 0 \\ I & 0 \end{bmatrix}, \bar{B}_i = \begin{bmatrix} B_i \\ 0 \end{bmatrix}, \bar{D}_i = \begin{bmatrix} D_i \\ 0 \end{bmatrix}.$$

In the case $B_1 = B_2 = ...B_r = B$, the TS model of the system with an integral action can be written in the augmented form:

$$\dot{\bar{X}}(t) = \sum_{i=1}^{r} h_i(z(t))[G_i\bar{X}(t) + \bar{D}_i w(t)]. \tag{10}$$

where

$$G_i = \bar{A}_i - \bar{B}\bar{K}_i, \quad \bar{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}.$$

3 H_∞ Tracking Control Design

The design purpose of this study is how to specify a fuzzy controller in (8) for the augmented system (10) with the guaranteed H_∞ tracking performance for $w(t)$, and the output of system can follow the desired variables $x_d(t)$. The presence of disturbances $w(t)$ will deteriorate the tracking control performance of the fuzzy control system. In order to minimize the effect of $w(t)$ on the control system, H_∞ tracking performance related to tracking error has been considered [20], [21].

$$\int_0^\infty \bar{X}(t)^T\bar{X}(t)dt \leq \gamma^2 \int_0^\infty w^T(t)w(t)dt \tag{11}$$

Where γ is a prescribed value which denotes the worst case effect of $w(t)$ on $\bar{X}$ and P is some symmetric positive de?nite weighting matrix. The following lemma gives the results of the H_∞ norm bounded.

Lemma 1. *If there exists a symmetric and positive de?nite $X^T = X$ such that the following matrix inequality are satis?ed*

$$\begin{bmatrix} \bar{A}_iX + X\bar{A}_i^T - \bar{B}M_i - M_i^T\bar{B}^T & \bar{D}_i & X \\ \bar{D}_i^T & -\gamma^2 I & 0 \\ X & 0 & -I \end{bmatrix} < 0. \tag{12}$$

Then for a prescribed γ, H_∞ tracking control performance in (6) is guaranteed via the TS fuzzy model-based state feedback controller . The control gain are given by:

$$\bar{K}_i = M_i X^{-1} \tag{13}$$

Proof. See Appendix A In what follows, the virtual desired variables x_d and cobtrol law $u(t)$ are to be discussed.

4 Virtual desired and control law design

The remaining design for the tracking control is to determine $x_d(t)$ and then obtain the practical controller input $u(t)$. To this end, we use the facts:

$$g(x) = \sum_{i=1}^{r} h_i(z(t))B_i, \quad A(x) = \sum_{i=1}^{r} h_i(z(t))A_i.$$

and rewrite (4) as the following compact form:

$$g(x)(u(t) - \tau(t)) = -A(x)x_d(t) + \dot{x}_d(t)). \tag{14}$$

where $A(x) = \sum_{i=1}^{r} h_i(z(t))A_i$. From (14), the existence of the control input $u(t)$ depends on the form of $g(x)$. Here, the input matrix $g(x)$ is assumed with full-column rank. If necessary, we can rearrange the coordinate frame in (1) such that:

$$\begin{cases} g(x) = \begin{bmatrix} 0_{n-m} \\ --- \\ B(x) \end{bmatrix}, \\ A(x) = \begin{bmatrix} A_{n-m} \\ --- \\ A_m \end{bmatrix}, \\ x_d(x) = \begin{bmatrix} x_{d_{n-m}} \\ --- \\ x_{d_m} \end{bmatrix}. \end{cases} \tag{15}$$

Where $0_{n-m} \in R^{(n-m)\times m}$ is a zero matrix and $B(x) \in R^{(m\times m)}$. Consequently, the condition (13) is with the following form:

$$A(x) = \sum_{i=1}^{r} h_i(z(t))A_i.$$

From (14), the existence of the control input $u(t)$ depends on the form of $g(x)$. Here, the input matrix $g(x)$ is assumed with full-column rank. If necessary, we can rearrange the coordinate frame in (1) such that:

$$\begin{bmatrix} 0_{n-m} \\ --- \\ B(x)(u-\tau) \end{bmatrix} = \begin{bmatrix} \dot{x}_{d_{n-m}} - A(x)_{n-m}x_d(t) \\ \dot{x}_{d_m} - A(x)_m x_d(t) \end{bmatrix}. \tag{16}$$

From (16), we can write:

$$u = B^{-1}(x)[\dot{x}_{d(t)_m} A_m(x)x_d] + \tau. \tag{17}$$

Where

$$\tau = -\sum_{i=1}^{r} h_i(z(t))F_i\tilde{x} - \sum_{i=1}^{r} h_i(z(t))F_i\tilde{x}_I\hat{.}$$

Replacing τ by its value in (17):

$$u = -\sum_{i=1}^{r} h_i(z(t))K_i\tilde{x} - \sum_{i=1}^{r} h_i(z(t))F_i\tilde{x}_I$$
$$+B^{-1}(x)[\dot{x}_{d(t)_m} A_m(x)x_d]. \tag{18}$$

Also from (18), the practical control input:

$$u = -\sum_{i=1}^{r} h_i(z(t))\bar{K}_i\bar{X} + B^{-1}(x)[\dot{x}_{d(t)_m} A_m(x)x_d]. \tag{19}$$

5 Application: Speed Control of a Permanent Magnet Synchronous Machine (PMSM)

5.1 T-S fuzzy model of the PMSM

By considering the classical simplifying assumptions, the dynamic model of the Permanent Magnet Synchronous Machine, in the synchronously $d-q$ reference frame, can be described as [22]:

$$\begin{cases} \dot{\omega}(t) = \frac{3p}{2J}\varphi_v i_q - \frac{B}{J}\omega - \frac{1}{J}C_r. \\ \dot{i_q}(t) = -\frac{R}{L}i_q + p\omega i_d + \frac{1}{L}u_q. \\ \dot{i_d}(t) = -\frac{R}{L}i_d + p\omega i_q + \frac{1}{L}u_d. \end{cases} \tag{20}$$

where $\omega(t)$ is the rotor speed, (i_q, i_q) are the $d-q$ axis stator currents, (u_q, u_d) are the $d-q$ axis stator voltages, C_r is the load torque (C_r is a known step disturbance). The motor parameters are: J the moment of inertia of the rotor, R the sator winding resistance,($L_d = L_q = L$) the $d-q$ axis inductances, B the friction coefficient relating to the rotor speed, φ_v the flux linkage of the permanent magnets and p the number of poles pairs.

The nonlinear model of the PMSM can be written as the state space form:

$$\dot{x}(t) = A(\omega)x(t) + Bu(t) + DC_r(t). \tag{21}$$

With

$$x = [\omega \;\; i_q \;\; i_d]^T, \;\; u = [u_q \;\; u_d]^T.$$

$$A(\omega) = \begin{bmatrix} -\frac{B}{J} & \frac{3p\varphi_v}{2J} & 0 \\ -\frac{p\varphi_v}{L} & -\frac{R}{L} & -p\omega \\ 0 & p\omega & -\frac{R}{L} \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ \frac{1}{L} & 0 \\ 0 & \frac{1}{L} \end{bmatrix},$$

$$D = \begin{bmatrix} -\frac{1}{J} \\ 0 \\ 0 \end{bmatrix}.$$

To transform the nonlinear model of the machine into a fuzzy linear Takagi-Sugeno model, the adopted method is to use a transformation on functions of one variable, as follows [17]:

Lemme1.

for $x \in [-b, a]$, a, $b \in R^+$, conseder $f(x)$:$R \longrightarrow R$ a bounded function, then there are always two functions $w_1(x)$ and $w_2(x)$ and two scalars α and β verifying the following properties:

$$\begin{cases} f(x) = \alpha w_1(x) + \beta w_2(x), \\ w_1(x) + w_2(x) = 1, \quad w_1(x) \geq 0, \quad w_2(x) \geq 0. \end{cases} \tag{22}$$

Proof. Considering that $f(x)$ bounded as:

$$f_{min} \leq f(x) \leq f_{max}.$$

With

$$\alpha = f_{max}, \beta = f_{min}, w_1 = \frac{f(x) - f_{min}}{f_{max} - f_{min}}, w_2 = \frac{f_{max} - f(x)}{f_{max} - f_{min}}.$$

By applying this decomposition to the PMSM, we obtained local models defined by equation (2). The model state (21) of the machine has three variables, their products with the state variable vector generates the nonlinearity of the model machine. In our study, we have a linearity, resulting from the product of ω with d-q axis currents. Thus the vector of premise variables may be defined by:

$$z(t) = z_1(t) = \omega(t).$$

with

$$\omega_{min} \leq \omega(t) \leq \omega_{max}.$$

From lemma.2, we can write:

$$z(t) = F_{11} z_{max} + F_{12} z_{min}.$$

$$F_{11} = \frac{(z(t) - z_{min})}{(z_{max} - \omega_{min})}, F_{12} = \frac{(z_{max} - z(t))}{(z_{max} - z_{min})}.$$

We can introducing plant rule of fuzzy model:

Plant Rule 1: IF $z_1(t)$ is F_{11}THEN

$$x?(t) = A_1 x(t) + B_1 u(t) + D_1 C_r(t).$$

Plant Rule 2: IF $z_2(t)$ is F_{12} THEN

$$x?(t) = A_2 x(t) + B_2 u(t) + D_2 C_r(t).$$

Where F_{11} and F_{12} are fuzzy sets. The corresponding membership functions are:

$$F_{11} = h_1, F_{12} = h_2.$$

The subsystem matrices are:

$$A_1 = \begin{bmatrix} -\frac{B}{J} & \frac{3p\varphi_v}{2J} & 0 \\ -\frac{p\varphi_v}{L} & -\frac{R}{L} & -p\omega_{max} \\ 0 & p\omega_{max} & -\frac{R}{L} \end{bmatrix},$$

$$A_2 = \begin{bmatrix} -\frac{B}{J} & \frac{3p\varphi_v}{2J} & 0 \\ -\frac{p\varphi_v}{L} & -\frac{R}{L} & -p\omega_{min} \\ 0 & p\omega_{min} & -\frac{R}{L} \end{bmatrix}.$$

$$B_1 = B_2 = B = \begin{bmatrix} 0 & 0 \\ \frac{1}{L} & 0 \\ 0 & \frac{1}{L} \end{bmatrix}.$$

$$D_1 = D_2 = \begin{bmatrix} -\frac{1}{J} \\ 0 \\ 0 \end{bmatrix}.$$

The virtual desired variables x_d are needed to satisfy (14), which is rewritten below:

$$g(x)(u(t) - \tau(t)) = -A(x)x_d(t) + \dot{x}_d(t).$$

Then we obtain the following matrix form

$$\begin{bmatrix} 0 & 0 \\ \frac{1}{L} & 0 \\ 0 & \frac{1}{L} \end{bmatrix} (u(t) - \tau(t)) = - \begin{bmatrix} -\frac{B}{J} & \frac{3p\varphi_v}{2J} & 0 \\ -\frac{p\varphi_v}{L} & -\frac{R}{L} & -p\omega \\ 0 & p\omega & -\frac{R}{L} \end{bmatrix} \times$$

$$\begin{bmatrix} \omega_d \\ i_{qd} \\ i_{dd} \end{bmatrix} + \begin{bmatrix} \dot{\omega}_d \\ \dot{i}_{qd} \\ \dot{i}_{dd} \end{bmatrix}. \tag{23}$$

The second index word d denotes the desired states. According to the first equation of (23), it follows that:

$$\dot{\omega}_d = -\frac{B}{J}\omega_d + \frac{3p\varphi_v}{2J} i_{qd}.$$

which induces that:

$$i_{qd} = (\dot{\omega}_d + \frac{B}{J}\omega_d)\frac{2J}{3p\varphi_v}.$$

$$\dot{i}_{qd} = (\ddot{\omega}_d + \frac{B}{J}\dot{\omega}_d)\frac{2J}{3p\varphi_v}.$$

Then, from the second and the third equations of (23), we can obtain the control input:

$$\begin{cases} u_q = p\varphi_v\omega_d + Ri_{qd} + L\dot{i}_{qd} + Lp\omega i_{dd} + \tau_q \\ u_d = -pL\omega i_{qd} + Ri_{dd} + L\dot{i}_{dd} + \tau_d \end{cases}$$

Note that for synchronous machines there is no need for a flow model. Consequently, the rotor position is the angle of reference. Moreover, as we have a smooth poles machine, the best choice for its operation is obtained for a value where the internal angle of the machine is equal to $\frac{\pi}{2}$ that is to say $i_{dd} = 0$. We obtained the following form:

$$\begin{cases} u_q = p\varphi_v\omega_d + R_s i_{qd} + L_s\dot{i}_{qd} + \tau_q \\ u_d = -pL_s\omega i_{qd} + \tau_d \end{cases}$$

where τ_q,τ_d are new controller to be designed via LMIs approach. The inferred output of the PDC controller is determined by the summation:

$$\tau = -\sum_{i=1}^{2} h_i(z(t))K_i\tilde{x} - \sum_{i=1}^{2} h_i(z(t))F_i\tilde{x}_I.$$

Where

$$\tau = [\tau_q \quad \tau_d]^T, \quad x = [\omega \quad i_q \quad i_d]^T, \quad x_d = [\omega_d \quad i_{qd} \quad i_{dd}]^T.$$

$$\tilde{x}(t) = x(t) - x_d(t), \quad \tilde{x}_I(t) = \int_0^{t_f} \tilde{x}(t)dt.$$

5.2 Simulation Results

The proposed feedback control law has been first tested in simulation, The closed-loop simulation results are shown in Figs.2-7. Using the LMI approach, the application of lemma.1 to calculate the control gains of the fuzzy control law (19) gives the followings result:

$$K_1 = \begin{bmatrix} 30.4 & 19.988 & 0.19162 \\ -0.42351 & -0.084928 & 9.0015 \end{bmatrix},$$

$$K_2 = \begin{bmatrix} 30.388 & 19.987 & -0.0026286 \\ 0.96489 & 0.27894 & 9.005 \end{bmatrix},$$

$$F_1 = \begin{bmatrix} 69.025 & 2.9217 & 2.546 \\ -0.95528 & -0.053095 & 30.088 \end{bmatrix},$$

$$F_2 = \begin{bmatrix} 68.998 & 2.9733 & -3.022 \\ 2.1594 & -0.4187 & 30.04 \end{bmatrix}.$$

Figures (1) to (7) illustrates the dynamic behaviour of the Permanent Magnet Synchronous Machine for desirted $\omega_d = sin(2t)$ and initial state $x(0) = [0 \ \ 0 \ \ 0]^T$. At $t = 1s$ a load torque $C_r = 5Nm$ is applied. By analysis the simulation results, a less tracking errors is observed for the speed and the currents in spite of the load torque is applied, that show the good performance of the fuzzy control law proposed in terms of track and disturbance rejection. The results show that the closed-loop system with the synthesized fuzzy controller has a good deportment: indeed, the measured speed and dq-axis currents track well the trajectory of reference one with good reliability over the whole speed range. Furthermore, the dq-axis currents responses are satisfying. From the results we can conclude the performance of the proposed control scheme.

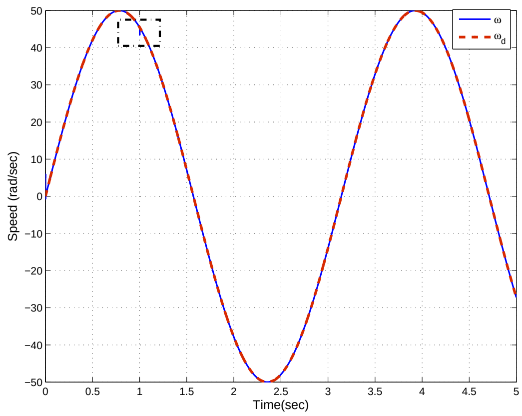


Fig. 2. Speed and its reference

6 Conclusion

A fuzzy tracking controller for nonlinear systems with disturbances has been proposed. To this end, an integral control and tracking fuzzy control with H_∞ performancd are combined to design a robust tracking controller. Suf?cient conditions are derived for stabilization by Lyapunovs method. the concept of virtual reference has been used to simplify the design of model reference. The controller has been tested in simulation with success on a Prmanent Magnet Synchronous Machine. Simulation results show that the desired performance for nonlinear systems can be achieved via the proposed fuzzy tracking control method.

7 appendices

Proof of lemma.1 Consider the Lyapunov function $V(\bar{X}(t)) = \bar{X}^T(t)P\bar{X}(t)$ where $P = P^T > 0$ the common positive matrix. In order to establish to the

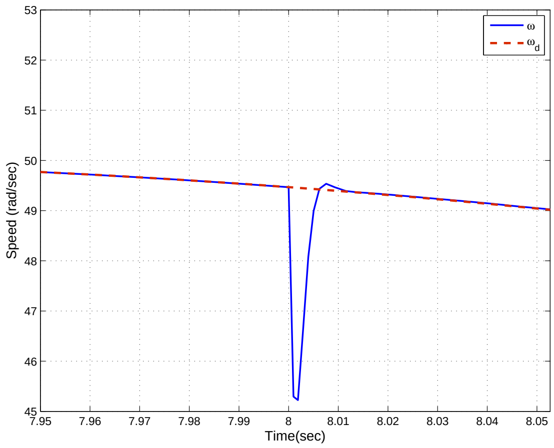


Fig. 3. Zoom window on speed and its reference

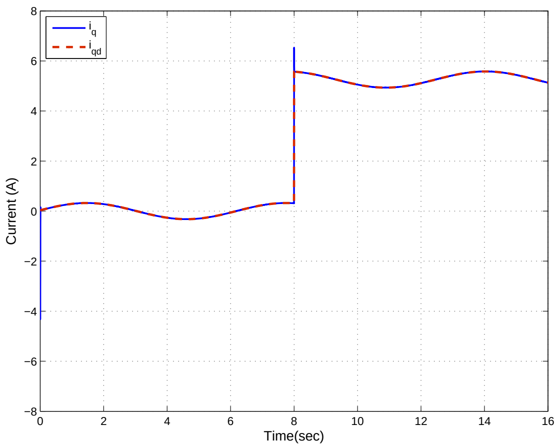


Fig. 4. q-axis current and its reference

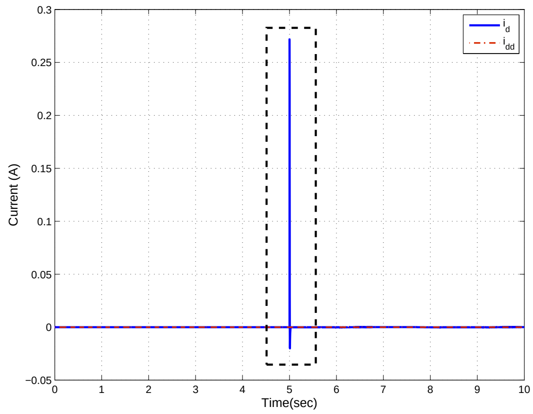


Fig. 5. d-axis current and its reference

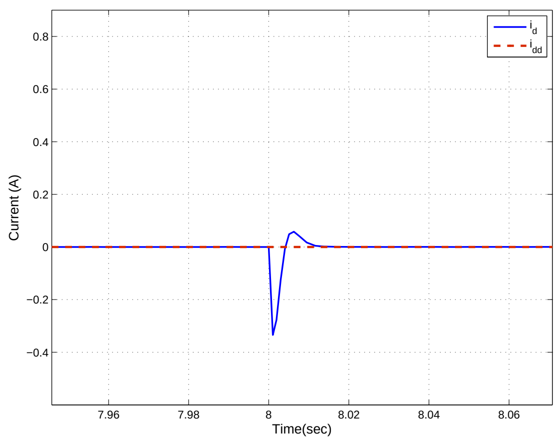


Fig. 6. Zoom window on d-axis current and its reference

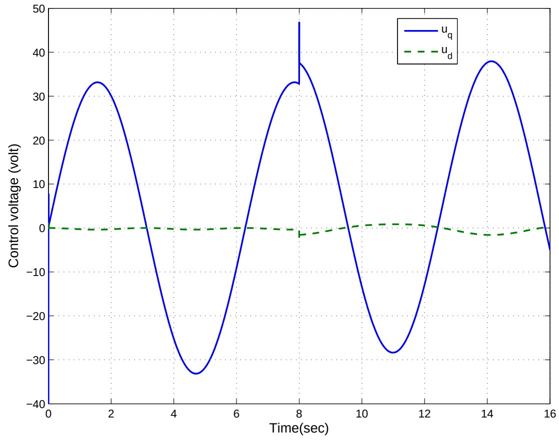


Fig. 7. Control voltage

asymptotic stability of the inequality (11), the time derivative of $V(\bar{X}(t))$ has to satisfy the following condition:

$$V(\bar{X}(t)) < 0 \tag{24}$$

In order to achieve the H_∞ tracking performance related to the tracking error, (24) becomes:

$$\dot{V}(\bar{X}(t)) + \bar{X}^T(t)\bar{X}(t) - \gamma^2 w^T(t)w(t) < 0 \tag{25}$$

Replacing $V(\bar{X}(t))$ by its value $\bar{X}^T P \bar{X}$ in (25), the last equation can be written as the following form LMI:

$$\sum_{i=1}^{r} h_i(z(t)) \left\{ \bar{X}^T (G_i P + P G_i^T) \bar{X} + w^T \left[\bar{D}_i^T \; P \right] \bar{X} \right\}$$

$$+ \sum_{i=1}^{r} h_i(z(t)) \bar{X}^T \left[P \; \bar{D}_i \right] w - \gamma^2 w^T w < 0.$$

$$\sum_{i=1}^{r} h_i(z(t)) \left[\bar{X}^T \; w^T \right] \begin{bmatrix} G_i P + P G_i^T & \bar{D}_i \\ \bar{D}_i^T & -\gamma^2 I \end{bmatrix} \times$$

$$\begin{bmatrix} \bar{X} \\ w \end{bmatrix} < 0.$$

$$\begin{bmatrix} \sum_{i=1}^{r} h_i(z(t))(G_i P + P G_i^T) + I & P \sum_{i=1}^{r} h_i(z(t)) \bar{D}_i \\ \sum_{i=1}^{r} h_i(z(t)) \bar{D}_i^T P & -\gamma^2 I \end{bmatrix}$$

$$< 0.$$

$$\begin{bmatrix} \sum_{i=1}^{r} h_i(z(t))(G_i P + P G_i^T) & P \sum_{i=1}^{r} h_i(z(t)) \bar{D}_i \\ \sum_{i=1}^{r} h_i(z(t)) \bar{D}_i^T P & -\gamma^2 I \end{bmatrix}$$

$$+\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} < 0.$$

$$\begin{bmatrix} \sum_{i=1}^{r} h_i(z(t))(G_i P + P G_i^T) & P \sum_{i=1}^{r} h_i(z(t)) \bar{D}_i \\ \sum_{i=1}^{r} h_i(z(t)) \bar{D}_i^T P & -\gamma^2 I \end{bmatrix}$$

$$+\begin{bmatrix} I \\ 0 \end{bmatrix} \begin{bmatrix} I & 0 \end{bmatrix} < 0. \tag{26}$$

Using the Schurs complement, (26) is equivalent to:

$$\begin{bmatrix} \sum_{i=1}^{r} h_i(z(t))(G_i P + P G_i^T) & P \sum_{i=1}^{r} h_i(z(t)) \bar{D}_i & I \\ \sum_{i=1}^{r} h_i(z(t)) \bar{D}_i^T P & -\gamma^2 I & 0 \\ I & 0 & -I \end{bmatrix}$$

$$< 0. \tag{27}$$

From (27), we can write:

$$\begin{bmatrix} (G_i P + P G_i^T) & P \bar{D}_i & I \\ \bar{D}_i^T P & -\gamma^2 I & 0 \\ I & 0 & -I \end{bmatrix} < 0. \tag{28}$$

After congruence with $diag(\begin{bmatrix} P^{-1} & I & I \end{bmatrix})$, inequality (28) becomes:

$$\begin{bmatrix} P^{-1} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} (G_i P + P G_i^T) & P \bar{D}_i & I \\ \bar{D}_i^T P & -\gamma^2 I & 0 \\ I & 0 & -I \end{bmatrix} \times$$

$$\begin{bmatrix} P^{-1} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} < 0. \tag{29}$$

Developed the last equation, (29) can be written as:

$$\begin{bmatrix} P^{-1}(G_i P + P G_i^T) P^{-1} & P^{-1} P \bar{D}_i & P^{-1} \\ \bar{D}_i^T P P^{-1} & -\gamma^2 I & 0 \\ P^{-1} & 0 & -I \end{bmatrix} < 0. \tag{30}$$

Considering $X = P^{-1}$ and $M_i = K_i X$, we obtain the same matrix as in (12):

$$\begin{bmatrix} \bar{A}_i X + X \bar{A}_i^T - \bar{B} M_i - M_i^T \bar{B}^T & \bar{D}_i & X \\ \bar{D}_i^T & -\gamma^2 I & 0 \\ X & 0 & -I \end{bmatrix} < 0.$$

The PMSM considered in the paper has the following data and parameters:
Rated power : 300 (W).
J : 6.36×10^{-4} $(Kg.m^2)$.
R_S : 4.55 (Ω).
R_r : 4.55 (Ω).

L_S : 11.6 (mH).
L_r : 11.6 (mH).
φ_v : 0.317 $(V.sec/rad)$.
B : 6.11×10^{-3} $(N.m.sec/rad)$.
p : 2.

The premise variables is bounded as:

$$\omega \in \begin{bmatrix} -50 & 50 \end{bmatrix} (rd/s).$$

The sets of the matrixes A_i de?ned in the 2^1 If-then fuzzy rules are given:

$$A_1 = \begin{bmatrix} -9.6069 & 1495.3 & 0 \\ -54.655 & -392.24 & -100 \\ 0 & 100 & -392.24 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} -9.6069 & 1495.3 & 0 \\ -54.655 & -392.24 & 100 \\ 0 & -100 & -392.24 \end{bmatrix}.$$

References

[1] T. Takagi and M. Sugeno, Fuzzy identi?cation of systems and its applications to modeling and control, IEEE Trans. Syst. Man, Cybern., vol. SMC-15, no. 1, pp. 116132, Jan. 1985.

[2] C. T. Sheng and L. Peter, Robust output tracking control for discrete-time nonlinear systems with time-varying delay: Virtual fuzzy model LMI-based approach, Expert systems with applications , vol. 39, no. 9 , pp. 8239-8247, Jui. 2012.

[3] A. M. Jesus, C. G. Julio and C. Bernardino, Exact Output Regulation for Nonlinear Systems Described by Takagi-Sugeno Fuzzy Models, vol. 20, no. 2, pp. 235-247, Apr. 2012.

[4] Z. Xihai and Z. Ming , Multi-Objective Control of Spacecraft Attitude Maneuver Based on Takagi-Sugeno Fuzzy Model, Control engineering and applied informatics, vol. 14 , no. 1 , pp. 31-36 , Mar. 2012.

[5] R. Ketata, Y. Rezgui, N. Derbel , Stability and robustness of fuzzy adaptive control of nonlinear systems , Applied Soft Computing, vol. 11, pp. 166178 , 2011.

[6] M. Cococcioni, B. Lazzerini, F. Marcelloni, On reducing computational over-head in multi-objective genetic TakagiSugeno fuzzy systems , Applied Soft Computing , pp. 675688, 2011.

[7] C.-W. Chen , Modeling and fuzzy PDC control and its application to an oscillatory , Mathematical Problems in Engineering 2010 (2010) 13.

[8] M.-L. Lin, C.-W. Chen, Q.-B. Wang, Y. Cao, J.-Y. Shih, Y.-T. Lee, C.-Y. Chen, S. Wang , Fuzzy model-based assessment and monitoring of deserti?cation using MODIS satellite imagery , Engineering Computations: International Journal for Computer-Aided Engineering, pp. 745760 , 2009.

[9] R. Rajesh, M. Kaimal, T-S fuzzy model with nonlinear consequence and PDC controller for a class of nonlinear control systems, Applied Soft Computing, vol. 7, pp. 772782, 2007.

[10] X.J. Ma, and Z.Q. Sun, Output tracking and regulation of nonlinear system based on Takagi-Sugeno fuzzy model, IEEE Trans. Systems, Man, and Cybernetics Part B: Cybernetics, vol. 30, 2000, pp. 47-59..

[11] J. Zhang, M. Fei, T. Yang, and Y. Tan, Robust fuzzy tracking control of nonlinear systems with uncertainty via T-S fuzzy model, Lecture Notes in Computer Science, vol. 4223, pp. 188-198, Springer-Verlag, Berlin, 2006.

[12] C. S Tseng, B. S. Chen, and H. J. Uang, Fuzzy tracking control design for nonlinear dynamic systems via T-S fuzzy model, IEEE Trans. Fuzzy Systems, vol. 9, pp. 381-392, 2001.

[13] H. J. Uang and G. S. Huang, A robust fuzzy model following observer*based control design for nonlinear system, Proc. IEEE Int. Conf. Control Applications, Taiwan, pp. 171-176, 2004.

[14] K. Y. Lian and J. Liou, Output Tracking Control for Fuzzy Systems Via Output Feedback Design, IEEE Trans. Fuzzy Syst, vol. 14, no. 5, pp. 381392, Oct. 2006.

[15] Tseng C, Bor sen Chen, and Uang H.J. Fuzzy tracking control design for non linear dynamic system via ts fuzzy model. IEEE Trans fuzzy system, 9:381392, 2001.

[16] B. S. Chen, C. S. Tseng, and H. J. Uang, Mixed H_2/H_∞ fuzzy output feedback control design for nonlinear dynamic systems: An LMI ap-proach, IEEE Trans. Fuzzy Syst., vol. 8, no. 3, pp. 249265, Jun. 2000.

[17] P.-F. Toulotte, S. Delprat, T.-M. Guerraa and J. Boonaert, Vehicle spacing control using robust fuzzy control with pole placement in LMI region, Engineering Applications of Arti?cial Intelligence, no. 21, pp. 756768, 2008.

[18] D. Khiar. Robust takagi-sugeno fuzzy control of a spark ignition engine. Control Engineering Practice, 2007

[19] F. Zheng, Q-G. Wang, and T. H. Lee. Output tracking control of mimo fuzzy nonlinear systems using variable structure control approach. IEEE Trans. Fuzzy system, 10(6), 2002.

[20] Chong. Lin, Q-G. Wang, and T. H. Lee. Output tracking control for nonlinear via t-s fuzzy model approach. IEEE Trans. systems. Cybernetics, 36(2), 2006.

[21] B. S. Chen, C. S. Tseng, and H. J. Uang, Robustness design of nonlinear dynamic systems via fuzzy linear control, IEEE Trans. Fuzzy Syst., vol. 7, no. 5, pp. 571585, Oct. 1999.

[22] Z. Boulbair, Mise eouvre d'une commande sans capteur d'une machine synchrone á aimants permanents, Mémoire de Magistére, Ecole Polytechnique de l'Université de Nantes, 2002.

[23] Y. Morre, Mise en oeuvre de lois de commande pour les modéles flous de type Takagi Sugeno, Thése de Doctorat, Université de Valenciennes et du Hainaut Cambrésis, 2001.

Table of Contents

PROCEEDINGS ICSENT'12

International Conference on Software Engineering and New Technologies

Edited by: **Mohamed Ridda LAOUAR**