People's democratic republic of Algeria
Ministry of higher education and scientific research

University Blida 1
Faculty of Sciences
Computer Science Department

Final thesis to obtain the Master's degree

Specialty: Computer Science
Option: Software Engineering

Implementation of Organizational Knowledge Acquisition services in Cloud Computing environment based on knowledge engineering

Case Study: MOBILIS ATM - ALGERIA

Presented by:

- AFIR Sofiane
- BOUMECHTA Med Issam Eddin

Infront of juries composed of:

- Mrs Abed (President)
- Mrs Berramdane (Examinator)
- Mme CHIKHI Imane (Promotrice)

2020/2021

Appreciation

First of all we thank ALLAH, the Almighty for giving us the strength and courage to accomplish this modest work of research.

We warmly thank our supervisor Mme CHIKHI Imane for entrusting us this work first and for her support, her precious advice, her encouragement and her great patience throughout the period of this work.

Our sincere thanks also go to our co-supervisor Mr Amine Ouaoua for his help, his great availability and his contribution with MOBILIS database.

We thank the members of the jury, who did us the honor of participating at the judgment of this work.

We also wish to thank all of our computer science teachers for all the knowledge they gave us all these years of study.

A very warm thank you to our parents for their support and advice throughout our studies as well as all family members, our friends and all those, near or far, have brought us their contribution, to accomplish this work.

ملخص

المعرفة قوة. تحاول كل منظمة إدارة معارفها قدر الإمكان لاستخدامها من أجل تحديد أهدافها وتحقيقها. يعتبر اكتساب المعرفة من أهم العمليات في إدارة المعرفة. يسمح باكتساب المعرفة التنظيمية الضمنية والصريحة داخل الشركة. الهدف من هذه العملية هو اكتساب المعرفة من أصحاب المعرفة مثل خبراء المجال والموظفين والمديرين ... ومن مصادر المعرفة الصريحة غير المتجانسة مثل الوثائق وقواعد البيانات.... الهدف الرئيسي من عملنا المقترح في هذه الأطروحة هو تعريف عملية اكتساب المعرفة ومناقشة تأثير نموذج الحوسبة السحابية لتحقيق العملية داخل المنظمة. يعتمد النهج المقترح بشكل أساسي على هندسة المعرفة والحوسبة السحابية. قمنا بتطوير نظام اكتساب المعرفة لاكتساب المعرفة الضمنية والصريحة. يعتمد اكتساب المعرفة الضمنية على أساليب اكتساب المعرفة الضمنية الحالية ونمذجة المعرفة. يعتمد اكتساب المعرفة الصريحة على نموذج اكتشاف البيانات وتقنيات التنقيب في البيانات حيث نركز على المعرفة المضمنة في قواعد البيانات التشغيلية. نحن نتحقق من صحة النظام المقترح من خلال دراسة حالة.

الكلمات المفتاحية: إدارة المعرفة ، نماذج المعرفة ، عملية اكتساب المعرفة ، هندسة المعرفة ، استخراج البيانات ، الحوسبة السحابية

ABSTRACT

Knowledge is power. Each organization try to manage their knowledge as much as possible to use it in order to determine and achieve their objectives. The knowledge acquisition is one of the most important process in Knowledge Management. It allows to acquire crucial tacit and explicit organizational knowledge within the company. The objective of this process is to acquire knowledge from knowledge holders such as domain experts, employees, managers... and from heterogenous explicit knowledges sources such documents, databases…. The main goal of our work which is proposed in this thesis is the definition of the Knowledge Acquisition process and discuss the impact of Cloud Computing paradigm to achieve the process within an organization. The proposed approach is mainly based on knowledge engineering and Cloud Computing. We developed a Knowledge Acquisition System for tacit and explicit knowledge acquisition. Tacit knowledge Acquisition is based on existing Tacit Knowledge Acquisition methods and knowledge modeling. Explicit Knowledge Acquisition is based on the Knowledge Discovery form Data process and Data Mining technics as we focus on knowledge embedded in operational databases. We validate the proposed system through a case study.

Key words: Knowledge Management, Knowledge Models, Knowledge Acquisition Process, Knowledge Engineering, Data Mining, Cloud Computing.

Résumé

La connaissance est un pouvoir. Chaque organisation vise à gérer au maximum ses connaissances pour les utiliser afin de déterminer et atteindre ses objectifs. L'acquisition de connaissances est l'un des processus les plus importants de la Gestion des Connaissances de l'entreprise. Il permet d'acquérir les connaissances cruciales de l'entreprise, tacites et explicites. L'objectif de ce processus est d'acquérir des connaissances auprès des détenteurs de connaissances tels que les experts de domaine, les employés, les directeurs... et à partir de sources de connaissances explicites hétérogènes telles que les documents, les bases de données…. L'objectif principal de notre travail proposé dans ce mémoire est de définir le processus d'acquisition de connaissances et discuter l'impact du paradigme du Cloud Computing pour réaliser le processus au sein d'une organisation. L'approche proposée est principalement basée sur l'ingénierie des connaissances et le Cloud Computing. Nous avons développé un système d'acquisition de connaissances pour l'acquisition de connaissances tacites et explicites. L'acquisition de connaissances tacites est basée sur les méthodes existantes d'acquisition de connaissances tacites et la modélisation de connaissances. L'acquisition de connaissances explicites est basée sur le processus de découverte de connaissances à partir de données et les techniques de Data Mining, car nous avons pris le cas des connaissances formalisées dans les bases de données opérationnelles. Nous avons validé notre système par une étude de cas.

Mots clés : Gestion des connaissances, Modèles de connaissances, Processus d'acquisition de connaissances, Ingénierie des connaissances, Data Mining, Cloud Computing.

Table of contents

List of figures

List of tables

Lists of acronyms

KE: Knowledge Engineer
KA: Knowledge Acquisition
KM: Knowledge Management
KDD: Knowledge Discovery in Databases
DM: Data Mining
KDP: Knowledge Discovery Process

GENERAL INTRODUCTION

In the business and organizational domain, knowledge is one of the most valuable assets. It is used for many purposes such as innovation, development and decision-making.

With sudden and unexpected force, Knowledge Management (KM) has established itself in companies as a major issue. A set of strategic elements contributes to this emergence, like knowledge is an economic capital, knowledge is a strategic resource, knowledge is a factor of stability of the company, knowledge brings a decisive competitive advantage, which the main goal is providing the business value. The strategic vision that business leaders can now have on their knowledge heritage, leads them to define global objectives to the best manage this resource. These goals always revolve around three key points: Capitalize (Knowing where you come from, knowing where you are, to better know where you are going), Share (Go from individual intelligence to collective intelligence), Create (Create, innovate to survive). Knowledge management is a considerable managerial challenge that is part of a long-term change and new visions of the organization.

In order to achieve successful implementation of knowledge management in organizations, this must be aligned with the organization's business objectives as well as the quick evolving of technology. In addition, organizations tend to evolve virtually and globally. Therefore, the classical approaches proposed for knowledge management systems architectures presented some shortcomings in terms of flexibility, security, consistency... The traditional approaches for acquiring knowledge have become inadequate and ineffective. The Cloud Computing provides IT resources as a service, delivered on demand via a computer network including Internet. It reduces the cost of using materials and resources.

In many organizations when comes to knowledge management we can clearly notice that they mainly are facing these three problems obsolete technology, employee lack of motivation, information hard to be find.

The main goal of our work which is proposed in this thesis is the definition of the Knowledge Acquisition process and discuss the impact of Cloud Computing paradigm to achieve the process within an organization. The proposed approach is mainly based on knowledge engineering and Cloud Computing. We developed a Knowledge Acquisition System for tacit and explicit knowledge acquisition. Tacit knowledge Acquisition is based on existing Tacit Knowledge Acquisition methods and knowledge modeling. Explicit Knowledge Acquisition is based on the Knowledge Discovery in Databases (KDD) process and Data Mining (DM) technics as we focus on knowledge embedded in operational databases. We validate the proposed system through a case study.

We organized our thesis as following:

In the **first chapter** we introduce the concept of knowledge in the organization and the Organizational Knowledge Management domain.

The **Second chapter** defines the organizational knowledge acquisition process including its two sub-processes, tacit knowledge acquisition process, and explicit knowledge acquisition process. We analyze proposed tacit knowledge acquisition methods in the literature and discuss the applicability of the Cloud Computing services in the case of each method. The definition of explicit knowledge acquisition process is based on proposed KDD process models.

In the **third chapter** we present the Cloud Computing paradigm and definitions related to Knowledge Engineering.

The **fourth chapter** is related to the conception of our Knowledge Acquisition system and the definition of used knowledge models, using UML modeling language.

The **fifth chapter** presents the implementation of the proposed Knowledge Acquisition system and its validation using a Case study.

Chapter 1: Organizational Knowledge Management

1.1 Introduction

At this first chapter, we start by defining the three similar concepts (Data, information and Knowledge) and compare them, then we move to knowledge typologies with definitions and types. Last but not least, we introduce the knowledge management concept, objectives, approaches and its process.

1.2 Organizational Knowledge

1.2.1 Definitions

The word Knowledge means different things to different people. Some of them see Knowledge as a synonym with information and data. They seem to be the same and generally they are used as synonyms but they denote different concepts.

Data is the lowest form and refers to unstructured facts and figures, which lack any kind of organization.

Information refers to data that has been organized condensed Contextual life and so on the data now has a Direction and becomes a message with a purpose and meaning.

Knowledge '**mental evaluation'** based on experience and rooted in context according to Davenport and prusak [1] where they have defined it as a fluid mix of framed experience, values, contextual Information, expert insight that provides an environment and framework for evaluating and incorporating new experiences and Information.

	characteristic	**example**
Data	Uninterpreted raw	...-...
Information	Meaning attached to data	SOS
Knowledge	attach purpose and competence to information potential to generate action	emergency alert start rescue operation

Table 1.1: Distinctions between data, information, and knowledge [2].

In another way as shown in figure 1.1, **data** is formed of individual facts, which can be transformed to an information by structuring and organizing these facts. **Information** becomes individual **knowledge** when it is accepted and retained by an individual as being a proper understanding of what is true [3].

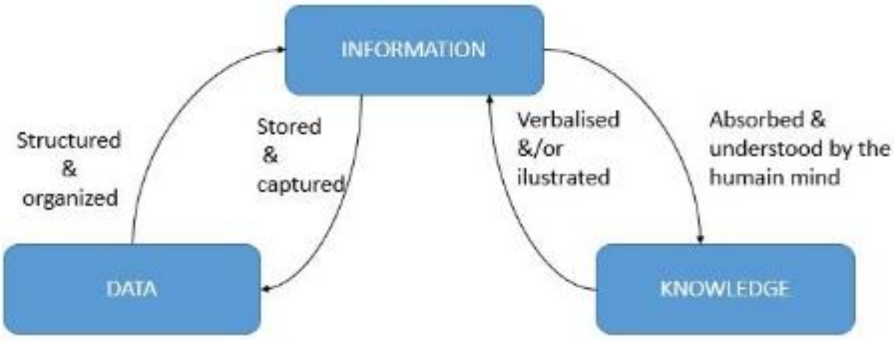


Figure 1.1: Relationship amongst knowledge, information and data. [148]

1.2.2 Organizational Knowledge typologies

Over decades many attempts have been made to classify Organizational knowledge.

According to Nonaka 1994 [4], There are two crucial types of knowledge. The first one is **tacit knowledge**. It cannot be expresses by words or formula also it cannot be articulated because it is experience based and personal in nature [5]. The second type is **explicit knowledge**, which is easy to articulate, identify and share because it is written-down in many formats (CD, Documents, Diagrams, Electronic, Documents ...), accessible (easy to attend) and mainly it can be stored.

Nonaka and Takeuchi (1995) have built a whole theory about knowledge and its creation, on the basis of this distinction between tacit and explicit knowledge. **Figure 1.3,** explains the four transformations of knowledge:

- **Socialization:** from tacit to tacit knowledge. This can happen by the discussion, explanation and understanding of an idea.
- **Externalization:** from tacit to explicit knowledge, report an idea by creating a declarative knowledge and write it in a document.
- **Combination:** from explicit to explicit knowledge. Transform the document in to another format to share it (e.g., electronic document), also combining different pieces of explicit knowledge.
- **Internalization:** from explicit to tacit knowledge. Performing a task frequently leads to a personal state where we can carry out a task successfully without thinking about it.

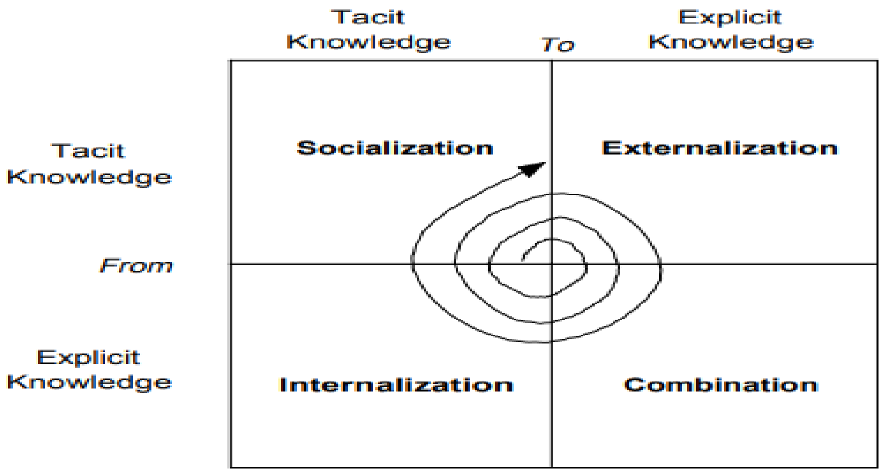


Figure 1.3: Nonaka's model of the dynamics of knowledge creation [5].

H.Zack **[6]** defined another knowledge typology that consists of:

- **Procedural knowledge (knowledge-how):** Represents the comprehension of how to execute a specific procedure. It also refers to the ability to perform a specific set of actions. It's reverberated on motor or manual skills and mental skills, we think, we play, we dance we read customer's faces and moods.
- **Declarative knowledge (knowledge-about):** It is about descriptions of facts, rules, law, methods and procedures, declarative knowledge has much in common with explicit knowledge.
- **Causal knowledge (knowledge-why):** It refers to know-when and know-why, concerning the moment and the context in which the procedure can be performed.
- **Strategic Knowledge:** refers to a reader's ability to formulate an appropriate purpose and/or logical goal for reading. It also refers to a reader's capacity to flexibly apply select strategies based upon the specific purpose for reading a text.

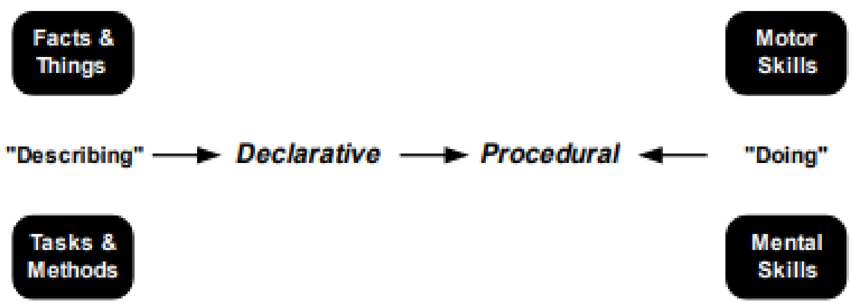


Figure 1.4: Declarative and procedural knowledge. [150]

According to Alquier [7], organizational knowledge include:

- **Collective knowledge:** represents normalized knowledge and skills contributed by individuals to an organization.
- **Departmental knowledge:** refers to an organization department, which contains a group of people with same objectives and same language.
- **Individual knowledge:** it is specific for a determined and decision-making workplace. It often has degrees of significant complexity and it comes from individual or small group (restricted number), where they are autonomous with the definition of their work.
- **Cooperative knowledge:** the quasi-decomposition between the collective, departmental and individual knowledge allowed the minimizing of exchanges between them, which are essential. They give back to the juxtaposition of the whole organization, the required integration to re institute the global knowledge system of the organization. The articulation of this knowledge is in their cooperation, in their work in coordination, in the linguistic communication.

[8] propose an organizational knowledge typology which includes:

- **Local knowledge:** which is important to accomplish a specific task by individual or group of individuals.
- **Product Knowledge:** that concerns a product across their entire lifecycle: documents associated to the product, skills used for product design, fabricate, sale and maintain. It includes also other important information, which are not articulate such as errors, modification reasons.
- **Organizational knowledge:** it is used by the organization direction to attain their objectives and their strategies to provide business value.

According to pomian [9], there is another important typology:

- **Descriptive knowledge:** take into consideration the description of themes and subjects of person's interest, group or even the organization itself. it allows to contextualize the knowledge related to the organization business area.
- **Deductive knowledge:** it linked to a logical thinking based on a general idea to reach a specific conclusion.it differs from domain knowledge by their degree of precision and dependence on the expert point of view.
- **Documentary knowledge:** represents articulated knowledge in documents, which are

produced or used by the organization.

The knowledge typology adopted in the rest of this thesis is the typology proposed by Nonaka [4] including tacit knowledge and explicit knowledge. It is largely the most used typology in KM researches.

1.3 Organizational Knowledge Management

1.3.1 Definitions

KM is managing the corporation's knowledge through a systematically and organizationally specified process for acquiring, organizing, sustaining, applying, sharing, and renewing both the tacit and explicit knowledge of employees to enhance organizational performance and create value **[128]**.

KM is a collaborative and integrated approach to the creation, capture, organization, access, and use of the enterprise's intellectual assets [**10**].

KM is the creation, extraction, transformation, and storage of the crucial knowledge and information in order to design better policy, modify action, and deliver results [**11**].

KM is the process of managing intellectual capital available in any organization, to achieve the organizations objectives [**12**]. It is the conscious process of defining, structuring, retaining and sharing the knowledge and experience of employees within an organization. The main goal of KM is to improve an organization's efficiency and save knowledge within the company **[11] [12]**.

1.3.2 Knowledge Management objectives

KM improve an organization in several ways. It will ensure that the specialized knowledge of employees does not leave with them when they retire or quit, or go unutilized by other employees who would benefit from that knowledge. It boosts the efficiency of an organization's decision making. We can summarize them in these 6 points:

- More efficient workplace.
- Faster, better decision making.
- Increased collaboration.
- Building organizational knowledge.
- Employee onboarding and training process is optimized.
- Increased employee happiness and retention, due to valuing of knowledge, training, and innovation.

Knowledge management is an important tool in any company that wants to increase its market.

[151]

The primary role of KM is to connect to "knowledge nodes" both the knowledge providers and the knowledge seekers. The knowledge of the mind of one provider may thus be ultimately transferred to the mind of someone who seeks that knowledge, so that a new decision can be made or situation handled. KM provides a means of capturing and storing knowledge and brokering it to the appropriate individual.

Based on a study of 31 KM projects in 24 companies, four business objectives that fulfill this primary role were identified in [13]:

- **Capturing Knowledge:** This goal can be achieved by creating KM repositories. Those will consist of structured documents with knowledge embedded in them (memos, reports, presentations, articles) stored in a way that they may be easily retrieved.

- **Improving Knowledge Access:** To facilitate the processes of knowledge transfer between individuals and between organizations.

- **Enhancing the Knowledge Environment:** by proactively facilitating and rewarding knowledge creation, transfer and use.

- **Managing Knowledge as an Asset:** some companies are including their intellectual capital in the balance sheet; others are leveraging their knowledge assets to generate new income from or to reduce costs with their patent base.

1.3.3 Knowledge Management approaches

Various classifications of KM approaches have been proposed in the literature. [14] Define two important approaches to KM:

- **The information-oriented approach:** this approach focuses on management improvement and information exchange, with trying to avoid organizational and professional borders. It frequently involves the use of information technologies to enhance the quality and speed of knowledge distribution in the organizations. These technologies may include intranets, data warehousing, knowledge repositories, decision support tools, and groupware. This approach took a part of a lot researches in certain domains such as organization theory and interaction homme machine. The focus of this approach is on building the social environments or communities of practice necessary to facilitate the sharing of tacit knowledge. It also allows the sharing of explicit knowledge using workflow or document management tools.

- **Oriented knowledge approach**: which is linked to researches in certain domains such as knowledge engineering. It is based on capitalization phase including knowledge identification and modeling [16].

[15] defined three fundamental approaches to KM:

- **Social and cooperative approach:** it is related to the study of structure interactions taking a place inside a group, in order to offer tools and structuring methods which permit better value implementation of exchanged knowledge and guaranteeing easy reuse.

- **Descendant approach:** which the domain knowledge modelization is in the middle of the conception of the knowledge management system. This modelization is used to map the domain knowledge, which needed to be capitalized. The system or the cogniticians interact with knowledge holders to extract information that they needed.

- **Ascendant approach:** which the domain knowledge modelization is in the middle of knowledge management process. We are not looking to structure the information based on hypothesis related to cooperatives activities of the actors but we are looking to identify the structure of the concepts and the domain reasoning. In this approach, we take in consideration the information's relatives to a domain (rapport, curriels...) to extract knowledge.

According to [17] there is an important two approaches to KM:

- **Codification approaches:** known as knowledge capitalization approaches. They focus on the transformation of implicit knowledge into the explicit form.
- **Personalization approaches:** aim at the exchange of knowledge through direct interactions between organization actors. They consider that the knowledge is linked to their holders besides that it is hard to be articulated.

Managerial approach refers to centralize the exchanges, knowledge sharing and the cooperatives behavior of the organization actors in the middle of KM process. The main goals of this approach are to focus on the objectives which consist technical aspects which is possible to associate the efficiency gains and the value enhancement. [18]

Technological approach is used a lot in the domain of knowledge engineering according to [151], which defined in this domain as the process of acquisition and stored the knowledge of all knowledge typologies founded in the organization then their distribution where it is ready to product benefits [19]. He proposed tools for storing knowledge, tools to traditional search engine for sharing knowledge and tools for extracting knowledge.

1.3.4 Knowledge Management process

As shown in the table 1.2, Many studies have addressed the KM processes. They divide KM into several processes and numerous KM process models have been proposed in the literature.

KM process Model	KM Sub processes				
[20]	knowledge creation	knowledge transfer	knowledge embedding		
[22]	knowledge goals, identification	acquisition	development,	distribution, preservation	use and measurement
[23]	knowledge creation or import	capture	retrieval or access	use	
[24]	knowledge relate/value (identify, verify, filter, select)	acquire (formalize, codify, represent, format, map)	organize (store, transform)	enable reuse (adapt, create)	transfer (share, distribute, forward, link to people) and use (apply, integrate, learn)
[25]	knowledge planning	creating knowledge	integrating knowledge	organizing knowledge and transferring knowledge	maintaining knowledge and assessing knowledge
[26]	knowledge discovery	capture	sharing and application		
[28]	learning opportunities (training available, technical expertise, and knowledge level)	knowledge acquisition	knowledge application	knowledge sharing	
[29]	knowledge development (acquisition, capture)	knowledge (re)combination (assembly, sharing, integration)	knowledge use (leverage, exploitation)		
[30]	knowledge storage	Knowledge retrieval			
[31]	knowledge	knowledge	knowledge		

	acquisition	sharing	utilization		

Table 1.2: Proposed KM process models

The KM process adopted in the rest of this thesis is the one proposed by [46] and presented in the Figure 1.5. It takes into consideration benefits and common points of the existing KM process models. The derived KM process includes the following sub processes: (1) Knowledge Spotting, (2) Knowledge Acquisition, (3) Knowledge Formalization, (4) Knowledge Sharing, (5) Knowledge Utilization, (6) Knowledge Update and (7) Organizational Environment related sub process. Each sub process includes knowledge activities.

Figure 1.5: The adopted KM process (adapted from [46])

1.4 Conclusion

By the end of this chapter, we introduced the distinction between data, information and knowledge besides the potential classification of the organizational knowledge (typologies), then we introduced Knowledge management concept, objectives, approaches and its process. Finally, we depict the adopted KM process in the rest of this thesis. In our work we focus on the Knowledge Acquisition process of the KM process which we define in the next chapter.

Chapter 2: Organizational Knowledge Acquisition

2.1 Introduction

In this chapter, we start by defining the organizational knowledge acquisition with the three different participant roles, and then we will define the knowledge acquisition process, furthermore we describe its two sub-processes including tacit knowledge acquisition process and explicit knowledge acquisition process.

2.2 Definitions

Knowledge acquisition is the process of extracting, structuring and organizing knowledge from knowledge sources, usually human experts, so it can be used in software such as an Expert System (ES). [154]

Knowledge acquisition is the process used to define the rules and ontologies required for a knowledge-based system. The phrase was first used in conjunction with expert systems to describe the initial tasks associated with developing an expert system, namely finding and interviewing domain experts and capturing their knowledge via rules, objects, and frame-based ontologies. [155]

The principal objective of knowledge acquisition is to transfer expertise. This objective aims at the establishment of methods and techniques to provide knowledge from experts, documents or manuals to computer system.[145][146][147]

Knowledge acquisition is not an easy task. It includes knowledge identification, knowledge representation in a proper format and knowledge transferring to a machine. The knowledge acquisition process can be greatly influenced by the roles of three major participants: **knowledge engineer**, **domain expert**, and the **end user**. The interrelationships between these participants are described in [66] and presented in Figure 2.1. The domain experts should take a very active role in the creation of a knowledge base. The knowledge engineer should act like a teacher of knowledge structuring, a tool designer, and a catalyst at the interface between the expert and the end users. This approach could minimize problems such as inter-human conflicts, knowledge engineering filtering, and end-user acceptance of the system. Also, knowledge maintenance problems can be reduced. [67] analyzed the participants roles involved in knowledge acquisition. They suggested that the participants play one or more roles, including acting as knowledge sources, agents, and targets for knowledge acquisition processes.

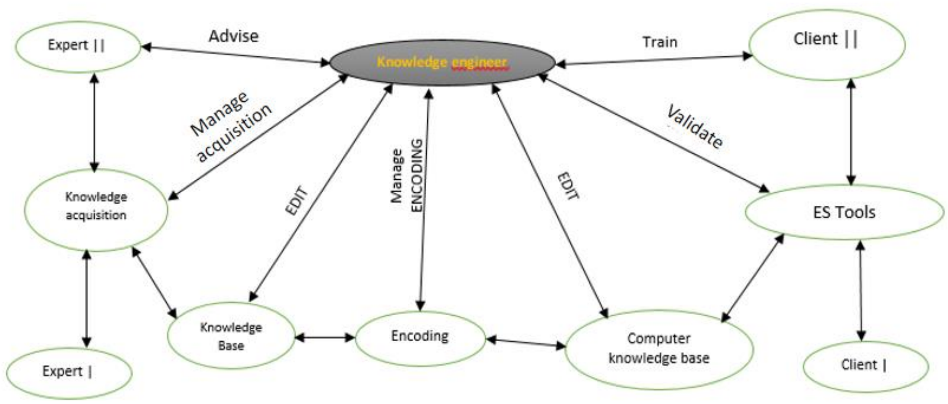


Figure 2.1: Knowledge Engineer's Roles in Interactive Knowledge Acquisition.[165]

2.3 Tacit knowledge Acquisition

We present in this section major tacit knowledge acquisition methods and technics found in the literature. The figure 2.2 presents a taxonomy of tacit knowledge acquisition methods and some particular techniques within each category.

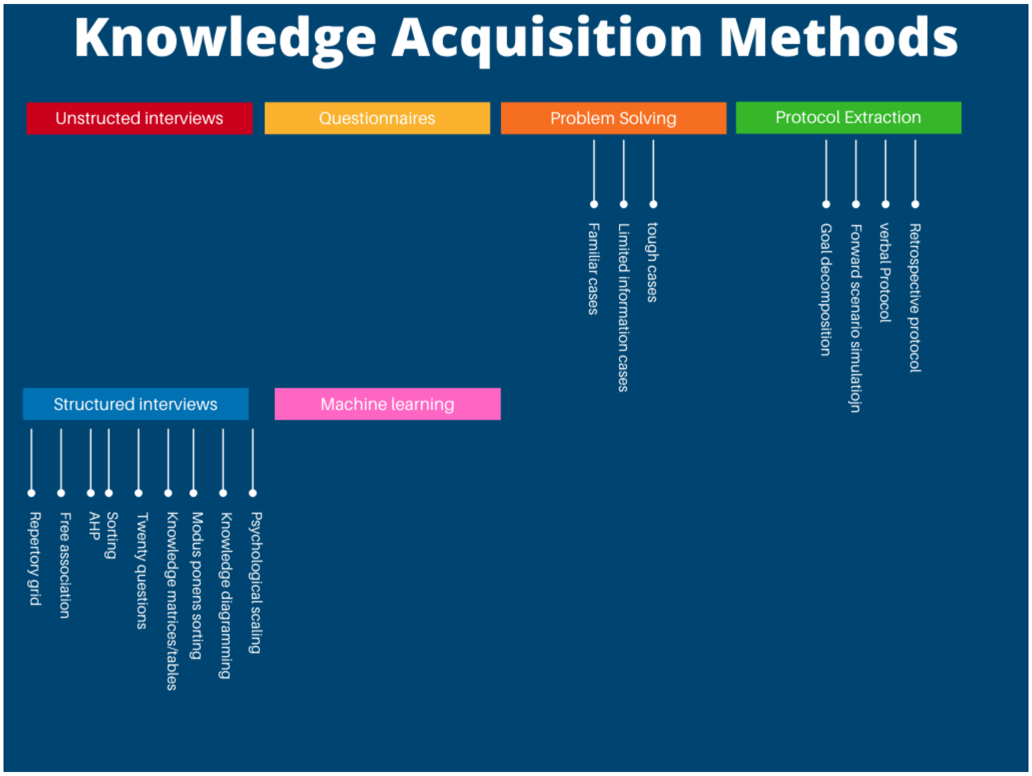


Figure 2.2: A taxonomy of Tacit knowledge Acquisition methods (Adapted from [68]).

2.3.1 Tacit knowledge Acquisition Manual methods

Manual methods are basically structured around an interview of some kind. The knowledge engineer elicits knowledge from the expert or other sources and then codes it in the knowledge base. The two major manual methods are interviewing, protocol Extracting, and observation problem solving. Manual methods are slow, expensive, and sometimes inaccurate.

2.3.1.1 Interviews

According to [68] Interviews are the most commonly used form of knowledge acquisition. It involves a direct dialog between the expert and the knowledge engineer. There are two basic types of interviews: unstructured (informal) interviews and structured interviews.

2.3.1.1.1 Unstructured Interviews

This kind of interviews are characterized by a lack of organization. The expert responds to the knowledge engineer's questions. Questions follow no preset or designed format and topics are pursued in whatever order or at whatever length seems best to the knowledge engineer. This type of interview may be useful during first phases of knowledge acquisition. Their lack of structure permits a sort of free association dialogue that may illuminate many of the major issues that are important that returns to unstructured interviews are more flexible as questions can be adapted and changed depending on the respondents' answers. The interview can deviate from the interview schedule, but they also have some drawbacks. They are by nature unstructured and, hence, may be very inefficient at collecting knowledge because of redundancies and omissions. To work effectively, the interviewer must be a very skilled communicator and have an ability to identify key issues through incisive questioning besides the ability to establish rapport and knowing when to probe. This level of interviewer talent implies, instead, some unconscious systematic methodology by the interviewer, which means that he or she may, in essence, be conducting a structured interview [68].

2.3.1.1.2 Structured interviews

A structured interview is a quantitative research method where the interviewer presents a set of prepared closed-ended questions in the form of an interview schedule, which he/she reads out exactly as worded. Within the general realm of structured interviews, a large number of methods may be utilized [68]. A fundamental intent of these methods is to provide the expert with guidance for his or her responses, thereby increasing interview efficiency. Explicit guidance allows an expert to focus on the subject matter rather than on how responses should be formatted. The following paragraphs briefly describe a number of these techniques.

a. Free association

Free association technic is suggested by [69]. In this technic several concepts in the subject area are posed to the expert and she/he will respond and react with other concepts, which are related to each of them. By applying this technic, the knowledge engineer will easily construct a matrix or graphical road map of terms and concepts in the subject area. Free association technic is based on theory named spreading activation theory, which is proposed by [69]. This theory consists that one thinks of a particular concept, all concepts that are associated to the particular concept can easily be activated. The Figure 2.3 presents a terminology of three defoliators of red pine which are linked with terms that are related to these insects. RPS, EPS, and RHPS refer to red pine sawfly, European pine sawfly, and redheaded pine sawfly, respectively. Associations among these concepts can then be represented graphically or in a tabular manner using a matrix.

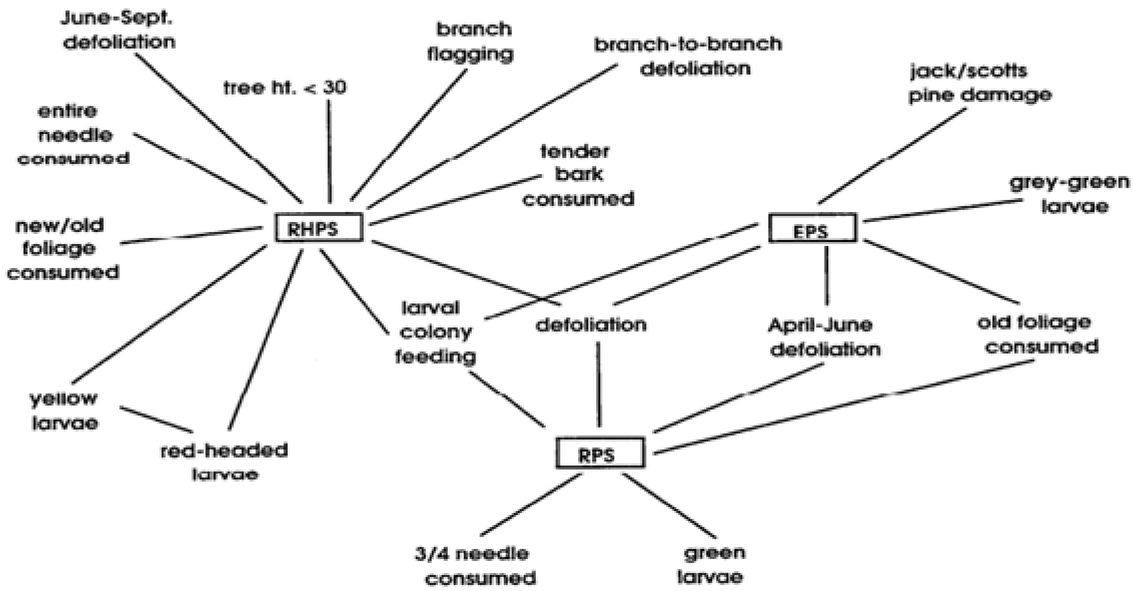


Figure 2.3: Terminology of Three defoliators of red pine.**[68]**

b. Psychological Scaling

In addition to associations among concepts in memory, there also exist strengths of association, which have relation to cognitive proximity. Therefore, the stronger an association is, the closer, or more related, two concepts are. From lecture above, we know that free association provides a matrix of related concepts but scaling methods has one more step further consists of having values from numerical scale to assign the relationships to indicate relatedness between those concepts. This association matrix called also distance matrix can be converted into a special representation by using multidimensional scaling [76], a hierarchical representation by using cluster analysis [77] or a network representation [78] of an expert's knowledge in a subject area.

As an example, the free association example presented in Figure 2.3 extended. It is converted into matrix with numerical values as shown in the figure 2.4. Ten concepts were selected from the free-association graph and values (0-9) have been entered for two rows of the matrix. Because the matrix is symmetrical, it can be displayed as upper or lower triangular Distances between concepts can be used to calculate clusters of similar concepts.

	green larvae	yellow larvae	red-headed	new/old foliage	old foliage	entire needle	branch flagging	grey-green larvae	bark consumed	3/4 needle consumed
green larvae	0									
yellow larvae		0								
red-headed			0							
new/old foliage				0						
old foliage					0					
entire needle						0				
branch flagging	8	3	3	2	8	2	0			
grey-green larvae	2	6	8	9	4	3	8	0		
bark consumed							1	9	0	
3/4 needle consumed							9	7		0

Figure 2.4: Example of a distance matrix. [68]

c. Anthropological Methods

In addition to psychology research, a number of useful elicitation techniques have also been developed in anthropology [79]. To understand the language and thought patterns of vastly dissimilar cultures, anthropologists have been forced to develop numerous methods to describe the acquisition of personal/cultural knowledge; Knowledge that is quite difficult for indigenous peoples to impart to others from outside their particular culture. Such cultural roadblocks are analogous to the situation of a domain expert and his or her subject matter specialty, which can also be viewed as a culture of sorts. Many of these methods are detailed in a book by [71]

d. Twenty questions

The knowledge engineer collects several task scenarios prior to the interview. Then, during the interview, the expert is instructed to ask questions about a particular task much like would transpire in an actual problem-solving situation. The expert also supplies rationale for asking each question. The questions that the expert asks and the rationale provided for that questioning can elicit decision factors, their relationships, and how those factors are used in problem solving. While the 20-question format is very natural for the expert, the knowledge engineer is forced to record session dialog, note the questions, and analyze the transcripts later for knowledge content that can help to determine key concepts, properties and can structure the domain [68].

e. Modus-Ponens Sorting

In this technic, an expert is asked to list important factors to making a decision in the subject area and also to list all possible outcomes (final decisions). Then an expert must connect, in the form of if-then rules, factors to each other, outcomes to each other, and factors to outcomes. This is a sorting method based on if-then relationships, hence the use of the term modus ponens sorting. This technique includes lexical analysis (lists of terms are created), syntactic analysis (factors are linked to each other), and semantic analysis (factors are linked to outcomes, i.e., decisions, via tactical rules). If the number of pertinent factors to each solution is small, it may be possible for the knowledge engineer to construct all “reasonable” combinations of factors for each solution and present these to an expert for critiquing. [90].

f. Knowledge Diagraming

Graphing domain concepts and the relationships between those concepts, as revealed by an expert, can be a useful visual aid. This knowledge diagram creates a sort of road map of an individual’s cognitive structures. Then, in subsequent interviews, the interviewer and expert have a record of where they have been and what topics may need to be expanded further [68]. Knowledge diagraming has also been used with question probes by [72], [73] and also by [81] as part of multi expert elicitation. Actual diagram creation is highly recommended even if the knowledge diagram is not utilized in successive interview sessions. The activity of creating a diagram helps the knowledge engineer visualize and better understand the domain’s knowledge structure.

Figure 2.3 can be expanded into a knowledge diagram by adding connecting phrases to the links and by expanding and grouping various concepts as shown in Figure 2.5. For example, the relationship between “larval colonies” and “defoliation” could include the connector, as in “larval colonies” “defoliation.” Also, the different types of needle consumption, e.g., “entire needle,” “¾ths of needle,” “branch to branch,” might be grouped under the concept “needle consumption.” Explicit labeling of terms and their relationships creates both lexical and syntactic knowledge [68]. A portion of the free-association graph of Figure 2.3 appears with different concepts connected by labeled links. This graph provides more structure and more detail than the free-association graph.

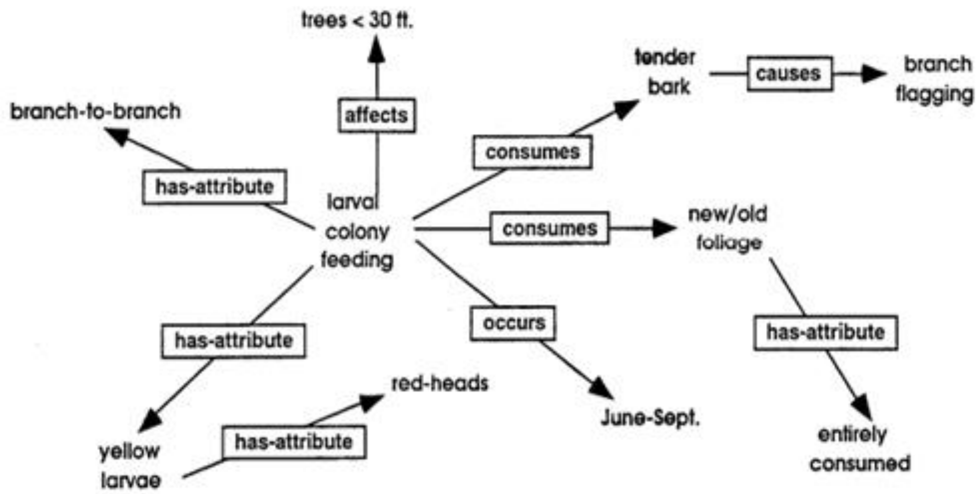


Figure 2.5: Example of Knowledge diagraming. [68]

g. Analytic Hierarchy Process

Another aid to the analysis of a decision process is the analytic hierarchy process developed by [73]. It allows persons with decision-making expertise to structure a complex problem in the form of a hierarchy. The process requires an ability to enumerate all possible decisions, i.e., alternative solutions, a priori. Then criteria are established to evaluate those decisions. Likewise, there may also be criteria to evaluate each of the previous criteria which forms a hierarchy. At each level, pair-wise comparisons are made regarding the relative likelihood, relative preference, or the relative importance of each criterion versus each of the other criteria at the same level. [68]

For the example presented in figure 2.6, erosion hazard rating, establishment likelihood, political/social impacts, and downstream values would be compared in a pairwise manner. Similarly, runoff potential and soil cover would be compared, as would on-site seed viability and historic success, etc. In the analytic hierarchy process, these comparisons can then be converted into numbers that represent each criterion's contribution to the overall decision. At each of the leaf nodes, a score is recorded by an expert for each possible decision with respect to the criteria immediately above that node.

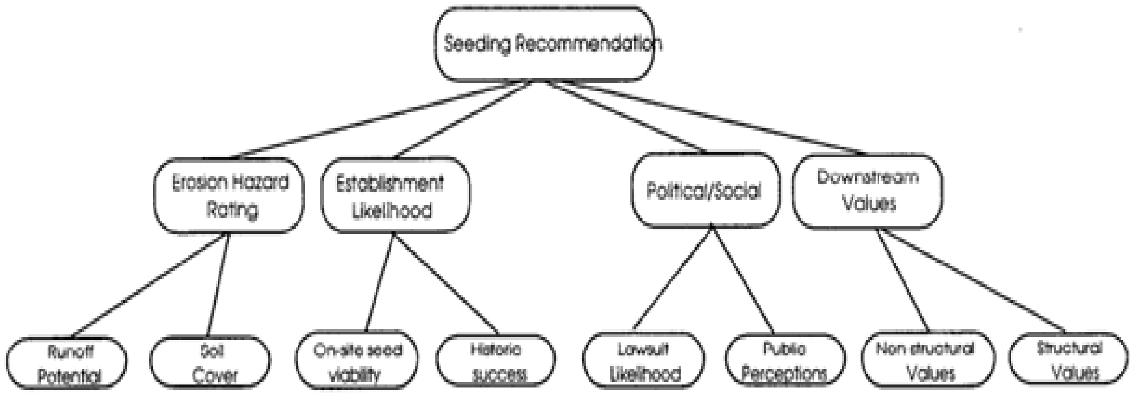


Figure 2.6: An analytic hierarchy process dissecting a decision process into a hierarchy of criteria. [68]

2.3.1.2 Questionnaires

Occasionally, it may be difficult to physically meet with an expert or it may be necessary to extract very detailed and specific knowledge about some topic. Questionnaires can be quite useful in both circumstances ([82],[83]). Also, when several experts are to be interviewed, the time and effort of the knowledge engineer can be reduced if questionnaires can be used in place of face-to-face meetings. This situation was encountered with the Knowledge Based Systems (KBS) development effort for seeding recommendations. The content of questionnaires may be very specific and require only short answers or they may be more general and intended to elicit longer prose.

2.3.1.3 Protocol Extraction

According to [68], Protocol Extraction method encouraging an expert to explicitly detail how either typical or specific problems are solved can elucidate many specifics of his or her decision-making process. Explication of an expert's problem-solving protocol identifies factors that are important, relationships among them, hypotheses that can be inferred, and strategies of how and when these factors are applied. A number of different protocol extraction methods have been utilized:

a. **Goal Decomposition** is one of the most basic methods of formalizing problem-solving strategies. An expert is asked to enumerate the steps to be followed (sub goals) as a problem is solved. A particular problem scenario maybe used as an example or general types of problems may be discussed more in an abstract manner. These sub goals can then be used in recursive goal decomposition until the sub goals become simple and readily accomplished tasks. This approach may be used in combination with knowledge diagraming or other methods presented above. [68]

b. **Forward scenario simulation** is almost identical to goal decomposition but prompts an expert for additional information, such as decision factors and explanations, in addition to sub goals. It is referred to as a simulation because an expert does not actually solve a problem scenario, but only describes how it might be done. Forward scenario simulation represents a very general protocol method because it elicits both decision factors and problem-solving strategies. [72] [84]

c. **Verbal protocol:** an expert is asked to solve a particular problem in the domain and to verbalize his or her rationale at the same time. Suggests the use of three different types of tasks for verbal protocol: (1) typical and familiar tasks, (2) limited information tasks, and (3) rare or tough cases. [85]

2.3.1.4 Observed Problem Solving

An expert may work differently when he or she is not required to justify problem-solving steps. To avoid overly self-conscious and, hence, unnatural decision-making situations, an expert maybe asked to solve problems without providing explanations. Observed problem solving may occur either in its natural environment, i.e., on the job, or in an artificial situation. Different types of tasks, i.e., familiar cases, limited information cases, and tough cases, may also be used with this method. Because problem-solving steps are not made explicit by the expert when using this knowledge acquisition method, the knowledge engineer must infer implicit strategies that are employed to solve various types of problems [72].

2.3.1.5 Other manual knowledge Acquisition methods

There is another manual knowledge acquisition methods:

a. **Case analysis:** Experts are asked how they handled specific cases in the past. Usually, this method involves analyzing documentation. In addition to the experts, other people (e.g., managers, users) may be questioned.

b. **Critical incident analysis:** In this approach, only selected cases are investigated; usually those that are memorable, difficult, or of special interest. Both experts and non-experts may be questioned.

c. **Discussions with users:** Even though users are not experts, they can be quite knowledgeable about some aspects of the problem. They can also indicate areas where they need help. The expert may be unaware of some of the users' needs.

2.3.2 Tacit knowledge Acquisition Semiautomatic methods

Knowledge acquisition can be supported by computer-based tools. These tools provide an environment in which knowledge engineers or experts can identify knowledge through an interactive process.

2.3.2.1 Repertory Grid Analysis

The repertory grid technique is presented in [86] according to a theory of personal constructs. This theory posits that each of us operates somewhat like a "personal scientist," i.e., we attempt to organize, predict, and control our own world by categorizing and classifying our experiences. This is similar to a scientist who develops and tests theories about the physical world. A repertory grid is a clinical technique developed to identify and analyze personal constructs, i.e., mental models. [87],[72] automated and applied this technique to the acquisition of expert knowledge.[87] process attempts to fill in a matrix, where each column corresponds to an element that is to be discriminated (i.e., final solutions), and each row represents a personal construct (decision factor) that differs across several of the column elements. Each construct has a diametric description, e.g., short/ tall or good/bad. Then each column element (solution) can be assigned a value for each construct (decision factor) that indicates the association between it and each diametric construct.

The repertory grid has the nice advantage of forcing an expert to create a tabular representation of his or her internal concepts about a subject area. Decision factors are enumerated and discrimination rules are created as the grid is constructed. Then the grid can be used for syntactic analysis of knowledge structure or for semantic analysis and actual rule construction [74]. [72] notes, however, that it is limited in application to declarative types of knowledge (declarative knowledge, in this case, includes tactical knowledge). Procedural, strategic, and causal knowledge are difficult to represent with this technique.

In the Figure 3.7, the repertory grid technique can effectively discriminate among a set of elements (solutions) based on various bipolar constructs. All the solutions are scored for each construct on some arbitrary scale. [68].

Elements			Constructs
RHPS	RPS	EPS	
2	5	4	new/old foliage
2	4	4	spring/summer
5	4	2	% needle consumption
3	1	1	no flagging/flagging

Figure 2.7: Example of repertory grid. [68]

2.3.2.2 Sorting

The sorting method uses a stack of cards, each with a term/concept from the domain written on it. The process of generating these terms constitutes factor knowledge analysis. An expert is instructed to sort the cards into different piles using whatever criteria seem appropriate. Following a sort, the expert is asked to provide a verbal description of the sort criteria used. When the cards contain only atomic decision factors and intermediate hypotheses, then the subject matter expert can group cards to indicate composite terms, i.e., additional intermediate hypotheses, or can group cards based on concept proximity. In this manner, some lexical knowledge analysis occurs after the original cards have been created, but in most cases the act of sorting produces syntactic knowledge only by using 3 cards trick, which consists first to pick three cards and identify which two cards are the most similar then create new term by writing down the similarity and create another new term by writing down the difference between the two previous cards and the 3rd one.

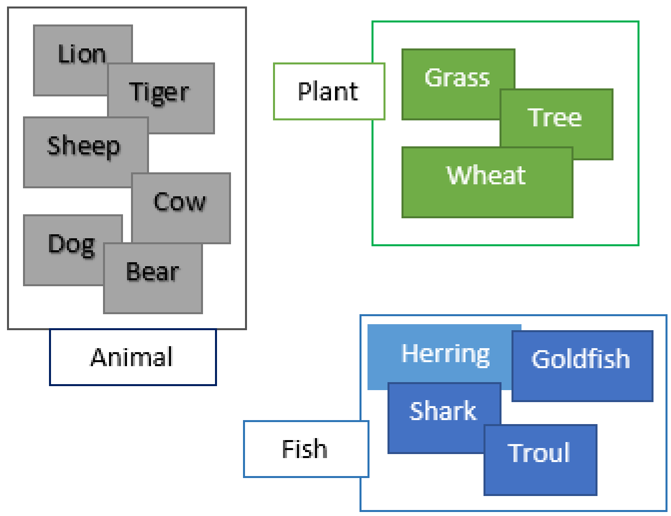


Figure 2.8: Example of several stacks of cards each with a concept.

2.3.3 Analyze of Tacit knowledge Acquisition methods

We discuss in the table below the advantages, disadvantages and the Cloud Computing Applicability for the tacit knowledge acquisition methods presented above.

Tacit Knowledge Acquisition Method	Manual / Semiautomatic	Advantages	Disadvantages	Cloud Computing Applicability
Unstructured (informal) interview	Manual	-Useful during initial stages of Knowledge acquisition. -Their lack of structure permits a sort of free association dialogue that may illuminate many of major issues that are important. -They are more flexible as questions can be adapted and changed depending on the respondents' answers. -knowledge engineer's questions follow no preset or designed format. -Topics are pursued in whatever order or at whatever length that seem best to the knowledge engineer.	-Very inefficient at collecting knowledge because of redundancies and omissions. -the interviewer must be a very skilled communicator and have an ability to identify key issues through incisive questioning besides the ability to establish rapport and knowing when to probe. -Possibility of not using expert's time efficiently.	Informal interview can be done using cloud-based virtual meeting
Free association (Structured	Manual	-The knowledge engineer will easily construct a matrix or		

interviews)		graphical road map of terms and concepts in the subject area.		As structured interviews are conducted around a set of prepared closed-ended questions in the form of an interview schedule, to guide expert in providing response, it is possible to use: -Cloud-based virtual meeting for expert interviewing. -Cloud services to create questionnaire for expert interrogating. -Cloud services to construct graphical representation of domain conceptual models, domain problems hierarchies and domain problems resolution technics once defined by expert.
Psychological Scaling (Structured interviews)	Manual	-Assign numerical values to relationships among concepts to indicate relatedness between those concepts. -The distance matrix generated can be converted into different representations such as: multidimensional scaling, hierarchical representation by using cluster analysis, etc. -Possibility to calculate clusters of similar concepts.		
Anthropological Methods (Structured interviews)	Manual	-Inspired from developed methods in anthropology that describe the acquisition of Knowledge that is quite difficult for indigenous peoples to impart to others from outside their particular culture.		
Twenty questions (Structured interviews)	Manual	-Help to determine key concepts and properties to structure a domain.	- the knowledge engineer is forced to record session dialog with domain expert, note the questions, and analyze the transcripts later.	
Modus ponens sorting (Structured interviews)	Manual	-It is a sorting method for a specific subject area. -It is based on if-then rules, connecting important factors to making decision, to each other, outcomes (final decisions) to each other, and factors to outcomes.		
Knowledge diagraming (Structured interviews)	Manual	-Creates a sort of road map of an individual's cognitive structures. -Provides more structure and more detail than the free-association graph. -Explicit labeling of terms and their relationships creates both		

		lexical and syntactic knowledge. -Can be used as part of multi expert elicitation. -Helps the knowledge engineer to visualize and better understand the domain's knowledge structure.		
Analytic Hierarchy Process (Structured interviews)		-Allows persons with decision making expertise to structure a complex problem in the form of a hierarchy.		
Questionnaires	Manual	-Can be used in place of face-to-face meetings. -Can be used in place of face-to-face meetings. -Useful then when it is difficult to physically meet with an expert. -Can reduce the time and effort of the knowledge engineer when several experts are to be interviewed.		Using cloud services to create questionnaire for expert interrogating
Goal Decomposition (Protocol Extraction)	Manual	-The sub goals or steps to be followed for a problem solving are recursively decomposed until they become simple and readily accomplished tasks. -May be used in combination with knowledge diagraming or other methods.		-Using cloud-based virtual meeting for expert interviewing. -Cloud services to create questionnaire around problem solving steps and their decomposition, decisions factors and explications related to problem-solving strategies.
Forward scenario simulation (Protocol Extraction)	Manual	-Describes how a problem solving might be done. -Compared to Goal Decomposition, it prompts an expert for additional information, such as decision factors and explanations. -It elicits both decision factors and problem-solving strategies.		
Verbal protocol	Manual	-Asks an expert to verbalize	-The expert is	Cloud messaging service

(Protocol Extraction)		his or her rationale at the same time when solving a particular domain problem using three types of tasks: typical and familiar tasks, limited information tasks, and rare or tough cases.	required to justify problem-solving steps while working.	allowing experts to verbalize their problem-solving strategies.
Observed Problem Solving	Manual	-Asks expert to solve problem without providing explanations. -It avoids overly self-conscious and, hence, unnatural decision-making situations.	-The knowledge engineer must infer implicit strategies that are employed to solve various types of problems as problem-solving steps are not made explicit by expert.	As observed problem solving may occur either in its natural environment or in an artificial situation, it may rely on cloud-based video conferencing.
Repertory Grid Analysis	Semiautomatic	-Forces an expert to create a tabular representation of his or her internal concepts about a subject area. -Decision factors are enumerated and discrimination rules are created as the grid is constructed. -The grid can be used for syntactic analysis of knowledge structure or for semantic analysis and actual rule construction.		-Using cloud-based virtual meeting for expert interviewing. -Cloud services to construct graphical representation of domain conceptual models once defined by expert.
Sorting	Semiautomatic	-Uses a stack of cards, each with a term/concept from the domain written on it. -Expert can group cards based on concept proximity or to indicate composite terms. -Expert is asked to provide a verbal description of the used sort criteria.		-Using cloud-based virtual meeting for expert interviewing. -Cloud messaging service allowing communication among knowledge engineering and experts, and providing a verbal description of the sort criteria by experts.

Table 2.1 Analyze of Tacit knowledge Acquisition methods.

As shown in the figure below, based on the analyze of different Tacit knowledge acquisition methods proposed in the literature, we define the principal tasks occurring in the Tacit knowledge acquisition process and its main actors.

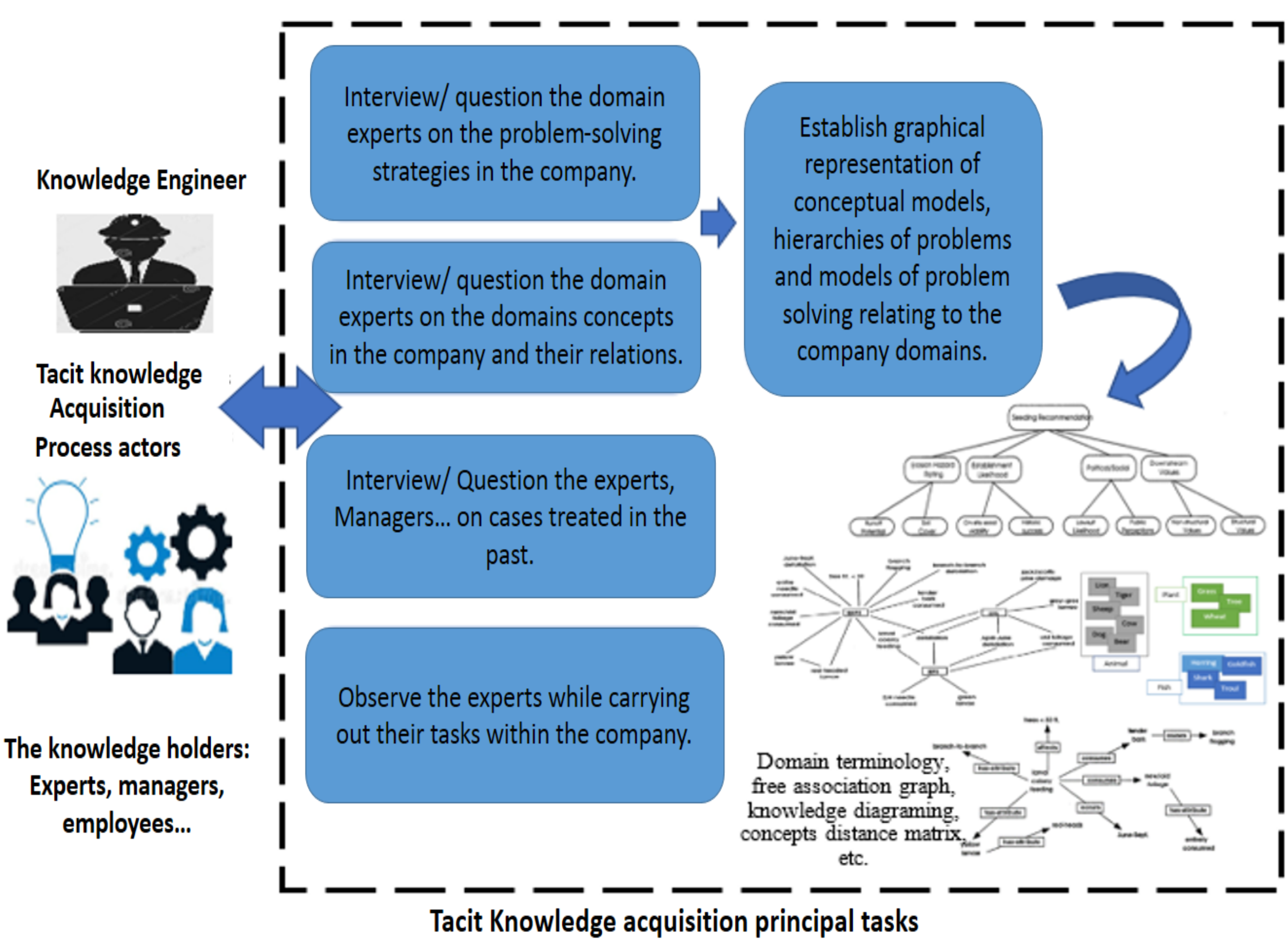


Figure 2.9: Tacit Knowledge Acquisition Process

2.4 Explicit knowledge Acquisition

These days the organizations are facing a surcharge in the amount of data they possess, including structured data not yet explored, knowledge that is "hidden" in the data and information that cannot be "discovered" by traditional means or by using just the human factors. This is where data mining and text mining comes in the play. [166]

Data mining sometimes called knowledge discovery in databases (KDD) has been defined as the automatic analysis of large and complex data sets. It is one of the most recent technologies for data analysis. Data Mining main objective is the development of new significant patterns or trends as without using this technique could remain unidentified. Because of recent research in Data Mining have developed more efficient methods for finding these new patterns, as well as huge volumes of data into knowledge based on effective methods of classification to clustering, analysis of frequent patterns, sequential or structural. Text Mining, also known as text data mining, is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights. [166]

The following sections are related to the acquisition of organizational knowledge embedded in data bases.

2.4.1 Data Mining and Knowledge Discovery in databases (KDD)

Confusion remains to this day between the terms Data Mining, and (KDD). Indeed, for many researchers and practitioners, the term DM is used as a synonym for KDD in addition to being used to describe one of the steps in the KDD process [141]. KDD is a non-trivial process that identifies, in data, patterns that are ultimately understandable, valid, new and potentially useful [142].

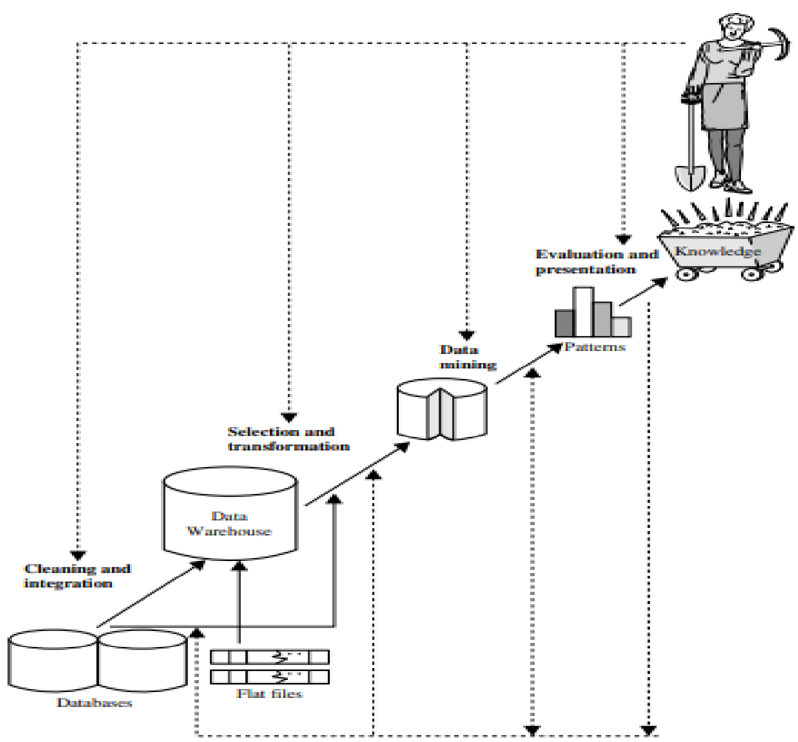


Figure 2.10: Data mining as a step in the process of knowledge discovery. [119]

Datamining is the process of applying computational methods to large amounts of data in order to reveal new non-trivial and relevant information. Data mining is not only used for finding interesting patterns from the data but also for exploring large data sets, for building models that describe the relevant properties of data, and for making predictions based on the data [122].

[152] They believe that "data mining allows analysts and store managers to find the answers to company data, which they have not even put". Data mining has been described as "the science of extracting useful information from large volumes of data and database". Data mining in relation to economic resources planning is the statistical analysis and logical data volumes of transactions, looking for patterns that can help decision-making process. The data mining process is extracting means from knowledge bases or data warehouses, knowledge previously unknown, valid and operational at the same time. Data mining seeks not only verify the hypotheses, but aims at discovering new knowledge, information totally unknown until then. Thus, the results are very valuable.

2.4.2 KDD process models

The KDD process aims to transform data (large, multifaceted, stored in different formats on media that can be distributed) in knowledge. This knowledge can be expressed in the form of general concepts that enrich the field semantics of the user in relation to a question that concerns him. It can take the form a report or graph. They can express themselves as a mathematical or logical model for decision making. The knowledge extracted must be as intelligible as possible for the user. They must be validated, formatted and arranged [153].

The main motivation for formally structuring KDD as a process is an observation of the problems linked to a blind application of data mining methods on the data in hand. This activity, rightly criticized by statisticians, is called "Data dredging” can lead to the discovery of unnecessary patterns [142] Therefore, the main objective to define and implement process models for KDD is to ensure that the end result is useful to the user [142].

Another important factor, and one which is often underestimated by KDD researchers, is to support multiple actors in an KDD process (management problem). Indeed, in most cases, data analysis projects involve several experts (domain experts, data mining experts, database experts, etc.) who work together, and therefore require careful collaboration and planning of tasks. This arouses the need to define models and methodologies for the field of KDD.

The proposed categories for Knowledge Discovery Process (KDP) modeling include:

- **Traditional KDP Approach:** This approach is widely used by most of KDP modeling innovators. Starting with [142] KDD process modeling, many of KDP modeling used the same process flow including most of the following steps: business understanding, data understanding, data processing, data mining/modeling, model evaluation, and deployment/visualization.

- **Ontology-based KDP Approach:** This approach is the integration of ontology engineering and traditional KDP approach steps. Three directions were identified in this approach: Ontology for KDP, KDP for Ontology, and the integration of both previous directions [156].

- **Web-based KDP Approach:** This approach mainly deals with web log analysis. It is mainly similar to traditional KDP approach, but it has some unique steps to deal with log web data [157],[158]

- **Agile-based KDP Approach:** This approach is the integration between agile methodologies and KDP traditional methodologies [159].

We review in the Table 2.1 different proposals for KDD process models.

KDD Model	KDD Approach	Number of Steps	KDD Model Steps
FAYYAD ET AL. (1996)	Traditional KDD Approach	**5**	Data selection, Data preprocessing, Data Transformation, Data Mining, Interpretation/Evaluation
Adriaans and Zantinge 1996	Traditional KDD Approach	**6**	Data Selection, Cleaning, Enrichment, Coding, Data Mining, reporting
(Berry and Gordon 1997)	Traditional KDD Approach	**4**	Identifying the Problem, Analyzing the Problem, Taking Action, Measuring the Outcome
SEMMA BY SAS INSTITUTE (1997)	Traditional KDD Approach	**5**	Sample, Explore, Modify, Model, Assess
Feldens et al. (1998)	Traditional KDD Approach	**3**	Pre-Processing, Data Mining, Post-Processing
Cabena et al. (1998)	Traditional KDD Approach	**5**	Business Objectives Determination, Data Preparation, Data Mining, Domain Knowledge Elicitation, Assimilation of Knowledge

Edelstein (1998)	Traditional KDD Approach	**5**	Identifying the Problem, Preparing the Data, Building the Model, Using the Model, Monitoring the Model.
LEE and KERSCHBERG (1998)	Traditional KDD Approach	**6**	Plan for learning, generate and test hypothesis, discover knowledge, determine knowledge relevancy, evolve knowledge/data, critique by a panel of experts
COLLIER ET AL. (1998)	Traditional KDD Approach	**8**	Define the objectives, Select the relevant business data, Data quality analysis, Clean and transform data, Data Mining, Acquire knowledge, Evaluate results, Deploy results or reiterate
Buchner et al. (1999)	Web-based KDD Approach	**8**	Human Resource Identification, Problem Specification, Data Prospecting, Domain Knowledge Elicitation, Methodology Identification, Data Preprocessing, Pattern Discovery, Knowledge Post-processing
Reinartz (1999)	Traditional KDD Approach	**7**	Business Understanding, Data Understanding, Data Preparation, Data Exploration, Data Mining, Evaluation, Deployment.
Kopanakis and Theodoulidis (1999)	Traditional KDD Approach	**6**	Scrub, verify and summarize data, Selection of the training data sample, Model derivation algorithm, Verify and evaluate, Selection of most interesting models, Model usage + population shift monitoring and incremental learning.
Han and Cercone (2000)	Traditional KDD Approach	**5**	Original Data Visualization, Data Reduction, Data Preprocess, Pattern Discovery, Pattern Visualization.
Cios et al. (2000)	Traditional KDD Approach	**6**	Understanding the Problem Domain, Understanding the Data, Preparation of the Data, DM, Evaluation of the Discovered Knowledge, Using the Discovered Knowledge.
CRISP-DM (2000)	Traditional KDD Approach	**6**	Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment
Han and Kamber (2001)	Traditional KDD Approach	**9**	Learning the Application domain, Creating a Target Data Set, Data Cleaning and Preprocessing, Data Reduction and Transformation, Choosing Functions of Data Mining, Choosing the Mining Algorithm(s), Data Mining, Pattern Evaluation and Knowledge Presentation, Use of Discovered Knowledge
Klosgen and Zytkow (2002)	Traditional KDD Approach	**7**	Definition and Analysis of Business Problems, Understanding and Preparation of Data, Setup of the Search for Knowledge, Search for Knowledge, Knowledge Refinement, Application of Knowledge in Solving the Business Problems, Deployment and Practical, Evaluation of the Solutions

Hofmann (2003)	Traditional KDP Approach	**9**	Hypotheses/Objectives preparation stage (consists of business understanding, data understanding and hypotheses/objectives definition steps), Data preparation stage (Consists of select/ sample data, pre-process and transformation steps), Discovery and validation stage (Consists of Data Mining, evaluation and deployment steps)
Haglin et al. (2005)	Traditional KDD Approach	**7**	Goal Identification, Target Data Creation, Data Preprocessing, Data Transformation, Data Mining, Evaluation and Interpretation, Take Action steps.
Pabarskaite and Raudys (2007)	Web-based KDD Approach	**9**	Data collection, Data cleaning, User identification, Session identification, Feature selection, Data transformation, Data combination, Mining the data, Result visualization.
Li and Ruan (2007)	Traditional KDD Approach	**6**	Data Collection, Selection, Preprocessing, Transformation, Data Mining, Interpretation/Evaluation.
Gottgtroy (2007)	Ontology-based KDP Approach	**5**	Ontology Preparation, Ontology Analysis, Instance Preparation, Modeling, Evaluation
Rennolls and AL-Shawabkeh (2008)	Ontology-based KDD Approach	**5**	Data Collection and Processing, Data Understanding, Data Mining/Modeling, Knowledge Understanding, Business Understanding
Alnoukari et al. (2008)	Agile-based KDP Approach	**6**	Speculation (Includes business and data understanding, and data preparations including ETL (Extraction, Transformation, Load) operations), Collaboration, Learning (Testing and evaluating)

Table 2.2: KDD process Models.

2.4.3 Different tasks of KDD

The task represents the goal, or objective, of an KDD process. According to [142], in practice, we can distinguish two high-level primary tasks of datamining: **prediction** and **description**. Prediction consists of using variables or fields in the database to predict future or unknown values of other variables of interest. While the description focuses on finding patterns (patterns, diagrams or rules) describing the data and interpretable by the user. Although the limits between the prediction and description are not clear (indeed a predictive model can be descriptive, as it is understandable and gives an idea of the data, and vice versa), the distinction between these two tasks is useful for understanding the overall objective of the KDD process.

Prediction and description tasks can be performed using a wide variety of data mining methods. In the sections bellow we present a brief overview of some of these methods as well as the dependencies between them:

a. **Data description and summarization:** It is a task purely descriptive with the aim of concise identification of the characteristics of data, usually in elementary and aggregate form. It thus allows an analyst to have a synthetic understanding of all of your data [164]. Data description and synthesis techniques are often used for interactive exploration and analysis of data and for generation automatic reporting [142].

b. **Clustering:** It is a common descriptive task where one seeks to identify a finite set of categories or clusters to describe the data [160]. The members of each cluster share some significant characteristics. The categories can be mutually exclusive and exhaustive, or they consist of a richer representation such as hierarchical or overlapping categories. Clustering alone can be the task of an KDD process. Thus, the detection of clusters would be the main focus of data mining. However, clustering is often an intermediate step for carrying out other KDD tasks. In that case, the objective of clustering may be to keep data to a reasonable size or to detect homogeneous subsets of data that are easier to analyze. [164]

c. **Classification:** Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analysed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. [164]

d. **Regression:** Regression technique can be adapted for predication. It is similar to the classification. The only difference is that in a regression problem the target attribute (the class) is not a discrete but continuous qualitative attribute. Thus, the regression aims to make explicit a linear or non-linear relationship between a set of so-called explanatory variables (or exogenous) and a continuous real variable called dependent or to be explained (or endogenous). Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict.

Unfortunately, many real-world problems are not simply prediction. The same model types can often be used for both regression and classification. [164]

e. **Dependency analysis:** It consists of finding a model that describes the dependencies important (or associations) between data elements or events. The dependencies can be used to predict the value of a new data item knowing information on other items. Although dependencies are useful for predictive modeling, they are mainly used for the description and understanding the data. Dependencies can be strict or probabilistic. Dependency analysis has strong links with classification and clustering. Indeed, dependencies can be implicitly used for building predictive models. Clustering is sometimes used to detect homogeneous data segments on which it is more relevant to perform a dependency analysis. [164]

f. **Neural network:** Neural network is a set of connected input/output units and each connection has a weight present with it. During the learning phase, network learns by adjusting weights so as to be able to predict the correct class labels of the input tuples. Neural networks have the remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. Neural networks too can create both classification and regression models. [164]

2.4.4 Some KDD tools

In this section we present a comparative analysis of 13 tools for KDD. The choice of these tools is based on the fact that they suit the KDD task and that they are the most used by the community according to [161], [162],[163] and some Google search also confirms that these 13 tools are popular (mainly due to their long history of developing and updating level)

KDD Tool	Description	Features
Weka	Weka (Waikato Environment for Knowledge Analysis)is a workspace composed a collection of data mining algorithms and preprocessing and visualization tools(Witten et al., 2011), developed entirely in Java at the University of Waikato, New Zealand, and available under the GNU General Public License (GPL). Weka has a graphical interface that guides the user through the various	- platform independent - open source and free - different machine learning algorithms for datamining - easy to use - data processing tools - flexibility for scripting experiments - 3 graphical user interface

	exploration tasks of data. Algorithms can be applied directly from this interface graphical, as they can be called as API from a Java program. Weka contains preprocessing tools (especially data transformation filters),a set of methods to perform standard data mining tasks (classification, regression, clustering, extraction of association rules, and selection attributes), and tools for viewing the input data and extracted patterns.	
R-project	R (R Project for Statistical Computing)is a free software language and environment allowing you to perform statistical calculations and create graphs. R is similar to S language and its environment created at Bell Laboratories by John Chambers and his colleagues (Becker and Chambers, 1984). R can be considered as another implementation of S. There are some important differences, but a lot of code written for S is executed without modification under R. Although R is often associated with statisticians, it is actually very suitable for data mining. Indeed, R offers a wide variety of statistical and mining methods. Of data: linear and non-linear modeling, statistical tests, series analysis chronological, classification, clustering, dependency analysis, and techniques of graphic visualization. Among the major strengths of R, we cite its extensibility through packages and programs developed by the community, and the ease with which one can create graphics well designed containing mathematical symbols and formulas as needed. Only drawback, it works using a command interpreter, which requires some practice to really take advantage of it.	-open source -Strong Graphical Capabilities -Highly Active Community -A Wide Selection of Packages -Comprehensive Environment -Can Perform Complex Statistical Calculations -Distributed Computing -Running Code Without a Compiler -Interfacing with Databases -Data Variety -Machine Learning -Data Wrangling -Cross-platform Support -Compatible with Other Programming Languages -Data Handling and Storage -Vector Arithmetic -Generates Report in any Desired Format
RapidMiner	RapidMiner, successor of the YALE software (Yet Another Learning Environment) (Mierswa andal., 2006), is undoubtedly the leader in open source tools for data mining and predictive analytics since 2010 (according to the annual survey conducted by KDnuggets). RapidMiner has more than 1000 methods and techniques for: data mining(classification, clustering, associations ...), predictive analysis, ETL, reporting, generation of dashboards, visualization, evaluation and deployment. In particular it integrates all the algorithms and techniques of ECD available under Weka and R. RapidMiner has a graphical interface for designing EDC analyzes in the form of pipelines (also called an operator tree). In addition to the visual appearance of the process, an XML file is generated describing the steps and operators applied to the data. RapidMiner algorithms can also be run	-Easy to use visual environment for building analytics processes: -Graphical design environment makes it simple and fast to design better models -Visual representation with Annotations facilitates collaboration among all stakeholders -Every analysis is a process, each transformation or analysis step is an operator, making design fast, easy to understand, and fully reusable Guided process design leveraging the Wisdom of Crowds, i.e. the knowledge and the best practices of more than 200,000 users in the RapidMiner community -Operator recommender suggesting

	from the line command or from a program as an external API. RapidMiner is developed in Java and distributed in open source under the GNU AGPL license.	next steps -Parameter recommender indicating which parameters to change & to which values -Convenient set of data exploration tools and intuitive visualizations -Share your feedback within the product from the tutorials or help panel -More than 1500 operators for all tasks of data transformation and analysis -Support for scripting environments like R, or Groovy for ultimate extensibility -Seamlessly access and use of algorithms from H2O, Weka and other third-party libraries -Transparent integration with RapidMiner Server to automate processes for data transformation, model building, scoring and integration with other applications -Extensible through open platform APIs and a Marketplace with additional functionality -Powerful Global Search sifts through repositories to quickly retrieve anything, including processes, models, operators, extensions and even UI actions
Knime	Knime (Konstanz Information Miner) is an open source platform for integrating data, preprocessing, analysis and data mining, developed in Java at the University of Konstanz. It integrates different components for KDD and has a graphical interface for designing KDD processes as a pipeline. One of the keys to the success of the Knime tool is its modular approach to design and management an execution plan, which consists of documenting and recording the analysis process in the order in which it was designed and implemented, while ensuring that the intermediate results are still available. Knime also offers some reporting features. It integrates all analysis modules of Weka and allows you to generate scripts in R language.	-Scalability through sophisticated data handling (intelligent automatic caching of data in the background while maximizing throughput performance) -High, simple extensibility via a well-defined API for plugin extensions -Intuitive user interface -Import/export of workflows (for exchanging with other KNIME users) -Parallel execution on multi-core systems -Command line version for "headless" batch executions
Orange	Orange is an open source data visualization, machine learning and data mining toolkit. It features a visual programming front-end for exploratory data analysis and interactive data visualization. Orange is a component-based visual programming software package for data visualization, machine learning, data mining and data analysis. Orange components	-Show a data table and select features -Read the data -Compare learning algorithms and train predictors -Visualize data elements

	are called widgets and they range from simple data visualization, subset selection and pre-processing, to evaluation of learning algorithms and predictive modeling. Visual programming in orange is performed through an interface in which workflows are created by linking predefined or user-designed widgets, while advanced users can use Orange as a Python library for data manipulation and widget alteration.	
Python	Available as a free and open source language, Python is most often compared to R for ease of use. Unlike R, Python's learning curve tends to be so short that it becomes easy to use. Many users find that they can start building datasets and doing extremely complex affinity analysis in minutes. The most common business-use case-data visualizations are straightforward as long as you are comfortable with basic programming concepts like variables, data types, functions, conditionals and loops.	-Easy to Code and to read -free and open source - Robust Standard Library - Data exploration & analysis (Included here: Pandas; NumPy; SciPy; a helping hand from Python's Standard Library) - Data visualization (Included here: Matplotlib; Seaborn; Datashader) -Classical machine learning(Included here: Scikit-Learn, StatsModels) -Deep learning (Included here: Keras, TensorFlow, and a whole host of others.) -Data storage and big data frameworks' Apache Spark; Apache Hadoop; HDFS; Dask; h5py/pytables.)
Xplenty	provides a platform that has functionalities to integrate, process, and prepare data for analytics. Businesses will be able to make most of the opportunities offered by big data with the help of Xplenty and that too without investing in related personnel, hardware, and software. It is a complete toolkit for building data pipelines.	-Xplenty provides you with an visual, intuitive interface to design your ETL data flows. -Integrate semi-structured data with structured data. Our package designer makes it a snap for every data and BI user to write complex data flows for your flat files and JSON files on top of Hadoop without writing a single line of code. - Integrate semi-structured data with structured data. Our package designer makes it a snap for every data and BI user to write complex data flows for your flat files and JSON files on top of Hadoop without writing a single line of code.
Sisense	Sisense is extremely useful and best suited BI software when it comes to reporting purposes within the organization. It is developed by the company of same name 'Sisense'. It has a brilliant capability to handle and process data for the small scale/large	-Ad-hoc analysis of high-volume data - Handles data at scale on a single commodity server -Complex business queries without programming or SQL writing

	scale organizations.	-Integrates with external websites or web applications
Oracle Data Mining	It is a representative of the Oracle's Advanced Analytics Database. Market leading companies use it to maximize the potential of their data to make accurate predictions. The system works with a powerful data algorithm to target best customers. Also, it identifies both anomalies and cross-selling opportunities and enables users to apply a different predictive model based on their need. Further, it customizes customer profiles in the desired way.	- Option to Oracle Database Enterprise Edition. -In-Database analytics - Wide range of algorithms - Automatic data preparation - Score models at storage layer on Exadata -Easy to use GUI - Text mining - PL/SQL and Java APIs
Apache Mahout	Apache Mahout(TM) is a distributed linear algebra framework and mathematically expressive Scala DSL designed to let mathematicians, statisticians, and data scientists quickly implement their own algorithms. Apache Spark is the recommended out-of-the-box distributed back-end, or can be extended to other distributed backends.	-The algorithms of Mahout are written on top of Hadoop, so it works well in distributed environment. Mahout uses the Apache Hadoop library to scale effectively in the cloud. -Mahout offers the coder a ready-to-use framework for doing data mining tasks on large volumes of data. -Mahout lets applications to analyze large sets of data effectively and in quick time. -Includes several MapReduce enabled clustering implementations such as k-means, fuzzy k-means, Canopy, Dirichlet, and Mean-Shift. -Supports Distributed Naive Bayes and Complementary Naive Bayes classification implementations. - -Includes matrix and vector libraries
SSDT (SQL Server Data Tools)	SSDT is a universal, declarative model that expands all the phases of database development in the Visual Studio IDE. BIDS was the former environment developed by Microsoft to do data analysis and provide business intelligence solutions. Developers use SSDT transact- a design capability of SQL, to build, maintain, debug and refactor SQL Server Data Tools (SSDT) provides project templates and design surfaces for building SQL Server content types - relational databases, Analysis Services models, Reporting Services reports, and Integration Services packages.	
IBM SPSS Modeler	IBM SPSS is a software suite owned by IBM that is used for data mining & text analytics to build predictive models. It was originally produced by SPSS Inc. and later on acquired by IBM. SPSS Modeler has a visual interface that allows users to work with data mining algorithms without	- Support for many data sources. - Automatic data preparation. - A range of algorithmic methods. - Easy model deployment - Automated modeling

	the need of programming. It eliminates the unnecessary complexities faced during data transformations and to make easy to use predictive models.	
SAS Data Mining	Statistical Analysis System (SAS) is a product of SAS Institute developed for analytics & data management. SAS can mine data, alter it, manage data from different sources and perform statistical analysis. It provides a graphical UI for non-technical users.	-Interactive programming in a web-based development environment. - Embedded support for Python & R languages. -Model development with modern machine learning algorithms.
SPARK APACHE	**Apache Spark** is an open-source unified analytics engine for large-scale data processing. Spark provides an interface for programming entire clusters with implicit data parallelism and fault tolerance. Originally developed at the University of California, Berkeley's AMPLab, the Spark codebase was later donated to the Apache Software Foundation, which has maintained it since.	-Speed: Run workloads 100x faster.Apache Spark achieves high performance for both batch and streaming data, using a state-of-the-art DAG scheduler, a query optimizer, and a physical execution engine. - Ease of Use: Write applications quickly in Java, Scala, Python, R, and SQL. Spark offers over 80 high-level operators that make it easy to build parallel apps. And you can use it *interactively* from the Scala, Python, R, and SQL shells. - Generality: Combine SQL, streaming, and complex analytics. Spark powers a stack of libraries including SQL and DataFrames, MLlib for machine learning, GraphX, and Spark Streaming. You can combine these libraries seamlessly in the same application. - Runs Everywhere: Spark runs on Hadoop, Apache Mesos, Kubernetes, standalone, or in the cloud. It can access diverse data sources. You can run Spark using its standalone cluster mode, on EC2, on Hadoop YARN, on Mesos, or on Kubernetes. Access data in HDFS, Alluxio, Apache Cassandra, Apache HBase, Apache Hive, and hundreds of other data sources.

Table 2.3: KDD Tools.

2.4.5 Analyze of KDD process models

As show in the figure below, based on the study of KDD process models proposed in the literature, we define the principal tasks occurring in the explicit knowledge acquisition process mainly the knowledge formalized in data bases.

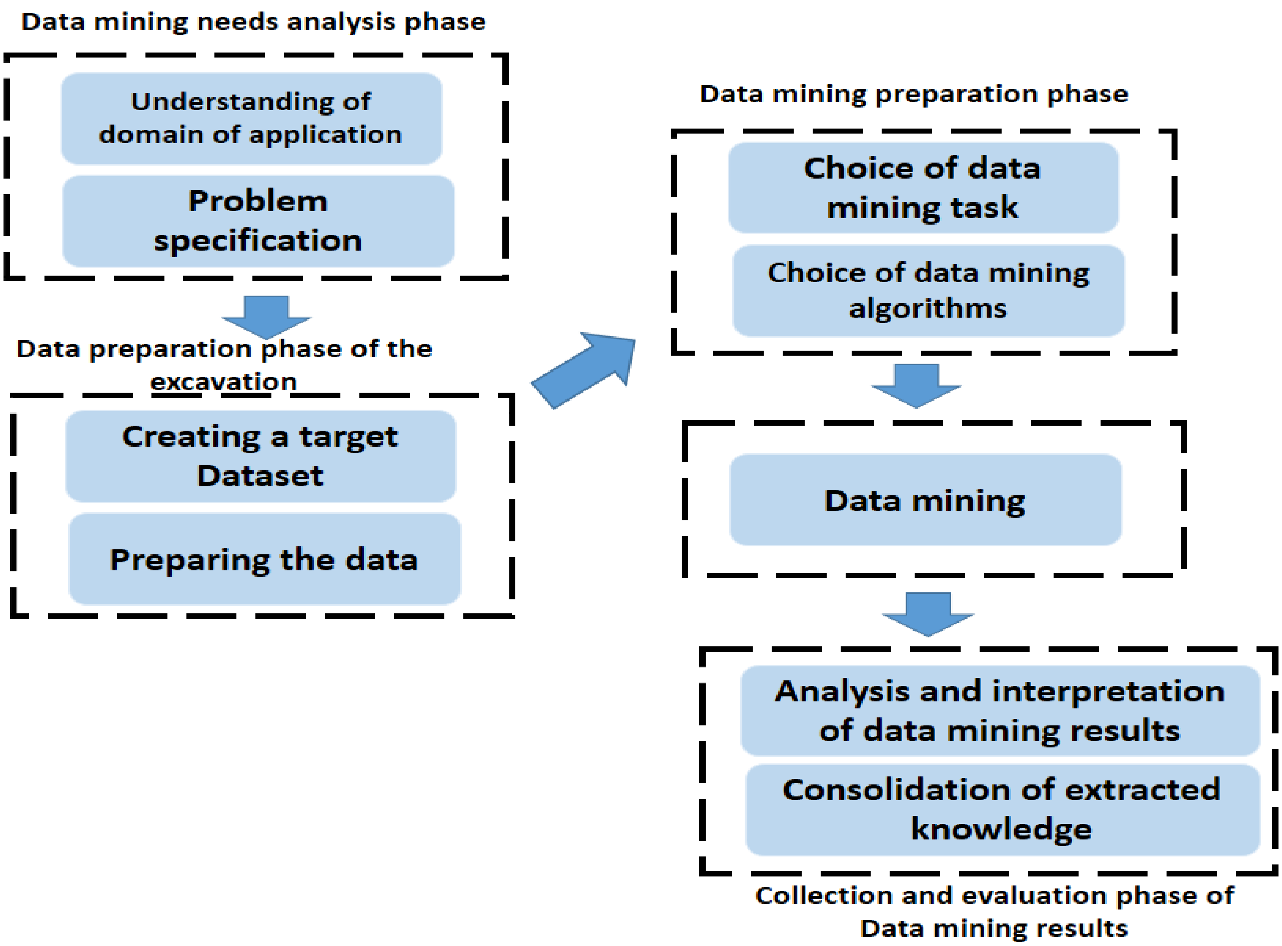


Figure 2.11: Explicit knowledge Acquisition process from operational databases.

2.5 Conclusion

At this chapter, we introduced the organizational knowledge acquisition and its objectives besides the different roles that can participate to this process. We divided the knowledge acquisition into two sub-process and define each of them. We started with tacit knowledge acquisition sub-process by defining massive number of techniques with examples and classifying it into Manual and semi-automatic techniques. Then we moved to explicit knowledge acquisition sub-process, we started with introducing the Knowledge Discovery in Database, Data Mining besides the confusion and the relation between DM and KDD. After that we defined steps of different KDD process models and KDD tools. Our Knowledge acquisition system is based on knowledge engineering and Cloud computing paradigm which are defined in the next chapter.

Chapter 3: Cloud Computing paradigm and Knowledge Engineering

3.1 Introduction

The proposed approach in this thesis for Organizational Knowledge Acquisition is based on Cloud Computing and Knowledge Engineering. Then, we define first in this chapter, the Cloud Computing paradigm and technology and its advantages for Organizational Knowledge Management. Second, we present some Knowledge Engineering definitions. n.

3.2 Cloud Computing

3.2.1 Definition

Cloud computing is a topic that received a great deal of attention by individuals and organizations from different disciplines in the last decade. This new environment implies great flexibility and availability of computing resources at different levels of abstraction at a lower cost. Cloud Service Providers (CSPs) (e.g., Google, Microsoft, Amazon) are vendors who lease to their customers cloud computing resources and services that are dynamically utilized based on customer's demand according to a certain business model [133]. General services in different application areas such as business, education and governance are provided to the customers online and are accessed through a web browser, while data and software programs are stored on the cloud servers located in the data centers [134].

3.2.2 Cloud Services types

Cloud services types are generally classified into three types known as cloud service models and are shown in figure below. Cloud service models are a Service-Oriented Architecture (SOA) that describe cloud services at different levels of abstraction.

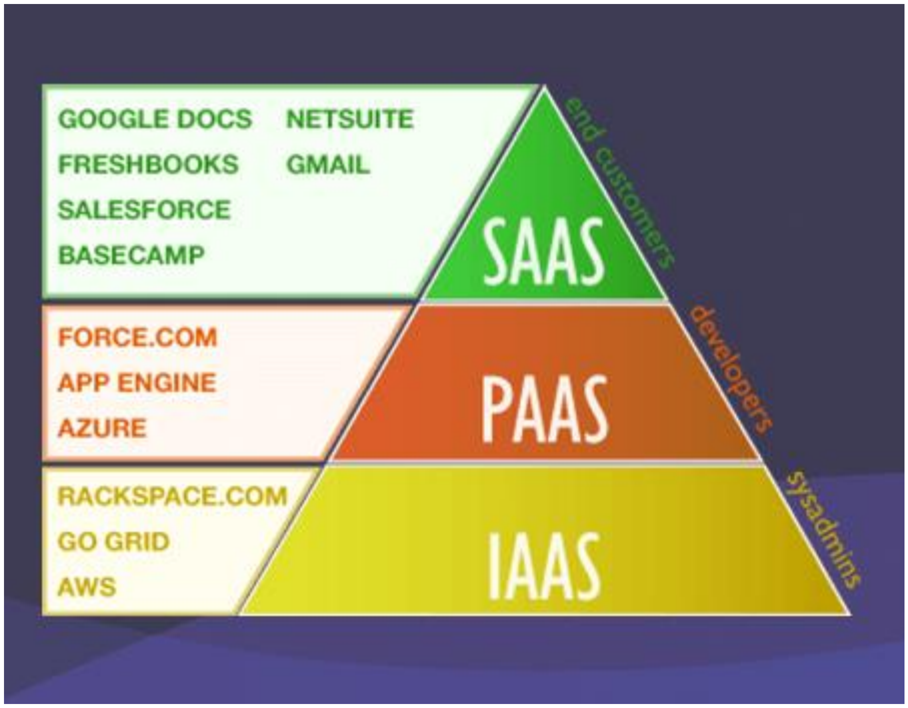


Figure 3.1: Services provided in cloud computing environment with their users [169]

3.2.2.1. Software as a Service (SaaS)

In this model, CSPs are responsible for running and maintaining application software, operating system and computing resources. The customer views the SaaS model as a web-based application interface where services and complete software applications are delivered over the Internet and are accessed via a web browser. Customers can access hosted applications such as Gmail and Google Docs through different client devices such as laptops, iPads and cell phones. Unlike traditional software, SaaS has the advantage that the customer does not need to buy licenses, install, upgrade, maintain or run software on his own computer [135]. It has also other advantages such as multitenant efficiency, configurability and scalability [136]. Examples of SaaS providers are Zoho, Google Apps and Salesforce.com.

3.2.2.2. Platform as a Service (PaaS)

In PaaS, a CSP provides, runs and maintains both system software (i.e., the operating system) and computing resources. The customer manages and runs the application software under the operating system and on the virtual resources provided by the CSP. The customer has little or no control over the operating system and hardware resources [135].Unlike SaaS that provides the customer with complete (ready to use) applications, PaaS gives him/her the opportunity to design, model, develop and test applications directly on the cloud; therefore, he/she can control the software lifecycle [136]. PaaS supports collaborative work between members of a project team. For instance, a number of users located in different countries can collaborate in developing a website using a PaaS cloud service. Examples of PaaS providers are windows Azure, Google Apps Engine and Aptana cloud.

3.2.2.3. Infrastructure as a Service (IaaS)

In this model, the CSP provides a set of virtualized computing resources (e.g., network bandwidth, storage capacity, memory, processing power) in the cloud. It is the responsibility of the customer to run and maintain the operating system and the software applications on these virtual recourses. IaaS uses virtualization technology to convert physical resources into logical resources that can be dynamically provisioned and released by customers as needed. Examples of IaaS providers are Drop Box, Amazon EC2 and Akamai.

3.2.3 Analysis of Cloud Computing Systems

Cloud computing systems are classified as public cloud, private cloud, community cloud and hybrid cloud. These classes are known as deployment models and they describe the scope of services offered on the cloud to the customers. Figure 2 below shows different cloud deployment models.

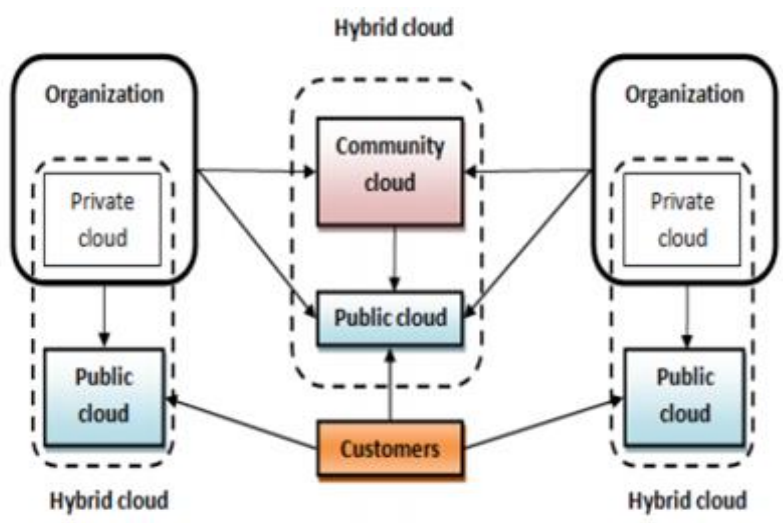


Figure 3.2: Cloud computing Deployment Models [140].

3.2.3.1 Public Cloud

In public clouds the infrastructure and other cloud services are made available to the general public over the Internet. The cloud is owned and managed by a CSP who offers services to consumers on a pay-per-use basis. Public cloud users are by default treated as untrustworthy; therefore, security and privacy are big concerns about this type of cloud [137]. Many popular cloud services are public including Amazon EC2, Google App Engine and Salesforce.com.

3.2.3.2 Private Cloud

In private clouds the computing resources are operated exclusively by one organization. It may be managed by the organization itself or a CSP. Private clouds are considered to be more secure than public clouds since their users are trusted individuals inside the organization. The other two deployment models, community clouds and hybrid clouds, fall between public and private clouds [137].

3.2.3.3 Hybrid Clouds

In hybrid clouds, the cloud infrastructure consists of a combination of two or more public, private or community cloud components. The cloud components are bound together by standardized technology and managed as a single unit, yet each cloud remains a unique entity [138] [139]. Hybrid clouds allow organizations to optimize their resources, so the critical core activities can be run under the control of the private component of the hybrid cloud while other auxiliary tasks may be outsourced to the public component. Figure 2 below shows different cloud deployment models.

3.2.4 Cloud Computing and Knowledge Management

Cloud Computing brings many advantages for organizational KM. KM activities and issues assisted and solved using technology, would be even more efficient in a Cloud Computing environment mainly knowledge storing, knowledge access, knowledge gathering and knowledge dissemination. In the context of these KM sub processes, Cloud Computing provides to organizations:

- Repositories for knowledge storing and integration which are virtual, have large storage capacity and can be easily and dynamically extended and replicated.

- Huge amount of knowledge assets available on real-time and coming from heterogeneous knowledge sources.
- Knowledge assets accessible via multiples platforms and devices. Powerful search engines enabling knowledge retrieval.
- Interesting tools facilitating collaborative work and communication between knowledge users, in order to share tacit knowledge within organization. [46]

In order to study the impact of the Cloud Computing on organizational KM process, [46] analysis existing Cloud based KM works and present in the Table 3.1 Cloud Computing advantages for Knowledge Management.

Knowledge Management based Cloud Computing works		**Context**	**Cloud Computing advantages for Knowledge Management**
CBPKM : Cloud based personal knowledge management platform [181]	The proposed platform provides: - Knowledge organizing, storage and internalization services based on cloud-based knowledge repository. - Knowledge search and retrieve services based on Cloud search engine. - Access service to domain experts and their opinions about shared knowledge for knowledge evaluation.	Personal Knowledge Management	- More sophisticated search engines enabling cloud users to search and retrieve necessary information anytime and anywhere. - A cloud-based knowledge repository is not just a storage space for cloud users to keep information. It enables them to internalize their personal knowledge. - Mobility and portability.
Five layer framework for KaaS platform [182]	The proposed architecture provides various knowledge services including knowledge resources collection (via direct or indirect participation of domain experts), organization of knowledge in virtual repository and the exploitation of knowledge by different level of users.	Knowledge as a Service platform	The knowledge resources repository is mainly a virtual repository, rather than a physical. It can be easily dynamically expanded and replicated.
Four layers Framework architecture for Cloud-based Knowledge	Cloud knowledge integration service layer provides knowledge-	Knowledge-based Collaborative Product	- Providing the meta-knowledge model. - Availability of large amounts

Integration of Collaborative Product Design [183]	based services including: - Knowledge integration service supporting knowledge mapping and merging - Knowledge search and sharing services based on knowledge evaluation. - Knowledge storage service based on cloud knowledge base.	Design	of knowledge of design knowledge in real-time and from multiple sources. - Increased accessibility across many platforms including mobile. - Improved efficiency from high utilization of sharing physical servers. - Improved reliability with replication of data within the systemand higher level of fault tolerant.
Knowledge Management as a Service: Cloud based architecture for Knowledge Management [184]	The proposed architecture includes the knowledge Finder, Access Control, Gathering, distribution, Inference, Storage and retrieval, Integration, Personalization, Expert mediation and Knowledge pusher services.	Organizational Knowledge Management	- keeping KM aligned to the new technological progresses; - Sharing and acquisition of knowledge in a highly distributed and dynamic environment. - Providing ways for intercommunications between public, private, community and hybrid clouds enabling controlled knowledge sharing among virtual organizations.
A Framework for Cloud Based KM [185]	The proposed Framework provides knowledge map, linking, Knowledge cycle, Rewards and ratings based on knowledge, Classification and knowledge packages creation, and Reporting services.	Organizational Knowledge Management	- Reduced licensing charges by using a single content management system. - Reduced ownership charges due to transition to cloud. - Enhanced efficiency and cooperation throughout enterprises via a unified system. - Reduced risks due to improved flexibility and scalability. - Worldwide user profiles consolidation and improved security due to centralization. - Navigation and branding consistency by using a single KM tool instead of multiple tools. - Enhanced content accessibility via a central data store. - Reduced the need for organizations to employ expert personnel.

Cloud Storage Application as Knowledge Management System [186]	KM activities include preservation, acquisition, distribution / sharing and knowledge development. These activities can be assisted by different cloud storage applications.	Knowledge Management System	- Facilitating knowledge acquisition activity in which a person could gain access to obtain knowledge through multiple devices. - Retrieving knowledge across multiple devices - Providing several sharing methods such as sharing links or sharing folders - Providing means to improve existing knowledge and create new knowledge

Table 3.1: Cloud Computing advantages for Knowledge Management [46]

3.2.5 Disadvantages of the Cloud

With all of the speed, efficiencies, and innovations that come with cloud computing, there are, naturally, risks. Security has always been a big concern with the cloud especially when it comes to sensitive medical records and financial information. While regulations force cloud computing services to shore up their security and compliance measures, it remains an ongoing issue. Encryption protects vital information, but if that encryption key is lost, the data disappears.
Servers maintained by cloud computing companies may fall victim to natural disasters, internal bugs, and power outages, too. The geographical reach of cloud computing cuts both ways: A blackout in California could paralyze users in New York, and a firm in Texas could lose its data if something causes its Maine-based provider to crash.[169]

3.3 Knowledge Engineering

Knowledge engineering is the field, which corresponds to the study of concepts, methods and techniques to model and / or acquire knowledge for systems carrying out or helping humans to carry out tasks that are formalized little or not a priori. [167].

Knowledge engineering tools and methods such as cognitive and conceptual modeling, representation and modeling languages, etc., have been developed to enable knowledge to be acquired from texts, to solve research problems information on the web, to implement management indicators in information systems or to manage and capitalize knowledge in business [168].

Knowledge engineering consists of articulating work on knowledge nature and representation in software and among users, on the analysis knowledge uses and sources and human-machine interactions. However, knowledge engineering is also about implementing software engineering approaches, with the aim of making knowledge available to users and within computer applications. Innovations in the field therefore include methods, software and interfaces to aid modeling, as well as conceptual or formal representations of knowledge [168].

3.4 Conclusion

We presented in this chapter, the Cloud Computing paradigm and technology. We started with giving Cloud Computing definition. We then moved to describe the three major types of Cloud services including IAAS, PAAS and SAAS. After that, we analyzed Cloud Computing Systems and their classification, which is composed of four crucial classes: public, private, community and hybrid Cloud. Some Cloud based KM works and the Cloud advantages for organizational KM are presented. Finally, we defined the knowledge Engineering filed and its main goals. In the next chapter, we present our Knowledge Acquisition approach and system, which are based on Cloud Computing and Knowledge engineering.

Chapter 4

Chapter 4: PRESENTATION OF OUR KNOWLEDGE ACQUISITION APPROACH AND SYSTEM

4.1 Introduction

We present in this chapter the conception of our Organizational knowledge acquisition system named akizitor. The system includes Cloud services enabling knowledge acquisition of an organization. The knowledge acquisition is based on knowledge engineering as the cloud services are based on knowledge models including the domain model and the problem resolution model. We describe then in this chapter these knowledge models and the establishment of the different tacit knowledge acquisition techniques presented previously in chapter 2 though UML diagrams. The main functionalities of the system and its potential actors are described.

4.2 Knowledge models

The sub-process of tacit knowledge acquisition aims at the organization's experts knowledge transfer to the knowledge Management System. This knowledge refers to the problems that they faced at the domain activities of the organization and methods used to solve it. The Cloud services of our akizitor system supporting the knowledge acquisition process are based on the following knowledge models: domain model and domain problems model.

[147] proposed some meta-models that are used for knowledge acquisition with human

experts as part of the design of Computer Environments for Human Learning. The definition of these meta-models is based on other methods of knowledge acquisition in particular the CommonKads method. We have not used all of the models offered because we focused only on the models that are similar to the knowledge acquisition process, the crucial point of the similarity here is problems, which means both of them aim to determine the different problems at the organization, so the chosen models below help to prepare the techniques of the Knowledge acquisition process as structured interviews with their different techniques (Free association, Psychological Scaling, twenty questions, Modus ponens sorting, Knowledge diagram, AHP, Sorting, Repertory Grid), protocol extraction, questionnaire and problem-solving techniques (Goal Decomposition, Forward Scenario Simulation).

We use the UML Class diagram as show in the figure below to present the different concepts of the two knowledge models.

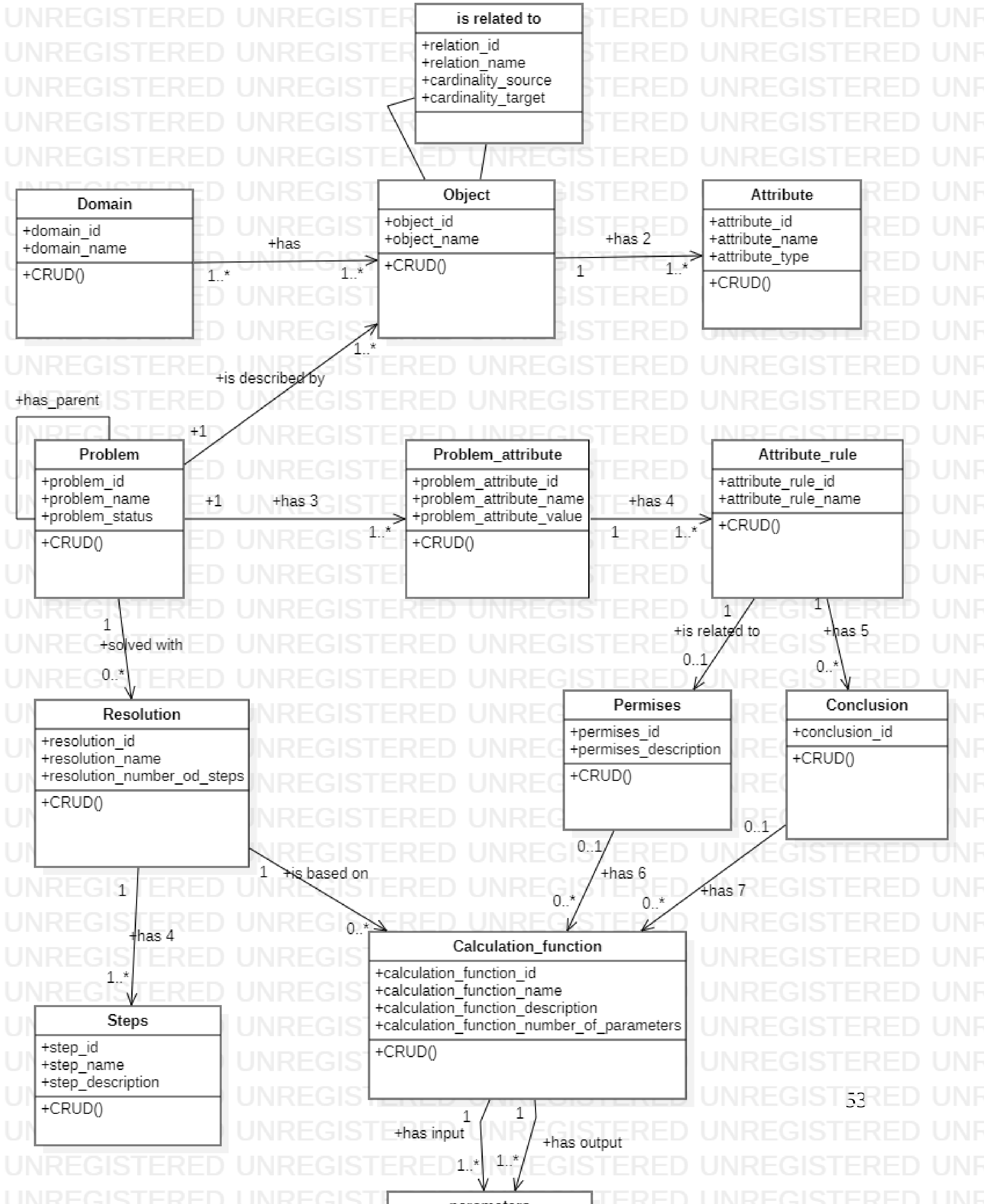

is related to
+relation_id
+relation_name
+cardinality_source
+cardinality_target
Domain
+domain_id
+domain_name
+CRUD()
+has
1..*
1..*
Object
+object_id
+object_name
+CRUD()
+has 2
1
1..*
Attribute
+attribute_id
+attribute_name
+attribute_type
+CRUD()
1..*
+is described by
+has_parent
+1
Problem
+problem_id
+problem_name
+problem_status
+CRUD()
+1
+has 3
1..*
Problem_attribute
+problem_attribute_id
+problem_attribute_name
+problem_attribute_value
+CRUD()
+has 4
1
1..*
Attribute_rule
+attribute_rule_id
+attribute_rule_name
+CRUD()
1
+solved with
0..*
1
+is related to
0..1
1
+has 5
0..*
Resolution
+resolution_id
+resolution_name
+resolution_number_od_steps
+CRUD()
Permises
+permises_id
+permises_description
+CRUD()
Conclusion
+conclusion_id
+CRUD()
0..1
0..1
+has 6
0..*
+has 7
0..*
1
+is based on
0..*
1
+has 4
1..*
Steps
+step_id
+step_name
+step_description
+CRUD()
Calculation_function
+calculation_function_id
+calculation_function_name
+calculation_function_description
+calculation_function_number_of_parameters
+CRUD()
1
+has input
1..*
1
+has output
1..*
UNREGISTERED

Figure 4.1: UML Class Diagram for Domain Model and Domain Problems Model with their Resolution

4.2.1 The domain model

The domain model defines the vocabulary of a given domain. This model is used for the description of the problems encountered in the domain. On the other hand, it allows the understanding of a given domain at any host organization.

This model may be used for the preparation of the interviews and questions to understand the application domain and the activity area of the business aimed by the knowledge management project. Each domain is described by the following characteristics:

- **Name**.
- **Object types** to describe the domain problems. Object types are characterized with:
  - **Identifier**
  - **List of characteristics:** each characteristic is defined with a name and a type (simple type like String, Boolean or numbers ...), or complexed type like an object, which represent a relation with another object.

4.2.2 The Domain problems model

The domain problems model describes the problems resolution within a domain including the resolution methods. The problem description for a certain domain is based on the defined vocabulary for the chosen domain described using the domain model. For each problem it must:

- Choose from object types of the domain vocabulary parts, which are important to describe the domain then instantiate them. (This is **objects problem**)
- Specify the **question to be solved** for this problem.

The **problem resolution** is based on the **problem classification** and the used tools for the resolution. First, identify the class, which contains the problem to facilitate the choice of a resolution technique. Second, we must define **hierarchical classes for the main problem.** This hierarchy should be presented as a classification Tree, in which class C2 is a sub class of class C1 and each problem of C2 is a problem of C1, the classes, which we can associate. A problem class is defined by:

- **Name**
- **Status** (operational or not operational)
- Attribute value, which discriminates the current class with its parent.
- **Discriminant attribute** permit to discriminate the sub-classes of this class.
- Other **problem attributes** if needed for problem resolution.

For each **attribute** of this hierarchical class, the author needs to define **if-then** rules to calculate the attribute value to implement the problem on the hierarchical classification. Each rule is defined by:

- **Name**
- **set of premises** relating to elements of the utterance, which means the different problem objects or constraints on those objects. Those premises can be related to the problem attributes, which also may require the application of rules to calculate their values. We can also call calculation functions. The represented knowledge within the premises should be treated with the logic operator 'and'.
- **Set of conclusions** allowing to calculate or modify values of the problem attributes. We can also call **calculation functions**.

Each **calculation function** is defined by:

- **Name**.
- Its **description** in natural language, we can specify the possible values of the functions.
- **Number of parameters** and their **types**.

- The **return types**.

For classification to be useful in problem solving, a solving technique must be associated with each operational class in the classification tree. Each **resolution technique** is defined by:

- **Name**.
- **Resolution planning**: it is an algorithmic description of the actions (steps of resolution) to be carried out to solve a problem. The analysis of this plan must make it possible to find the solution of the problem.
- **Calculation function** allows to determine the solution elements from the data related to the statement.

4.3 UML Use Case diagram of our system

The main functionalities of the proposed knowledge acquisition system are presented via the UML use case diagram shown in the figure below. The diagram represents the different actors of the system and their main roles.

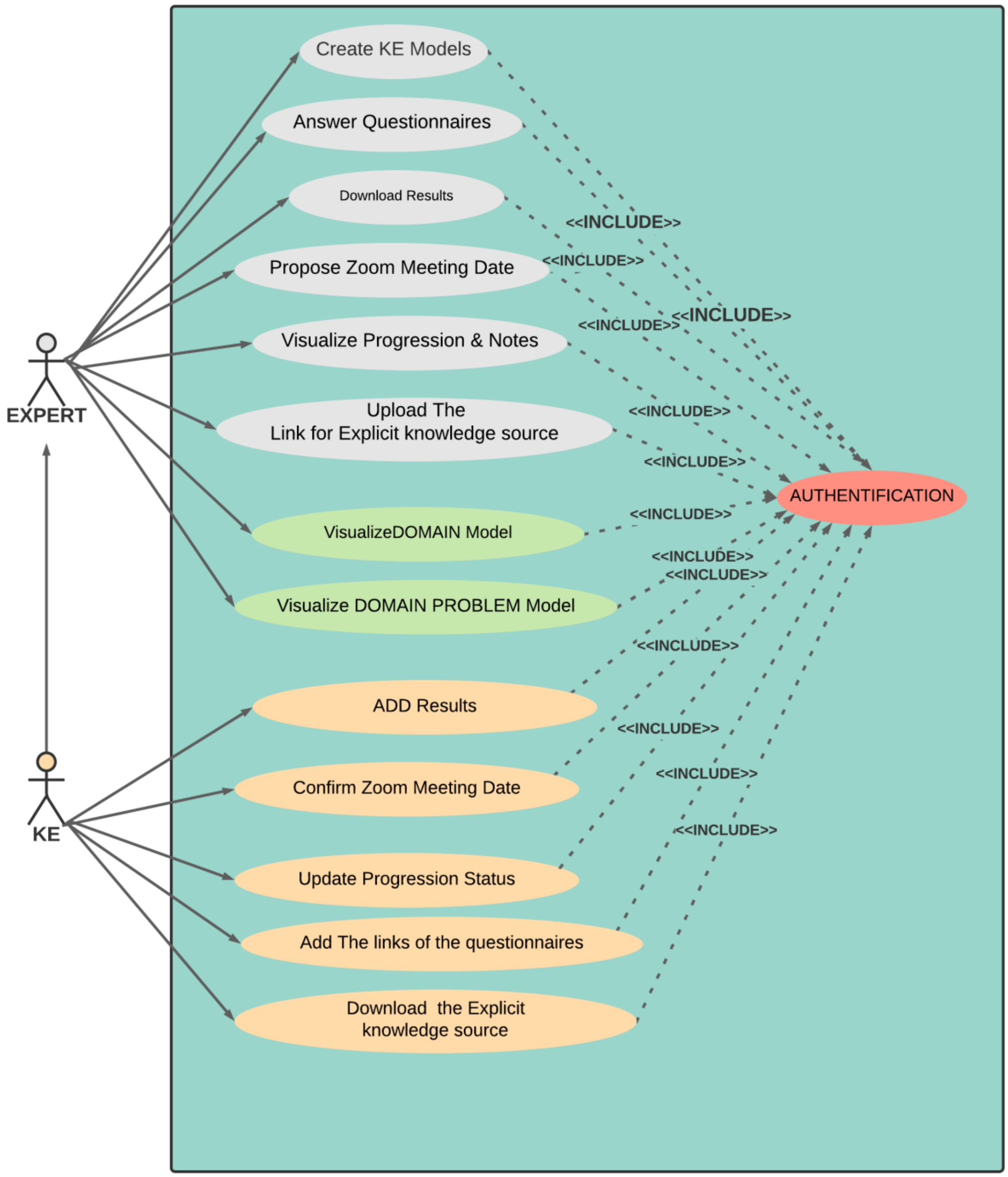


Figure 4.2: System use case with their principal actors

4.3.1 Presentation of our system actors

Two crucial actors for our knowledge acquisition system are distinguished. The first one is the knowledge Engineer as the admin of the website and the second is the expert as the end-user.

4.3.1.1 Knowledge Engineer (KE)

The person that carries out the process of knowledge acquisition on the system. It has the following roles within the system:

Knowledge Engineer Roles	Description
Confirming zoom meeting Date	After receiving a proposed date for meeting, the KE can confirm or deny it, if confirmed he will send the link to the meeting.
Update progression Status	after the expert complete a task or if a task is needed to be completed and the expert hasn't done it yet, the KE update the progression status.
Add a link to the questionnaire	After the expert create **Domain Model** & **Problem Domain Models**, the KE prepare a questionnaire using google form and add the link to it.
Explicit Knowledge acquisition	After the expert import the link of database, KE can access to it and download the Database.
Upload Results	After finishing KA process, The KE do the KDD process then imports the results PDF file to the website.
Visualize Models	The KA can visualize the Domain Model and Problem Domain Model of each expert.

Table 4.1: Knowledge Engineer Roles.

4.3.1.2 The KA process actors

The person who is going to interact with the system and put his knowledge to the use, he can be a simple employee or an expert in a specific domain, we also refer to him as the end user and can apply the following tasks and roles within the system:

Expert Roles	Description
Create Models	The expert can create models such as Domain Model and Problem Domain Models.
Answer questionnaires	The expert access to the questionnaires using the links shown on the site then answer them.
Propose a zoom meeting Date	The expert can propose a date for zoom meeting and wait for the KE to confirm it.
Import Database	The expert imports the database on google Drive, then upload the link in the system.
Download Results	after completing all the Knowledge Acquisition process, the expert can download the acquired knowledge as PDF file.
Progression and notes	The expert can visualize the progression of KA process and the status of each task.
Visualize Models	The expert can visualize the Domain Model and Problem Domain Model.

Table 4.2: Description of the Expert different Use cases.

4.4 UML sequence diagrams

In this section, we present some UML sequence diagrams to describe use cases presented above.

The **figure 4.3** below presents UML Sequence Diagram related to Creating Domain model use case. It presents the different Steps and functions to be followed by the expert to create a Domain Model.

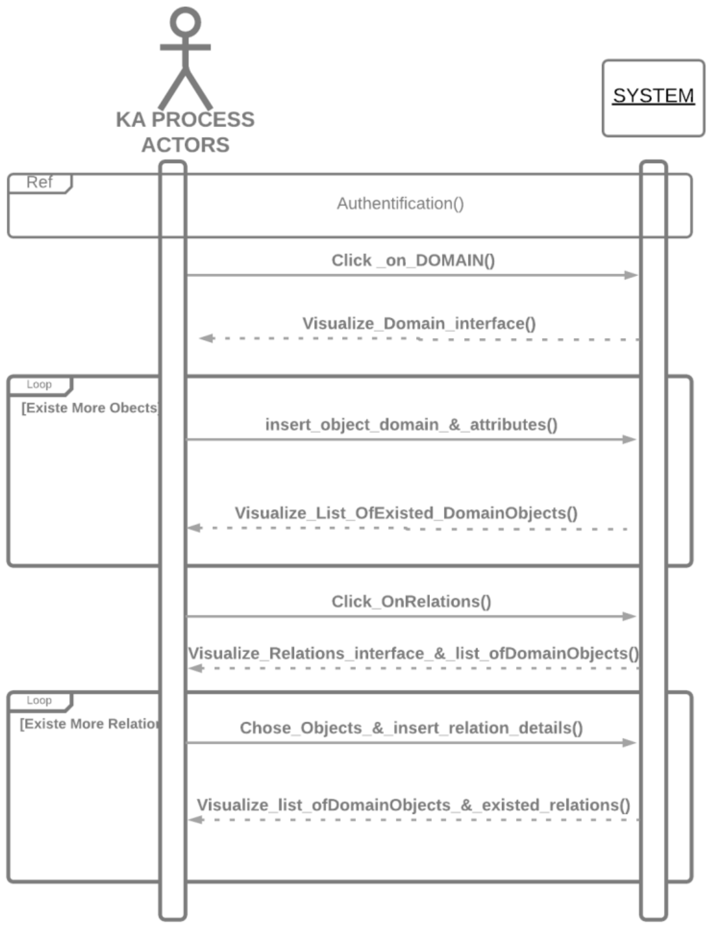


Figure 4.3: UML Sequence Diagram of Creating Domain Model use case.

The **figure 4.4** represents the UML Sequence Diagram describing the different Steps and functions to be followed by the EXPERT to Create Domain Problem Models

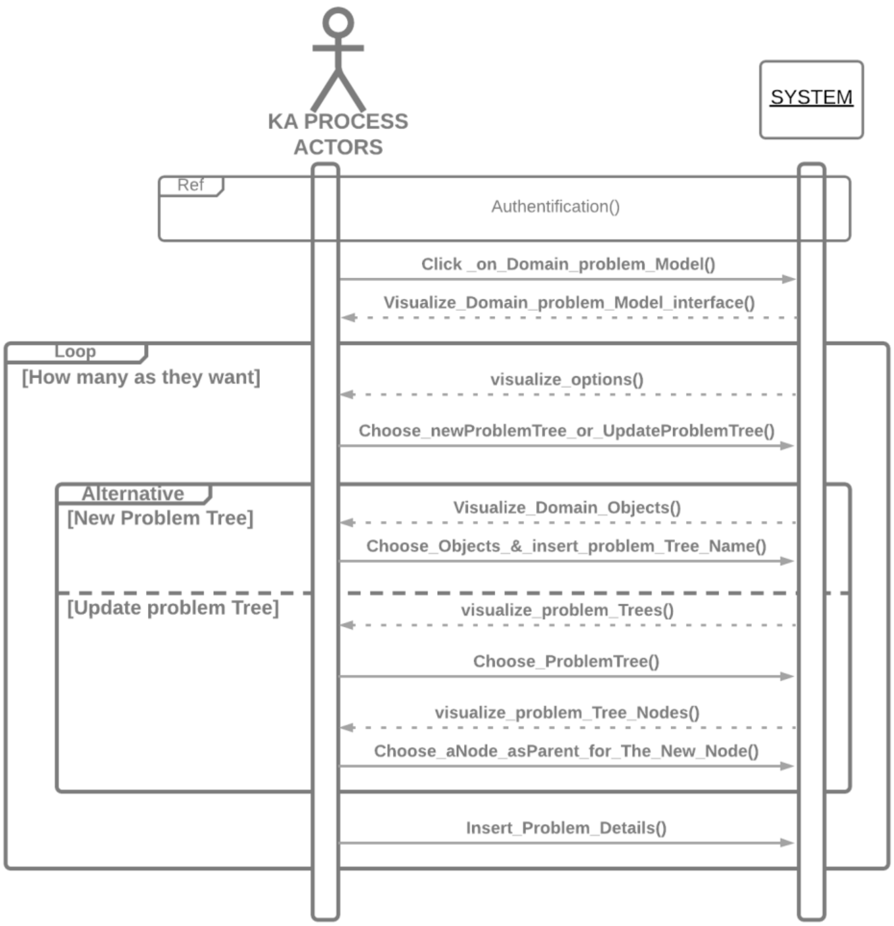


Figure 4.4: Sequence Diagram of Creating Domain Problem Model

The **figure 4.5** below represents an UML sequence Diagram describing how to create a problem resolution of a Domain Problem Model.

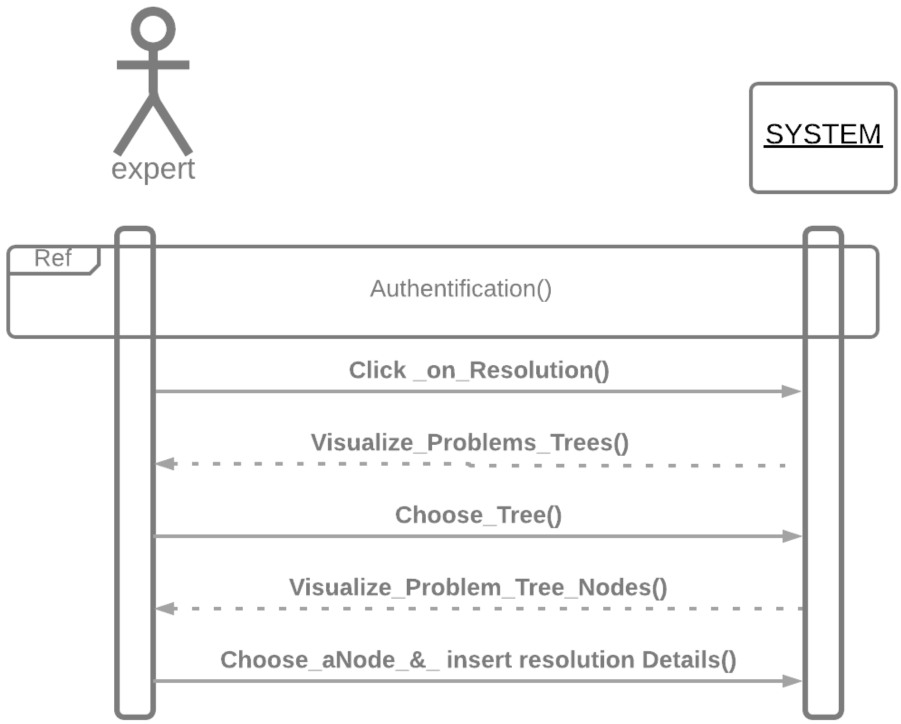


Figure 4.5: Sequence Diagram of Creating a Resolution for a Domain Problem Model use case.

The **figure 4.6** represents the UML Sequence Diagram of the Models(domain & problem domain) visualization from the Knowledge Engineer side.

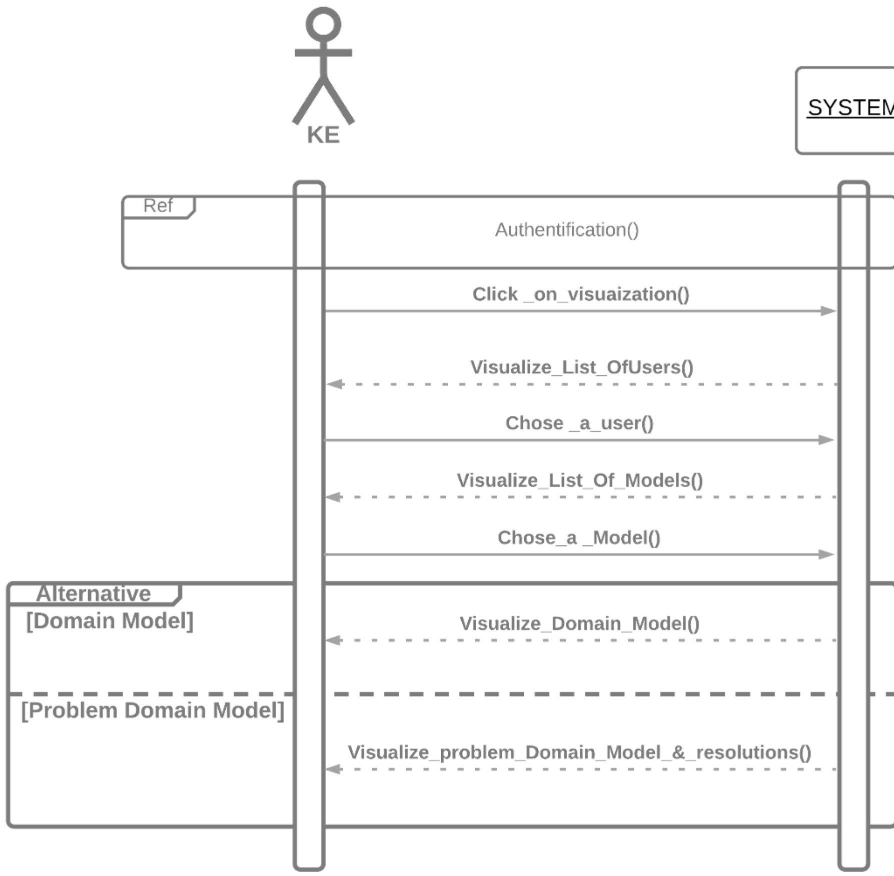


Figure 4.6: Sequence Diagram of Knowledge Models visualization use case.

The **figure 4.7** below represents an UML sequence Diagram describing how to access to the questionnaire and answer it.

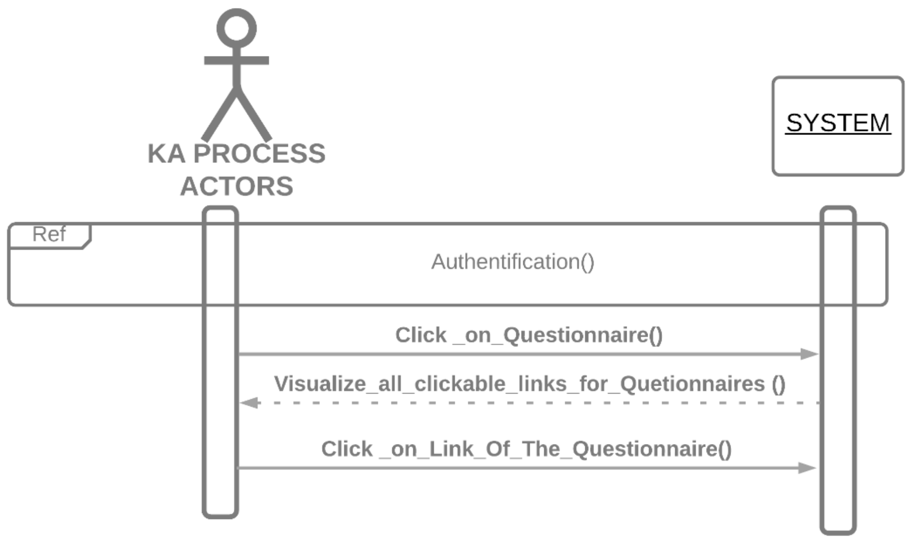


Figure 4.7: Sequence Diagram of KA Process ACTORS access to the Questionnaires.

The **figure 4.8** below represents an UML sequence Diagram describing how to upload the explicit knowledge resources on the system (name of the resource and the access link).

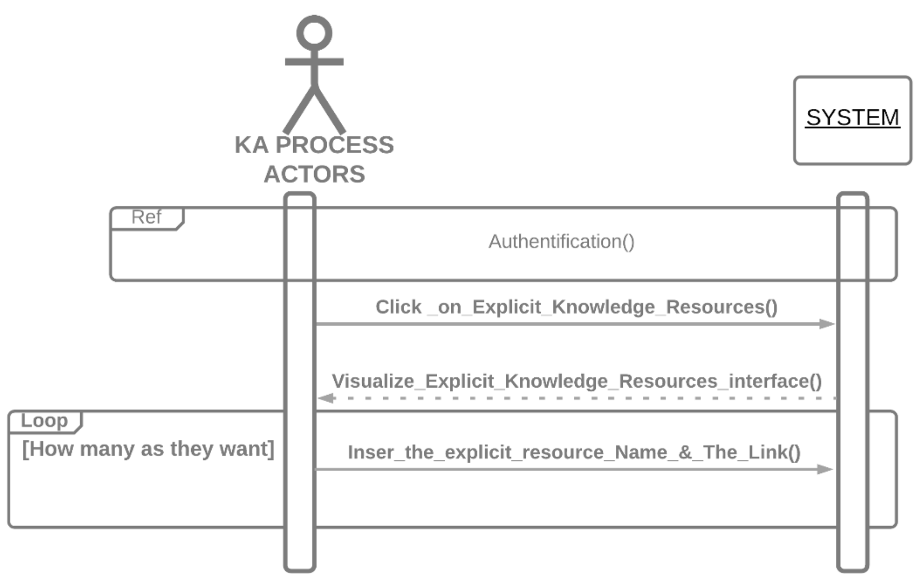


Figure 4.8: Sequence Diagram of KA Process ACTORS to upload the explicit Knowledge resources.

4.5. UML class diagram of the proposed system

The **figure 4.9** below illustrates the UML class Diagram of the Database for our proposed Knowledge Acquisition System "Akizitor".

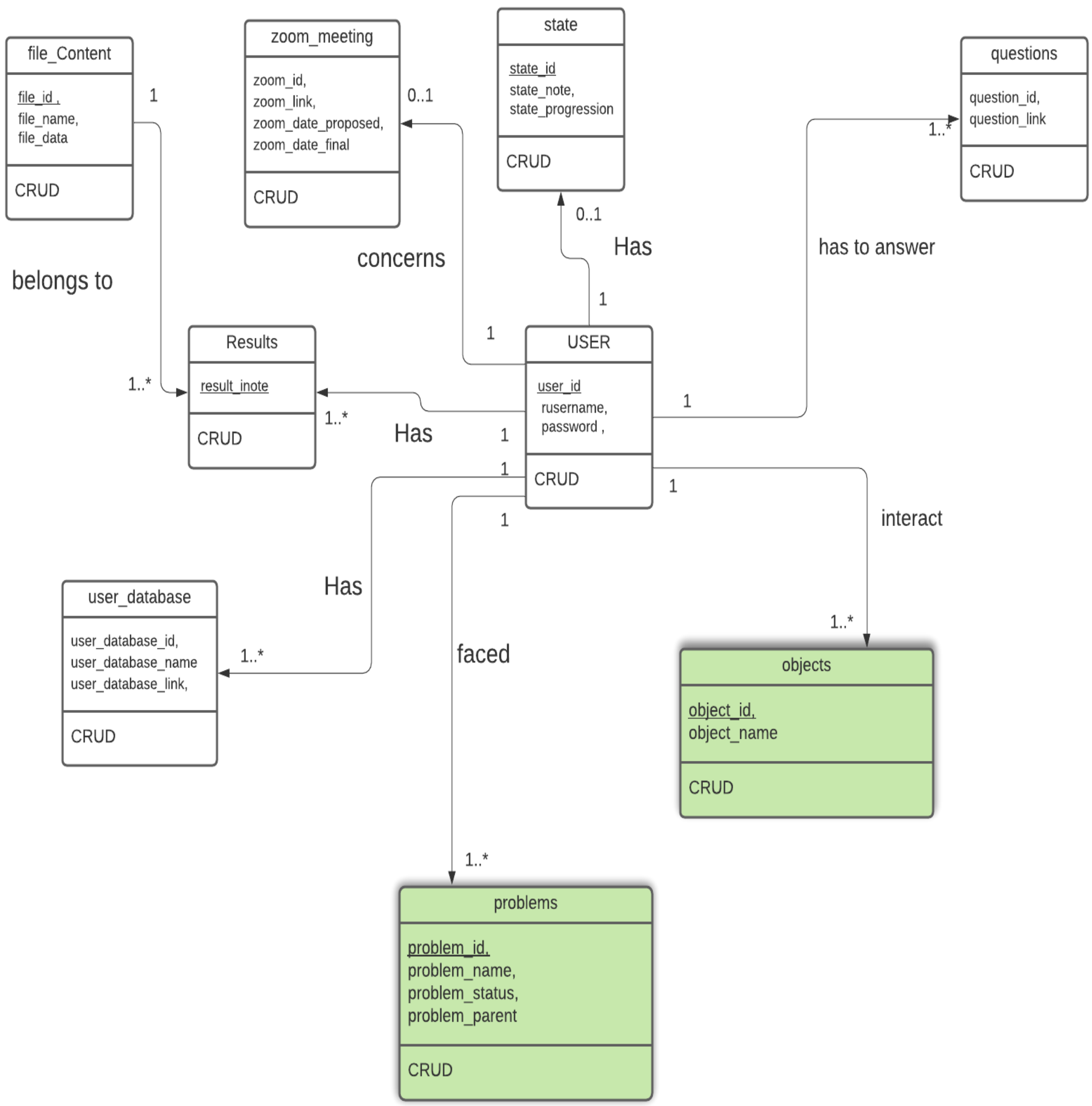


FIGURE 4.9: UML diagram class of our system "AKIZITOR".

The **Table 4.3** below illustrate the different classes of the diagram one by one besides their attributes within their description.

CLASS	DESCRIPTION	ATTRIBUTE	DESIGN	Attribute type
Problems	Contains all the possible problems from the domain problem Model	Problem_id	Each a problem has a unique ID	number
		Problem_name	Each problem has a name	String
		Problem_status	Each a problem has a status (operational, none operational)	String
		Problem_parent	To link it with its problem parent ID (except the head of the tree)	number
Objects	Contains all Domain Objects	Object_id	Each object has a unique ID	Number
		Object_name	Each object has a name	String
User	Contains KA process actors	User_id	Each actor has a unique ID	Number
		Username	Each actor has a name	String
		Password	Each actor has a password	String
User_database	Contains explicit knowledge resources	User_db_id	Each resource has a unique ID	Number
		User_db_name	Each resource has a name	String
		User_db_link	Each resource has a link	String
Zoom_meetin	Contains all meeting informations	Zoom_id	Each meeting has a unique id	Number
		Zoom_link	Each meeting has link	String
		Zoom_date_proposed	The proposed date for meeting	date
		Zoom_date_final	The final date for the meeting	date
State	The progression state for the KA process	State_id	Each progression state has a unique id	Number
		State_note	Note for the KA process actor to specify the next step to do	String
		State_progression	KA process	Number

			progression value	
Questions	Contains informations about questionnaires	Question_id	Each questionnaire has a unique id	Number
		Question_link	Each questionnaire has link	String
Results	Contains the result of KDD	Result_id	Each result has an id	Number
		Result_note	Each result has a note	String
File_content	Contains informations about files which conserve results.	File_id	Each file has an id	Number
		File_name	Each file has a name	String
		File_data	File content	Blob

Table 4.3: Description of Class Diagram of the Akizitor system.

4.6 Conclusion

This chapter is dedicated to the conception of our organizational knowledge acquisition system named Acquisitor. The tacit knowledge acquisition is based on knowledge models. We started by defining the knowledge models including the Domain model and the Domain problems resolution model through a UML class diagram. After describing the potential actors interacting with the system and their roles, we present the main functionalities of our system through UML use case and sequence diagrams. Finally, the UML class diagram of our system is presented.

Chapter 5: Implementation & validation

5.1 Introduction

In this chapter, we introduce the implementation and validation of our Knowledge Acquisition system. We present the programming languages and tools used for the system implementation, besides its architecture. Some interfaces are shown. The system is tested with a case study.

5.2. The Architecture of our Knowledge Acquisition System

The figure bellow depicts our proposed system architecture for Organizational knowledge acquisition. The architecture includes the basic components of the system, its major actors and their interaction.

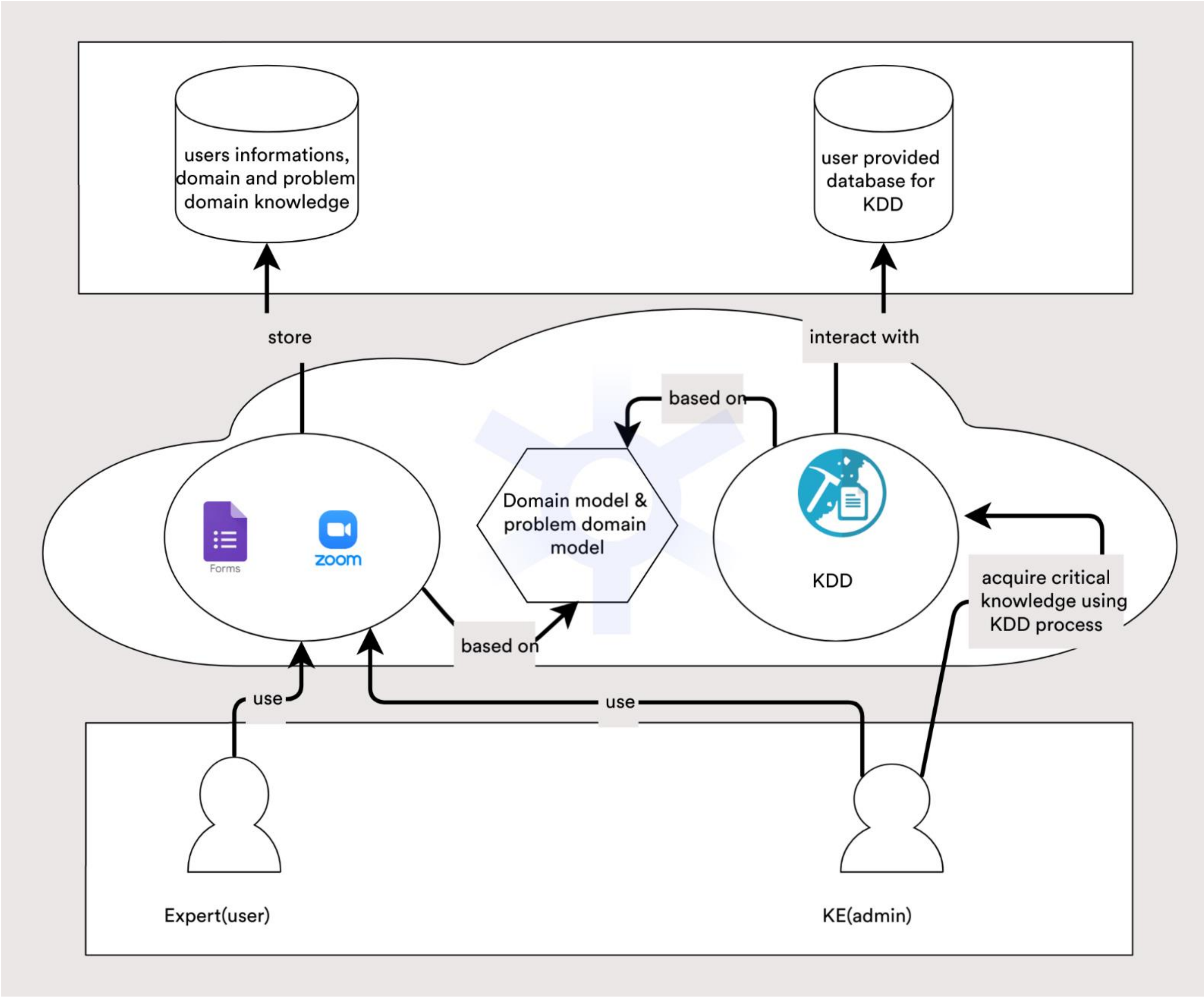


Figure 5.1: The architecture of the proposed Knowledge Acquisition System.

Figure 5.1 illustrates the architecture of the acquisitor, where we can easily detect the two major actors of the system: the KE and the Expert (user).

The KA process actor (user) interacts with the system interfaces to visualize the acquired knowledge and to complete the steps of the knowledge acquisition prosses which are required by the KE besides that he can provide the system with Databases to extract some useful knowledge.

The KE (admin) interacts with the system interfaces to visualize the user data and prepare the required steps for the user besides that he can access to the database which is provided by the

Expert. the KE apply a KDD process to acquire knowledge from the provided database.

The system interacts with the System Database and some computing services as google forms and zoom meeting for the purpose of applying the steps of tacit KA.

The acquisitor System Database stores the user Data, access link to external Databases and the results of Knowledge Acquisition prosses.

The KDD process is where the KE will do his part and acquire the knowledge from the database provided by the user.

The KDD process can be easily integrated on the system since it's also running with python language (flask framework). However, the lack of material resources can cause the system to run slow. We decided then to separate them and only display the result of the process as a downable PDF on the system.

Concerning the questionnaires, multiple experts can answer it.

5.3 Used Tools & languages for the system implementation

The system is built with python (flask framework). For the KDD process, we used some of the following python libraries:

- Pandas for data manipulation and analysis
- Seaborn for data visualization
- Numpy for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays
- Matplotlib is a plotting library
- Spark Apache framework for data processing and that can quickly perform processing tasks on very large data sets, and can also distribute data processing tasks across multiple computers

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools. [175]

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals. Its name is a play on the phrase "Python data analysis" itself. Wes McKinney started building what would become pandas at AQR Capital while he was a researcher there from 2007 to 2010. [176]

Seaborn is one of an amazing library for visualization of the graphical statistical plotting in Python. Seaborn provides many color palettes and defaults beautiful styles to make the creation of many statistical plots in Python more attractive.[177]

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. The ancestor of NumPy, Numeric, was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Numarray into Numeric, with extensive modifications. NumPy is open-source software and has many contributors.[178]

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK. There is also a procedural "pylab" interface based on a state machine (like OpenGL), designed to closely resemble that of MATLAB, though its use is discouraged.[179]

Apache Spark is an open-source unified analytics engine for large-scale data processing. Spark provides an interface for programming entire clusters with implicit data parallelism and fault tolerance. Originally developed at the University of California, Berkeley's AMPLab, the Spark codebase was later donated to the Apache Software Foundation, which has maintained it since. [180]

5.4. Presentation of the organization host for the case study

Mobilis, or Mobilis ATM (ATM acronym for Algeria Telecom Mobile), is an Algerian mobile telephone operator, a subsidiary of Algeria Telecom. It is one of the three major Algerian mobile operators. Became autonomous in August 2003, Mobilis offers its customers post and prepaid offers.

On December 15, 2004, Mobilis launched the first experimental UMTS (Universal Mobile Telecommunication System) network in Algeria. With its “Mobi +” GPRS offer, Mobilis is a multimedia operator in Algeria. It has launched a vast project to deploy its GSM network. Today, nearly the network covers 80% of the Algerian population.

Mobilis' subscriber base (GSM + 3G) stood at +18 million in July 2021.

In December 2019, Mobilis obtained a global telecommunications license (2G, 3G and 4G) to deploy in Mali2.

Mobilis has sought, since its inception, to define basic objectives, including: Providing the best services, taking good care of subscribers to ensure their loyalty, creativity and introducing the new in line with the technological developments and this enables it to achieve significant business numbers and reach, in a short time, to the inclusion of 18 million subscribers.

By choosing and adopting a policy of change and creativity, Mobilis always works to reflect a positive image, by ensuring the provision of a high-quality network and very efficient service to subscribers, in addition to diversification and creativity in the proposed offers and services.

Mobilis wanted to position itself as a dealer closer to its partners and customers, which made it stronger for its slogan "Wherever You Are". This slogan is a pledge of constant listening, and evidence of its commitment to play an important role in the field of sustainable development and its contribution to economic progress [141].

Figure 5.2: The organigram of MOBILIS ATM organization.

5.4.1 Presentation of the host department (NETWORK TECHNIQUE & SERVICES)

This Department is the responsible of creating the telecom system, which contains two important parts clouds telecom and servers. It is embodied in two crucial divisions, Design Division and Maintenance Division.

The first one is Design Direction, it mainly deals with design planning and how to add different cycles and Media gateways for the new project which refers how to prepare the network and telecommunication architecture. This Service is responsible also for installing and buying materials such as BSI antennas, MSI antennas, Routers, and Servers.

After accomplishing those activities, the team will consider the project as completed for them and the responsibility moves to the service below. after that The Design Team can start working on another project now.

Second division called Maintenance Direction, after the design Team complete their job in good conditions now comes turn to this service to make sure the perfect enrolment and functioning of those materials and they will be responsible of applying the technologies upgrades, which is proposed by the sellers.

This division contains the Operating Support System (OSS), which is our host service.
OSS is the intermediary between the clouds telecom and personals (clients) who are working on the clouds telecom where there is BSI antennas (knots), transmissions and so on.
Who is the intermediary between them? The servers are the responsible for the communication between the telecom clouds and the clients and those servers are under the responsibility of the OSS Service who are the administrators. The main task of OSS team is DATA collection, so they collect Data from different knots that exist at the clouds telecom and they stored it in databases, which are usually hard to understand. The Servers depends on the sellers, so each one has a specific server for example Ericson has a server, Hawaii has a different server and ZTE has another different server, it also depends on the technologies for example radio core has a server and transmission has another different server.

Usually a reporting tool (business Object) uses the stored Data (statistics values) and counters to provides the Radio experts with graphs, it exposes different details about the clouds telecom that can be analyzed. After analyzing it, the expert will know if there is a problem in a knot, it does not work perfectly or the antenna is broken and that helps him to make decisions.

The second task is assuring the perfect functioning of the servers and that can happen, only when the collected data is perfectly stored at the perfect time. Those two factors are too important in this domain and missing one of them leads to spark a failure. One of the potential failures that can happen when they are collecting the alarm data from a knot but there is a lag between collecting this data and storing it in the server.

The third task is how to detect the potential failures (the failure above), how to avoid and find solutions to it.

5.5 Presentation of the system

Scenario 1: login and registration

If the expert access to the system for the first time, he needs to register by completing a registration form (username and password). After that, he goes to the login page to access to his account using his information to login. For the admin he only has to login with the information given to him by the website owner.

In the interface below the KA process actor can login (after he registers).

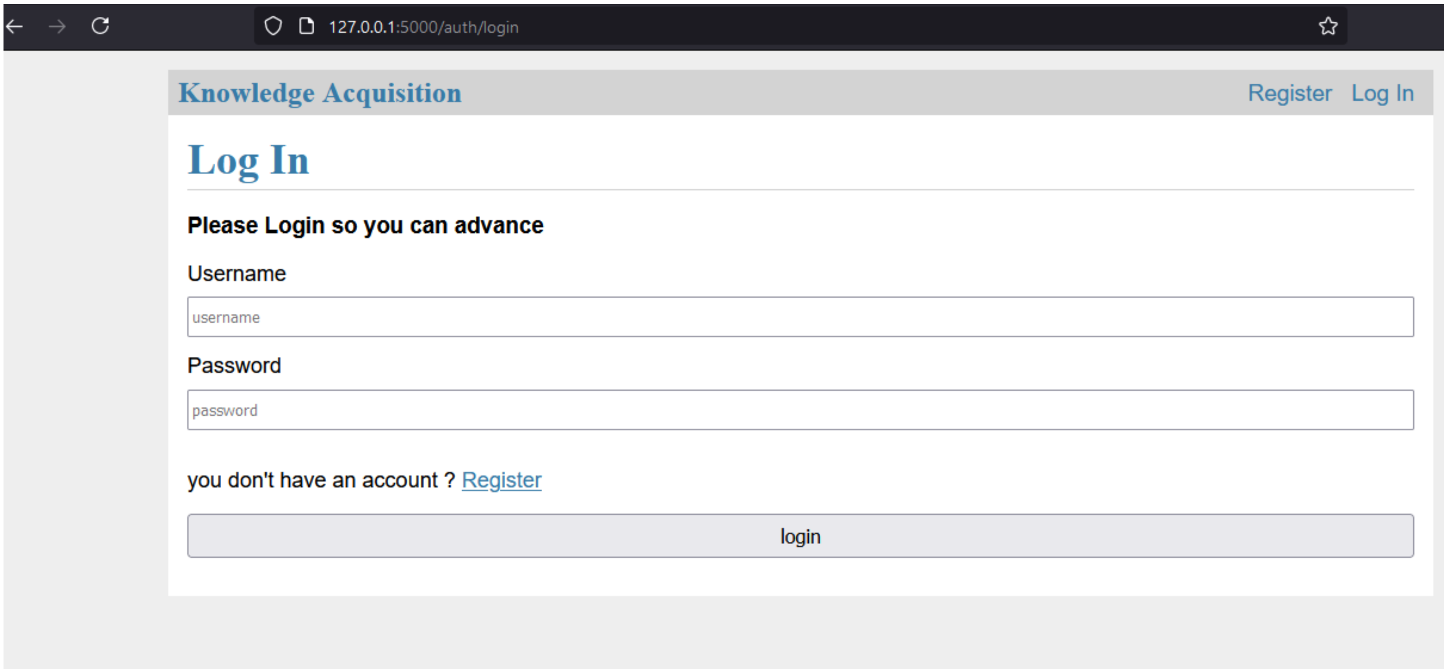


Figure 5.3: KA process actors Login interface

The next figure represents the interface of the KA process actor home page.

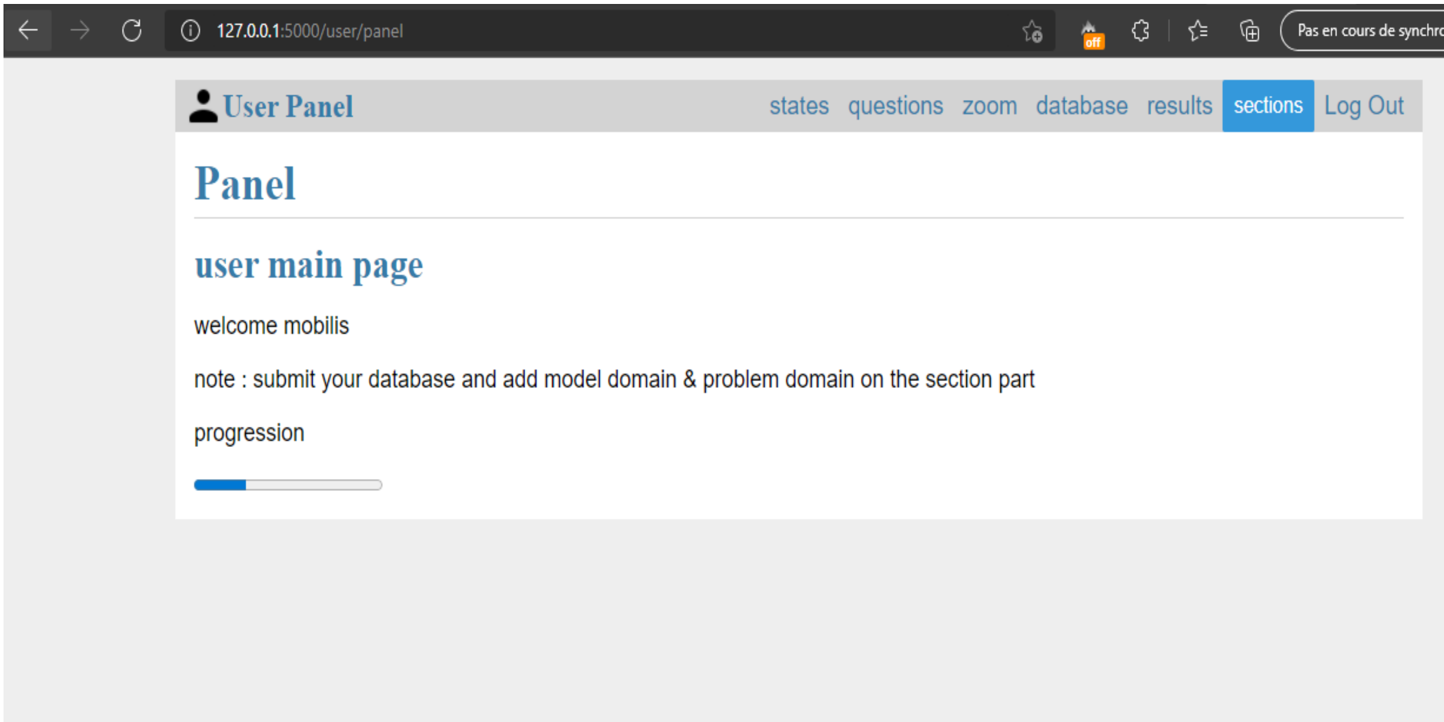


Figure 5.4: KA process ACTORS home page interface

The next figure represents an interface of the KE login page.

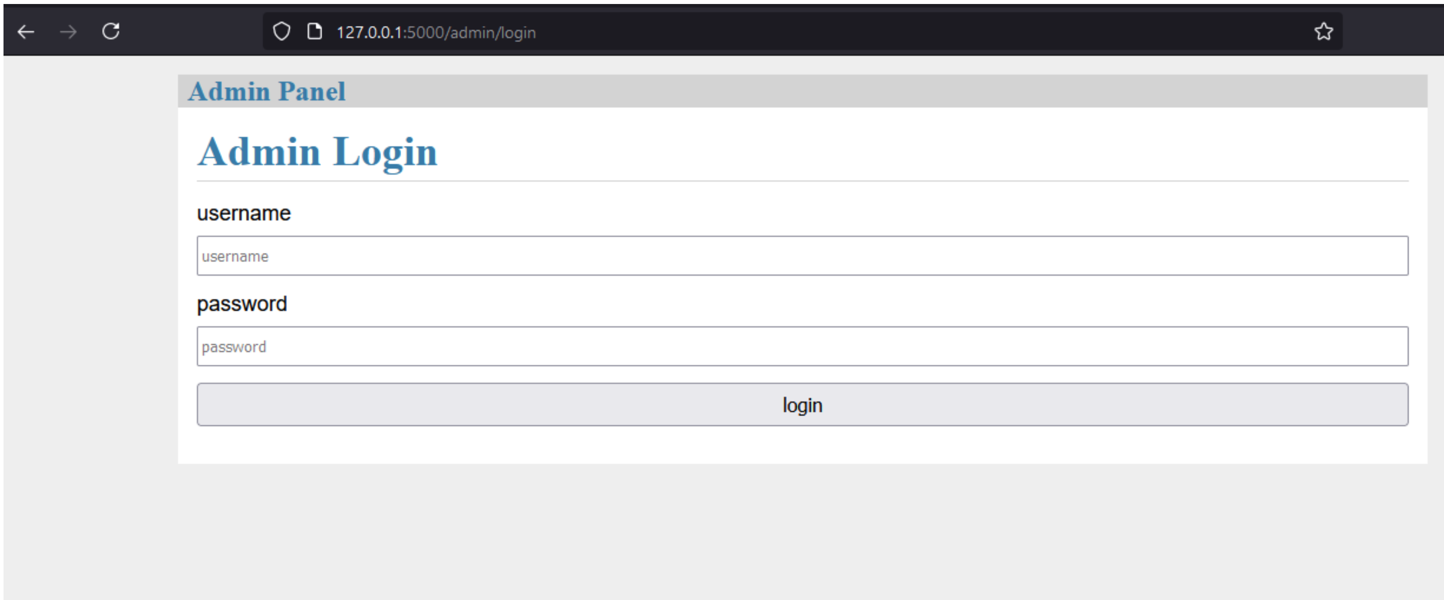


Figure 5.5: The KE login interface

Scenario 2: Questionnaire, zoom meeting & explicit Knowledge sources:

In this step the KE is going to update the user state and inform him to propose a zoom meeting date, figure 5.11.

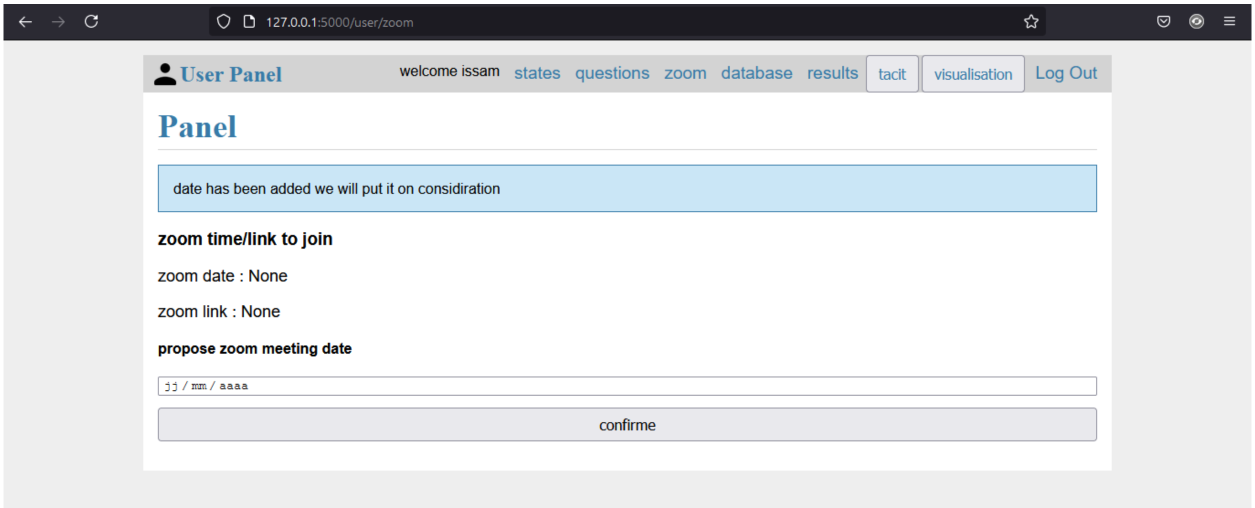


Figure 5.13: interface for proposing meeting date by KA process side.

The next figure represents the interface of the meeting confirmation date.

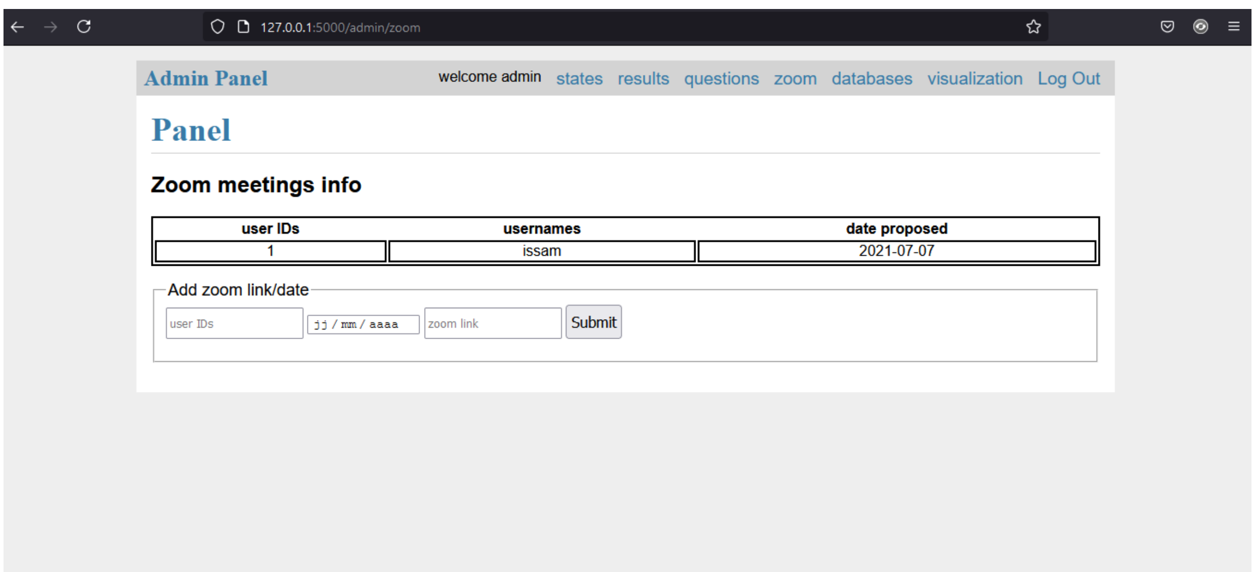


Figure 5.14: interface of the meeting date confirmation by the KE.

After the interview, the KE will prepare a questionnaire using google form and add its link for the expert in order to answer it as shown in the Figure 5.13.

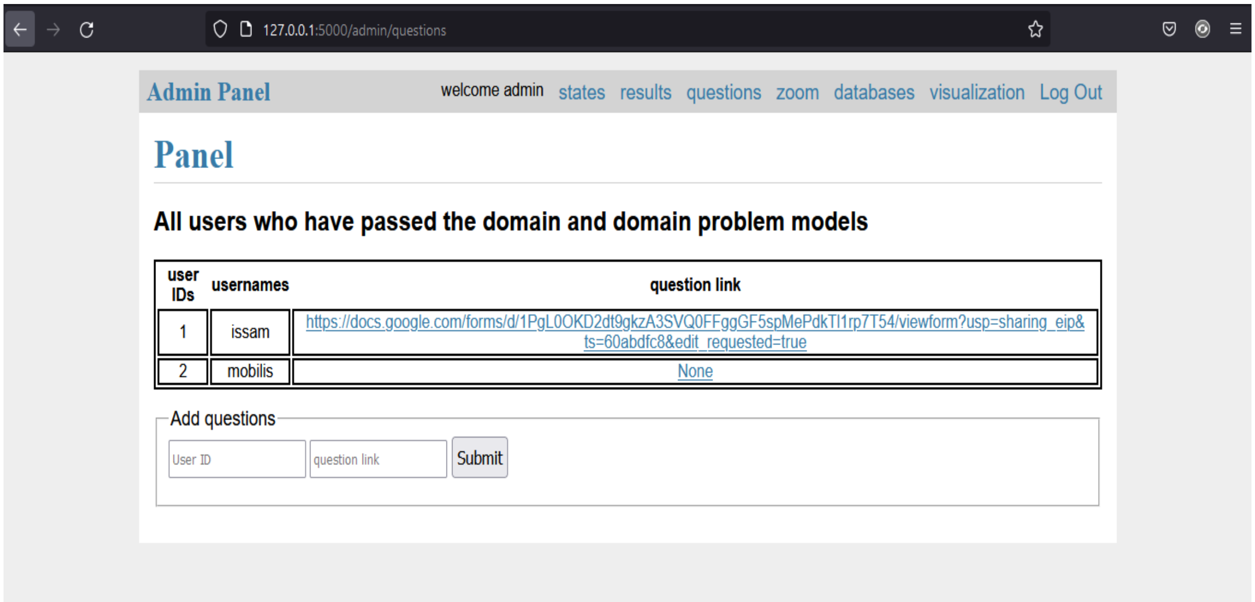


Figure 5.15: interface for uploading the link of the questionnaire by the KE.

The **figure 5.16** below will visualize all the questionnaires (the links) that the KA process actor should answer.

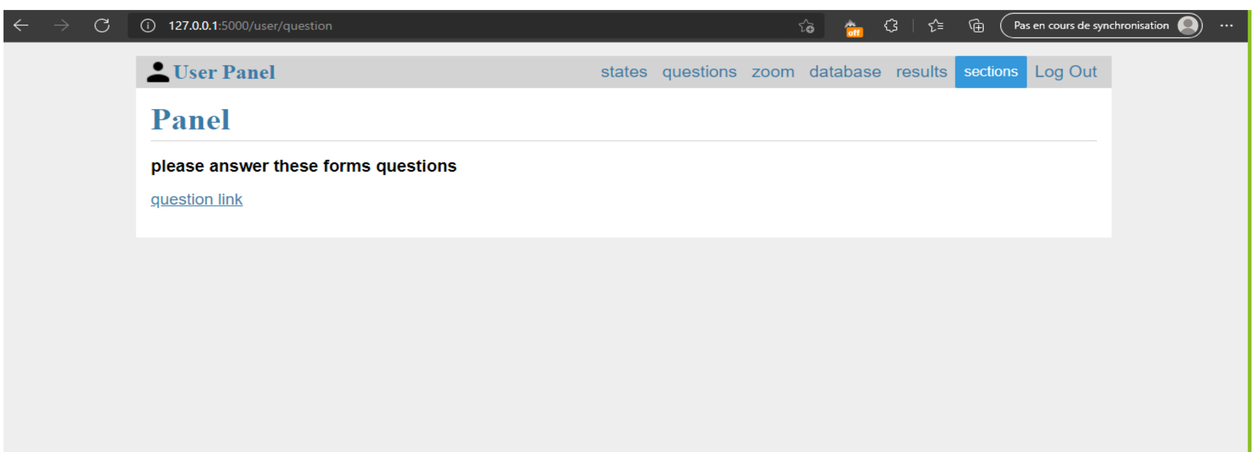


Figure 5.16: interface for visualizing all the questionnaires for the KA process actor.

The next figure represents one of our questionnaires which is uploaded on google service (google form).

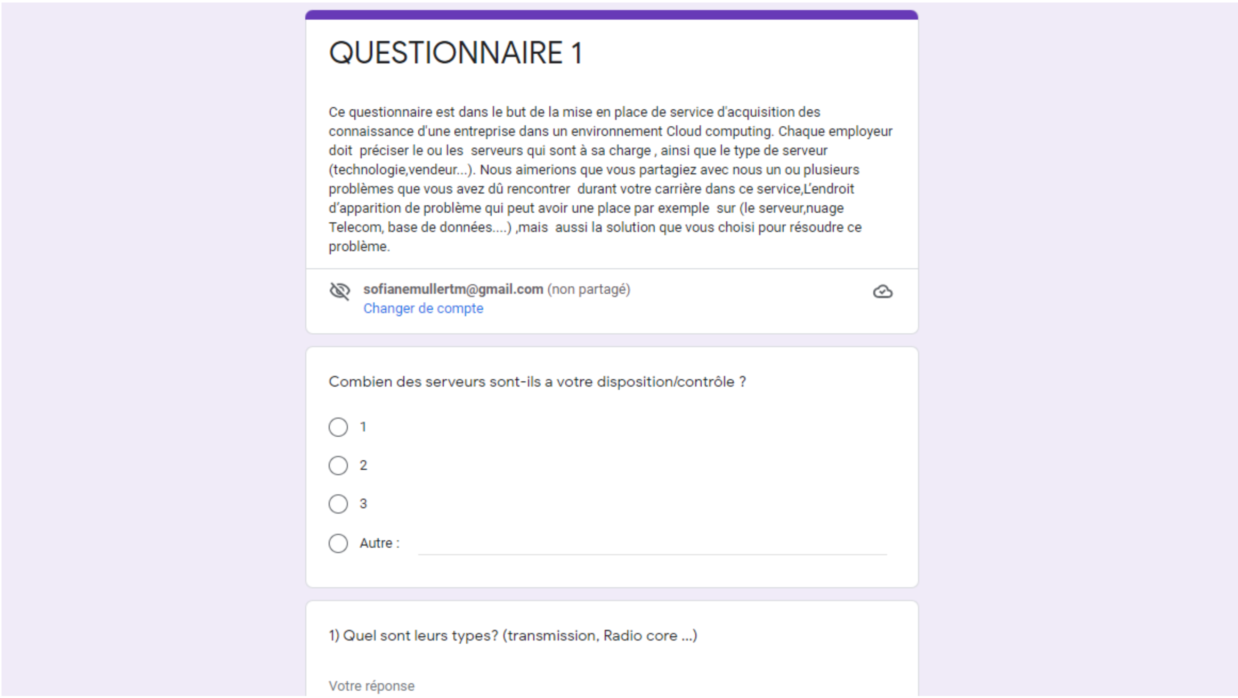


Figure 5.17: example of a questionnaire.

The expert is asked to push the Explicit knowledge sources to a google drive then linking it into the site, as shown in **figure 5.18**.

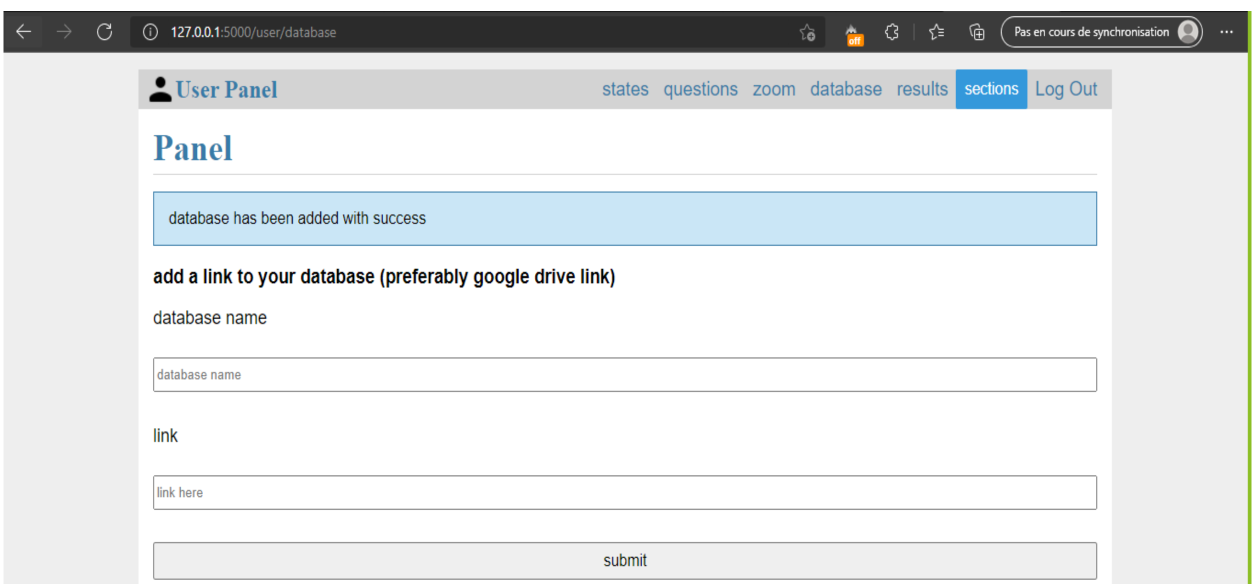


Figure 5.18: interface for uploading the link of the explicit Knowledge resources

Scenario 3: Creation of Domain and Domain problem models

After the expert has logged in, he waits for the admin to update his state.

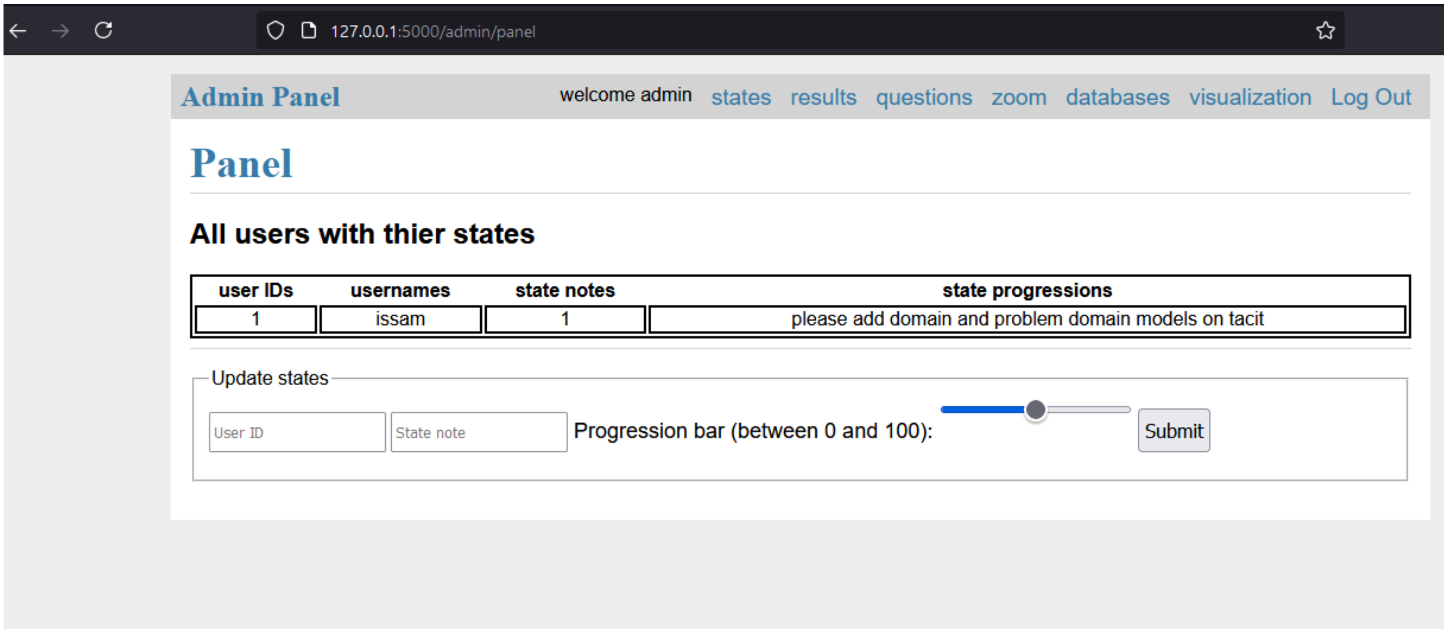


Figure 5.6: Interface of KA process progression (KE side).

Then the KA process actor follows the state given to him in this case.

First, the expert starts with the DOMAIN MODEL so, he adds all the objects and their related attributes with his expertise domain, as shown in the figure below.

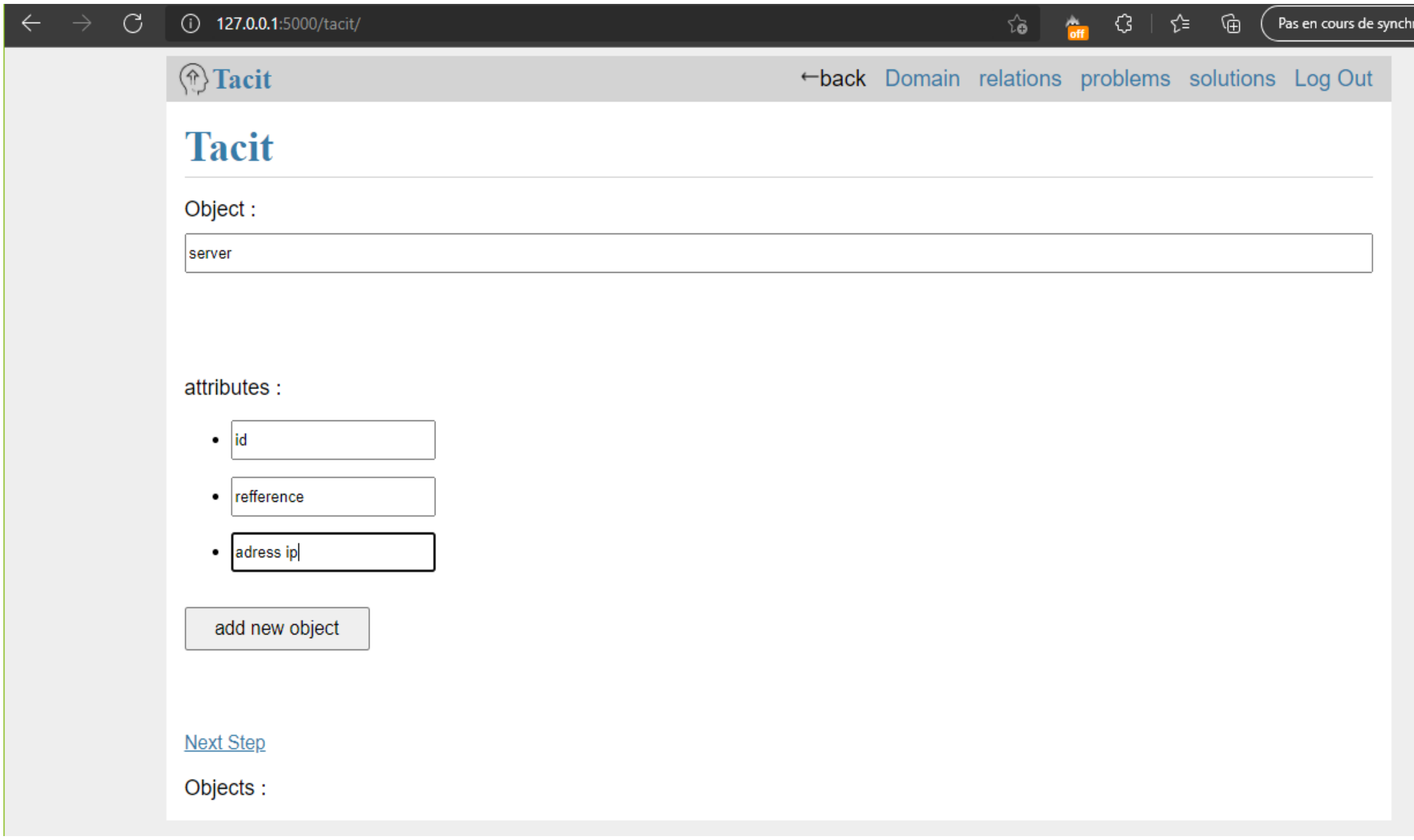


Figure 5.7: Domain model creation interface.

Then he inserts multiple relations between domain objects if exists, as represented in the figure below.

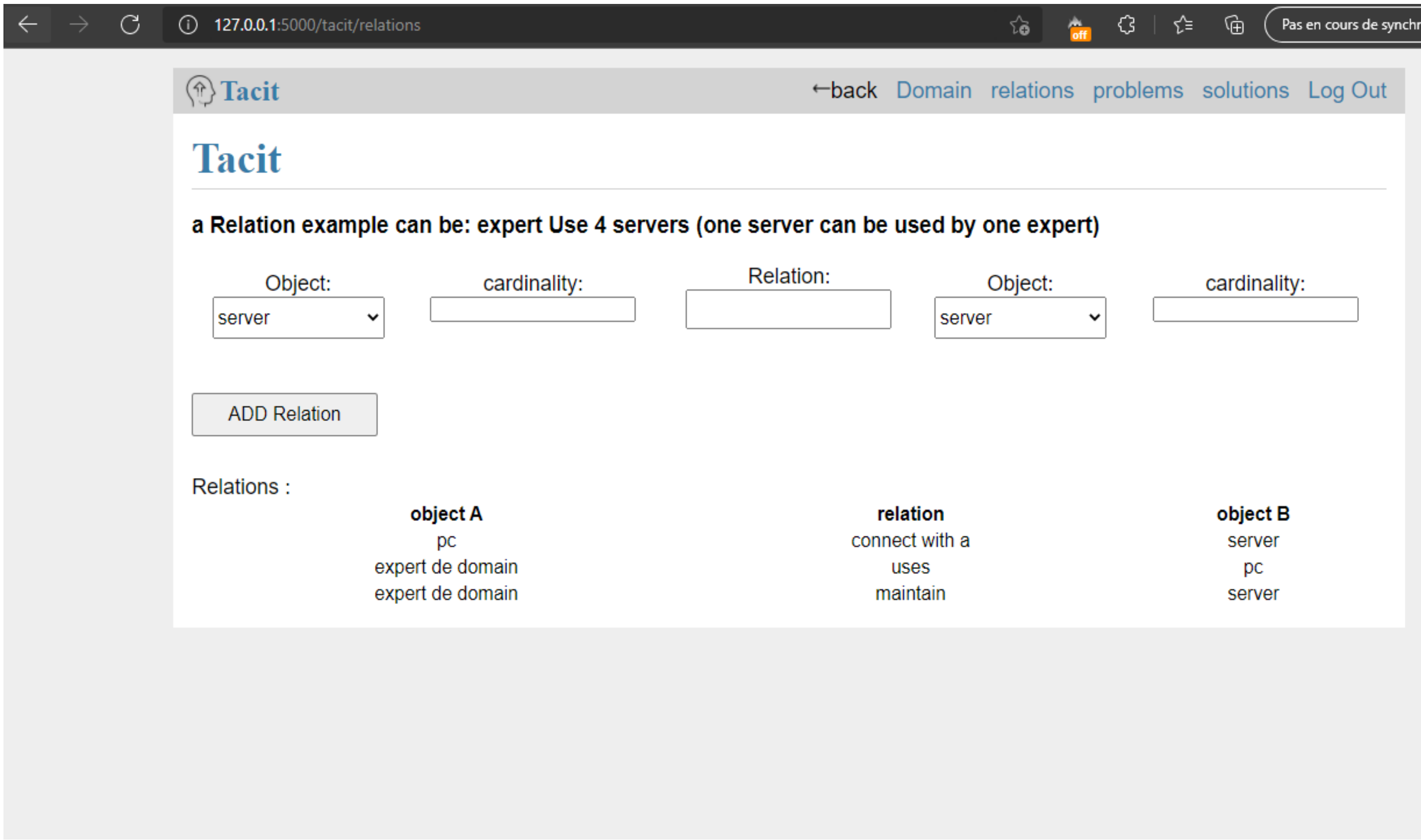


Figure 5.8: relation between objects (complex type) interface

Then he moves to PROBLEM DOMAIN MODEL.

The **Figure 5.7**, **5.8** and **5.9** represent the interface to create the Domain Problem model.

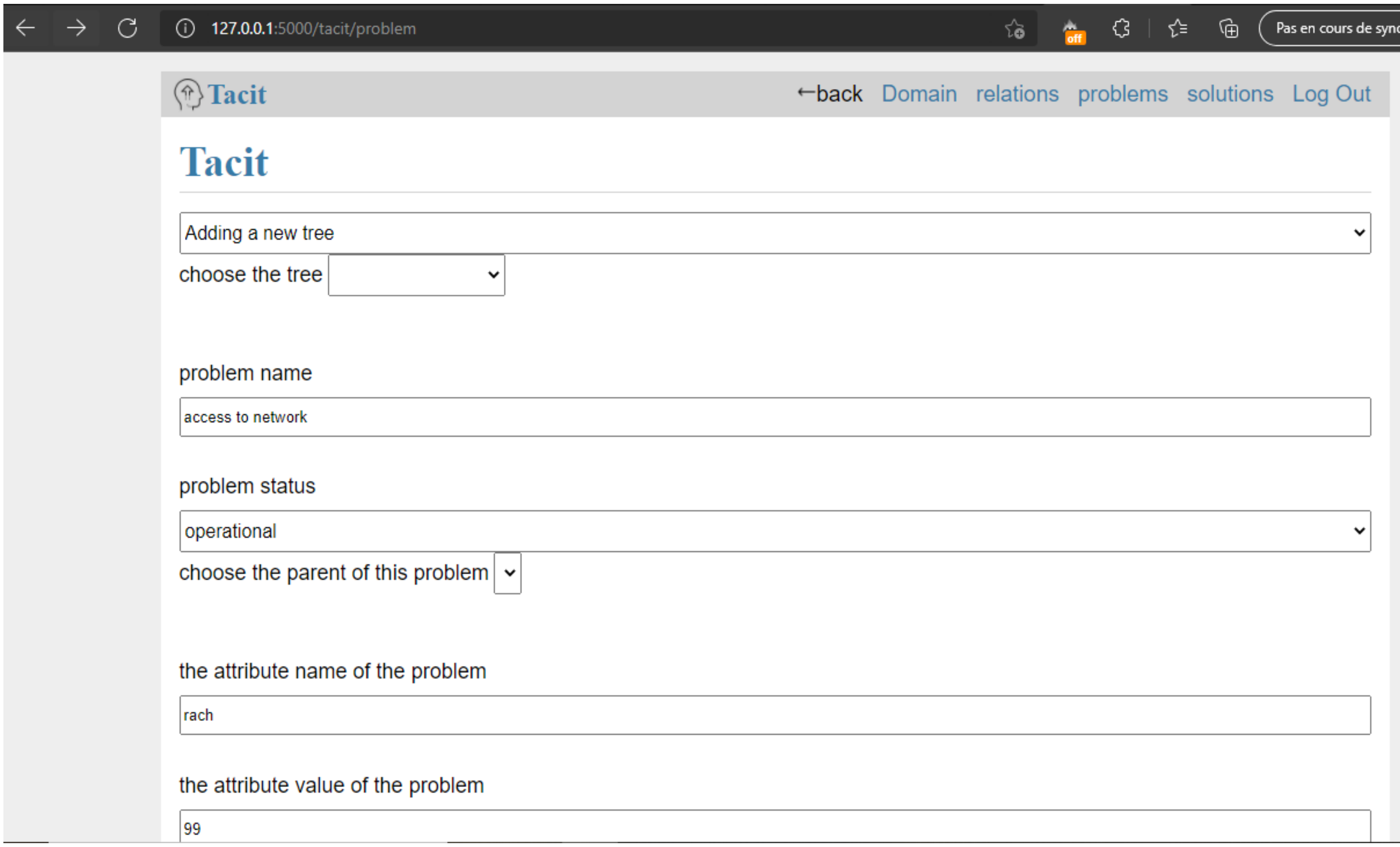


Figure 5.9: interface of the domain problem model creation. (Part 1)

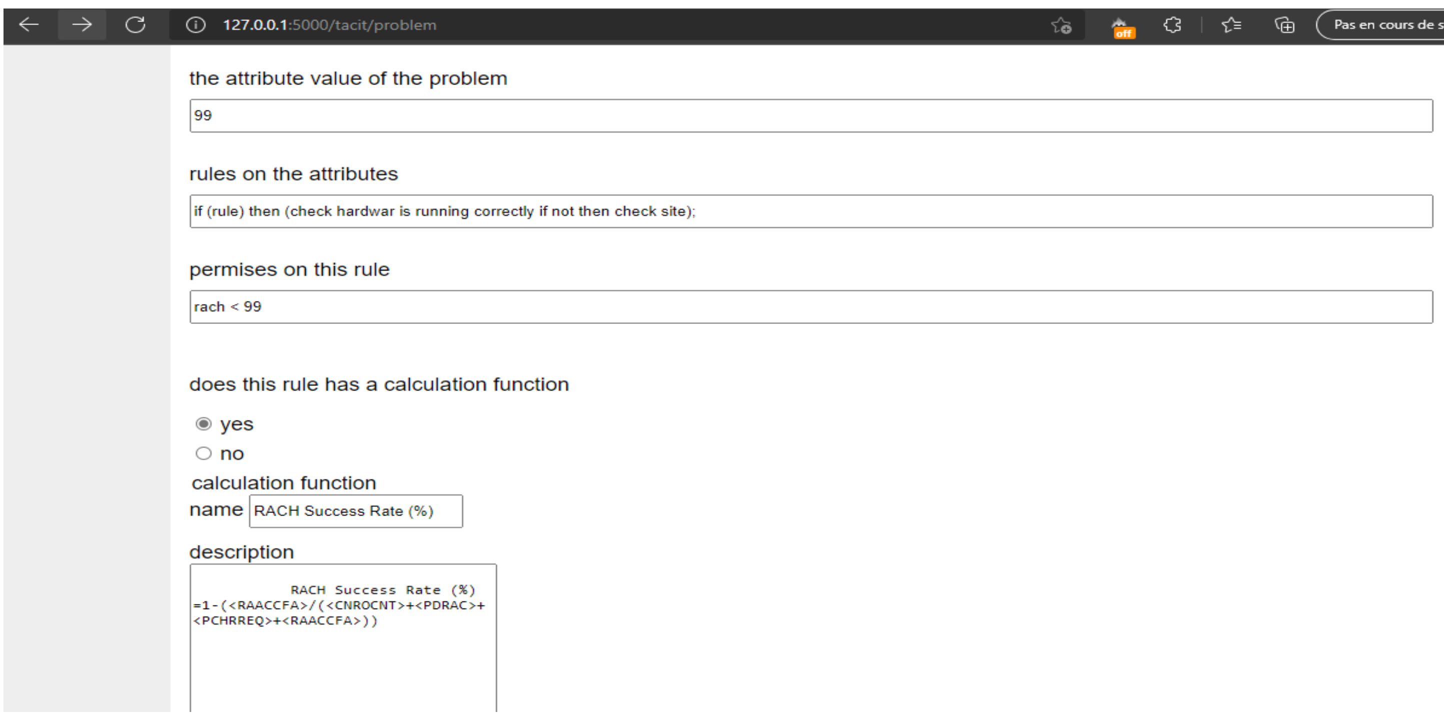


Figure 5.10: interface of the domain problem model creation. (Part 2)

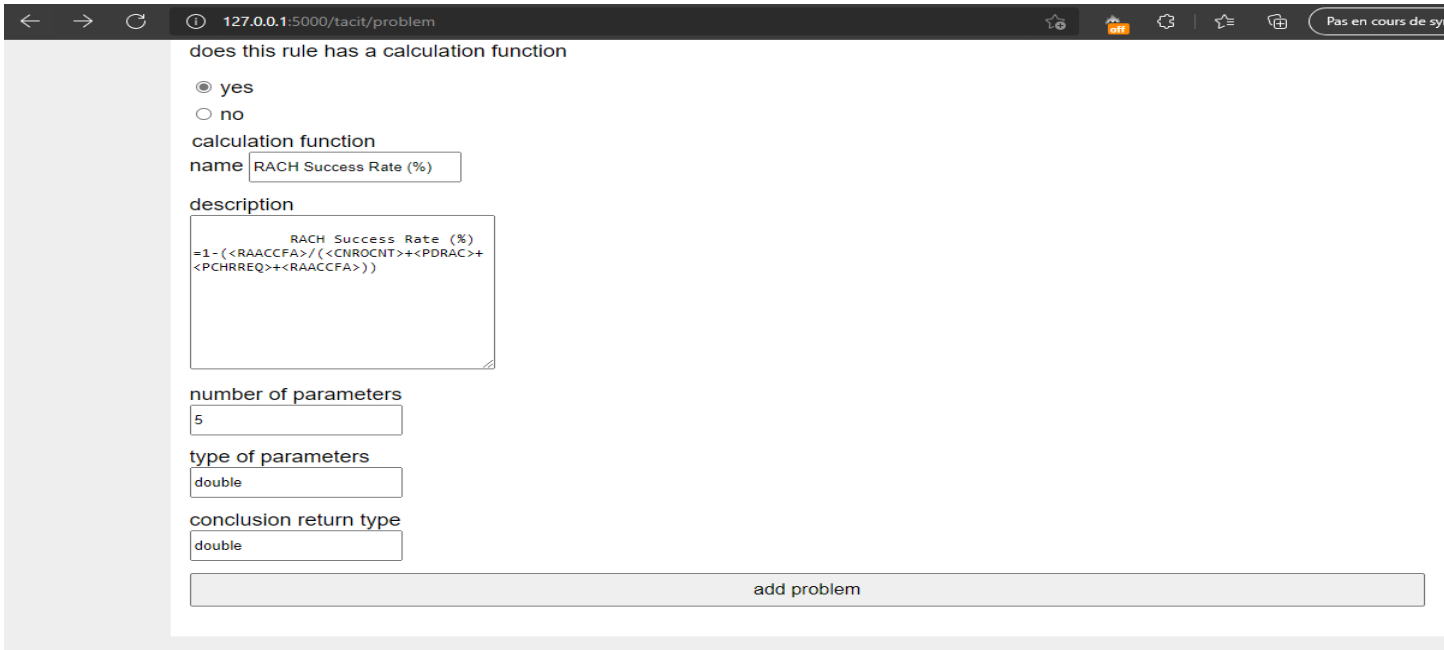


Figure 5.11: interface of the domain problem model creation. (Part 3)

Finally, he specifies solution to each problem with their steps, if the problem has a solution, as it is shown in the next Figure.

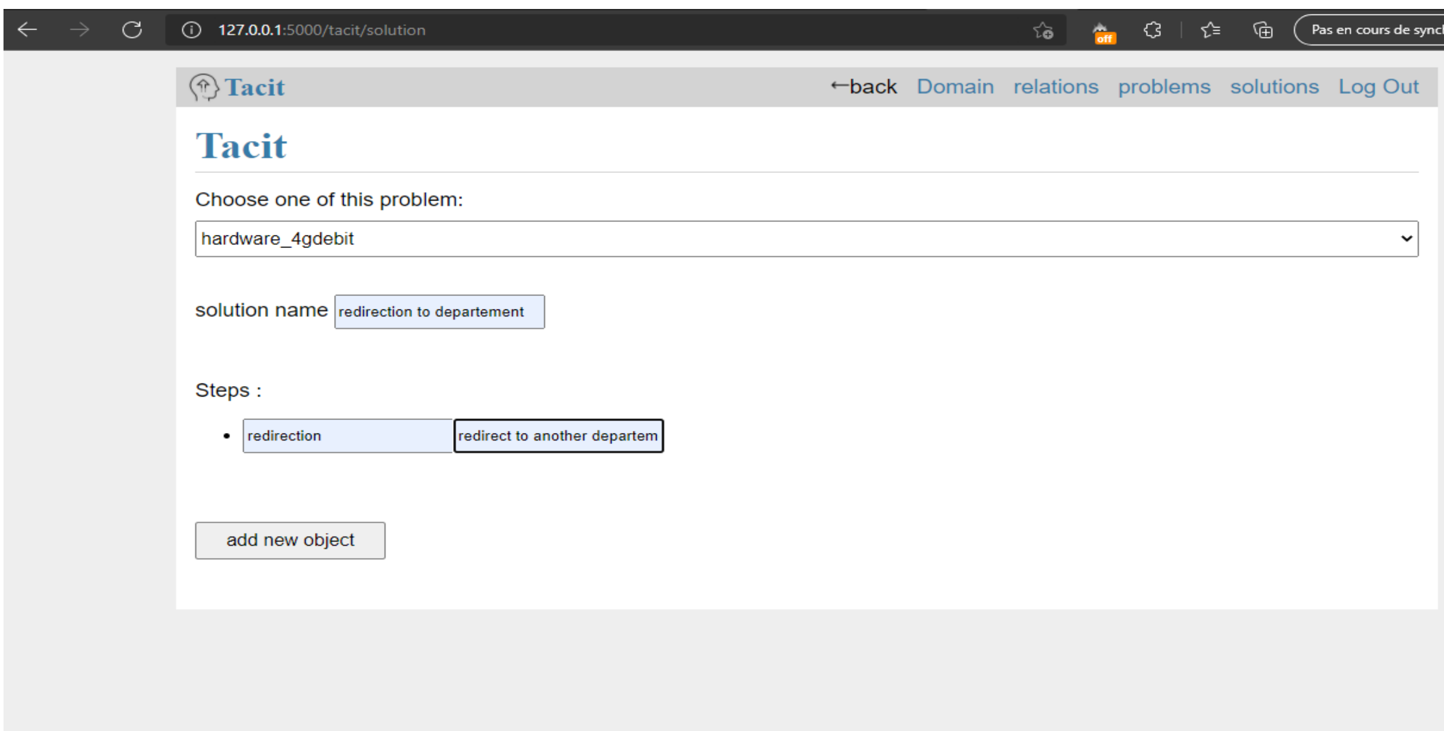


Figure 5.12: insert resolution interface

In case the KE was his role to do the creation of model then he uses the information taken from zoom meeting and the questionnaire answered by expert

Scenario 4: visualization & results

After the creation of the models, the expert can visualize them.

The next figure depicts the DOMAIN MODEL VISUALIZATION which is created by the KA process actor.

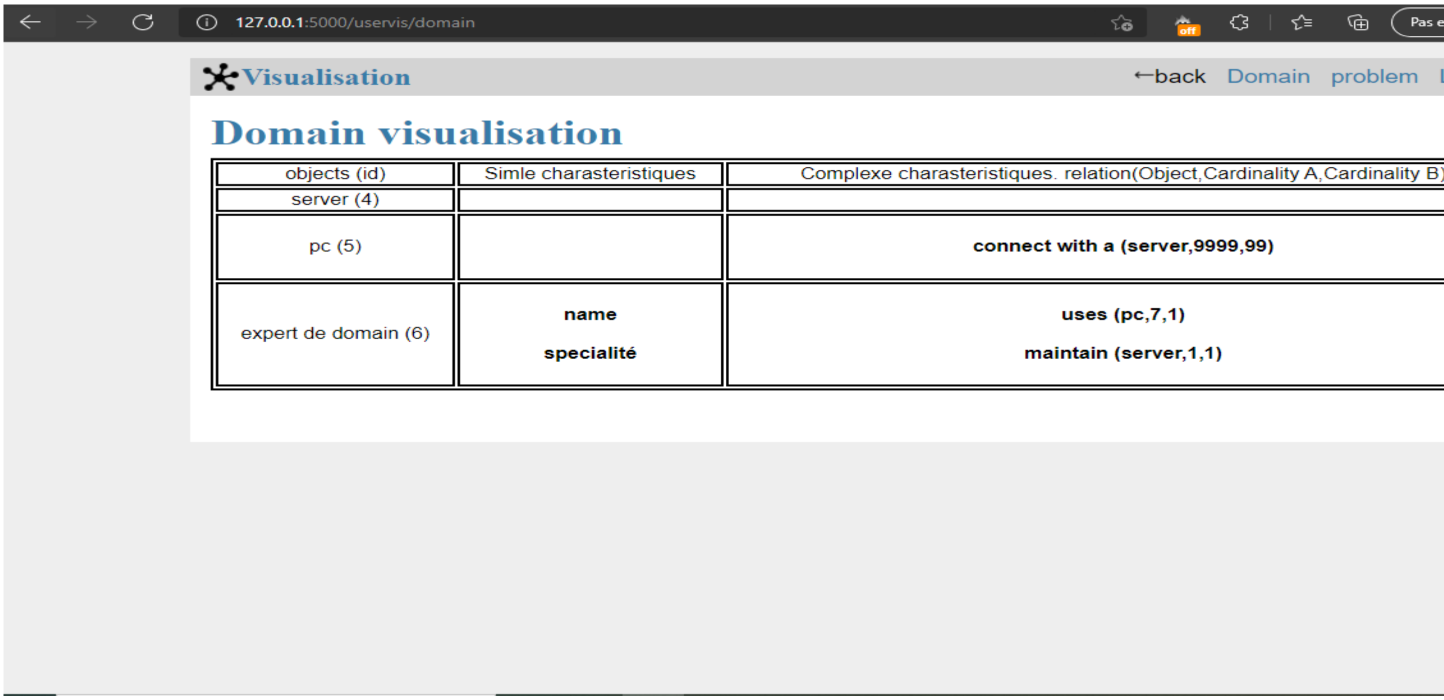


Figure 5.19: interface for the domain model visualization.

The figures below represent the PROBLEM DOMAIN VISUALIZATION steps. The **figure 5.20** represents the first step which is "Chose a problem tree"

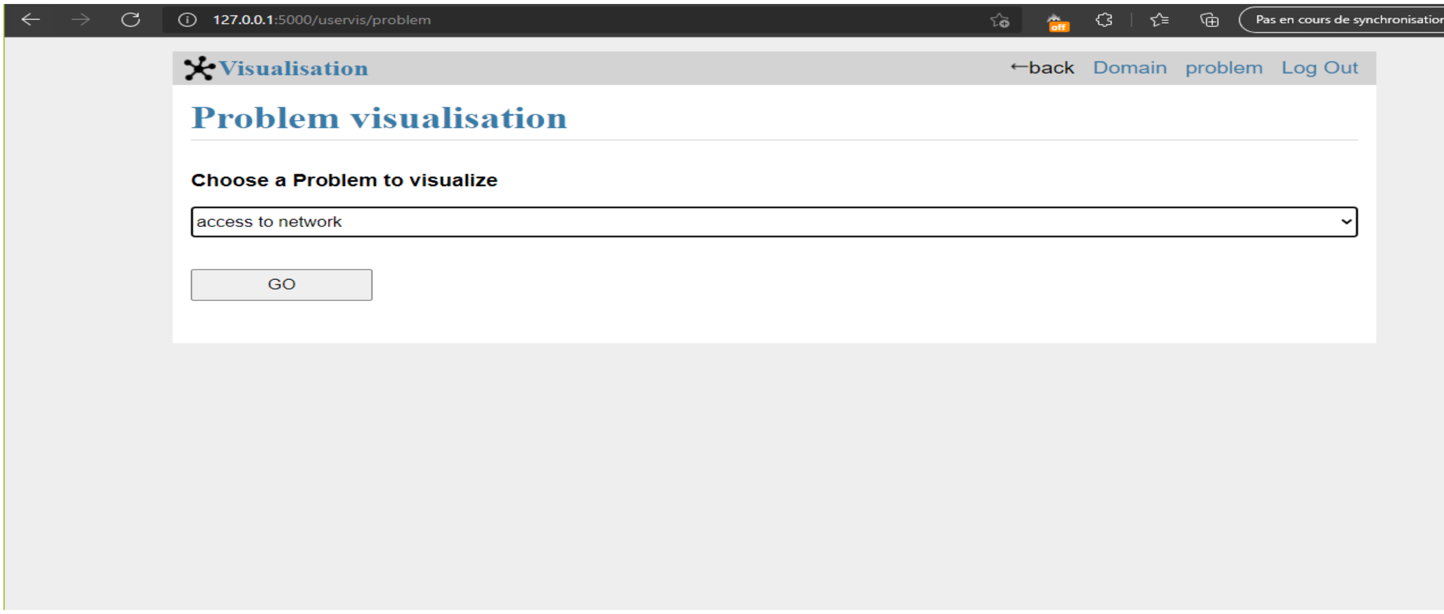


Figure 5.20: interface for the domain problem visualization. (Part 1)

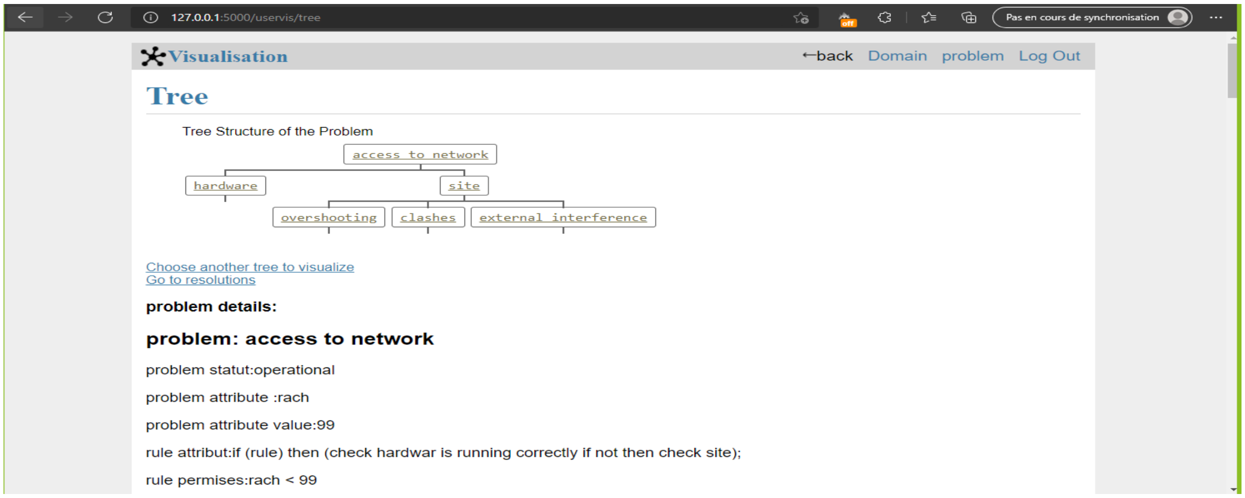
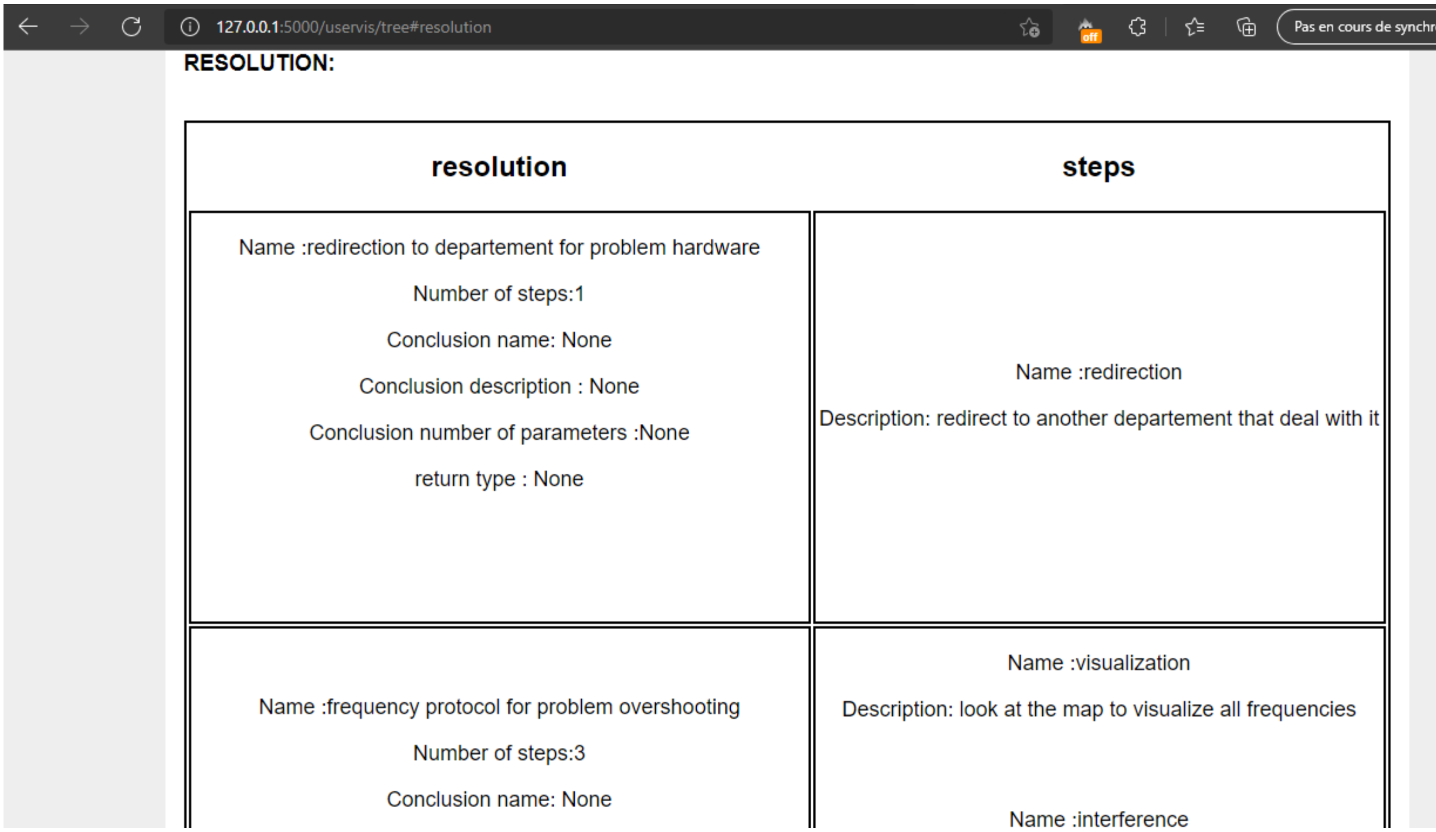


Figure 5.21: interface for the domain problem visualization. (Part 2)

For the VISUALIZATION OF DOMAIN PROBLEM RESOLUTION, you move to the resolution section as shown in the **figures 5.22, 5.23**.



RESOLUTION:

resolution	steps
Name :redirection to departement for problem hardware Number of steps:1 Conclusion name: None Conclusion description : None Conclusion number of parameters :None return type : None	Name :redirection Description: redirect to another departement that deal with it
Name :frequency protocol for problem overshooting Number of steps:3 Conclusion name: None	Name :visualization Description: look at the map to visualize all frequencies Name :interference

Figure 5.22: interface for domain problem model resolutions. (Part 1)

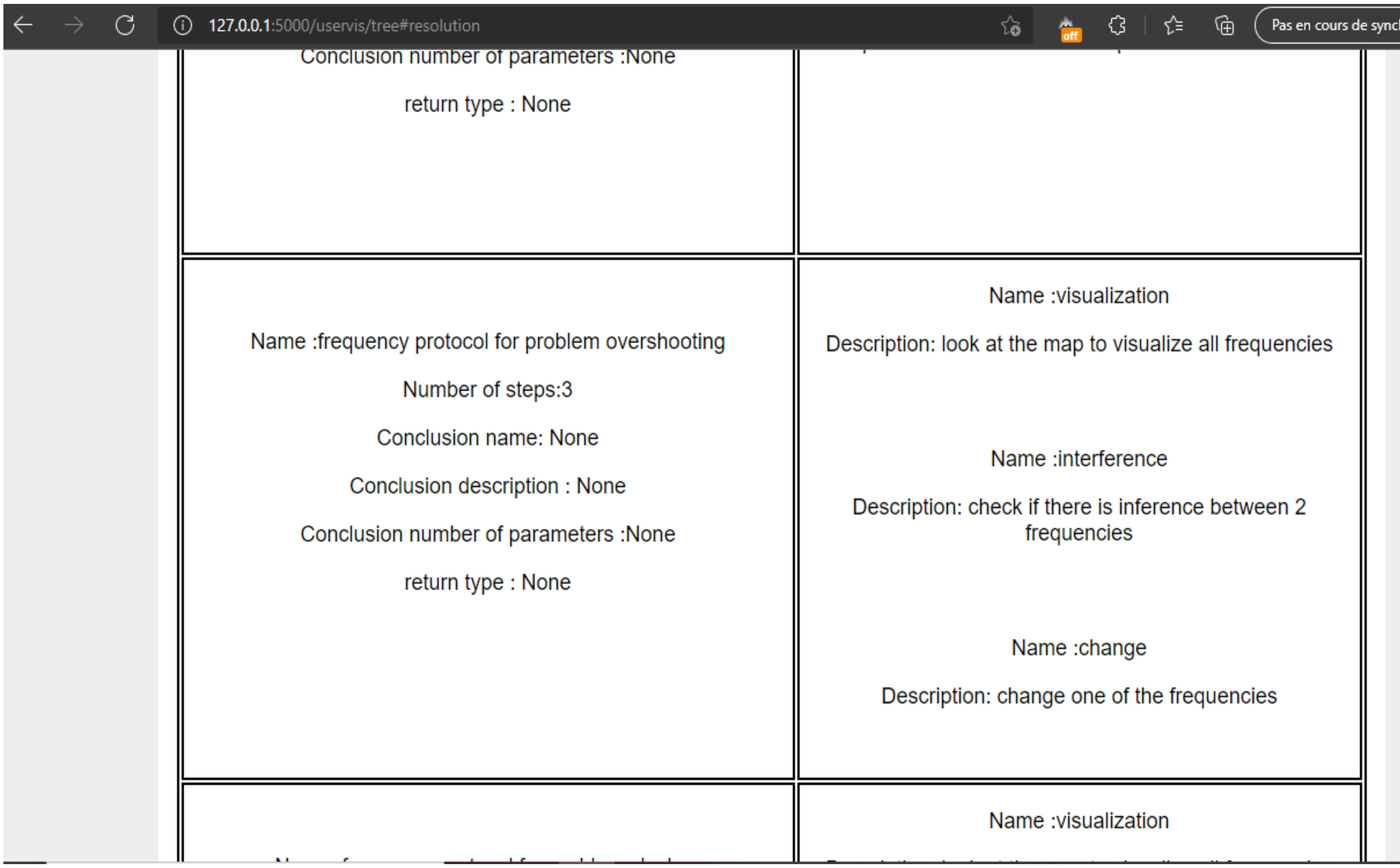


Figure 5.23: interface for domain problem model resolutions. (Part 1)

As shown in **Figure 5.24**, one the KE have done the KDD process, he will push the results (PDF) containing all knowledge acquired and some critical notes,

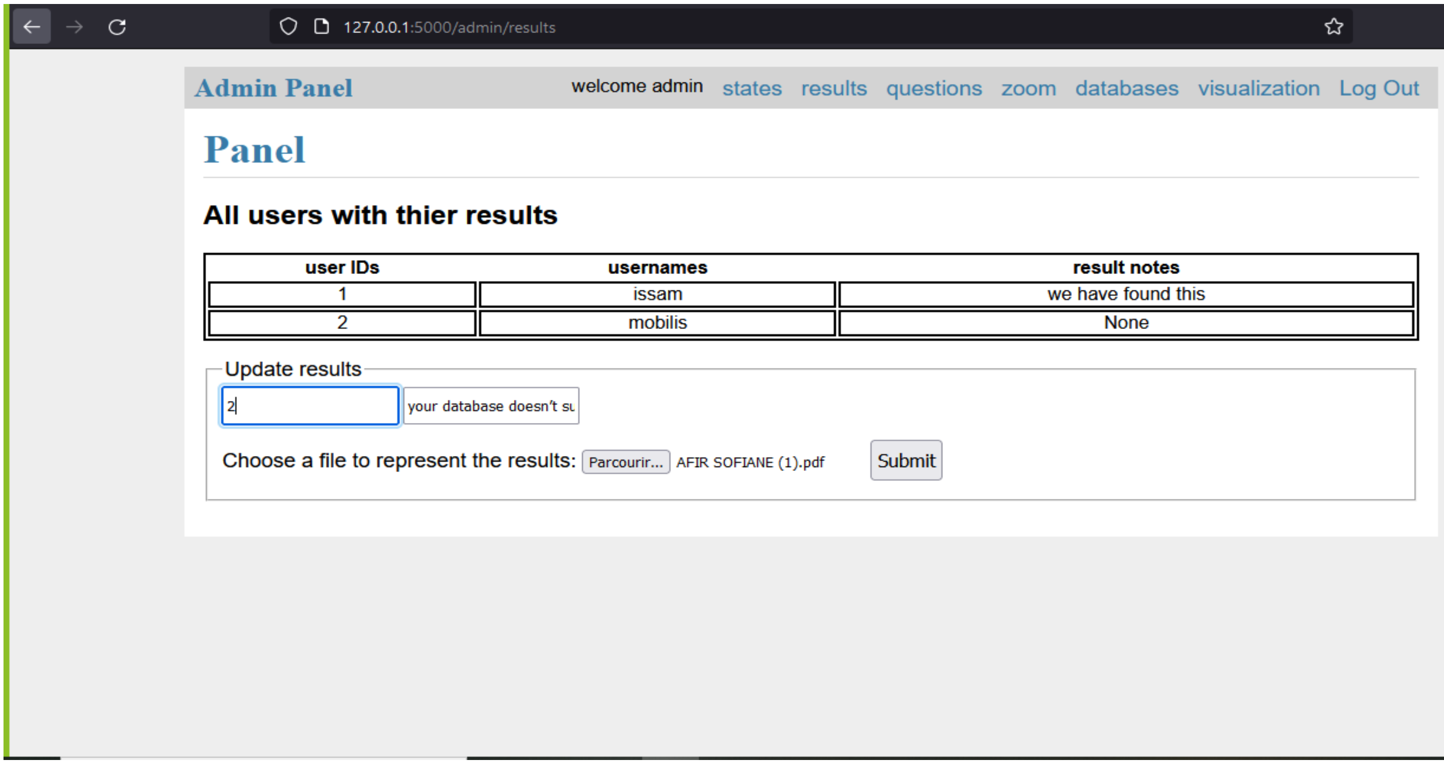


Figure 5.24: Interface for uploading the acquired knowledge.

When the KA process is completed, the figure below shows that the KA process actor has the possibility to download the acquired knowledge in a pdf file.

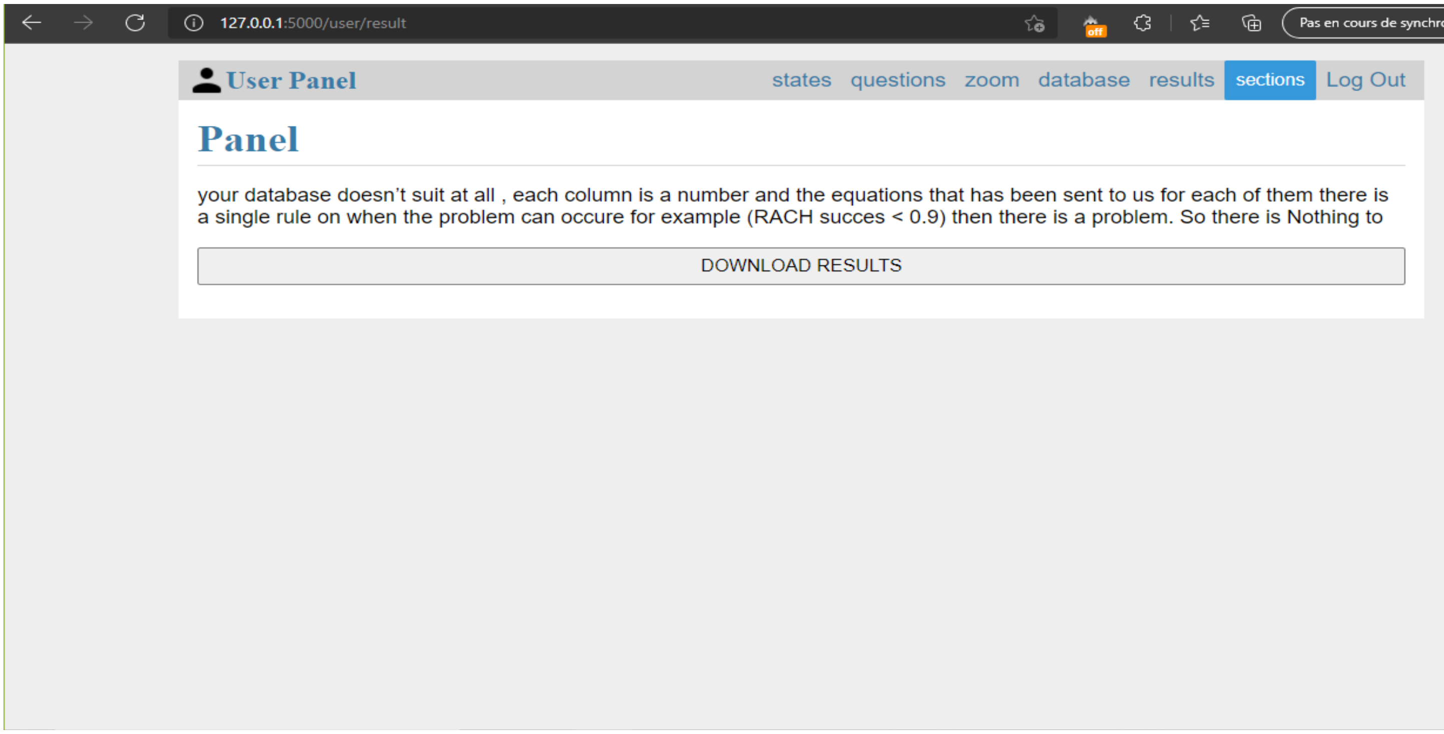


Figure 5.25: Interface for the ability to download the acquired knowledge by the KA process ACTOR.

5.6 Presentation of the KDD process for the Explicit knowledge Acquisition

Mobilis company has presented to us two databases one for 4G users and the other for 2G with a set of rules for each database:

Salam

KPI 2G	Formula
RACH Success Rate (%)	=1-(<RAACCFA>/(<CNROCNT>+<PDRAC>+<PCHRREQ>+<RAACCFA>))
TCH Drop Rate (%)	=(<TFNCEDROPSUB>+<TFNCEDROP>+<THNCEDROPSUB>+<THNCEDROP>)/(<TFNDROP>+<TFNDROPSUB>+<THNDROP>+<THNDROPSUB>+<DISNORM>)
Handover Success Rate (%)	=<HOVERSUC>/<HOVERCNT>
Subscriber Perceived TCH Congestion	=(<CNRELCONG>+<TFNRELCONG>+<TFNRELCONGSUB>+<THNRELCONG>+<THNRELCONGSUB>)/<TASSALL>

KPI 4G	Formula
Avg DL user throughput (Mbps)	=(<pmPdcpVolDlDrb>-<pmPdcpVolDlDrbLastTTI>)/<pmUeThpTimeDl>
Taux de coupure - PS	=(<pmErabRelAbnormalEnbAct> + <pmErabRelAbnormalMmeAct>) / (<pmErabEstabSuccInit> + <pmErabEstabSuccAdded>)
Taux d'échec E-RAB (all)	=1-(<pmErabEstabSuccInit>+<pmErabEstabSuccAdded>)/(<pmErabEstabAttInit>+<pmErabEstabAttAdded>)
Taux d'échec RRC	=1-<pmRrcConnEstabSucc>/(<pmRrcConnEstabAtt>-<pmRrcConnEstabAttReatt>)

https://drive.google.com/drive/folders/1HhedAbdJJoY6Qgim6y_4Bn0tGVt7tc9w?usp=sharing

Cordialement

Figure 5.26: an email including different set of rules sent by mobilis.

We started by importing the 2G database then showing the first 5 row

```
In [1]: import pandas as pd

        df = pd.read_csv('DC.DC_E_BSS_CELL_CS_DAY.csv')
```

```
In [2]: #get the first five rows
        df.head(5)
```

Out[2]:

	Date	Cell Name	Site Name	BSC Name	RAACCFA	CNROCNT	TFNCEDROPSUB	TFNCEDROP	THNCEDROP	THNCEDROPSUB	...	CNRELCONG	TFNREL
0	2020/06/17 00:00:00	021001A	021001A	BECHLFR	3.0	3108.0	0.0	0.0	0.0	0.0	...	0.0	
1	2020/06/17 00:00:00	021001B	021001B	BECHLFR	3.0	10214.0	0.0	27.0	23.0	0.0	...	1.0	
2	2020/06/17 00:00:00	021001C	021001C	BECHLFR	51.0	3571.0	0.0	3.0	1.0	0.0	...	0.0	
3	2020/06/17 00:00:00	021002A	021002	BECHLFR	12.0	18947.0	0.0	9.0	44.0	0.0	...	0.0	
4	2020/06/17 00:00:00	021002B	021002	BECHLFR	18.0	8237.0	0.0	21.0	1.0	0.0	...	0.0	

5 rows × 25 columns

Figure 5.27: interface of jupyter notebook showing 5 rows of the 2G database.

We needed more informations about the database columns.

```
In [7]: #get number of (row, column)
        df.shape

Out[7]: (463854, 25)
```

Figure 5.28: interface of jupyter notebook showing shape of the 2G database.

```
In [8]: #get info about the dataframe
        df.info

Out[8]: <bound method DataFrame.info of                    Date Cell Name     Site Name BSC Name  RAACCFA  \
        0       2020/06/17 00:00:00   021001A       021001A  BECHLFR      3.0
        1       2020/06/17 00:00:00   021001B       021001B  BECHLFR      3.0
        2       2020/06/17 00:00:00   021001C       021001C  BECHLFR     51.0
        3       2020/06/17 00:00:00   021002A        021002  BECHLFR     12.0
        4       2020/06/17 00:00:00   021002B        021002  BECHLFR     18.0
        ...                     ...       ...           ...      ...      ...
        463849  2021/06/11 00:00:00    34805B    BAB_ESSOUK   BEBBA1     29.0
        463850  2021/06/11 00:00:00    42218F           NaN  BECHLFR      0.0
        463851  2021/06/11 00:00:00    43675A  BENI_HAROUNE  JIJEVO2     16.0
        463852  2021/06/11 00:00:00    43675B  BENI_HAROUNE  JIJEVO2     25.0
        463853  2021/06/11 00:00:00    43675D        43675D  JIJEVO2     27.0

                CNROCNT  TFNCEDROPSUB  TFNCEDROP  THNCEDROP  THNCEDROPSUB  ...  \
        0        3108.0           0.0        0.0        0.0           0.0  ...
        1       10214.0           0.0       27.0       23.0           0.0  ...
        2        3571.0           0.0        3.0        1.0           0.0  ...
        3       18947.0           0.0        9.0       44.0           0.0  ...
        4        8237.0           0.0       21.0        1.0           0.0  ...
        ...         ...           ...        ...        ...           ...  ...
        463849   1202.0           0.0        1.0        0.0           0.0  ...
        463850      0.0           0.0        0.0        0.0           0.0  ...
        463851  18532.0           0.0       12.0       13.0           0.0  ...
```

Figure 5.29: interface of jupyter notebook showing info of the 2G database.

Out[9]:

	RAACCFA	CNROCNT	TFNCEDROPSUB	TFNCEDROP	THNCEDROP	THNCEDROPSUB	TFNDROP	TFNDROPSUB	THNDROP	THNDR(
count	453673.000000	453673.000000	453697.0	453697.000000	453673.000000	453673.0	453697.000000	453697.0	453673.000000	45
mean	248.302440	20426.890641	0.0	18.004100	38.139700	0.0	23.191399	0.0	49.159174	
std	1569.749425	19150.175230	0.0	23.963396	59.492315	0.0	46.116950	0.0	76.731541	
min	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	
25%	10.000000	7544.000000	0.0	4.000000	5.000000	0.0	6.000000	0.0	6.000000	
50%	30.000000	15157.000000	0.0	11.000000	20.000000	0.0	15.000000	0.0	26.000000	
75%	80.000000	27833.000000	0.0	24.000000	50.000000	0.0	30.000000	0.0	65.000000	
max	309253.000000	643257.000000	0.0	1592.000000	5028.000000	0.0	22019.000000	0.0	7036.000000	

8 rows × 21 columns

Figure 5.30: interface of jupyter notebook showing description of the 2G database.

As highlighted there are 453673 rows for RAACCFA column and when we printed the shape of the dataframe the number of rows was 463854, clearly some rows are empty so we had to do some cleaning by using this function.

In [5]:
```
def clean_dataset(df):
    assert isinstance(df, pd.DataFrame), "df needs to be a pd.DataFrame"
    df.dropna(inplace=True)
    indices_to_keep = ~df.isin([np.nan, np.inf, -np.inf]).any(1)
    return df[indices_to_keep].astype(np.float64)
```

Figure 5.31: interface of jupyter notebook showing a function to clean a dataframe.

We didn't have any information about the columns only that they served to calculat the rules given by mobilis, but incase we would miss an information we decided to visualize some columns.

In [15]:
```
#histogram of 3 diffrent column
df.hist(['RAACCFA','CNROCNT','TFNCEDROPSUB'])
```

Out[15]:
```
array([[<AxesSubplot:title={'center':'RAACCFA'}>,
        <AxesSubplot:title={'center':'CNROCNT'}>],
       [<AxesSubplot:title={'center':'TFNCEDROPSUB'}>, <AxesSubplot:>]],
      dtype=object)
```

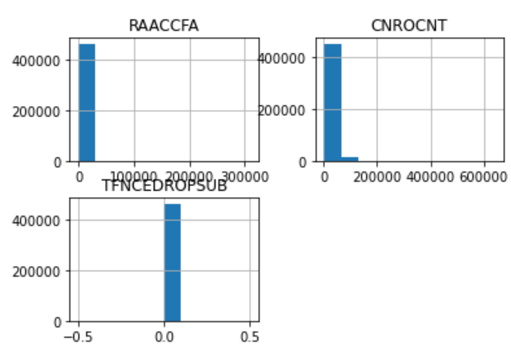


Figure 5.32: interface of jupyter notebook showing a histogram of some columns of the 2G

database.

```
In [16]: import seaborn as sns

In [25]: sns.boxplot(data = df[['RAACCFA','CNROCNT']])
Out[25]: <AxesSubplot:>
```

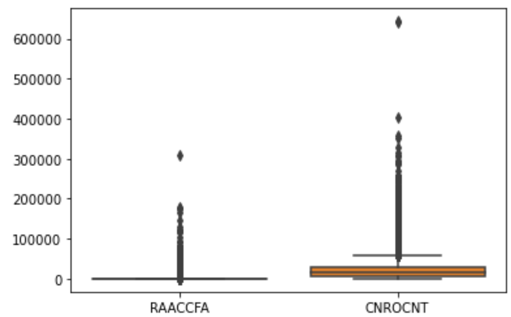


Figure 5.33: interface of jupyter notebook showing a plot of some columns of the 2G database.

As concluded before we didn't get any useful information so we moved to the next step. Before we do anything else, we had to calculate the rules and integrate them in the database.

```
In [2]: #create new columns with the rule given
#save df copy incase
X=df.copy()
#RACH_Success_Rate=1-(<RAACCFA>/(<CNROCNT>+<PDRAC>+<PCHRREQ>+<RAACCFA>))
df['RACH_Success_Rate'] = 1 - ( df['RAACCFA']/(df['CNROCNT']+df['PDRAC']+df['PCHRREQ']+df['RAACCFA']) )

#TCH Drop Rate (%)=(<TFNCEDROPSUB>+<TFNCEDROP>+<THNCEDROPSUB>+<THNCEDROP>)/(<TFNDROP>+<TFNDROPSUB>+<THNDROP>+<THNDROPSUB>+<DISNOI
df['TCH_Drop_Rate']=((df['TFNCEDROPSUB']+df['TFNCEDROP']+df['THNCEDROPSUB']+df['THNCEDROP'])/(df['TFNDROP']+df['TFNDROPSUB']+df['

#Handover Success Rate (%)	=<HOVERSUC>/<HOVERCNT>
df['Handover_Success_Rate']=df['HOVERSUC']/df['HOVERCNT']

#Subscriber Perceived TCH Congestion=(<CNRELCONG>+<TFNRELCONG>+<TFNRELCONGSUB>+<THNRELCONG>+<THNRELCONGSUB>)/<TASSALL>
df['Subscriber_Perceived_TCH_Congestion']=(df['CNRELCONG']+df['TFNRELCONG']+df['TFNRELCONGSUB']+df['THNRELCONG']+df['THNRELCONGSU
```

Figure 5.34: interface of jupyter notebook showing the creation of rule columns.

```
In [10]: df.head(5)
Out[10]:
```

TASSALL	PDRAC	PCHRREQ	HOVERSUC	HOVERCNT	RACH_Success_Rate	TCH_Drop_Rate	Handover_Success_Rate	Subscriber_Perceived_TCH_Congestion
1392.0	4354.0	44.0	3087.0	3097.0	0.999600	0.000000	0.996771	0.000000
4258.0	17297.0	41.0	6123.0	6131.0	0.999891	0.007852	0.998695	0.000235
1666.0	5471.0	44.0	4036.0	4036.0	0.994418	0.001467	1.000000	0.000000
7814.0	31821.0	150.0	15087.0	15106.0	0.999764	0.003806	0.998742	0.000000
2938.0	24999.0	48.0	2981.0	2995.0	0.999459	0.005062	0.995326	0.000000

Figure 5.35: interface of jupyter notebook showing the 5 rows of the updated 2G database with rule columns.

Since we replaced some missed values the result of some rules are 0. Now that we have the necessary columns we move to the datamining. We choose clustering since the majority of columns are numbers but for the sake of this study and implementing the maximum number of KDD task we choose to also do the regression.

We begin with regression, but before we start we have to clarify that it has no real impact since if we have those set of rules given by mobilis we don't have to predict those rules results because it will be calculated mathematically, the only scenario that regression is applied with real impact and can be useful is if the rule are precalculated and given with the database inputs, then in the next time and after the learning faze is done the user will give the columns except the rules columns, then the algorithm going to predict the values of those rules.

For our case we have choose Rach success rate rule to be predicted.

```
In [47]:
#DATA MINGIN
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
Y=df['RACH_Success_Rate']
Y=Y.fillna(0)

X_train, X_test, Y_train, Y_test= train_test_split(X,Y)
lg =LinearRegression()
lg.fit(X_train,Y_train)
lg.score(X_test,Y_test)

Out[47]: 0.9994609375807034
```

Figure 5.36: interface of jupyter notebook showing the training faze of the regression algorithm.

After the training we have a score of 0.99 that the result will be correct.

```
In [50]: Y_pred = lg.predict(X_test)
         print(Y_pred)

         [0.99919441 0.99902396 0.9980309  ... 0.99944827 0.99810921 0.9996041 ]
```

Figure 5.37: interface of jupyter notebook showing the predicted values of the training.

The clustering part is up next, we start by importing the KMeans library "an algorithm mean for clustering" then we save the old dataframe into a variable named X to clean it without worrying about the original one.

```
In [4]: from sklearn.cluster import KMeans
        import numpy as np

In [4]: def clean_dataset(df):
            assert isinstance(df, pd.DataFrame), "df needs to be a pd.DataFrame"
            df.dropna(inplace=True)
            indices_to_keep = ~df.isin([np.nan, np.inf, -np.inf]).any(1)
            return df[indices_to_keep].astype(np.float64)

In [6]: X=df
        X= X.drop(['Date','Site Name','Cell Name','BSC Name'],axis=1)
        clean_dataset(X)
        X = X.reset_index()
```

Figure 5.38: interface of jupyter notebook showing the cleaning of the data.

We have choosed RACH_Success_Rate and Handover_Success_Rate since they both represent success then we put the predicted cluster on column that is named cluster_success, the number of clusters are 2. we gave each cluster a color then we visualize it on plot to see the result clearly.

```
In [49]: km = KMeans(n_clusters=2, random_state=0)
         X['cluster_success']=km.fit_predict(X[['RACH_Success_Rate','Handover_Success_Rate']])
         # get centroids
         centroids = km.cluster_centers_
         cen_x = [i[0] for i in centroids]
         cen_y = [i[1] for i in centroids]
         ## add to X
         X['cen_x'] = X.cluster_success.map({0:cen_x[0], 1:cen_x[1]})
         X['cen_y'] = X.cluster_success.map({0:cen_y[0], 1:cen_y[1]})
         # define and map colors
         colors = ['#DF2020', '#81DF20', '#2095DF']
         X['c'] = X.cluster.map({0:colors[0], 1:colors[1]})
```

Figure 5.39: interface of jupyter notebook showing the steps of clustering.

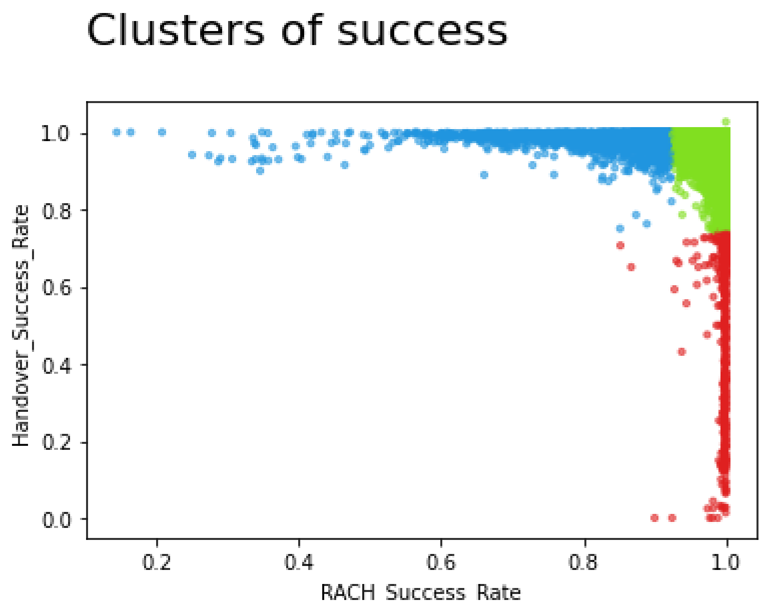


Figure 5.40: interface of jupyter notebook showing a plot of cluster of success.

We did the same thing for TCH Drop Rate and subscriber perceived TCH congestion and this was the result.

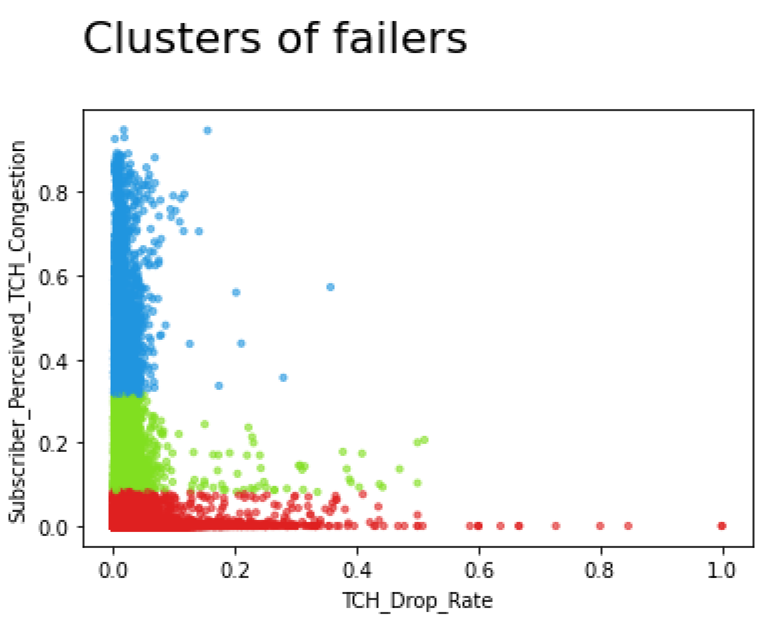


Figure 5.41: interface of jupyter notebook showing a plot of cluster of failers.

For more information we did some visualization, here are some of them.

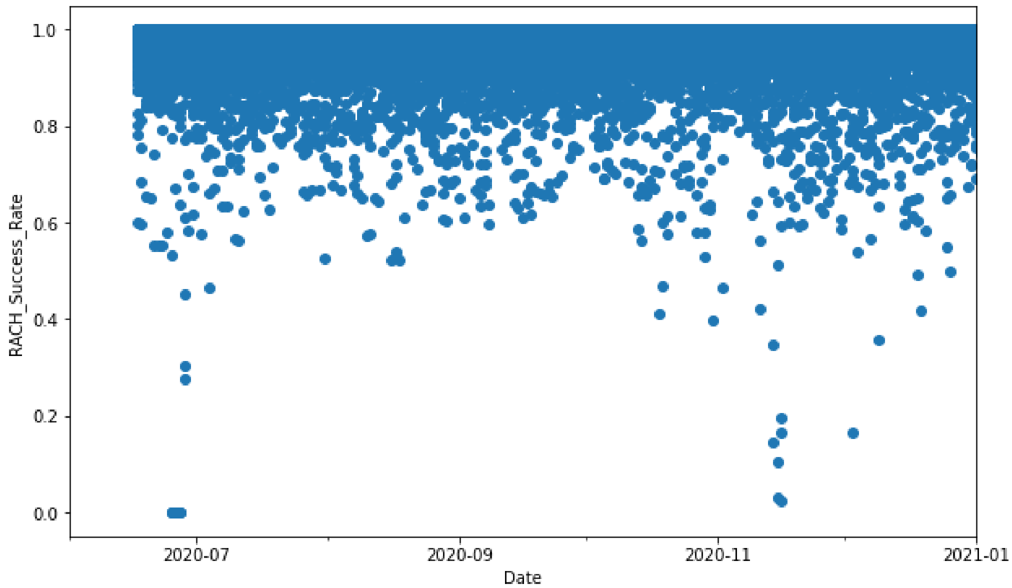


Figure 5.42: interface of jupyter notebook showing a plot of Rach success rate by Date.

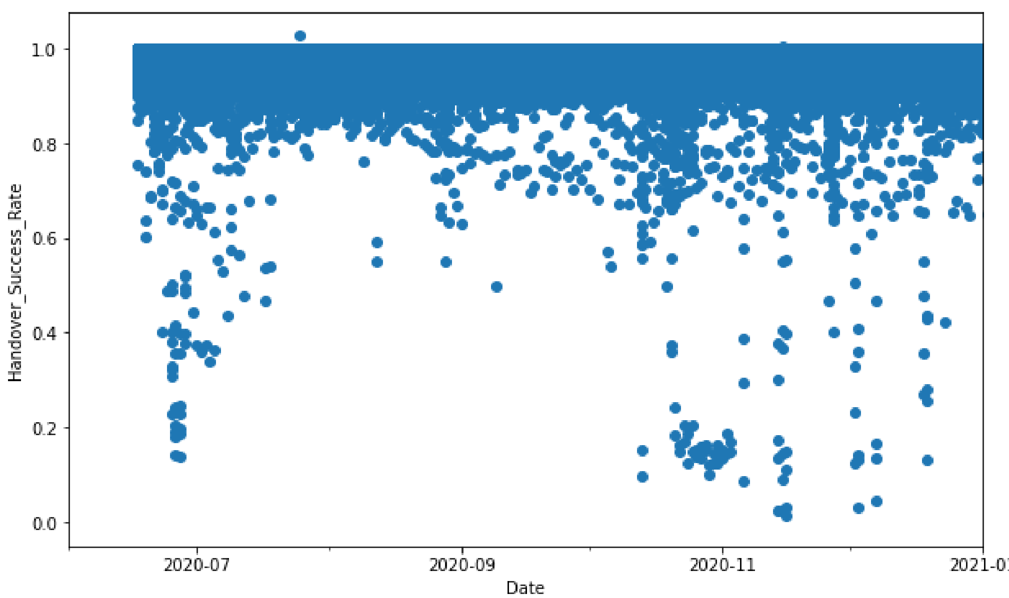


Figure 5.43: interface of jupyter notebook showing a plot of Handover success rate by Date.

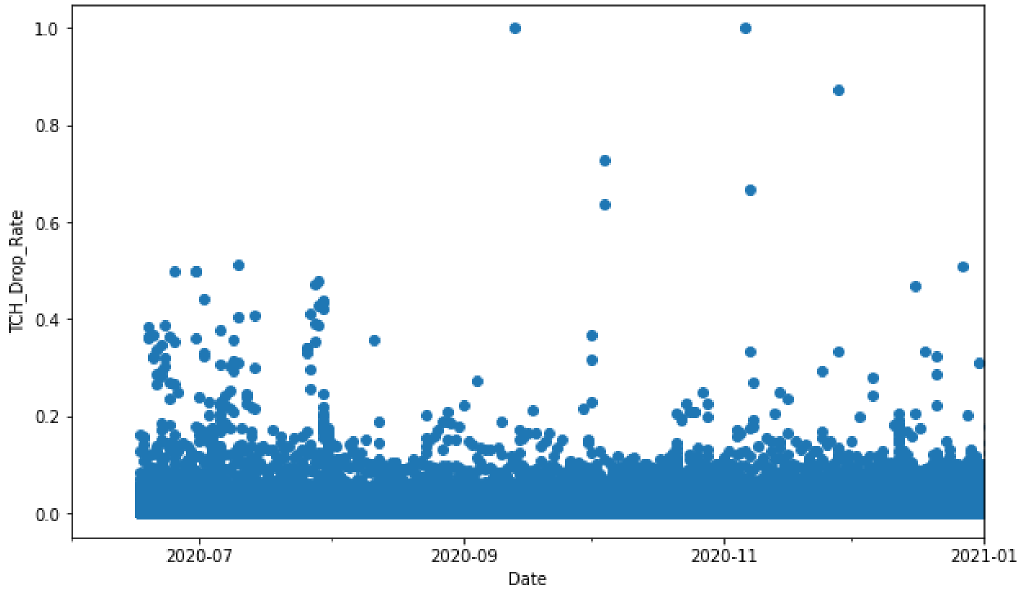


Figure 5.44: interface of jupyter notebook showing a plot of TCH drop rate by Date.

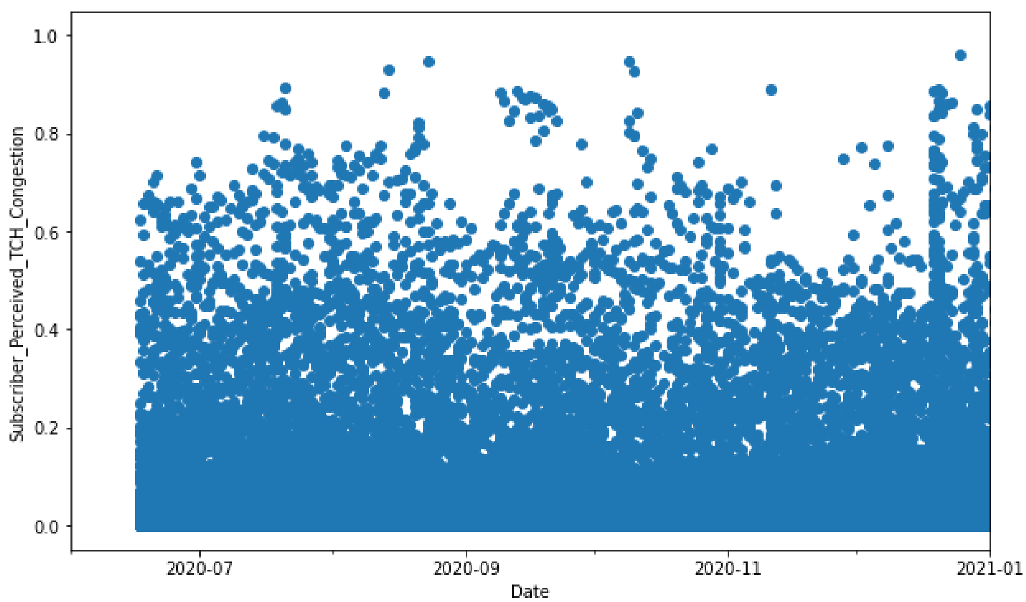


Figure 5.45: interface of jupyter notebook showing a plot of Subscriber perceived TCH congestion by Date.

5.7 Conclusion

We introduced in this last chapter the implementation and validation of the proposed system for tacit and explicit knowledge acquisition within an organization. First, we presented the programming languages and tools that we have used for the implementation besides the reasons of choosing each one. Second, we depicted the system architecture. Third, we described the organization host for the case study for the validation of the system, by providing its definition, organigram and objectives. Finally, we exposed some screenshots of the system interfaces illustrating our knowledge acquisition system, within the organization host.

General conclusion

In this thesis, we presented an approach for Organizational Knowledge Acquisition in Cloud Computing environment based on knowledge engineering.

For this, we started by defining organizational knowledge with their different typologies and the concept of organizational Knowledge Management including Knowledge Management approaches, objectives and processes.

After that, we presented the adopted Knowledge Management process in our work. We focus then on the Knowledge Acquisition sub-process. We divided the process of Knowledge Acquisition within the organization into two crucial sub-processes: Tacit Knowledge Acquisition process and Explicit Knowledge Acquisition process.

We reviewed proposed Knowledge Acquisition methods in the literature and presented an analysis of the applicability of each of them in Cloud Computing environment. The existing KDD process models allowed us to define the explicit Knowledge acquisition process in the case of knowledge embedded in operational databases within organization.

Our approach for Organizational knowledge Acquisition is based on knowledge engineering mainly knowledge modeling. We presented some Knowledge engineering definitions and the used knowledge models in our approach. The knowledge models include the domain model and the domain problems model. The knowledge acquired from those two models helps to prepare and complete some Knowledge Acquisition steps. They facilitate the interaction with the knowledge holders.

Finally, to realize our approach we created a web application to:

- ✓ interact with different knowledge holders including the knowledge engineer and domain expert for the purpose to acquire Tacit knowledge, based on the defined knowledge models.
- ✓ establish graphical representation of conceptual knowledge models.
- ✓ Visualize KDD and DM results for Explicit knowledge acquisition.

Our approach can be used in any organization which has tacit and explicit knowledge. We tested the developed system for the case of Mobilis ATM organization precisely at the department of NETWORK TECHNIQUE & SERVICES. We did the tacit knowledge acquisition part and the explicit knowledge acquisition part.

Our perspectives for the near future is:

- Complete the definition of the explicit knowledge acquisition sub process by taking into consideration other explicit knowledge sources of such as documents, using Text Mining.

- The formalization of the Domain Model and the Domain problem model using a knowledge representation language.

- Instead of using a database :
    - The formalization of the acquired tacit knowledge.
    - The formalization of knowledge resulting from Data Mining.

REFERENCES

[1] T. H. Davenport and L. Prusak, Working knowledge:' How organizations Manage what they know'. Harvard Business Press, 1998.

[2] G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N.R Shadbolt, W. Van de Velde, B. Wielinga, 'Knowledge Engineering and Management the CommonKADS Methodology ', 2001,P4.

[3] K. Lehrer, Theory of Knowledge (no. 167). Westview Press, 1990, pp. 265-266.

[4] Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. Organization Science, 5(1), 14–37.

[5] Polanyi, M. (1966). The Tacit Dimension. Garden City, NY: Doubleday.

[6] B. Wielinga, G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N.R Shadbolt, W. Van de Velde, B. Wielinga, 'Knowledge Engineering and Management the CommonKADS Methodology ', 2001, P70.

[7] Zack MH. Developing a Knowledge Strategy. California Management Review. 1999;41(3):125-145. doi:10.2307/41166000

[8] A. M. Alquier, "modélisation des systèmes d'information: Modèle Coopératif," Thèse d'habilitation à diriger des recherches, Université de Toulouse1, 1993.

[9] J. Barthes, "Capitalisation des connaissances et intelligence artificielle," Journées Franco-finlandaises, Tampere, 1997.

[10] Christopher grey ,'special section on critique and renewal in management education',university of leeds,1996.

[11] M. Horwitch and R. Armacost, "Be all it can be: Helping knowledge management," Journal of Business Strategy, vol. 23, pp. 26-31, 12/31 2002. **[12]** Souad, D. (2015). Knowledge Management and Intellectual Capital in an Enterprise Information System, European Conference on Knowledge Management, 213-221.

[13] T. H. Davenport, D.W. De Long, and M.C. Beers, "Successful KM Projects," Sloan Management Review, volume 39, number 2, Winter 1998, pp. 43-57.

[14] Caussanel J. et Chouraqui, E., ‘‘Informations et connaissances : quelles implications pour les projets de capitalisation de connaissances’’, In: Revue Document numérique - Gestion des documents et Gestion des connaissances, V. 3-4, (1999), 101-119.

[15] Boughzala, I., Zacklad, M. et Matta, N. : Gestion des connaissances dans une entreprise étendue - Mémoire d'entreprise et systèmes d'information coopératifs interentreprises. EGC, (2001), 259-270.

[16] Abecker, A., Bernardi, A., Hinkelmann, K., Kühn, O. et Sintek, M., ‘‘Towards a Technology for Organizational Memories’’, IEEE Intelligent Systems & Their Applications 13(3),1998.

[17] Hansen M.T., Nohria N., et Tierney T., ‘‘What’s your strategy for managing knowledge?’’ Harvard Business Review, V. 77, n° 2, (1999), 106-116.

[18] Barthelme-Trapp ’ Analyse comparée de méthodes de Gestion des connaissances pour une approche managériale’, 2001.

[19] Dominique Crié De l'extraction des connaissances au Knowledge Management, university of Lille, 2003

[20] C. Armistead,'Business process management: Exploring social capital within processes',1999.

[21] G. Probst, S. Raub, K. Romhardt,'Managing Knowledge: Building Blocks for Success',2000.

[22] Probst et al.'Managing knowledge: Building blocks for success',J. Wiley,2000.

[23] S. Staab, R. Studer, H.-P. Schnurr, and Y. Sure,'Knowledge Processes and Ontologies',2001.

[24] M. Lytras, N. Pouloudi, and A. Poulymenakou, "Knowledge management convergence: Expanding learning frontiers," Journal of Knowledge Management, vol. 6, pp. 40-51, 03/01 2002.

[25] H. Rollet,'Knowledge Management: Processes and Technologies',2003.

[26] I. Becerra-Fernandez, A. Gonzalez, R. Sabherwal,'Knowledge Management, Chalanges, Solutions and Technologies',2004.

[27] Nonaka, I. (1994) A Dynamic Theory of Organizational Knowledge Creation, Organization Science

[28] Lin, H.F., & Lee, G.G. (2005). Impact of organizational learning and knowledge management factors on e-business adoption. Management Decision, 43(2), 171-188. https://doi.org/10.1108/00251740510581902

[29] Paarup Nielsen, A. (2006), "Understanding dynamic capabilities through knowledge management", Journal of Knowledge Management, Vol. 10 No. 4, pp. 59-71. https://doi.org/10.1108/13673270610679363

[30] Franco, M. and Mariano, S. (2007), "Information technology repositories and knowledge management processes: A qualitative analysis", VINE, Vol. 37 No. 4, pp. 440-451. https://doi.org/10.1108/03055720710838515

[31] D. Nayir and Ü. Uzunçarsili, "A cultural perspective on knowledge management: The success story of Sarkuysan company," *J. Knowledge Management*, vol. 12, pp. 141-155, 04/04 2008. https://doi.org/10.1108/13673270810859578

[32] V. Supyuenyong, N. Islam, and U. Kulkarni, "Influence of SME characteristics on knowledge management processes: The case study of enterprise resource planning service providers," *J. Enterprise Inf. Management*, vol. 22, pp. 63-80, 01/01 2009.

[33] P. Sun, "Five critical knowledge management organizational themes," *Journal of Knowledge Management,* vol. 14, no. 4, pp. 507-523, 2010.

[34] K. Dalkir, Knowledge *Management in Theory and Practice*. 2013.

[35] Bukowitz W. R., Williams R. L., The Knowledge Management Fieldbook, Financial Time, Prentice Hall, London 2000.

[36] McElroy, M. W. (1999). "The Second Generation of KM," Knowledge Management (October, 1999), pp. 86-88**.**

[37] Meyer, M., & Zack, M. (1996). The Design and Implementation of Information Products. Sloan Management Review, 37(3), 43-59

[38] H. Rollett, "Knowledge Management: Processes and Technologies," 01/01 2003.

[41] Wee, J. C. N., & Chua, A. Y. K. (2013). The peculiarities of knowledge management processes in SMEs: the case of Singapore. Journal of Knowledge Management, 17(6), 958-972. https://doi.org/10.1108/JKM-04-2013-0163

[42] B. Bigliardi, F. Galati, and A. Petroni, "How to effectively manage knowledge in the construction industry," *Measuring Business Excellence,* vol. 18, no. 3, pp. 57-72, 2014.

[43] M. García-Fernández, "How to measure knowledge management: Dimensions and model," VINE, vol. 45, pp. 107-125, 02/09 2015.

[44] A. Kianto, M. Vanhala, and P. Heilmann, "The impact of knowledge management on job satisfaction," Journal of Knowledge Management, vol. 20, pp. 621-636, 07/11 2016.

[45] M. Yusr, S. Mohd Mokhtar, A. Othman, and Y. Sulaiman, "Does Interaction between TQM Practices and Knowledge Management Processes Enhance the Innovation Performance?," International Journal *of* Quality *&* Reliability Management, vol. 34, pp. 00-00, 06/30 2017.

[46] I. Chikhi and H. Bouarfa, "Knowledge Management Process Through a Cloud Computing Based Approach," in European Conference on Knowledge Management, 2019, vol. 1, pp. 238-247: Academic Conferences International Limited.

[47] A. Tuzhilin, "Knowledge management revisited: Old Dogs, New tricks," vol. 2, no. 3 %J ACM Trans. Manage. Inf. Syst., p. Article 13, 2011.

[48] Motta, E., Rajan, T. & Eisenstadt, M. (2015). Knowledge Acquisition as a Process of Model Refinement

[49] G. Huber, "Organizational Learning: The Contributing Processes and the Literatures," Organization Schiences, vol. 2, 02/01 1991

[50] W. Cohen and D. Levinthal, "Absorptive Capacity: A New Perspective on Learning and Innovation," Administrative Science Quarterly, vol. 35, pp. 128-152, 03/01 1990.

[51] R. M. Grant, "Toward a Knowledge-Based Theory of the Firm," Strategic Management Journal, vol. 17, pp. 109-122, 1996.

[52] C. W. Holsapple and K. D. Joshi, "A formal knowledge management ontology: Conduct, activities, resources, and influences," vol. 55, no. 7, pp. 593-612, 2004.

[53] Zahra, S. A. & George, G. (2002). Absorptive Capacity: A Review, Reconceptualization, and Extension, The Academy of Management Review, 27(2) 185-203.

[54] P. N. Ghauri and B. I. Park, "The Impact of Turbulent Events on Knowledge Acquisition: Comparison of Cross-border Acquisitions Formed Before and After the Crisis," MIR: Management International Review, vol. 52, no. 2, pp. 293-315, 2012.

[55] Kohli, Jaworski, B., & Kumar, 1993; Loon Hoe & McShane, 2010,'Structural and informal knowledge acquisition and dissemination in organizational learning: An exploratory analysis',2010.

[56] Akgun, Lynn, & Byrne,'Organizational Learning: A Socio-Cognitive Framework',2003.

[57] Naomi Zack,'Philosophy of Science and Race',2002.

[58] Reus and Lamont,'The double-edged sword of cultural distance in international acquisition',2009.

[59] D. McIver and C. Lengnick-Hall, "KNOWLEDGE MANAGEMENT PROCESSES AND UNIT PERFORMANCE: A CONTINGENCY PERSPECTIVE," Academy of Management Proceedings, vol. 2011, pp. 1-7, 01/01 2011.

[60] Lane, P. J., B. R. Koka and S. Pathak (2006). 'The reification of absorptive capacity: a critical review and rejuvenation of the construct', Academy of Management Review, 31, pp. 833–863.

[61] L. Zheng, T.-M. Yang, T. Pardo, and Y. Jiang, Understanding the "Boundary" in Information Sharing and Integration. 2009, pp. 1-10.

[62] C.-F. Chen and Y.-Y. Chang, "Airline brand equity, brand preference, and purchase intentions--The moderating effects of switching costs," Journal of Air Transport Management, vol. 14, pp. 40-42, 01/31 2008.

[63] U. Lichtenthaler and E. Lichtenthaler, "A Capability-Based Framework for Open Innovation: Complementing Absorptive Capacity," vol. 46, no. 8, pp. 1315-1338, 2009.

[64] R. Lubit, "The keys to sustainable competitive advantage: Tacit knowledge and knowledge management," Organizational Dynamics, vol. 29, pp. 164-178, 12/01 2001.

[65] DiBella, Anthony J. "How Organizations Learn: An Integrated Strategy for Building Learning Capability." Design Management Journal 9.2 (1998)..

[66] K. Sandahl, "Transferring knowledge from active expert to end-user environment," Knowledge Acquisition, vol. 6, no. 1, pp. 1-21, 1994/03/01/ 1994.

[67] W. P. Wagner and C. W. J. E. S. Holsapple, "An analysis of knowledge acquisition roles and participants," vol. 14, no. 1, pp. 3-14, 1997.

[68] D. L. J. A. A.-. Schmoldt, "Knowledge acquisition using linguistic-based knowledge analysis," 1998.

[69] S. Gordon and R. J. A. A. i. N. R. M. Gill, "QUESTION PROBES ASTRUCTURED METHOD FOR ELICITING DECLARATIVE KNOWLEDGE ," vol. 3, no. 2, pp. 13-20, 1989.

[70] A. A. Mitchell, "The use of alternative knowledge-acquisition procedures in the development of a knowledge-based media planning system," International Journal of Man-Machine Studies, vol. 26, no. 4, pp. 399-411, 1987/04/01/ 1987.

[71] Brokenshaw, D.M. Warren, and O. Werner, "Indigenous Knowledge Systems and Development". University Press of America, New York, 1980.

[72] Gordon, S.E., and R.T. Gill, "Question probes: A structured method for eliciting declarative knowledge". AI Applications 3(2): 13-20, 1989.

[73] T. L. Saaty and L. G. J. B. s. Vargas, "Hierarchical analysis of behavior in competition: Prediction in chess," vol. 25, no. 3, pp. 180-191, 1980.

[73] Graesser, A.C., and L.F. Clark, "Structures and Procedures of Implicit Knowledge"Ablex Publishing, Norwood, New Jersey, 1985.

[74] Boose, J.H, "Expertise Transfer for Expert System Design" Elsevier, New York,1986

[75] J. R. Anderson, "A spreading activation theory of memory," Journal of Verbal Learning & Verbal Behavior, vol. 22, no. 3, pp. 261-295, 1983.

[76] R. N. Shepard, "The analysis of proximities: Multidimensional scaling with an unknown distance function. Part I," Psychometrika, vol. 27, no. 2, pp. 125- 140,1962.

[77] S. C. Johnson, "Hierarchical clustering schemes," (in eng), Psychometrika, vol. 32, no. 3, pp. 241-54, Sep 1967.

[78] N. M. Cooke and J. E. McDonald, "The application of psychological scaling techniques to knowledge elicitation for knowledge-based systems," International Journal of Man-Machine Studies, vol. 26, no. 4, pp. 533-550, 1987/04/01/ 1987.

[79] R. Benfer and L. Furbee, "Knowledge acquisition in the Peruvian Andes," 1989.

[80] R. Schweickert, A. M. Burton, N. K. Taylor, E. Corlett, N. R. Shadbolt, and A. J. A.I. R. Hedgecock, "Comparing knowledge elicitation techniques: a case study," vol. 1, no. 4, pp. 245-253, 1987.

[81] Schmoldt, D.L., and W.G. Bradshaw, " A cumulative Delphi approach to knowledge acquisition". Pages 149-162 in: Third Annual Rocky Mountain Conference on Artificial Intelligence, Y. Wills, editor. Rocky Mountain Society for Artificial Intelligence, Boulder, Colorado,1988.

[82] J. R. OLSON and H. H. RUETER, "Extracting expertise from experts: Methods for knowledge acquisition," vol. 4, no. 3, pp. 152-168, 1987.

[83] D. L. Schmoldt and D. L. J. E. M. Peterson, "Applying knowledge-based methods to design and implement an air quality workshop," vol. 15, no. 5, pp.623-634,1991.

[84] M. D. Grover, "A Pragmatic Knowledge Acquisition Methodology," in IJCAI, 1983,vol. 83, pp. 436-438: Citeseer.

[85] R. R. Hoffman, "The problem of extracting the knowledge of experts from the perspective of experimental psychology," AI Magazine, vol. 8, no. 2, pp. 53- 67,1987.

[86] Kelly, G.A., “The Psychology of Personal Constructs”, Norton, New York, 1955

[87] J. H. Boose, "A knowledge acquisition program for expert systems based on personal construct psychology," International Journal of Man-Machine Studies,vol. 23, no. 5, pp. 495-525, 1985/11/01/ 1985.

[88] D. Sleeman and F. Mitchell, "Towards painless knowledge acquisition,"in International Conference on Knowledge Engineering and Knowledge Management, 1996, pp. 262-277: Springer.

[89] Walker, D.H., F.L. Sinclair, and G. Kendon, “A knowledge based systems approach to agroforestry research and extension”. AI Applications 9(3):61-72,1995

[90] Schmoldt, D.L. 1987. “Evaluation of an Expert System Approach to Forest Pest Management of Red Pine (Pinus Resinos)”. Ph. D. dissertation. University Microfilms International.1987

[91] M. M. Owrang O and F. H. Grupe, "Using domain knowledge to guide database knowledge discovery," Expert Systems with Applications, vol. 10, no. 2, pp.173- 180, 1996/01/01/ 1996.

[92] U. Hahn, M. Klenner, and K. Schnattinger, "A quality-based terminological reasoning model for text knowledge acquisition," in Advances in Knowledge Acquisition, Berlin, Heidelberg, 1996, pp. 131-146: Springer Berlin Heidelberg.

[93] Wang, F.-H., “On Acquiring Classification Knowledge from Noisy Data Based on Rough Set.” Expert Systems with Applications, Vol. 29, No. 1.2005

[94] R. Agrawal and M. Dhingra, "A detailed study on text mining techniques," 01/01 2013.

[95] D. S. Dang and P. H. Ahmad, “A review of text mining techniques associated with various application areas,” International Journal of Science and Research (IJSR), vol. 4, no. 2, pp. 2461–2466, 2015

[96] R. Steinberger, “A survey of methods to ease the development of highly multilingual text mining applications,” Language Resources and Evaluation, vol.46, no. 2, pp. 155–176, 2012.

[97] N. Zhong, Y. Li, and S.-T. Wu, “Effective pattern discovery for text mining,” IEEEtransactions on knowledge and data engineering, vol. 24, no. 1, pp. 30–44, 2012.

[98] B. Laxman and D. Sujatha, “Improved method for pattern discovery in text mining,” International Journal of Research in Engineering and Technology, vol.2,no. 1, pp. 2321–2328, 2013

[99] A. Henriksson, H. Moen, M. Skeppstedt, V. Daudaravicius, and ˇ M. Duneld, “Synonym extraction and abbreviation expansion with ensembles of semantic spaces,” Journal of biomedical semantics, vol. 5, no. 1, p. 1, 2014.

[100] A. M. Cohen and W. R. Hersh, “A survey of current work in biomedical text mining,” Briefings in bioinformatics, vol. 6, no. 1, pp. 57–71, 2005.

[101] E. A. Calvillo, A. Padilla, J. Munoz, J. Ponce, and J. T. Fernan- ˜ dez, “Searching research papers using clustering and text mining,” in Electronics, Communications and Computing (CONIELECOMP), 2013 International Conference on. IEEE, 2013, pp. 78–81.

[102] B. L. Narayana and S. P. Kumar, “A new clustering technique on text in sentence for text mining,” IJSEAT, vol. 3, no. 3, pp. 69–71, 2015.

[103] B. A. Mukhedkar, D. Sakhare, and R. Kumar, “Pragmatic analysis based document summarization,” International Journal of Computer Science and Information Security, vol. 14, no. 4, p. 145, 2016.

[104] C. P. Chen and C.-Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on big data,” Information Sciences, vol, 275, pp. 314–347, 2014.

[105] R. Al-Hashemi, “Text summarization extraction system (tses) using extracted keywords.” Int. Arab J. e-Technol., vol. 1, no. 4, pp. 164– 168, 2010.

[106] David Milward, "Linguamatics I2E and Machine Learning",2017

[107] Craig Stedman,"text mining", 2020 adapted from https://searchbusinessanalytics.techtarget.com/definition/text-mining

[108] Huang, A.: Similarity measures for text document clustering. In: Proceedings of the sixth

New Zealand Computer Science Research Student Conference (NZCSRSC2008), Christchurch, New Zealand, pp. 49–56, 2008

[109] Clifton, C., Cooley, R.: TopCat: Data mining for topic identification in a text corpus. In: European Conference on Principles of Data Mining and Knowledge Discovery, pp. 174–183. Springer, Heidelberg, 1999

[110] Han, E.H., Karypis, G., Kumar, V., Mobasher, B.: Clustering based on association rule hypergraphs. In: DMKD, 1997

[111] Irfan, R., King, C.K., Grages, D., Ewen, S., Khan, S.U., Madani, S.A., … & Tziritas, N.: A survey on text mining in social networks. Knowl. Eng. Rev. 30(2), 157–170, 2015

[112] Goh, D.H., Ang, R.P.: An introduction to association rule mining: An application in counseling and help-seeking behavior of adolescents. Behav. Res. Methods 39(2), 259–266, 2007

[113] Wong, P.C., Whitney, P., Thomas, J.: Visualizing association rules for text mining. In: 1999 IEEE Symposium on Information Visualization, 1999. (Info Vis’ 99) Proceedings, pp. 120– 123. IEEE, 1999

[114] Jayashankar, S., Sridaran, R.: Superlative model using word cloud for short answers evaluation in eLearning. Educ. Inf. Technol., 1–20, 2016

[115] DePaolo, C.A., Wilkinson, K.: Get your head into the clouds: using word clouds for analyzing qualitative assessment data. TechTrends 58(3), 38–44, 2014

[116] Sinclair, J., Cardew-Hall, M.: The folksonomy tag cloud: when is it useful? J. Inf. Sci. 34(1), 15–29, 2008

[117] Viegas, F.B., Wattenberg, M., Van Ham, F., Kriss, J., McKeon, M.: Manyeyes: a site for visualization at internet scale. IEEE Trans. Vis. Comput. Graphics 13(6), 1121–1128, 2007

[118] S. Dang, "A Review of Text Mining Techniques Associated with Various Application Areas," International Journal of Science and Research (IJSR), vol. 4, pp. 2461-2466, 02/28 2015.

[119] Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", published by Morgan Kauffman, 3rdi ed, 2006.

[120] L. Gohil, "Text Mining: Process and Techniques," vol. 3, 05/28, 2015.

[121] A. Jivani, "A Comparative Study of Stemming Algorithms," Int. J. Comp. Tech. Appl., vol. 2, pp. 1930-1938, 11/01, 2011.

[122] D. Hand, H. Mannila, and P. Smyth, Principles of Data Mining. 2001.

[123] A. Luca, L. Lupu, and I. Herghiligiu, "ORGANIZATIONAL KNOWLEDGE

ACQUISITION - STRATEGIC OBJECTIVE OF ORGANIZATION," CBU International Conference Proceedings, vol. 4, p. 126, 09/21 2016.

[124] Nonaka, I. and Takeuchi, H, "The knowledge-creating company". New York, Oxford:Oxford University Press, 1995

[125] W. Swap, D. Leonard, M. Shields, and L. J. J. o. m. i. s. Abrams, "Using mentoring and storytelling to transfer knowledge in the workplace," vol. 18, no. 1, pp. 95-114, 2001.

[126] R. Grant, "Toward A Knowledge-Based Theory of the Firm," Strategic Management Journal, vol. 17, pp. 109-122, 12/01 1996.

[127] J. Nahapiet and S. J. A. o. m. r. Ghoshal, "Social capital, intellectual capital, and the organizational advantage," vol. 23, no. 2, pp. 242-266, 1998.

[128] M. Shami, N. Sakhaee, and H. J. R. j. o. i. t. Shahbaznezhad, "Mechanisms of customer knowledge management in e-commerce websites," vol. 1, no. 2, pp. 86-93, 2009.

[129] I. Pinho, A. Rego, and M. P. J. J. o. k. m. e Cunha, "Improving knowledge management processes: a hybrid positive approach," 2012.

[130] S. Hoe and S. McShane, "Structural and informal knowledge acquisition and dissemination in organizational learning: An exploratory analysis," Learning Organization, The, vol. 17, pp. 364-386, 05/25 2010.

[131] Nieminen, H. (2007). Developing competences through inter-organizational knowledge acquisition, Esa Print Tampere,Tampere.

[132] M. T. Hansen, N. Nohria, and T. J. T. k. m. y. Tierney, "What’s your strategy for managing knowledge," vol. 77, no. 2, pp. 106-116, 1999.

[133] Radu Prodan and Simon Ostermann, “A Survey and Taxonomy of Infrastructure as a Service and Web Hosting Cloud Providers”, 10th IEEE/ACM International Conference on Grid Computing, 2009

[134] “Cloud computing”,july 2021.URL: http://en.wikipedia.org/wiki/Cloud_computing

[135] Andrew Joint and Edwin Baker, “Knowing the past to understand the present- issues in the contracting for cloud based services”, Computer Law and Security Review 27, pp 407-415, 2011

[136] Vania Goncalves and Pieter Ballon, “Adding value to the network: Mobile operators’ experiments with Software-as-a-Service and Patform-as-a-Service models”, Telematics and Informatics 28, pp 12-21, 2011

[137] W. Jansen and T.Grance "Guidelines on Security and Privacy in Public Cloud Computing", NIST Draft Special Publication 800-144, 2011. http://csrc.nist.gov/publications/drafts/800-144/DraftSP-800-144_cloud-computing.pdf

[138] P. Mell and T. Grance, "The NIST Definition of Cloud Computing" Recommendation of NIST, Special Publication 800-145, 2011. http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

[139] Dimitrios Zissis and Dimitrios Lekkas, "Addressing cloud computing security issues", Future Geberation Computer Systems 28, pp. 583-592, 2012.

[140] A. Youssef, "Exploring Cloud Computing Services and Applications," Journal of Emerging Trends in Computing and Information Sciences, 07/01 2012

[141] El Moukhtar Zemmouri. Représentation et gestion des connaissances dans un processus d'Extraction de Connaissances à partir de Données multi-points de vue. Apprentissage [cs.LG]. Ecole Nationale Supérieure d'Arts et Métiers - Meknès, 2013. Français.

[142] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "Knowledge discovery and data mining: towards a unifying framework," presented at the Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, 1996.

[143] R. J. Brachman and T. Anand, "The process of knowledge discovery in databases," in Advances in knowledge discovery and data mining: American Association for Artificial Intelligence, 1996, pp. 37–57.

[144] L. Kurgan and P. Musilek, "A survey of Knowledge Discovery and Data Mining process models," Knowledge Eng. Review, vol. 21, pp. 1-24, 03/01 2006.

[145] Nathalie Aussenac-Gilles, 'conception d'une méthodologie et d'un outil d'acquisition de connaissances expertes',PHD thesis, Toulouse3,1989.

[146] Jean Charlet,'l'ingerie des connaissances, entre science de l'information et sciencede gestion'.In Entre la connaissance et l'organisation,l'activité collective,2004.

[147] Awa Diattara,'Problematique de l'acquisition des connaissances dans ces envirenements informatiques fortement orientés connaissances :vers un outil auteur pour le projet AMBRE',Envirenements informatiques pour l'apprentissage Humain.Université grenoble Alpes,2017.

[148] A. Liew, "Understanding Data, Information, Knowledge And Their Inter-Relationships," Journal of Knowledge Management Practice, vol. Vol. 7, 06/01 2007.

[150] F. Nickols, "The knowledge management yearbook 2000-2001," 2000, pp. 12-21.

[151] **"**https://www.valamis.com/hub/knowledge-management?fbclid=IwAR0ZeRdbSIC7h14YD5ZyK49XAt9tBiQXwmpIqaeMupWY884ap_pFO8gfBI8**"**,May,29,2019

[152] Gray P. H.J. Watson (1996) The new DSS: Data Warehouses, OLAP, MDD and KDD, 1996

[153] Z. D. Abdelkader and R. Ricco, "Extraction de connaissances à partir de données (ECD)," (in fre), Techniques de l'ingénieur Bases de données, Article de base documentaire vol. base documentaire : TIB309DUO, no. ref. article : h3744, 2002.

[154] Jones, P.H. Knowledge Acquisition. In: Barrett, J.R. and D.D. Jones. Knowledge Engineering in Agriculture. ASAE Monograph No. 8, ASAE, St. Joseph, MI. 1989.

[155] Knowledge acquisition,3 June 2021,URL: https://en.wikipedia.org/wiki/Knowledge_acquisition

[156] Gottgtroy Paulo, "Ontology Driven Knowledge Discovery Process: a proposal to integrate Ontology Engineering and KDD" . *PACIS 2007 Proceedings*. 88. 2007

[157] Z. Pabarskaite and A. Raudys, "Raudys, A.: A process of knowledge discovery from web log data: Systematization and critical review. Journal of Intelligent Informatin Systems 28(1), 79-104," J. Intell. Inf. Syst., vol. 28, pp. 79-104, 02/08 2007.

[158] Buchner, A. G., Mulvenna, M. D., Anand, S. S. & Hughes, J. G. 1999. An Internet-enabled Knowledge Discovery Process, 13–27.

[159] Alnoukari, M., Alzoabi, Z., & Hanna, S. (2008). Applying adaptive software development (ASD) agile modeling on predictive data mining applications: ASD-DM Methodology. In IEEE Proceedings of International Symposium on Information Technology (pp. 1083-1087)

[160] Jain, A. K., & Dubes, R. C. Algorithms for Clustering Data. Upper Saddle River, NJ: Prentice-Hall, Inc. 1988

[161] Joel Syder," Top 5 free data mining tools to try for your business!",12,17,2018 URL: https://bigdata-madesimple.com/top-5-free-data-mining-tools-to-try-for-your-business/

[162] Anjali UJ," THE TOP 10 DATA MINING TOOLS OF 2018".09,17,2018 URL: h https://www.analyticsinsight.net/the-top-10-data-mining-tools-of-2018/

[163] Chandan goopta," Six of the Best Open Source Data Mining Tools",12,17,2018 URL: https://thenewstack.io/six-of-the-best-open-source-data-mining-tools/

[164] M. Bharati and B. Ramageri, "Data mining techniques and applications," Indian Journal of

Computer Science and Engineering, vol. 1, 12/01 2010.

[165] Efraim Turban, Jay E. Aronson, Ting-Peng Liang ,"Decision Support Systems and Intelligent Systems, knowledge Acquisition, Representation, and Reasoning ",2005

[166] DIANA ELENA CODREANU, DENISA ELENA PARPANDEL, IONELA POPA," EXTRACTING KNOWLEDGE FROM DATA - DATA MINING", Nicolae Titulescu University,2011

[167] Charlet J., ZA CK LAD M., KAS SE L G. & BOURIG AU LT D. Ingénierie des connaissances :recherches et perspectives. In Charlet et al. (2000b), chapitre 1, p. 1–22,2000

[168] CHA RLET J., KR IV INE J.-P. & REYN AU D C. (1996b). Causal model-based knowledge acquisitiontools : Discussion of experiments. International Journal of Human-Computer Studies,26(1), 318–336.1996

[169] Benjamin Crawford, "The 3 Types of Cloud Computing Services", 08/07/2019. URL: https://rsscloud.org/types-of-cloud-computing-services

[170] L. Edvinson, M. Malone : Le capital immatériel de l'entreprise, Ed. Maxima, Paris 1999.

[171] Paul A. Strassmann : The Value of Knowledge Capital, American Programmer, March 1998

[172] Stewart T : Intellectual Capital, the New Wealth of Organizations, Currency/Doubleday., New York, 1997

[173] D. Foray : L'économie de la connaissance, Repères, n° 298, Ed. La découverte, Paris, 2000

[174] OCDE, L'économie fondée sur le savoir, des faits et des chiffres , 1999, Mesurer le capital humain, vers une comptabilité du savoir acquis 1996.

[175] **"**Flask (web framework)",june 2021. URL: https://en.wikipedia.org/wiki/Flask_(web_framework)

[176] **"**pandas (software)",July 2021, URL: https://en.wikipedia.org/wiki/Pandas_(software)

[177] " Python seaborn Library",nd , Viewed on July 2021, URL: https://www.javatpoint.com/python-seaborn-library

[178] "NumPy", July 2021, URL: https://en.wikipedia.org/wiki/NumPy

[179] "Matplotlib", July 2021, URL: https://en.wikipedia.org/wiki/Matplotlib

[180] "Apache Spark", July 2021, URL: https://en.wikipedia.org/wiki/Apache_Spark

[181] Cheong, R. K., & Tsui, E, From Skills and Competencies to Outcome- based Collaborative Work: Tracking a Decade's Development of Personal Knowledge Management (PKM) Models. Knowledge and Process Management, 18(3): 175-193, 2011.

[182] Ju, D. & Shen, B., On Building Knowledge Cloud. IEEE, 2011.

[183] Bohlouli, M., Holland, A. and Fathi, M. "Knowledge Integration of Collaborative Product Design Using Cloud Computing Infrastructure", IEEE International Conference on Electro/Information Technology, pp. 1–8, 2011.

[184] Khoshnevis, S. and Rabeifa, F. "Toward Knowledge Management as a Service in Cloud Based Environments", International Journal of Mechatronics, Electrical and Computer Technology, Vol. 2, No. 4, pp. 88–110, 2012.

[185] Shahbazi, R., Sadeghzadeh, A., Haghshenas, M. and Nassiriyar, M."Adoption of cloud based Knowledge Management", International Journal of Engineering and Innovative Technology, Vol. 3, pp 324–329,2014.

[186] Gunadham, T. "Potential of Cloud Storage Application as Knowledge Management System", International Journal of Innovation, Management and Technology, Vol. 6, No. 2, April, pp 153–157, 2015.