

Knowledge Mapping Evolution Guided by Data Mining

Brahami Menaouer

University of Oran

BP: 1524 El M'naouer, 31000
Oran, Algeria

Atmani Baghdad

University of Oran

BP: 1524 El M'naouer, 31000
Oran, Algeria

Nada Matta

ICD/TechCICO

University of Technology of
Troyes, France

ABSTRACT

Preserve knowledge, retain know, these are the objectives of a scalable company. The knowledge mapping is graphical techniques which allows preserve and visualize the strategic heritage of the acquired know-how during the years. We use Data mining techniques to exploit the different sources of data and to improve the process of acquiring this knowledge. In this paper, we present a new knowledge mapping approach based on the one hand, on the Boolean modeling of mapping domains knowledge and on the other hand, the use of different data sources by data mining technique to improve the process of acquiring knowledge explicit. To evaluate our approach, we have initiated a process of mapping guided by machine learning which operates in two stages: a data mining and automatic mapping. Data mining is to run initially from an induction Boolean case studies (explicit knowledge). The mapping rules are then used to automatically improve the Boolean model of the knowledge mapping.

Keywords

Knowledge management; Knowledge mapping; Boolean modeling; Data mining; evolution of maps

1. INTRODUCTION

The capacity for innovation and the performance of activities currently becomes a major issue for the success of companies. The company exists now more than ever in a very competitive environment to place itself in good position on the globalized market. In terms of company strategy, knowledge management is sight as means to improve the sharing and the use of information and knowledge, as well as a way to capture the best practices to foster innovation and to assist in the decision making. We are interested in our study, on mapping of critical knowledge domains. Knowledge can be mapped to be acquired, created, shared and finally to capitalized and transferred. The method of knowledge mapping is used to represent and analyze knowledge of a company grouped by area and by viewing them as a map [5], [7], [17]. We are talking then of a intangible capital, which includes not only the tacit knowledge that are know-how and competence (the talent and trades secrets acquired with practice and experience) but also the existing knowledge and formalized - stored in databases or data warehouses, procedures, plans, models, summary documents etc. -

Among obstacles to successfully extract knowledge from data mining, we quote: the increasing amount of information generated and made available to the departments concerned the right information at right time. As a result of the arrival of these two fields of application (Data Mining and Data Warehouse), a new idea is obvious: « Why not combine all these techniques to create powerful methods for automated knowledge extraction to improve the mapping process,

including all stages from data collection to evaluation of knowledge gained». Thus was born the idea of a knowledge mapping approach guided by a process of knowledge discovery in data (KDD). Our contribution in this area consists to exploit the proven performances by the techniques of mathematical formal of cell machine CASI (Cellular Automata for Symbolic Induction) [3] in order to design and experiment a new approach of mapping of trades knowledge by automatic learning. The result of the mapping of critical knowledge domains, made by MASK II [5], [7] is refined by a symbolic automatic learning process graph-based induction. This refinement is done by the Boolean Inference Engine (BIE) who is attend of the symbolic induction to optimization of the induction graph and who going ensure, thereafter, the internal representation of the new map of critical knowledge domains.

2. RELATED WORK

During the last decade, the idea of “knowledge management” has brought to light significant challenges related to the roles of intellectual capital, to the innovation and the transfer of knowledge by different forms of tools and processes [12], [16]. Knowledge mapping, considered an instrument which indicates the places where is located the knowledge (among some citizens, organizational units or sources of information) and how to be conducted the acquisition, access to and transfer of this knowledge. According [17], Knowledge mapping techniques aim to track the acquisition and loss of information and knowledge. It explores personal and group competencies and illustrates how knowledge flows throughout an organization or ‘network’. In the other words, knowledge mapping is a process by which organizations can identify and categorize knowledge assets within their organization – people, processes, content, and technology. It allows an organization to fully leverage the existing expertise resident in the organization, as well as identify barriers and constraints to fulfilling strategic goals and objectives. It is constructing a roadmap to locate the information needed to make the best use of resources, independent of source or form [5], [7], [15], [16], and [17].

In paper [2], Knowledge mapping is an important practice consisting of survey, audit, and synthesis. It aims to track the acquisition and loss of information and knowledge. It explores personal and group competencies and proficiencies. It illustrates or “maps” how knowledge flows throughout an organization. Its main purpose consists in quickly showing the collaborators of an organization, a network or pathway, where is located the expertise sought. Similarly, it allows for the indication of the importance of knowledge that is at risk of being lost and that must be preserved [26].

Several approaches to evolution of the mapping have been proposed to organize the cognitive resources of a company.

Aubertin et al. [5] have proposed three different approaches for the realization of a mapping by functional classification, which respectively use the organization chart, a classification by process and a classification by domains. Matta et al. [16] have conducted a project for mapping the knowledge and the technical competence critical within the direction of innovation and research of the INRS. Ermine J.L [7] has completed a project in Chronopost International (observatory of trades) which relies on two objectives, firstly, identify the know-how of trades affected by the strategy and, secondly, consider the evolution of critical skills in the future. For this, Ermine J.L has elaborated the project in several phases: the first phase is the realization a mapping strategic of business actions formalized by the graphical model approach "a map of knowledge domains". The second phase consists of an analysis of the know-how of trades critical through the use of criticality criteria taking into account the specifics of Chronopost International. Chabot J.L [8] has proposed a complete mapping of the different areas of expertise to the company HYDRO-Quebec. However, the primary objective was to identify on the one hand, the areas of knowledge, the other hand, to do a study of criticality to bring out the critical knowledge domains with the help of the French Society Kadrant. Barroso et al. [2] have conducted a project in the center of radio pharmacy in Sao Paulo (IPEN). Since the nuclear domain suffers from problems related to this considerable accumulation of the knowledge: such as the risk of non-preservation, the difficulty of transfer, etc... They have developed the project in several steps by using a process approach. The process was described in a conventional manner in the form of flow diagrams linking activities in process. Sellin and Duzert, [20] have studied the usage of a type of Knowledge Management System (KMS) in particular (the knowledge mapping of "trades") in the context of human resources management. In addition, they have discussed the appropriateness of their use in organization. This study was conducted in the TOTAL group.

In addition, there is a very wide range of works that propose data mining techniques for the extraction of hidden new knowledge from a large mass of data. The main objective of this work is to analyze the content data to extract significant knowledge. This supports and enhances, well, the decision process of knowledge management and in particular the mapping of critical knowledge domains. Several types of cooperation between mining and knowledge management have been proposed to improve the knowledge already formalized in the organization. The problem of fusion of mining and of knowledge management is complex and expensive, several techniques have been proposed to exploit the specific data of the application domain in order to optimize the process of mapping domains of critical knowledge [15], [17].

Among the problems open from the knowledge mapping, include notably research dynamic of knowledge domains decisional in a knowledge map that is becoming increasingly complex, due to several parameters (*number of domains, criteria and degrees of criticality*) thereby the evolution of knowledge maps exploiting other sources of enterprise data. Then, it does not allow formalize the data and information in

real time. For this, the generalization of the extraction process of knowledge on all available knowledge in the organization (Database, reference documents, enterprise portal) is desirable and possible by the use the data mining techniques.

In this context, we have set ourselves as the objective of processes integration the data mining to support approach MASK II in the mapping of critical knowledge domains of the health SEMEP service «Services of Epidemiology and Preventive Medicine» of the city of Mostaganem in Algeria. The service SEMEP has as a task: the collection, the analysis and the interpretation of sanitary data for the description and observation contained of student health in order to facilitating the planning, the implementation of evaluation of interventions and school health programs.

3. APPROACH PROPOSED

The approach (**BKMDM: Boolean Knowledge Mapping** guided by **Data Mining**) that we have developed in this article is conducted into two major phases described in Figure 1. Firstly, in the first phase, the construction of the map of service SEMEP and evaluation of different domains of know how critical trades of epidemiologies basing on a extraction work of expert knowledge (Doctors, Health technicians, Psychologists, etc...). For this construction, we have used one of the methods of modeling and knowledge representation, the method MASK II [5], [7]. After getting the global mapping of domains of critical knowledge of epidemiological [13], [14], we conducted the Boolean modeling inspired by the cellular machine CASI [3]. In the second phase, by the techniques of data mining, we exploit the different data sources to improve the mapping of critical knowledge domains. A contribution in this area consists of conceiving and experimenting a new Boolean approach of critical knowledge mapping domains. The result of the mapping of critical knowledge by MASK II is refined through a symbolic automatic learning process graph-based induction.

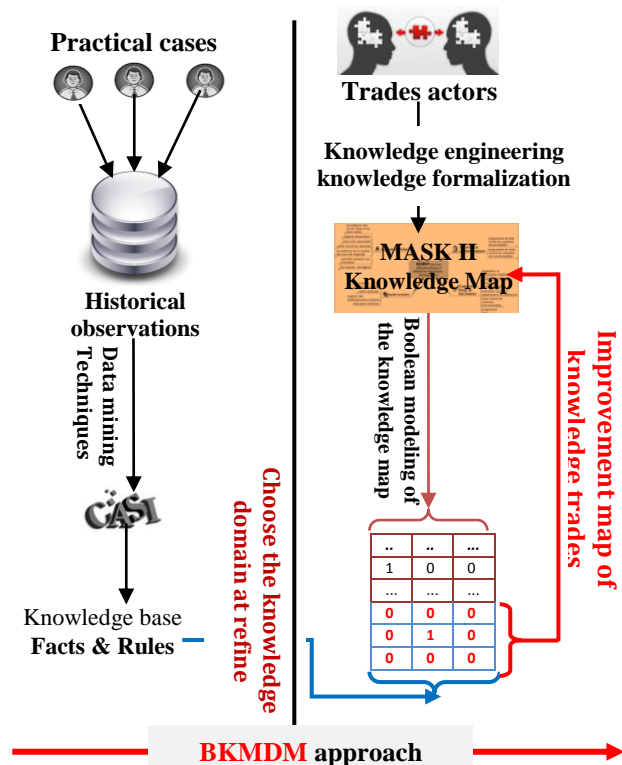


Fig 1: General architecture of the approach "BKMDM" proposed

3.1 Cellular Automaton “CASI”

CASI (Cellular Automata for Symbolic Induction) is a cellular method of generation representation and a means to optimize induction graphs generated from a set of learning examples [2]. This Cellular system is organized into cells where each cell is connected only with its neighbors (subset of cells). All cells obey in parallel to the same rule, which is called the “local transition function” This results in an overall transformation of the system. CASI is composed of three modules: **COG** (Cellular Optimization and Generation), **CIE** (Cellular Inference Engine), and **CV** (Cellular validation) (shown in Fig 2).

- **COG (Cellular Optimization and Generation) module:** Using a cellular automaton and cooperating with an induction graph (SIPINA method), COG module will extract new knowledge from training data. Two finite layers of finite automata represent the knowledge that is generated.
 - **CIE (Cellular Inference Engine) module:** The CIE module, heart of the cellular machine (CASI), simulates the functioning of basic cycle of an inference engine using two layers of finite automata finite. The first layer, CELFACT, for the basis of facts and, the second layer, CELRULE, for the basis of rules. The states of the cells consist of three parts: EF, IF and SF, respectively ER, IR and SR are the input, internal state and output of a cell of CELFACT, respectively a cell of CELRULE. The internal state, IF of a cell of CELFACT indicates the role of fact: in our graph IF=0 corresponds to a fact type top (si), IF=1 corresponds to a fact type attribute= value (Xi = value) [2]. For defining the vicinity of cells, we are using the two incidence matrices of the (RE) input and the (RS) output of the automaton. RE and RS represents the input / output relationship of facts and are used in forward chaining: of the root to the leaves. We can also use (RS) as input relation and (RE) as the output relationship to launch an inference in backward chaining: the leaves to the roots. Finally, the dynamics of CASI for simulate the operation of CIE inference engine uses two transition functions δ_{fact} and δ_{rule} where δ_{fact} is the phase of evaluation, selection and filtering, and δ_{rule} corresponds to the execution phase and that all the cells in parallel to obedient the same branch called local transition function, which a as consequence in a global transformation synchronous of the system [2].
- **CV (Cellular Validation) module:** After the rules have been generated by the SIPINA method, which has been coupled along with the CASI machine, validation of this knowledge could be done using the CV module.

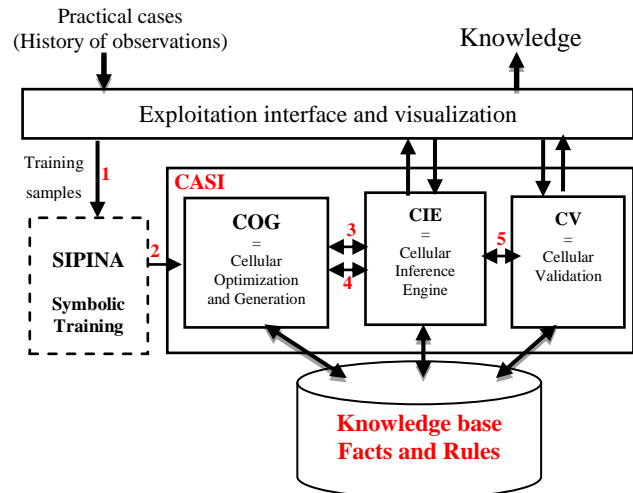


Fig 2: General diagram of the cellular machine (CASI) [2]

3.2 Boolean modeling of knowledge mapping (Phase I)

Our project is registered in the context of National Program of Research (NPR)¹, launched in collaboration between the health sector SEMEP of Mostaganem and our research team SIF «Simulation, Integration and Data mining» of the laboratory LIO «Computer Science Laboratory of Oran». Phase I (see Fig. 1) of our approach proceeds in two steps: a graphical representation of the knowledge map using the graphical model of the method MASK II [5], [7] and a second step of Boolean modeling knowledge mapping.

3.2.1 Definition of knowledge map

Experience has demonstrated that the implementation of an effective knowledge management goes through the establishment of a knowledge mapping. According Speel et al [17], knowledge mapping is defined as the process, methods and tools for analyzing knowledge areas in order to discover features or meaning and to visualize them in a comprehensive, transparent form such that the business-relevant features are clearly highlighted. Of more, she allows to have a fine understanding, by a criticality analysis, domains of knowledge where efforts should be made in terms of capitalization, sharing or innovation [7]. Moreover, the field of the knowledge mapping presents itself as a research area into emerging where work identified through the literature are few and few thorough [15]. There are various approaches to organizing the cognitive resources of an enterprise or agency. We can distinguish however two types of approaches cartographic the first focuses on the classification by process [15] and the classification by domains[7] which organizes the knowledge around the subjects, objects or purposes of common interest which are identified in the organization. The Method of mapping critical knowledge, named MASK II is classified in the latest approach.

In this context, we realized knowledge mapping of the SEMEP service in order to visually represent and of manner synthetic the knowledge domains. We recall, in this step that the realization of the map was based on analysis of references documents (organizational chart, description of the distributions of activity services, directory of staff activities,

¹ The project is entitled « A Service Oriented Architecture for the Expanded Program on Vaccination », Code 12/u310/65.

plan medium term (vaccination), syntheses epidemiological, studies, balance sheet of vaccination, etc.) and interviews with business experts (Doctors, Health technicians, Psychologists, midwives, etc.) and the responsible of the health sector.

The result is a map of knowledge domains or know-how map trades / strategic. This mapping is a description of level a

meta-knowledge [7] of SEMEP know-how. It provides an addressing system of know-how, to facilitate access. The result is a map of knowledge domains or map of know-how trades / strategic (Fig. 3).

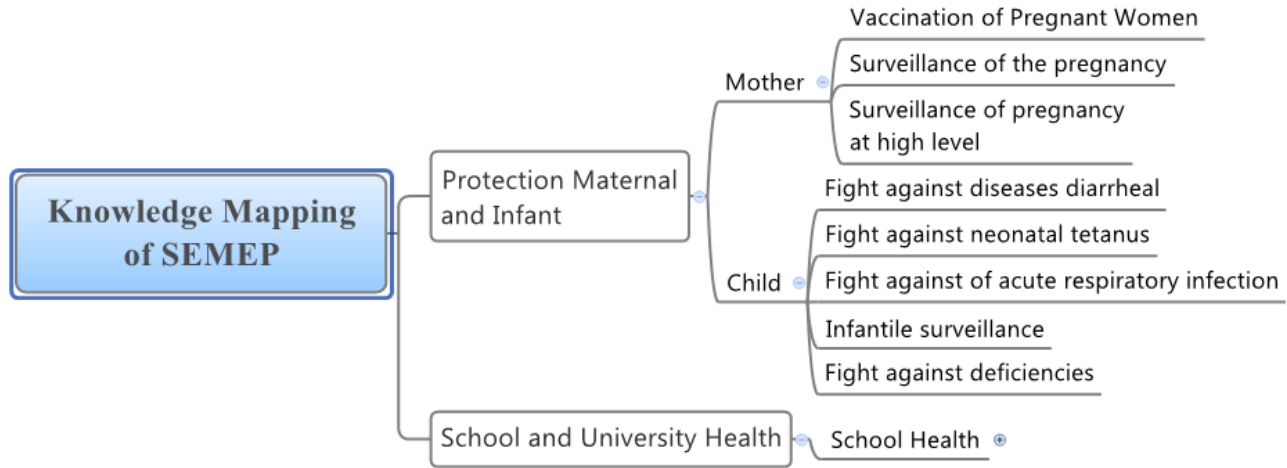


Fig 3: Knowledge mapping of the SEMEP

3.2.2 Boolean modeling of knowledge map

After getting the knowledge mapping of SEMEP, we conducted the Boolean modeling. We recall that the process of Boolean modeling graph (BMG) is structured in three phases (BOG (Boolean Optimization and Generation) is to optimize and generate the Boolean model from a graph and generate a Boolean knowledge base (BKB), BIE (Boolean Inference Engine) is able, from a Boolean model of a graph (BMG), to simulate the operation of the basic cycle of an inference engine using two layers of finite state machines; CELSUMMIT, for the summits and CELARC, for arcs. The states of cells are composed of three parts: ES, IS and SS, respectively, EA, IA and SA, which designate the input, the internal state and the output of a cell CELSUMMIT respectively a cell CELARC, and BV (Boolean Validation) is devoted to the validation process of the model developed[3], [13]. In this context, we consider the graph shown in Fig 4, extract from the knowledge mapping of SEMEP. Fig. 4 shows the axe: "Infant Health" with its themes and domains by the CARTOCEL system [13], [14]:

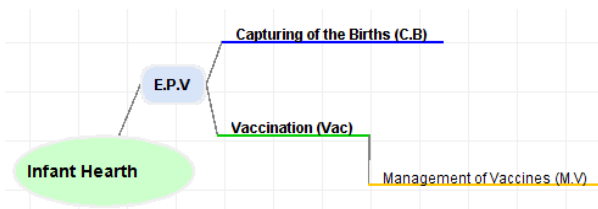


Fig 4: Knowledge map of the axe: "Infant Health" by CARTOCEL

The CARTOCEL system was developed within the framework of a research project in collaboration between the health sector SEMEP of Mostaganem in Algeria and our research team SIF «Simulation, Integration and Data mining» of the laboratory LIO «Computer Science Laboratory of Oran», which allows a dynamic visualization of domains of knowledge guided by machine learning.

The figure 5 represents the map obtained by using the partitions S_0 (partition Core activity) = (s_0), S_1 (partition Axes) = (s_1), S_2 (partition Themes) = (s_2), S_3 (partition Sub-themes) = (s_3) S_4 (partition Domains) = (s_4).

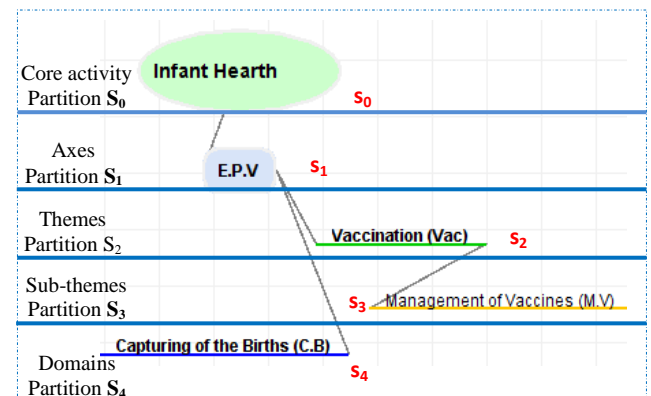


Fig 5: Generating partitions S_0 , S_1 , S_3 and S_4

After transformation of the knowledge domains mapping in a directed graph generated by partitions, four (04) rules have been deduced using the BOG module [13] in the form: « initial summit → final summit ». Table 1 describes the symbolic representation of the graph.

Table 1. Symbolic representation of the partitions illustrated by Figure 5

Initial summits	Attributes	Final summits
s_0	Expanded Program of Vaccination (E.P.V)	s_1
s_1	Vaccination (Vac)	s_2
s_2	Management of Vaccines (M.V)	s_3
s_3	Capturing of the Births (C.B)	s_4

The Table 2 shows how the knowledge base extracted from this graph is represented by layers CELSUMMIT and CELARC of Boolean inference engine (BIE). The internal state of an IS a cell from CELSUMMIT indicates the role of fact: in our graph IS = 0 corresponds to a fact of type summit (si), IS = 1 corresponds to a fact of type attribute=value (Xi = value).

Table 2. G_0 Initial configuration

Summits	ES	IS	SS
s0	1	0	0
Axe= P.E.V	0	1	0
s1	0	0	0
Theme= Vac	0	1	0
s2	0	0	0
Sub-theme= M.V	0	1	0
s3	0	0	0
Domain= C.B	0	1	0
s4	0	1	0
CELSUMMIT			

Arcs		EA	IA	SA
Arc 1	A1	0	1	1
Arc 2	A2	0	1	1
Arc 3	A ₃	0	1	1
Arc 4	A ₄	0	1	1
		CELARC		

The two incidence matrices R_E input and R_S output, developed in detail in [3], represent the relation of input / output summits and are used in forward and backward chaining (Table 3).

Table 3. The impact of input matrices R_E and Output R_S

R_E/R_S	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
s_0	1	1	0	0	0	0	0	0
Axe=P.E.V	0	0	0	0	1	0	0	0
s_1	0	0	0	0	1	0	0	0
Theme=Vac	0	0	0	0	0	1	0	0
s_2	0	0	1	1	0	1	0	0
Subtheme=M.V	0	0	0	0	0	0	1	0
s_3	0	0	0	0	0	0	1	0
Domain=C.B	0	0	0	0	0	0	0	1
s_4	0	0	0	0	0	0	0	1

The following table presents the forward chaining according to the synchronous mode width, using the same principle of inference engine (BIE)[3], [13], where all arcs (axes, themes, sub-themes, domains and sub-domains) of our candidates map are provoked.

Table 4. Forward chaining according of the width synchronous mode

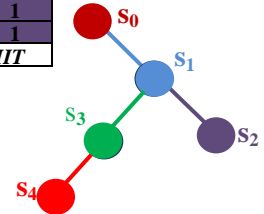
Cycle	Set fact base evolution	Rules
1	$\{s_0\} \cup \{(Axe=E.P.V), s_1\}$	1
2	$\{s_0, (Axe=E.P.V), s_1\} \cup \{(Theme=Vac), (Domain=C.B), s_2, s_4\}$	2, 4
3	$\{s_0, (Axe=E.P.V), (Theme=Vac), (Domain=C.B), s_1, s_2, s_4\} \cup \{(Sub-theme=M.V), s_3\}$	3

We consider the initial configuration of the cellular automata (see Table 5) and, $\Delta = \delta_{summit} \circ \delta_{arc}$ the global transition function: $\Delta(G_0) = G_1$. After the application configurations we obtain G_1 , G_2 and G_3 . In the following we present the final configuration G_3 : values

Table 5. The final configuration G_2 obtained after two synchronous iterations

Summits	ES	IS	SS
s_0	1	0	1
Axe= P.E.V	1	1	1
s_1	1	0	1
Theme= Vac	1	1	1
s_2	1	0	1
Sub-theme= M.V	1	1	1
s_3	1	0	1
Domain= C.B	1	1	1
s_4	1	0	1
CELSUMMIT			

Arcs	EA	IA	SA
A1	1	1	0
A2	1	1	0
A3	1	1	0
A4	1	1	0
CELARC			



Thus, we can, by modeling MBG, make inferences in backward chaining to go from the sub-domains to the axes and the core activity. Simply switch between R_E and R_S in the transition functions.

Moreover, we can deduce using the BOG module (in cooperation with BIE) [3], from the axe «*Infant Health*», five mapping rules that are of the form: If *Condition* Then *Conclusion*. *Condition* is a logical expression composed of conjunction which will be called *Premise* and *Conclusion* of knowledge domain (§ sub-section 3.1) in fact described by the conclusion (as an example for the conclusion rule 4 is «the target knowledge domain s_4 =Management of Vaccines»)(see Table 6).

Table 6. Conjunctive rules produced

MAPRULE 1	<i>if</i> (Axe=Expanded Program of Vaccination) <i>Then</i> the knowledge of target s_1
MAPRULE 2	<i>if</i> (Axe=Expanded Program of Vaccination <i>and</i> Theme= Vaccination) <i>Then</i> the knowledge of target s_2
MAPRULE 3	<i>if</i> (Axe=Expanded Program of Vaccination <i>and</i> Domain = Capturing of the births) <i>Then</i> the knowledge of targets s_3
MAPRULE 4	<i>if</i> (Axe=Expanded Program of Vaccination <i>and</i> Theme= Vaccination <i>and</i> Sub-theme= Management of Vaccines) <i>Then</i> the knowledge of target s_4

3.3 Towards a mapping guided by Data mining

The mapping process guided by machine learning, (see Fig. 1) of our project, proceeds in two steps: a data mining and automatic mapping. Data mining consists to launch initially the symbolic induction from case studies using CASI [3]. The mapping rules obtained are used to automatically improve the Boolean model of knowledge mapping already carried (see Fig. 1).

In [12], the process of Knowledge Discovery in Databases comprises 5 stages: selection, pre-processing, transformation, data mining and interpretation. In this framework, the data we used are the sheets of psychologist's SEMEP. In addition, they are very diverse and are not necessarily all usable by the data mining techniques. Most techniques used deal only with the tables of data in rows /columns traditional. The objective is to prepare the tables rows/columns, in other terms, individual tables/variables, obtained by the following steps (see Fig. 6):

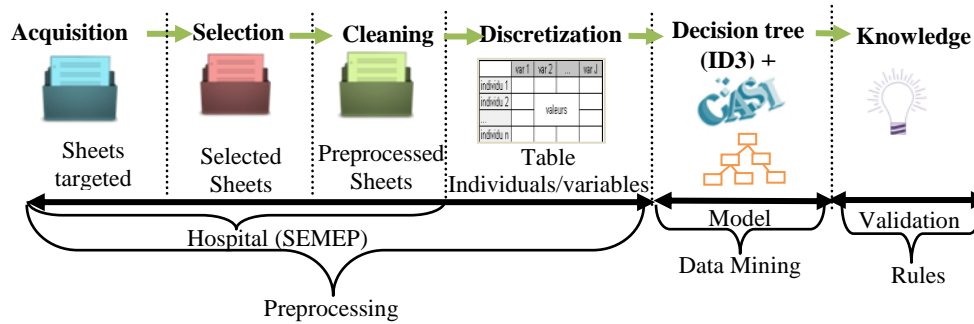


Fig 6: The process of extracting knowledge from psychological cards (personality test)

3.3.1 The preprocessing

The preprocessing is a key preliminary step in the process of knowledge discovery in databases (KDD), the results generated at the end of this phase depend in large part the quality of data used [9], [19].

The preprocessing steps include access to data in order to construct two-dimensional tables, called table individual variables, which include observations (explicit data). Based on the type of data (e.g. numeric, symbolic), methods of preprocessing format the data, the clean, handle missing data and select the attributes or observations (individuals) when there are too many: the choice of attributes most informative in the first case, sampling in the second case. This phase is important because it is the one that will determine the quality of the models developed in data mining. Indeed, these choices are intended to bring out the information contained within this mass of data [18].

In our case and to complete this step we had to understand the principle of the tests (the psychological tests «Personality Test- Beck-II») [1], extract the variables (attributes) in order to launch the qualitative phase data mining and generating knowledge.

3.3.2 Data mining by decision tree

Data mining is defined as the art of knowledge discovery in databases (KDD). The techniques of data mining (the segmentation algorithms, the association rules, the decision trees, the neural networks...) are proposed according on problem at solve. It may to act methods of structuring or classification, explanation or prediction [10], [19]. Among these techniques, we are interested to decision trees in the framework of this article.

3.3.2.1 Construction of a decision tree and rules generations

In the literature, decision trees are among the classification techniques of data mining the most popular, fastest and easiest to use. According to [4], decision trees are learning tools that produce rules such as « If condition then conclusion » in which « condition » means a disjunction of conjunctions of logical propositions of type « attribute, value » [10]. The set of rules is thus the prediction model. They use as input a set of individuals (n-uplets) described by variables (attributes). Each individual belongs to a class, the classes being mutually exclusive. To construct a decision tree, it is necessary to have learning population (table or view) consists of individuals whose class is known. The learning process then consists in determining the class of any individual based on the value of its variables [4], [9].

To illustrate this notation, consider the problem of psychological tests in particular the personality tests (BDI-II).

Table 7 shows our learning sample of 14 patients (the students in secondary schools). Each example or patient is described by four descriptive variables (Sense of failure, Sense of guilt, Sadness, and Pessimistic) « in statistics called exogenous variables » and associate a particular attribute (depression).

Table 7. Example for learning sample-patients

Ω_u	X_1 Sense of failure	X_2 Sense of guilt	X_3 Sadness	X_4 Pessimistic	X_5 Depression
1	yes	yes	yes	yes	Major
2	no	yes	yes	yes	Dysthymia
3	no	no	no	yes	Dysthymia
4	no	yes	no	yes	Dysthymia
5	yes	can be	no	yes	Major
6	no	yes	yes	no	Major
7	yes	yes	no	no	Major
8	yes	no	yes	no	Major
9	no	yes	no	no	Major
10	no	no	yes	can be	Dysthymia
11	no	yes	yes	can be	Dysthymia
12	yes	yes	no	can be	Major
13	no	yes	no	can be	Major
14	no	yes	no	can be	Major

In the following table, we have summarized an example of exogenous variables out of our learning sample. The value taken by $X_j(w)$ is called the term or value of the variable X_j for each individual (patient). We denote by L_j the number of different modalities assigned to the variable X_j .

Table 8. Table captions should be placed above the table

Variables and (l_j)	Signification	Values
$X_1(l_1=2)$	Sense of failure	yes ; no
$X_2(l_2=3)$	Sense of guilt	yes ; no ; can be
$X_3(l_3=2)$	Sadness	yes ; no
$X_4(l_4=3)$	Pessimistic	yes ; no ; can be
$X_5(l_5=2)$	Depression	Major ; Dysthymia

In this case, the patient population affected by the problem of learning is a set of tuples consisting of the four predictor variables X_1, X_2, X_3, X_4 and their classes (Major depression and Dysthymia), from these examples, we construct a tree said decision as:

- Each node corresponds to a test of the value $X_j(w)$ of an attribute X_j which l_j has possible values $(x_j^1, \dots, x_j^{l_j})$;
- Each branch leaving a node corresponds to a value x_j^v of the test on x_j with $v=1, \dots, l_j$;

- Each leaf is associated c_k value of the target attribute Y .

Suppose our learning sample is composed of 14 patients. s_0 the initial partition has a single element denoted s_0 , which includes all the learning sample with nine (09) individuals (patients) belonging to the class «major depression» and five (05) belonging to the class «dysthymia».

For the construction of the decision tree, we used the algorithm of ID3 method [10]. Starting from the root of the tree (Fig. 6) the partition $S_1 = \{s_1, s_2\}$ is generated by the variable X_1 with:

- $s_1 = \{\omega \in \Omega \alpha / X_1(\omega) = \text{yes}\}$;
- $s_2 = \{\omega \in \Omega \alpha / X_1(\omega) = \text{no}\}$;

Just as the node s_0 , we distinguish into s_1, s_2 the individuals of class 1 and 2 of the partition S_1 , the process is reiterated in search of a partition S_2 which would better according to the chosen gain (see Fig. 7). Thus, the decision tree can be then exploited to: extract the classification rules concerning our target attribute «depression».

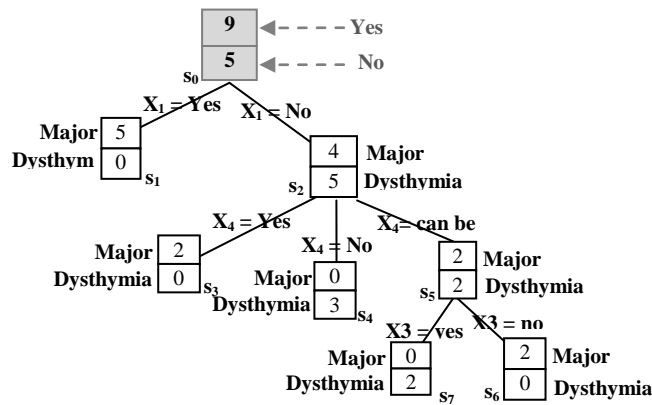


Fig 7: Example of a figure captiondecision tree obtained by ID3 on the example « Test of BECK ».

Finally, we exploited the tree before (Fig. 7) to extract five rules $R_1, R_2 \dots$ and R_5 of inductions psychological also on the target attribute «depression» useful and bearing knowledge critical have not been explicit in advance and which are of the form: **If Condition then Conclusion**.

1. **If** (sense of failure = yes) **then** major depression.
2. **If** (sense of failure = no and pessimistic = no) **then** major depression.
3. **If** (sense of failure = no and pessimistic = yes) **then** dysthymia.
4. **If** (sense of failure = no and pessimistic = can be and sadness = no) **then** major depression.
5. **If** (sense of failure = no and pessimistic = can be and sadness = yes) **then** dysthymia.

3.3.3 Exploitation of the rules of induction

In this section, we launch the validation phase across the CV module (Cellular validation) on induction rules psychological (Personality Tests «Test of BECK») presented in the previous section.

For the experimentation phase we used the platform WS4KDM [6] for extraction and the Boolean modeling the rules of psychological prediction. The objective and the

automatic improvement the Boolean model of knowledge mapping critical epidemiological guided by data mining. WS4KDM2, by applying the principle Boolean of the cell machine, takes as input the learning sample as a table individuals /variables in order to supply at output a basis of classification rules. The result of the knowledge mapping of SEMEP obtained by CARTOCEL [14] is illustrated in Fig. 8.

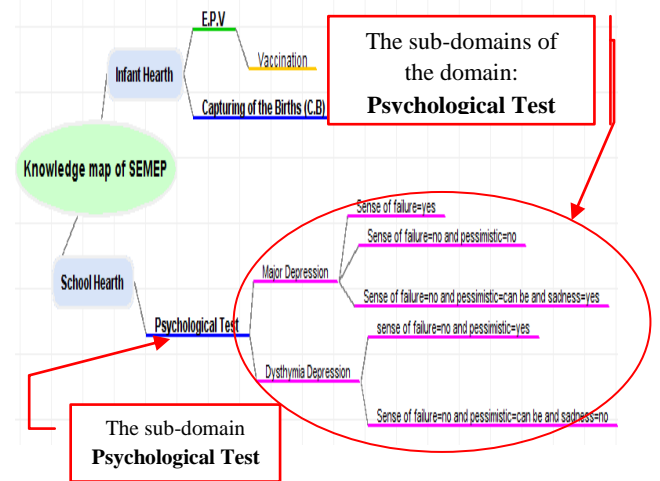


Fig 8: Knowledge domains mapping of SEMEP refined by a process of inducing rules «test of Beck»

Moreover, we find that the Boolean modeling of knowledge mapping in the CARTOCEL system to be one of our strong interests. It resides in the possibility of transforming the map into a set of rules, called *mapping rules*, by translating the set of paths leading from the root (heart of a business) at each leaf (axes, themes, and domains of knowledge). All of which contributes to organizational learning and can ensure a better exploitation of domains of knowledge by experts, since it shall make available the know-how of actors trades (doctors, psychologists, midwives) with the ability to improve practices in work situations. When it is used frequently, this contribution can support the intelligence of visual representation.

4. CONCLUSIONS

Competing motivations have led us to propose this principle cellular for optimization, generation, representation and use of Boolean knowledge mapping. Indeed, we have, not only wished to have a knowledge map optimal but also, we have, too, wanted to improve the construction and the visualization of this map by proposing a new process of knowledge mapping guided by data mining. The advantages of our approach based on the principle of Boolean modeling to render mapping of critical knowledge more flexible and scalable may be recapitulated as follows:

- The representation of knowledge well as his control are simple, in the form of binary matrices its requiring the minimum preprocessing;
- The results of the mapping are simple to be reorganized and used by and for a process of data mining;
- The incidence matrix, R_E , facilitates the transformation of links between partitions in

expressions Boolean equivalent, which allows us to use the Boolean elementary algebra to examine other simplifications.

- The ease of navigation and dynamic search of knowledge decisional by using the transition functions and production rules.

5. ACKNOWLEDGMENTS

The authors also acknowledge the service team SEMEP and the psychologists in secondary schools for her assistance on to finalize this project and the team of the TechCICO laboratory.

6. REFERENCES

- [1] A. T. Beck, and B. A. Alford. (2009). *Depression: Causes and treatment*. 2d edition Philadelphia: University of Pennsylvania Press, ISBN-13: 9780812219647, P.459
- [2] A.Barroso, R. Ricciardi, and J-L Emine. (2005). Knowledge Evaluation for Knowledge Management Implementation - the Case Study of the Radiopharmaceutical Center of IPEN, 1st international Conference on Nuclear Knowledge Management, INAC 2005, Santos (Brazil), 29st august 2005, P. 1-14
- [3] B. Atmani, B. Beldjilali, Knowledge Discovery in Database: Induction Graph and Cellular Automaton, Computing and Informatics Journal, Vol.26, N°2 (2007), P.171-197, 2007.
- [4] D.A. Zighed, R. Rakotomalala, Graphs of induction, Training and Data Mining, Hermes Science Publication (Edition Hermès Sciences), 2000, P. 21-23.
- [5] G. Aubertin. (2010). Knowledge Mapping: A Strategic Entry Point to Knowledge Management, In : Trends in Enterprise Knowledge Management, Edited by Imed Boughzala, Jean-Louis Ermine, Published Online: 19 JAN 2010, Chap11, P.197-216
- [6] H.Kadem and B. Atmani, (2010) "Conception d'une Plateforme Cellulaire Open Source d'Extraction et de Gestion des Connaissances : WS4KDM". 7ème Séminaire National en Informatique BISKRA (SNIB'2010), Université Mohamed Khider - Biskra, Algérie, 02-04 Nov 2010.
- [7] J-L, Ermine. (2010). Knowledge Crash and Knowledge Management. International Journal of Knowledge and Systems Science (IJKSS), Volume 1, Issue (4), P.79-95
- [8] J.-L. Chabot. (2006). The development of expertise – strategic stakes, 8th Symposium on Humain Resource Management – CESS, September 15, Quebec, Canada, P.1-29
- [9] J. Han, M. Kamber, and J. Pei. (2011). Data Mining : Concepts and Techniques, The Morgan Kaufmann Series in Data Management Systems, University of Illinois at Urbana-Champaign (Canada), 3rd Edition, July 6, 2011, ISBN-10: 0123814790, P.744
- The mapping system is a cellular model composed of a simple set of transition functions and production rules, which allow ensure the evolution of knowledge maps by exploiting other sources of the company data.
- [10] J. R. Quinlan, (1986), "Induction of decision trees", Machine Learning, Volume1: Issue 1, P: 8-106.
- [11] I.Nonaka, H.Takeuchi, (1995). The Knowledge-Creating Company, Oxford University Press, Oxford, New York.
- [12] U. Fayyad, G.P. Shapiro, P. Smyth, (1996). "The KDD process for extraction useful knowledge from volumes data", In: Communication of the ACM, Vol. 39, Nov.1996, P. 27-34.
- [13] M. Brahami, B. Atmani, Vers une cartographie des connaissances guidée par la fouille des données. 2ème Conférence Internationale CIA'09, Université de Saida, Algérie, Publier dans <http://CEUR-WS.org>, ISSN: 1613-0073, 2009.
- [14] M. Brahami, B. Atmani, M. Mokaddem, CARTOCEL : Un outil de cartographie des connaissances guidée par la machine cellulaire CASI, In 10ème Conférence Internationale Francophone sur l'Extraction et Gestion des Connaissances, EGC'2010, RNTI (E-19), Edition Cépaduès, ISSN : 1764.1667.
- [15] Grundstein, Michel . Three Postulates That Change Knowledge Management Paradigm. Hou, Huei-Tse. New Research on Knowledge Management Models and Methods. . 2012. pages 1-22.
- [16] Matta N.; Ducellier G.; Charlot Y.; Beldjoudi R.; Tribouillas F.; Higon E. (2011). Traceability of Design Project Knowledge using PLM, In IEEE proceedinds of International Conference on Cooperation Technologies and sciences, Philadelphia, May, 2011.
- [17] P.H. Speel, N. Shadbolt, W. De Vies and P.H. VanDam, O'hara K, Knowledge mapping techniques within the construction industry: An exploratory study, in: CIB W102-Information and knowledge management in Buildings, 3th - 4th June 2008, Finland.
- [18] R. Rakotomalala, Induction Graphs, Thesis for the obtaining of the Diploma PhD, University of Claude Bernard-Lyon 1, 1997, France.
- [19] P. Flach. Machine Learning, Editor: Cambridge University Press; Edition: 1 (November 29, 2012), ASIN: B009ZRNT0C, pp.5-50.
- [20] Sellin K. A. Duderzert. (2010). knowledge Maps for a pilotage of human resources and HR process, published in "Trends in HR and Management: the contribution of prospective: Essays in honor of Luc Boyer (2010), P.1-21