From the Business Decision Modeling to the Use Case Modeling in Data Mining Projects

Oscar Marban[1], José Gallardo[2], Gonzalo Mariscal[3] and Javier Segovia[1]
[1]Universidad Politécnica de Madrid
[2]Universidad Católica del Norte
[3]Universidad Europea de Madrid
[1,3]Spain
[2]Chile

1. Introduction

In the current arena where companies face extreme competiveness and continuous changes, a rapid and flexible capability to respond to market dynamism is a key factor for the success or failure of any organization. In this context, the development of efficient strategic and operational decision-making support systems is essential for guaranteeing business success and survival. Nowadays, data mining systems are an effective technology for supporting organizational decision-making processes.

From the viewpoint of data mining development, the year 2000 marked the most important milestone: CRISP-DM (CRoss-Industry Standard Process for Data Mining) was published (Chatam, et al., 2002), (Piatetsky-Shaphiro, 2000). CRISP-DM is the most used model for developing data mining projects (Kdnuggets, 2007). Data mining had been successfully applied to many real problems. As a result, data mining has been popularized as the business intelligence tool with the greatest growth projection. In recent years, data mining technology has moved out of the research labs and into companies on the 'Fortune 500' list (Kantardzic & Zurada, 2005).

Even so, the scientific literature is dotted with many examples of failed projects, project planning delays, unfinished projects, or budget overruns (Eisenfeld et al., 2003), (Meta Group Research, 2003), (Maciaszek, 2005). There are two main reasons for this. On the one hand, there are no standard development processes to implement an engineering approach in data mining project development (Marbán, 2008). On the other hand, requirements are not properly specified. One of the critical success factors of data mining projects is the need for a clear definition of the business problem to be solved, where data mining is considered to be best technological solution (Hemiz, 1999). This indicates the need for a proper definition of project requirements that takes into account organizational needs based on a business model.

Historically, research in data mining has focused on the development of algorithms and tools, without any detailed consideration of the search for methodological approaches that ensure the success of a data mining project.

In this paper, we propose a methodological approach to guide the development of a business model of the decision-making process within an organization. The business decision-making model (represented in i* notation) is translated into use cases based on

heuristics. In this way, the functional requirements of a data mining project can be associated with organizational requirements and the organization's strategic objectives.

This chapter is structured as follows. Section 2 summarizes the key dimensions to be considered in a requirements specification process and previous work related to requirements engineering. Section 3 presents a review of and concludes with a comparison of notations used in business modeling. Section 4 presents a business modeling process that takes into account the understanding of the business domain and the generation of a business decision-making model. Section 5 describes the process for creating the requirements model from the business decision-making model. In Section 6, the proposed methodology is applied to a case study. Finally, in Section 7 we summarize the conclusions of the presented research.

2. Requirements engineering and business modeling

Requirements engineering is a process covering all the activities involved in discovering, documenting and maintaining all the system requirements (Kotonya & Sommerville, 1998) (Sommerville, 2002), (Medina, 2004). Not only must a good requirements discovery process elicit what the customer wants, but it must also consider the analysis and understanding of the application and business domain in which the system will be used. The main dimensions that a requirements specification process should cover (Kotonya & Sommerville, 1998) are:

- Understanding of the application domain: This determines the minimum knowledge required about the domain in which the system is to be implemented.
- Problem understanding: This involves understanding the details of the business problem that the system is to solve.
- Business understanding: This means understanding how project development affects business components and what contribution it makes to achieving organizational goals.
- Understanding of stakeholder needs: This accounts for stakeholder needs, particularly regarding the work processes that the system is to support.

Taking into account the obvious need to apply requirements engineering in product development life cycles, many requirements engineering process models have been proposed in different areas of engineering. In software engineering, research has focused on requirements elicitation and monitoring for the design and implementation of software systems ((Cysneiros & Sampaio, 2004), (Gacitúa, 2001), (Gorschek & Claes, 2006)), since there is a broad consensus on the essentiality of the requirements elicitation phase in software development. There are several proposals ((Kotonya & Sommerville, 1998) (Davyt, 2001) (Sommerville, 2002)) that vary as to form and the emphasis they place on specific activities.

Rilston (Rilston et al. 2003) proposed the DWARF model (Data WArehouse Requirements deFinition) for OLAP and data warehouse development projects. DWARF supports management planning, specification, validation and requirements management. Bruckner et al. (Bruckner et al., 2001) define and discuss three abstraction levels for data warehouse requirements (business, user and system requirements), and show the requirements definition of a data warehouse system. Dale (Dale, 2004) presents a case study where the requirements definition process for developing a data warehouse is based on a modified Six Sigma Quality methodology.

In the e-commerce area, the E3-Value methodology (Gordijn et al., 2000) (Gordijn, 2003) has been proposed. It helps to systematically discover, analyze and evaluate e-business ideas. Requirements engineering is also used in the automotive industry. Weber and Weisbrod claim that the costs, sophistication and complexity involved in electrical and electronic automotive system development (telematics, interior and passenger comfort, driving assistance and safety-critical systems) are growing (Weber & Weisbrod, 2003).

Lately, the telecommunications field has expanded significantly. These advances have led to complex communications systems with different technologies. Therefore, the development of new high-quality services is a challenge for telecom operators. Gerhard presents the RATS (Requirements Assistant for Telecommunications Services) methodology (Gerhard, 1997).

There are many other areas where requirements engineering methodologies, techniques or activities are proposed and applied: control systems for nuclear power plants, avionics systems, lighting control, real-time embedded systems, complex systems (SCR method, Software Cost Reduction), intrusion detection systems, (Heitmeyer & Bharadwaj, 2000) (Heninger at al., 1978) (Heninger, 1980) (Slagell, 2002). There is also abundant research on security requirements engineering (Knorr, K. & Rohrig, 2005) (Leiwo et al., 2004). Firesmith defines different types of security requirements and gives some examples and guidelines associated with engineer training for specifying security requirements without unduly limiting security versus architecture (Firesmith, 2003).

In the case of data mining projects, there is not much research about applying requirements engineering techniques to requirements discovery and specification, as such projects are exploratory and return different types of results.

There is widespread agreement about the importance of business understanding in the requirements elicitation process (Kotonya & Sommerville, 1998). There are many references in the scientific literature to business process modeling for different purposes, such as a better organizational understanding, analysis and innovation, management or re-engineering (Martyn, 1995), (Koubarakis & Plexousakis, 1999), (Vérosle et al., 2003), (Gordijn & Akkermans 2007). However, there is little research on integrating the organizational model with definite requirements engineering activities (Mylopoulos et al., 2002), (Santander & Castro, 2002). Additionally, data mining projects are mainly developed to turn up knowledge or information to support decision-making at the strategic levels of an organization. This is an important issue because high-level decision-making processes are not structured, and hence are very difficult to model.

In this sense, business model development is essential for the requirements discovery and specification process in data mining projects, as the business model takes into account the main dimensions described above and relates project development to organizational strategic goals and objectives. To sum up, a better understanding of the business philosophy will help managers to make better decisions (Hayes & Finnegan, 2005).

3. Business modeling notations

There are different system specification techniques or methodology notations, where the term 'specification' is construed as the process of describing a system and its properties. A system specification can be described by a written document, a formal mathematic model, or a graphical representation. Each technique has advantages and disadvantages. Text descriptions are simpler and more flexible, but they are usually ambiguous, unclear or

include many hard-to-process documents. Formal languages do not have these drawbacks, but they are complex, less flexible than written text; also their use requires additional effort (training) (Sommerville, 2002). On the other hand, group of stakeholders can easily and quickly understand the system described by graphical models, as they represent visual scenarios (Berenbach, 2004) (Sommerville, 2005). On the downside, the model becomes more and more complex as the amount of information it has to represent grows. In complex systems, a combination of natural language and graphical models can be a good choice.

Some standard graphical notations used for business modeling are UML, BPMN (Russell et al., 2006), (Wilcox & Gurau, 2003), (Agulilar-Saven, 2004), (Berenbach, 2004), (Vérosle et al., 2003), (Wohed et al., 2006), (White, 2004), and i* (Yu, 1995), (Yu, 1996), (Yu & Mylopoulos, 1997), (Alencar et al., 2000), (Castro, et al., 2001), (Santander & Castro, 2002). i* is described below. Taking into account the advantages of graphical system specification notations over other text (natural language) or formal notations (mathematical language), we opted to use a graphical notation in the research described in this chapter. After analyzing the described notations (UML, BPMN, framework i*), we selected the i* framework. The main reasons for this decision are:

1. The i* framework is more than just a modeling language (like UML and BPMN); it is also a modeling technique, as it defines a process for developing an organizational model.
2. The model is easy and intuitive to understand, as it is possible to build a multilevel view of the process to be modeled.
3. Not only are organizational requirements linked to the system functionalities under development, but data mining projects, real-time systems, or process control systems are also strongly related to global requirements such as reliability, security, etc. Requirements play a key role in these kinds of systems. UML or BPMN do not explicitly take into account these types of requirements. In contrast, the i* framework provides primitives related to non-functional requirements.
4. The i* framework explicitly states organizational goals, tasks and resources, and the network of SD relationships among various actors needed to achieve strategic goals and understand the reasons behind the decision-making processes.
5. The i* framework brings the requirements specification gradually closer to organizational requirements in a straightforward manner.

3.1 I* framework

The i* framework technique was proposed by Eric Yu (Yu, 1995) and has the following features:

- It helps organizations to represent the actors involved in the process.
- It helps to represent dependencies explicitly among different organizational actors.
- It helps to build a simplified view of the business to be represented, showing the actors, dependencies, resources and operations to achieve the defined business goals.
- It employs graphics with a small number of primitives.

The i* framework consists of two models: the **strategic dependency model (SD)** and the **strategic rationale model (SR)**. Both models are complementary and they are composed of a set of primitive actors, related by a dependency.

An *SD* model describes a network of dependency relationships among various actors in an organizational context. The actor is usually identified within the context of the model. This

model shows who an actor is and who depends on the work of an actor. An SD model consists of a set of nodes and links connecting the actors. Nodes represent actors and each link represents a dependency between two actors. Nodes represent actors and each link represents a dependency between two actors. In the context of the i* framework, actors refer to generic entities that have intentionality. To reflect different degrees of concreteness of agency, the concepts of roles, positions and agents are defined as specializations of actors.

An *SR* model (Yu, 1996) is useful for modeling the motivations of each actor and their dependencies, and provides information about how actors achieve their goals and soft goals. This model only includes elements considered important enough as to have an impact on the results of a goal. The *SR* model (Yu, 1996) shows the dependencies of the actors by including the SD model. According to these dependencies, the SR model specifies achievement goals, soft goals, tasks and resources. Compared with the SD model, SR models provide a more detailed level of modeling. Intentional elements (achievement goals, soft goals, tasks, resources) appear in the SR model not only as external dependencies, but also as internal elements linked by *means-ends* relationships and *task-decompositions*. The *means-end* links provide understanding about why an actor would engage in some tasks, pursue a goal, need a resource, or want a soft goal; the *task-decomposition* links provide a hierarchical description of intentional elements that make up a routine.

There are three different types of actors (Alencar et al., 2000):

1. The depending actor is called *depender*
2. The actor who is depended upon is called the *dependee*.
3. The *depender* depends on the *dependee* for something to be achieved: the *dependum*.

There are four types of dependency (Alencar et al., 2000), classed according to the type of freedom allowed in the relationship:

1. *Resource Dependency*: In a resource dependency, an actor depends on another for the availability of some entity. Resources can be physical or informational.
2. *Task Dependency*: In a task dependency, an actor depends on another to carry out an activity. The task specification prescribes how the task is to be performed.
3. *Goal Dependency*: In a goal dependency, an actor depends on another actor to bring about a certain state or condition in the world. The *dependee* is given the freedom to choose how to do this.
4. *Soft-Goal Dependency*: A soft-goal dependency is similar to a goal dependency except that there are no a priori, sharply defined success criteria. The meaning of the soft goal is elaborated on and clarified between the *depender* and the *dependee* in terms of the methods that might be used to address it.

4. Business modeling

The process of building a business model that could result in the requirements of a data mining project is divided into two phases. The first step of the process is to elicit information to gain a proper understanding of the business domain. In the second phase, the business decision-making model is then developed based on the collected information.

4.1 Business domain understanding

The business domain of an organization is usually fairly complex and should be fully understood before initiating the development of any project. The success of a data mining project will largely depend on the correct understanding of the project goals and

requirements from a business or institutional viewpoint. So the objective here is to define a process to develop the task of understanding the business domain and establish a common vision with future users about the key project goals.

For the first phase of the data mining project life cycle (*Business Understanding*), the CRISP-DM development guide (Chapman, et al., 2000) proposes, among other things, two key tasks:

1. *Determine business objectives*: The first objective of the data analyst is to thoroughly understand, from a business perspective, what the client really wants to accomplish and which are the important factors that can influence the outcome of the project in the first instance.

2. *Assess situation*: This task involves more detailed fact-finding about all of the resources, constraints, assumptions and other factors that should be considered in determining the data analysis goal and project plan.

The above tasks are not easy to perform, and CRISP-DM does not suggest any methodological process for this purpose. Considering how important these tasks are for developing the business model, we present a methodological process that is designed to give project participants a better understanding of the organizational structure, strategic objectives, processes, business logic and other elements of the organization for which the future data mining system is to be developed.

The development of the proposal presented below involves identifying the relevant information to be elicited (based on the description of the essential components of a business model (Osterwalde et al., 2005), (Lagha et al, 2004)). This will give an overview of the business, identify the sources of such information and apply techniques and tools for requirements elicitation. The process of discovering and later specifying information related to the business domain is divided into two steps:

1. First, information is gathered to gain a view of the current company scenario (static vision of the business), that is, understand the components (tasks, organizational goals or requirements, organizational structure, products / services, market) that defines the organization and its environment at the start of the project.

2. Second, we have to elicit some information (see Figure 1) related to factors that influence or affect the achievement of objectives such as resources, capabilities, restrictions, SWOT analysis, etc. The achievement of organizational goals is in itself a definition of the future scenario of the organization.
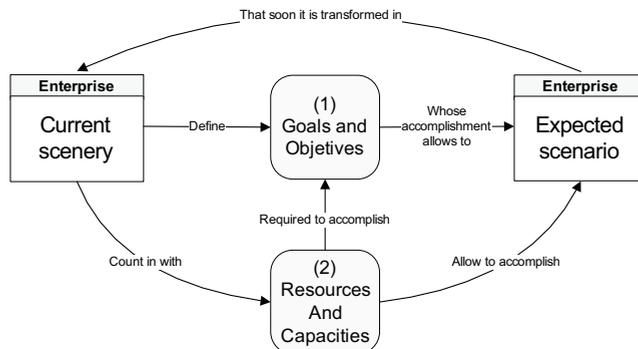

Fig. 1. Information about the business domain

Figure 2 shows a summary concept map. It includes a definition of the information that must be elicited in the above steps.

The aim of this concept map is to show, at a high level of abstraction, the key information (and its relevance) to be elicited to understand the business domain in the first instance.
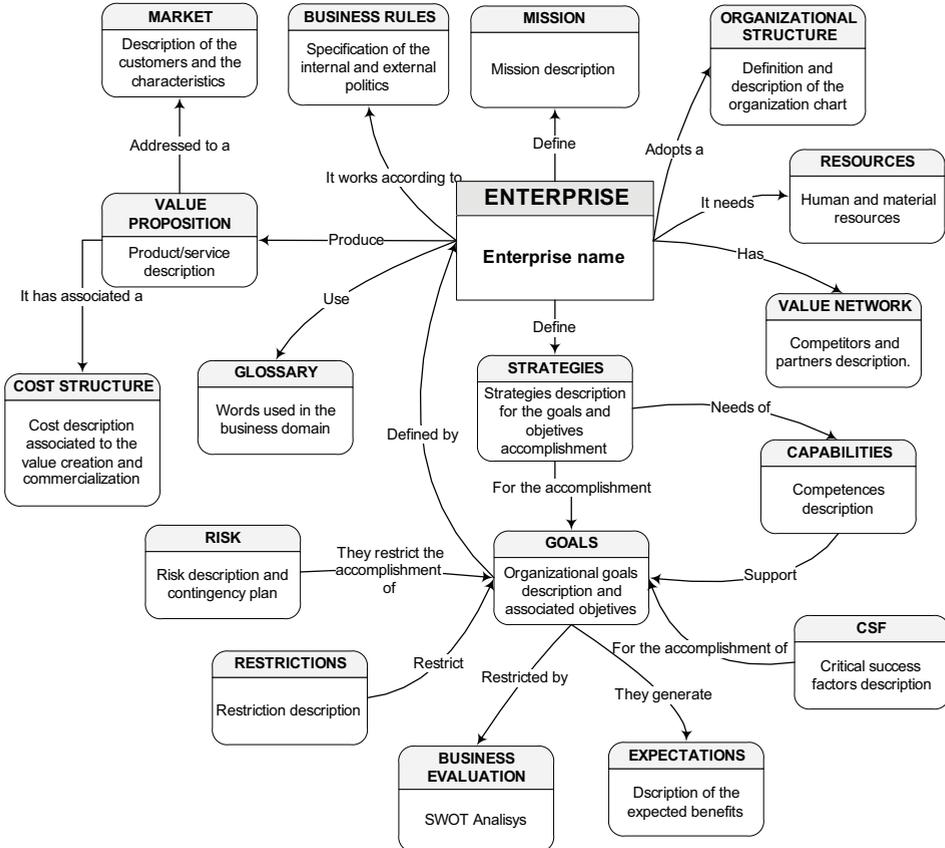

Fig. 2. Concept map that describes the business domain (based on (Ochoa, 2006))

4.2 Business decision-making model

After completing the business domain understanding phase and having identified the organizational goals, it is time to model the decision-making process. We propose a sequence of steps or stages for enacting this process (Figure 3).

The information required for each step can be elicited by requirements engineering and knowledge engineering techniques, such as interviews and questionnaires, JAD techniques, protocol analysis or laddering. A description by steps is given below.

4.2.1 Defining the initial goal of the decision-making process

This first step should identify the strategic goal underlying the decision-making process to be modeled from the organizational goals discovered in the business domain understanding
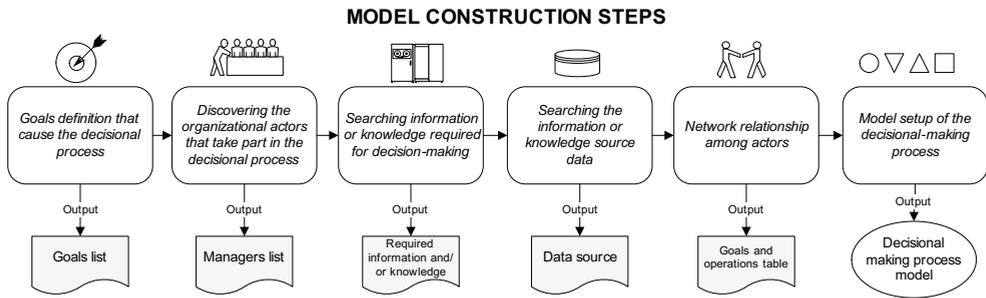
MODEL CONSTRUCTION STEPS

Fig. 3. Modeling process

step. Later, this goal will be the main objective to be achieved. It will be divided into a series of lower-level goals. These goals could then be further divided into a series of tasks developed by the organizational actors.

4.2.2 Discovering the organizational actors that take part in the decision-making process

Note, firstly, that there are different organizational actors inside an organization. These actors take part in the decision-making process at different levels of the organizational pyramid (Laudon & Laudon, 2004). The decisions made at the operational and knowledge levels are structured, that is, they are decisions based on procedures in place inside the organization. These decisions are not innovative and tend to recur. However, as the process progresses towards the strategic levels, decision-making becomes non-structured with less certain outcomes that affect the whole organization. Once all these considerations have been taken into account, the actors that make strategic decisions ('primary actors') must be identified at the respective level (Laudon & Laudon, 2004). Additionally, some actors that do not make decisions but do take part in the decision-making process also have to be identified ('secondary actors').

4.2.3 Eliciting the information or knowledge needed to make decisions

This third step should discover the information or knowledge to be gathered or assimilated for decision making by the primary organizational actors. This discovered information or knowledge will constitute potential goals to be achieved at the lowest level of the model on the way towards achieving the general goal underlying the decision-making process. In this step, there is an additional challenge for the knowledge modelers or engineers. This is to discover, as far as possible, all the factors that are not explicitly defined and could be unconsciously considered by the decision maker.

4.2.4 Determining useful data to be used as information or knowledge sources

The objective of this step is to determine all necessary data sources from which the decision makers can gather information or knowledge. At this point, it is important to consider that data are not necessarily available inside the organization. That is, data could be merged from diverse, both internal and external, organizational sources. Finally, it is important to discover and consider all necessary resources since data are the raw material of a data mining project.

4.2.5 Defining the dependency network among the different organizational actors

The purpose of this step is to determine how to establish the dependency network between the different actors involved in the decision-making process underlying the project. That is, the objective is to define how the actors are related to each other, and how responsible they are for the tasks required to achieve the established goals.

A refined goal tree can be built from the information elicited in the above steps (Glinz, 2000) (Martinez et al., 2002). In this goal tree, the highest-level goal is divided into achievement goals, operations and actors. All the recorded information will ultimately constitute the basic information for building a graphical representation of the decision-making process. Table 1 describes the notation used to define the refined goal tree.

SYMBOL	DEFINITION	SYMBOL	DEFINITION
GG	General Goals	AO	Associated Operations
AG	Achievement Goals	RA	Responsible Actors

Table 1. Acronyms to be used

4.2.6 Building the decision process model

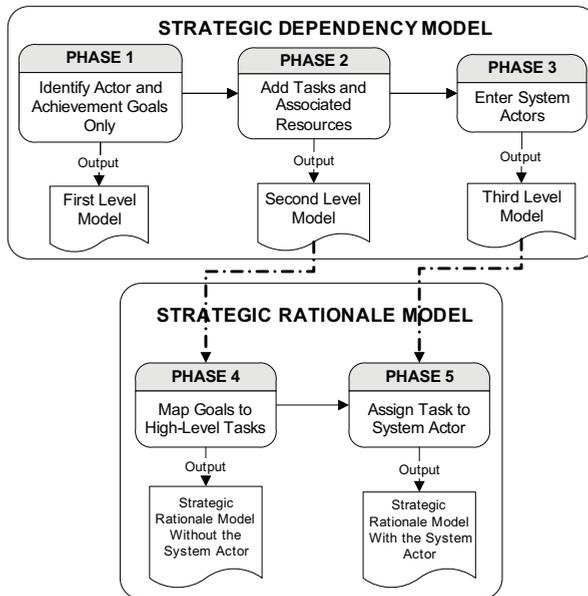Our proposal is to model the decision-making process in a five-step sequence (Figure 4).

Fig. 4. Modeling Process

This is the last modeling process step and consists of applying the i* framework in order to graphically represent the information that was captured in the previous steps. To do this, we use the two complementary models defined by the technique: the *SD model* and the *SR model*.

a.   *SD model*

This is the highest-level model, and its objective is to represent the actors who take part in the decision-making process and their dependency links. These links can be achievement goals, soft goals, tasks, or required and/or generated resources, necessary to achieve the general goals, previously identified in the first step of the process. A three-step incremental development is proposed:

Step 1.   Develop a very basic preliminary model (first-level model) with a high level of abstraction. The objective of this model is for all the stakeholders to be able to understand what the model represents in the simplest way at an achievement goal level. This preliminary model only identifies actors that take part in the process or in the dependency network. The dependencies are defined in the refinement goal tree.

Step 2.   Develop a second model (second-level model) with a higher level of detail. This second model represents the tasks originated by each achievement goal, the resources to achieve these goals and resources produced while developing process tasks or operations.

Step 3.   Define a third model (third-level model) to complete the construction of the SD model), adding the "system" actors. The highest-level achievement goals initially defined in the first-level model are now divided into simpler tasks or operations. Subsequently, we analyze which tasks could be automated or which activities require the support of a software system. This information should be previously entered in the refinement goal tree. The tasks identified for automation are used by the system and converted at this point into new achievement goals that are linked to the system actor. These new goals will are potential use cases in the future system requirements model.

b.   *SR model*

The second model's aim is to more explicitly represent the resources and the granular events (scenario) that originate the required activities to accomplish the achievement goals. The model can also be developed incrementally to give a better process understanding. The number of steps depends on many factors, such as the organizational complexity, modelers' experience, experts' business domain knowledge, and stakeholders' knowledge of the i* framework.

Step 4.   For simplicity's sake, modeling is initially based on the second-level SD model that was output in Step 2. This does not include the system actor. In this step, the achievement goals from the SD model (which were part of the dependency model network) are mapped to high-level tasks. These tasks can be divided by the *task-decomposition* constructor into less complex tasks or into elementary operations that can, depending on their complexity level, be developed by some particular actor. The '*means-end*' constructor is used to represent more than one alternative to achieve a goal or task. Note also that every resource involved in the process involves the specification of the resource's sending and receiving operations in the '*dependee*' actor and in the '*depender*', respectively.

Step 5.   To finish the modeling process, we have to take into account that some of the tasks that have to be developed by some organizational actors need software system support in order to process data and information. For that reason, we add the *system* actor, and consequently, this model has to be built depending on the third-level SD model (Step 3). All the tasks to be automated that were represented as achievement goals in the SD model are assigned to the *system* actor. Then, the

achievement goals of the third-level SD model, which are mapped into high level tasks, are divided into more specific tasks depending on what resources there are and the specific problem's specialization level. Later, this decomposition of high-level tasks (achievement goals in the SD model) into simpler tasks and associated resources is destined to represent the use case scenario in the requirements model.

5. Deriving requirements

In this section, we describe how to build the use case model from the previously built business decision-making model (figure 5).
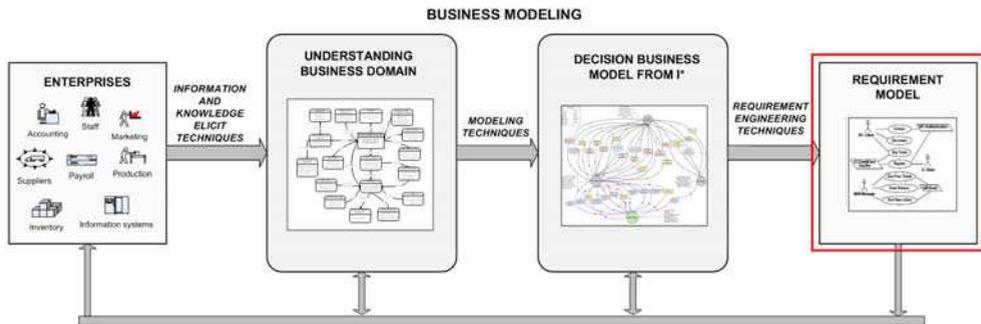

Fig. 5. Requirements modeling process

There are several approaches for extracting software requirements from a business model, like (Ortín et al., 2000), (Alencar et al., 2000) or (Santander & Castro, 2002). The transition guide used in this work is based on Santander and Castro's approach (Santander & Castro, 2002). In (Santander & Castro, 2002), transition from an organizational model to the requirements model is divided into three consecutive steps (see Figure 6). Unlike the approach presented in (Santander & Castro, 2002), use case definition in this research is restricted to the use cases that are derived from the achievement goal dependencies. This restriction is derived from the fact that, in a data mining system, a use case must basically represent the achievement goal that the user intends to achieve using the information or knowledge that the data mining system provides and never a task or goal as in a development-related process. In (Santander & Castro, 2002), use cases can be mapped from a goal dependency, a task dependency or a resource dependency.

Taking into account that we built an organizational model using the i* framework, identifying the actors that participate in the decision-making process, and the main achievement goals are part of a more general strategic goal, the main inputs for building the use case model will be the SD and SR models.

Step 1. *Identifying system actors*. There are several actors that participate in a business decision-making process (i* SD model). They are identified and placed in the refined goal tree; however use case actors are (directly or indirectly) linked to system actors by some kind of achievement goal dependency. Actors that are independent of the system actors cannot be considered use case actors. Additionally, if there are two or more actors that share any dependency linked by an is–a relationship in the i* model, they must be mapped as individual actors in the use case model, linked by a new generalization relationship.
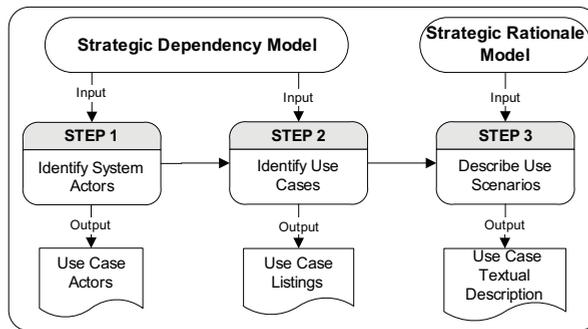
Fig. 6. Requirements modelling activities (based on (Santander & Castro, 2002))

Step 2.   *Identifying Use Cases.* Taking the SD model as input, every goal dependency (dependum) must be identified. Actors identified in Step 1 play the dependee role or the depender role in the dependency relationship.

- An actor that plays the *dependee* role in a goal relationship must provide an informational resource and the dependency object (*dependum*) and the system will generate a use case as output.
- An actor that plays the depender role in a dependency with the system must provide knowledge and information in order to achieve the goals related to the identified use case. In this case, the dependency object (dependum) is again turned into a use case.

Step 3.   *Describing use case scenarios.* The description of the use case scenario related to a specific actor is the third requirements modeling step. Therefore, it includes a description of the event sequence for carrying out tasks and getting resources related to goal achievement. The SR model graphically represents this event sequence. The description of the scenarios associated with use cases is no more than the textual description of a use case, including the field of action of every actor that participates in the achievement of a goal, task or resource represented in the SR model.

6. Case study

In this section, we describe a case study to illustrate the proposed methodology. This case study addresses the creation of a new program of studies at a technical-vocational training institute. The purpose of the case study is to assess the proposed procedure in a real situation in order to illustrate its use and get conclusions for refining the proposed methodology. The scope of developing a case study is limited to modeling the business decision-making process related to the creation of a new program of studies at a technical-vocational training institute (TVTI). This model will then be used to output a requirements model for a data mining system that supports the decision-making process. The assessment method consists of comparing the case study results with other project results in which the methodology was not used.

6.1 Understanding the business domain
This is the first step for modeling business decision-making. It consists of eliciting as much information as possible about the organization. This information will help stakeholders to

assimilate all elements that they need to know to understand the organizational problems to be solved. For example, the elicited information is as follows:

The institution helps to provide people with technical-vocational training that is certified, whenever possible, by accredited bodies. To do this, it teaches the outcomes and values required by the region's industrial and service sectors. The organizational structure and decision-making levels at the institution are set out in the articles of the institution and in TVTI Council meeting minutes. The Council and Executive Committee are in charge of defining the organizational structure and decision-making levels.

In this case study, the head teacher is in charge of proposing the opening of new programs of study to the Council. The Council is responsible for studying and validating the information given by the Executive Director to support the proposal of opening a new degree. Finally, the Executive Director is assisted by a group of collaborators that advise and support the Executive Director, and they must collect and process the information related to the proposal.

On top of this, the institution has the mechanisms and structure required to identify and select the facilities and equipment needed to implement the program. Taking into account the knowledge and outcomes to be learned by the students, the Curricular Committee has to estimate the costs and investment needs for each program to decide whether existing equipment is to be used or it has to be renewed.

6.2 Modeling the decision-making process

In this section we describe the application of the guide proposed in Section 4.2 in order to obtain the decision-making process model of the case study.

6.2.1 Identifying the underlying goal of the decision-making process

From the information elicited in the understanding of the business domain step, we deduce that the underlying goal of the decision-making process is to materialize one of the strategic goals set by the organization: 'provide ongoing technical-vocational training programs that continuously meet the demands of the region's industrial and service sectors'.

6.2.2 Identifying organizational actors that participates in the decision-making process

In this step, we have to identify the actors that participate in the decision-making process. Table 2 shows the identified actors involved in opening the new degree process.

ROLE	ACTOR TYPE	ROLE	ACTOR TYPE
Executive Director	Primary	Consultants	Secondary
Council	Primary		

Table 2. Stakeholders

6.2.3 Eliciting information or knowledge needed to make decisions

According to the elicited information, the decision about whether or not to open a new degree depends on the following information or knowledge:

1. There is a market as companies are demandinmg technical professionals with the profile that the new degree offers, and there are future students interested in taking the new degree

2.   Feasibility study
3.   Support throughout time

6.2.4 Identifying data sources

Table 3 shows the data sources that are available at the beginning of the study. However, Table 3 also shows unavailable data sources that will be required later.

INFORMATION	DATA SOURCES	STATE	DESCRIPTION
Market	Sale management (training)	Available	Data about taught training courses
	Innovation	Available	Data derived from the use of new technologies or new procedures introduced by companies.
	…	…	…
Feasibility study	Utilities	Not Available	Data about utilities that the new degree offers.
	Cost structure	Not Available	Involved cost.
	…	…	…
Support	Student behaviour	Available	Useful data to predict drop-out and graduation rates.
	…	…	…

Table 3. Data sources (partial)

6.2.5 Defining the dependency network between different organizational actors

This step consists of refining the goals included in the goal tree (see Table 4) that defines the general goals, including derived operations, participant actors, and dependency level among actors. Column 1 in Table 4 shows a hierarchy of the identified goals. The general goal is located at the top. The next level contains the existing achievement goals. Finally, the derived operations are entered. Column 2 shows the type of the goal, operation, or available resource. Column 3 shows the actors involved in the process of achieving established goals and executing operations.

6.2.6 Decision process modeling

SD model

Step 1.   This step builds the first-level model. The input is the refined goal tree (Table 4). The model includes the organizational actors that participate in the process and in the goal dependency network only. Figure 7 illustrates this first model for this case study. In the refined goal tree, the first actor in the actor column represents the *depender* and the second actor represents the *dependee* for each goal. The model shows, for instance, that the Executive Director actor (*depender*) depends on whether or not the Council actor (*dependee*) has validated the Reports (*dependum*) submitted by the Executive Director in the development of achievement goal 4.

GOAL NAME	TYPE	ACTORS
New program establishment	GG	Executive Director, Consultants, Executive Council
1. Market knowledge	AG	Executive Director - Consultants
1.1. Data supply	AO	Consultants - Executive Director
…	…	…
2. Feasibility	AG	Executive Director - Consultants
2.1. Data supply	AO	Consultants - Executive Director
…	…	…
3. Support	AG	Executive Director - Consultants
3.1. Data supply	AO	Consultores - Executive Director
…	…	…
4. Report validation	AG	Executive Director - Executive Council
4.1. Send report	AO	Executive Council - Executive Director
…	…	…

Table 4. Goal and operation table for the case under study (partial)
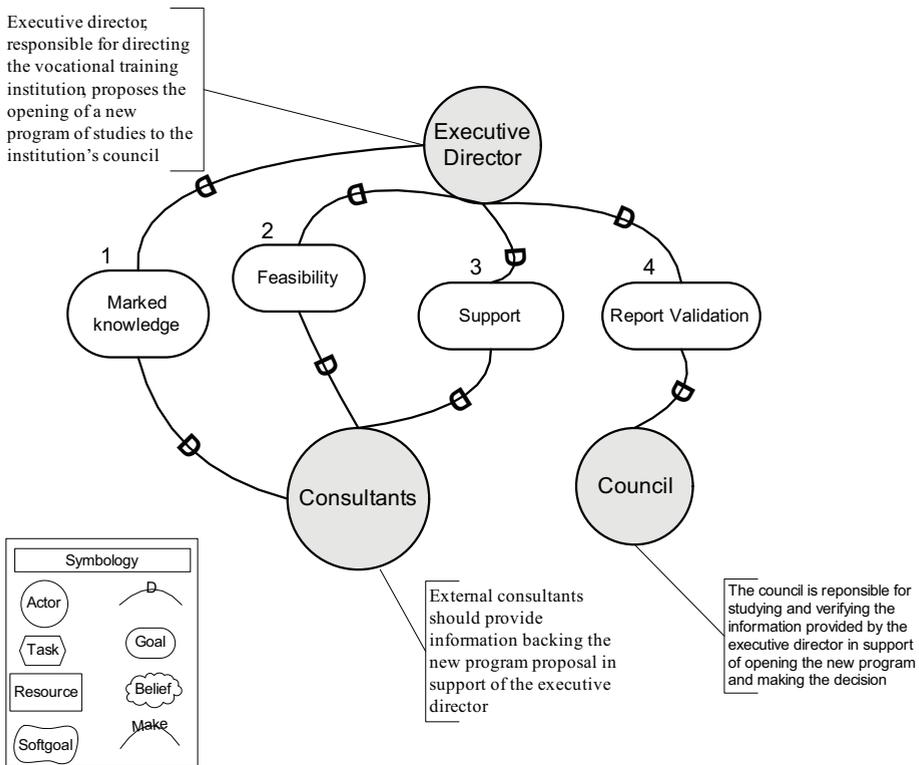
Fig. 7. First-level SD model

Step 2. This second model includes the tasks and resources needed by each achievement goal. Figure 8 shows the model output after completing this second step. For instance, this model specifies the tasks to be performed in order to achieve achievement goal 1 (market knowledge):

1. Executive Director actor (dependee in the task dependency relationship) must send required data to Consultants actor (depender)
2. Consultants actor (dependee) must develop the market analysis.
3. Consultants actor (dependee) must send analysis results to Director actor (depender).
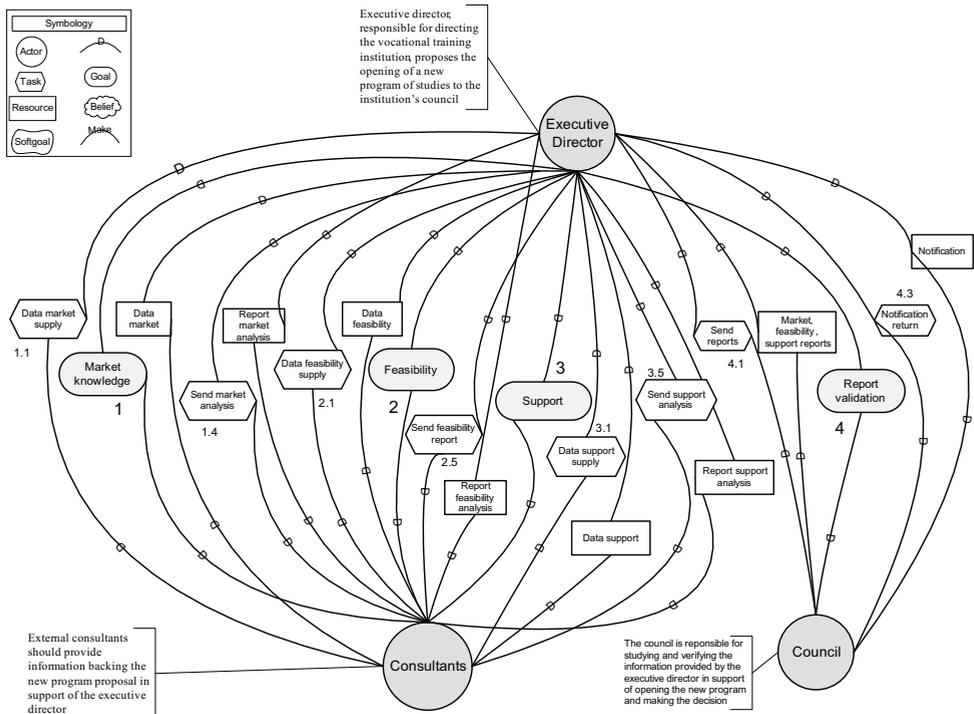
Fig. 8. Second-level SD model

Step 3. The final SD model (Figure 9) includes the system actor that is the responsible for the tasks and activities for which a software system is needed. In this third model, the Consultants actor delegates the data analysis involved in market, feasibility and support analysis to the system actor. Then, these tasks become achievement goals linked to the system actor, and the system actor becomes a *dependee* actor in the dependency relationship with the Consultants actor. It is important to include the system actor in this model since it will be easier to identify the main use cases later on.

SR Model

Following the modeling process, it is now time to build the SR model that justifies the dependencies between the different organizational actors in more detail.
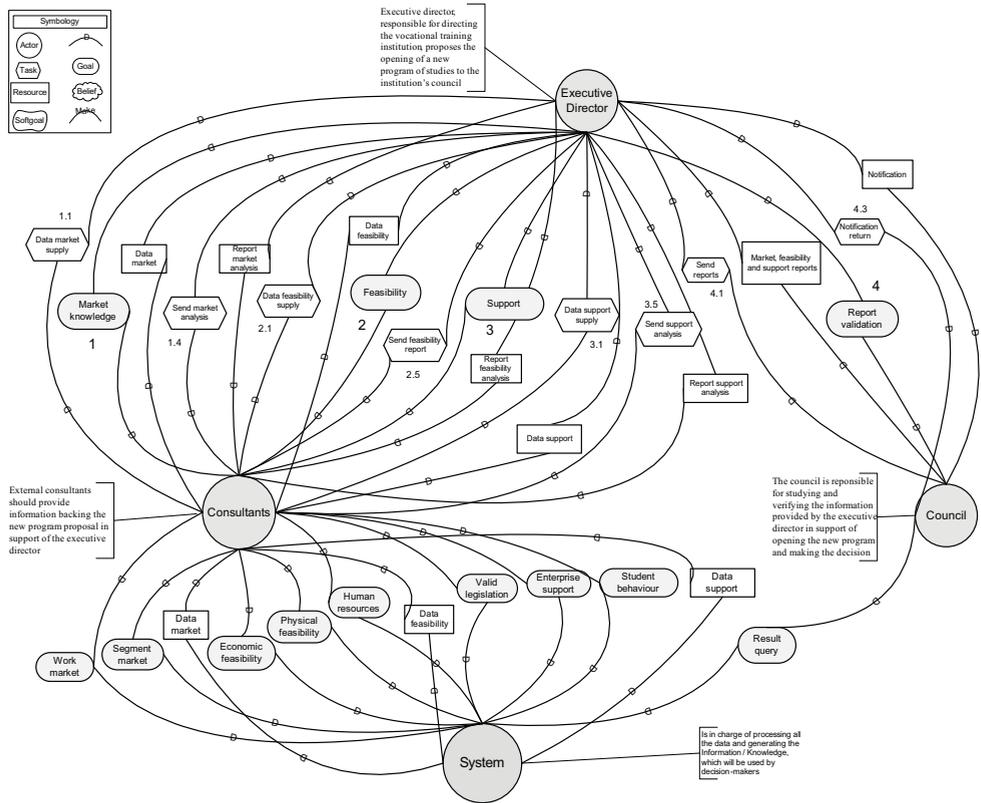
Fig. 9. Third-level SD model

Step 4.   The second-level SD model will be used for modeling to gain a higher level vision. The general goal in this case study is *open a new program of study*. This general goal triggers a sequence of goals, such as *market, feasibility and support studies*. These studies are achievement goals (*dependum*) in the SD model, and they are linked to the *Executive Director* (*depender*) and *Consultants* (*dependee*) actors (Figure 10). These studies involve three high level tasks: *market study query*, *feasibility study query* and *support study query*. Consequently, the *Executive Director* delegates these studies to the *External Consultants* actor. Resource dependency networks are developed when the *Consultants* actor starts these tasks, and these tasks can be divided into smaller tasks such as writing and sending reports.

Step 5.   The modeling process ends with the development of the SR Model (Figure 11), including the System actor. Achievement goals in the third-level SR model are converted into third-level tasks (business domain analysis, market analysis, economic feasibility analysis, etc.). At the same time, each task is divided into more specific tasks depending on the type of available resources and how specialized each problem is. For instance, the *surveys* subtask that uses survey and questionnaires data applied to several companies.
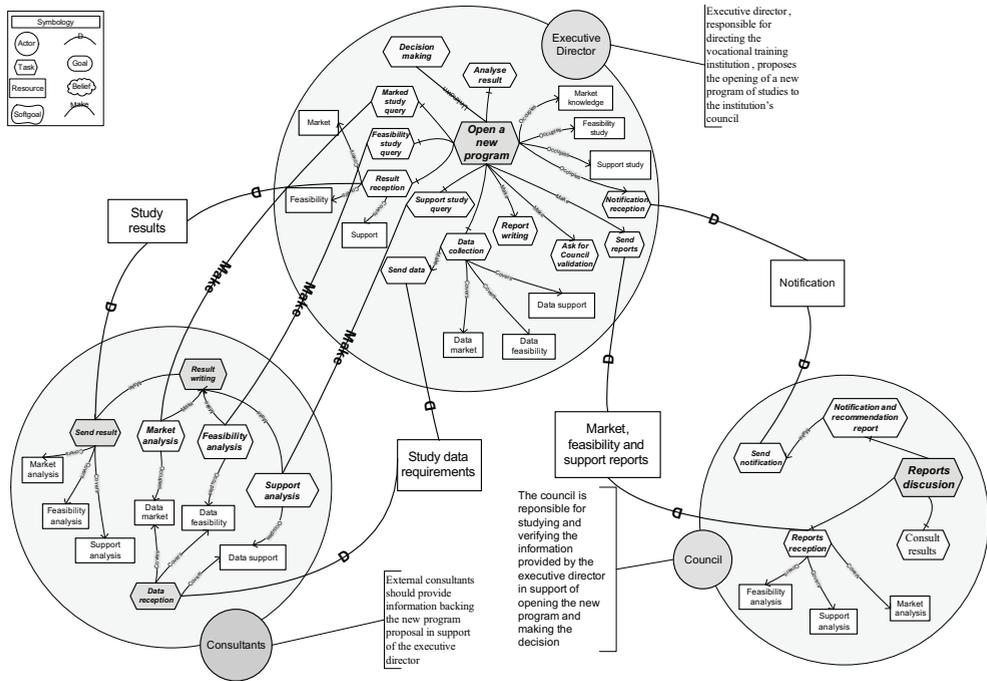
Fig. 10. SR Model

Another feature to be taken into account is the possibility of detailing the different alternatives to be weighed up to achieve a goal or perform a task. In this case study, for instance, tasks involved in the domain analysis can be achieved by validation, discovery and modeling.
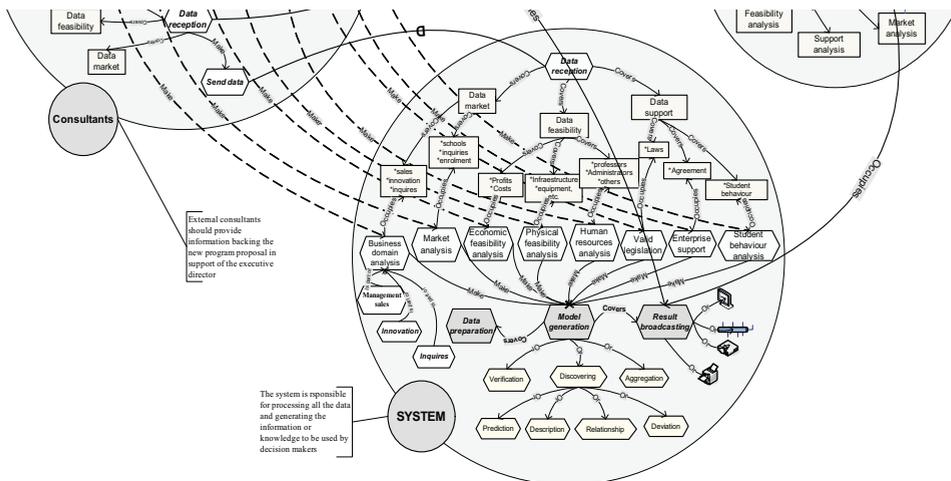

Fig. 11. SR Model with System actor (partial)

6.3 Modeling requirements

In the case under study, after applying the guidelines proposed in Section 4, they are as follows.

Step 1. Defining Use Case Studies:

Starting with the SD model output in the case under study (Figure 9) and applying the first step of the guidelines proposed in Section 4, three potential use case actors are identified: *Executive Director*, *External Consultants* and *Council*.

Looking at the model, we find that the *Council* actor has a dependency network with *System* actor by '*result query*' goal. The '*result query*' goal is a consequence of the '*report study*' achievement goal between *Council* and *Executive Director* actors. Therefore, *Council* actor is a potential actor.

The *External Consultants* actor is also a potential actor since there are dependency networks between *External Consultants* actor and *System* actor through several achievement goals.

There is no dependency network between *Executive Director* actor and *System* actor, therefore *Executive Director* is not considered an actor in the use case model.
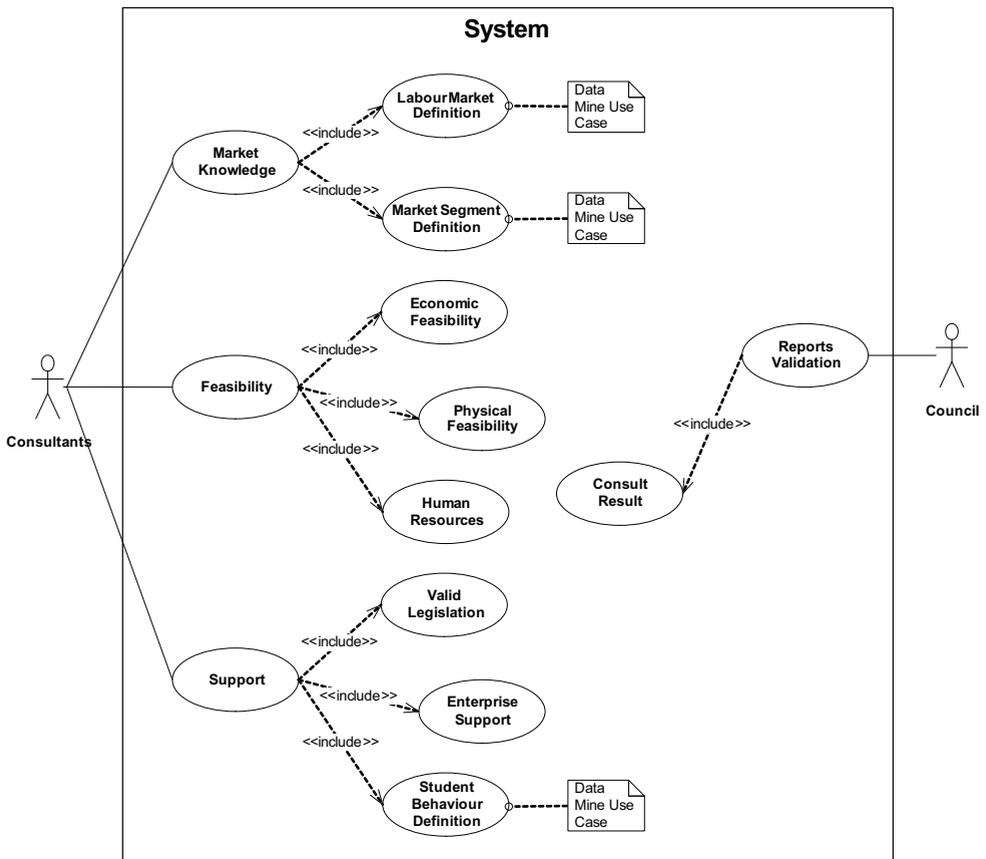

Fig. 12. Use case diagram for the system under study

*Step 2. Defining Use Cases***:**

Taking the SD model (Figure 9) as input, we can identify, according to Step 2 of the guidelines, the following use cases

1. Taking into account the identification of goal dependency networks in which previously identified actors play *a dependee* role in the dependency network, the following elements are identified:

    Actor: *Council* / Dependency objects: *Report Validation* / Actor: *Consultants* / Dependency objects: *Market knowledge, Feasibility, Support.*

2. Taking into account the identification of goal dependency networks, in which actors play a *depender* role in the dependency network with the System Actor, the following elements are identified:

    Actor: *Council* / Dependency objects: *Result query* / Actor: *Consultants* / Dependency objects: *Business domain, Market segment, Economic Feasibility, Physical Feasibility, Human Resources, Current Law, and Student Behavior*.

In the case under study, the use case diagram (Figure 12) obtained from i* models is based on standard UML notation. The graphical representation of the model shows the existing relationship between organizational actors and the system. The subsequent model description shows the event sequence that is triggered after an actor runs a use case.

Step 3. Describing the use case scenario

A description model is written for every identified use case to complete use case modeling. To do this, one of the templates proposed in the scientific literature (Robertson & Robertson, 1999) (Larman, 2003) can be used. The description must include actor intentions, and the responsibilities associated with the system.

Use Case	**Market Segment Definition**
	ID: RQ_006
	Type: DM
	Creation Date: 22/11/07
Description	This use case involves the data analysis of a specific market segment. It can predict potential students for a new program of studies.
Primary Actor	External Consultants
Assumptions	All the data required to run the study are available.
Resources	1. Regional school data 2. Results of surveys and questionnaires 3. Statistics related to Enrolment in Institutes and Universities
Steps	1. Provision of all the market data received by the head teacher by consultants 2. Definition of the data required for the study. 3. Data preparation 4. Definition of the models for developing the study 5. Results presentation

Table 5. Textual description of 'Market Segment Definition' use case

In this way, we will detail when actors ask for services or provide resources to the system and when the system gives information to the system actor. The information required to describe the use cases can be obtained from the SR Model (Figure 12). Table 5 shows an example of the description process for the 'Market Segment Definition' use case. After

finishing the descriptions of all data mining use cases, the data mining requirements specification is developed.

7. Evaluation of the proposed methodology

In order to establish an evaluation baseline and estimate the benefit of the proposed methodology, we compare two instances in which a data mining project was developed. In both cases, the problem domain is the same and the objectives are similar; however, the users and the data analysts are different in both cases because the two project instances were developed at different times. The results are compared in Table 6.

CRITERIA	PROJECT A: NOT APPLYING THE METHODOLOGY	PROJECT B: APPLYING THE METHODOLOGY
Goal achievement	The objectives were established informally and therefore when the project ended it was difficult to establish if the business objectives were fulfilled.	The project's objectives were well-established and were clearly aligned with the business objectives.
User participation	User participation dropped significantly during the project development time.	Users participated actively in all stages of the project.
Development time	The development time was greater than initially planned (the project took over 30% longer)	The development time was as initially scheduled.
Effort	Workload was greater than initially planned (over 40% more than the initially estimated person/hours).	The development effort was as initially planned.

Table 6. Comparing two projects to report benefits of the methodology

Project A (developed without the implementation of the proposed methodology) tried to identify relevant factors, to discriminate between higher education students that failed and students who successfully completed their undergraduate programs. The search was based on information about the university students' income. Descriptive models were used to build predictive models to forecast the likelihood of a new student enrolling for an undergraduate program successfully completing the program.

Project B (developed using the proposed methodology) tried to support a specific and strategic goal of the organization, which is "to improve the academic work assessment and academic support systems, and the management control of financial resources and organizational materials".

In general, from the development of the data mining project applying the methodology, we found that the methodology actually brings together a number of relevant aspects that should be considered at the beginning of the development of a data mining project, such as the requirements to be met by the project, the necessary resources, the project risks and constraints and, generally, all important aspects to be taken into account and that emerge from the business domain understanding phase outlined by the CRISP-DM standard.

8. Conclusions

In the presented work, we proposed a new methodology that consists of a sequence of steps for developing a business model of a decision-making process in a company or organization. The methodology uses the business model is as input to get organizational requirements and use cases applied to data mining projects. A decision-making process model is useful for defining what tasks of the strategic decision-making process can be supported by a data mining project and, also, outputs the initial project requirements.

A requirements model ensures that data mining project results will meet users' needs and expectations, effectively supporting the decision-making process involved in achieving the organizational goals. The construction of the organizational model, which will be used to model the requirements, is based on an incremental and iterative process that provides a better understanding of the business. Additionally, it is useful for reaching agreement, negotiating and validating that the model faithfully represents the organization's decision-making process and checking that the business problem really requires the support of a data mining system.

Note that not only can the requirements model, output based on the organization's business model, identify data mining use cases; it can also pinpoint other functionalities that can be implemented by other conventional software systems that work together with the data mining systems in order to achieve organizational goals.

As regards the modeling technique used in this research, we have shown that the i* framework has valuable features that make a decision-making process model easier to develop. Another important idea is the fact that the i* framework is useful for explicitly representing non-functional requirements associated with functional requirements in the organizational business model. This can be done using the soft goal dependency objects.

9. References

Aguilar Savén R.S. (2004), "Business process modelling: Review and framework", International Journal of Production Economics. Vol. 90, No. 2, p. 129 – 149.

Alencar F., et al. (2000), "From Early Requirements Modeled by the i* Technique to Later Requirements Modeled in Precise UML", III Workshop de Engenharia de Requisitos.

Berenbach B. (2004), "Comparison of UML and Text based Requirements Engineering", 19th annual ACM SIGPLAN Conf. on Object-oriented programming systems, languages, and applications (OOPSLA'04).

BPMI. (2004), Business Process Management Initiative: Business Process Modeling Notation (BPMN), Specification Version 1.0, May 3.

BPMN_OMG. (2006), "Business Process Modeling Notation Specification", OMG Final Adopted Specification, OMG.

Bruckner R., et al. (2001), "Developing Requirements for Data Warehouse Systems with Use Cases", Seventh Americas Conference on Information Systems, pp. 329-335.

Castro, J., et al. (2001), "Integrating Organizal Requirements and Object Oriented Modeling", Proceedings of the 5th IEEE International Symposium on Requirements Engineering

Chapman P., et al. (2000), "CRISP-DM 1.0 step-by-step data mining guide", Technical report, 2000.

Chatam B., et al.(2002), "CMR´s future: Humble growth through 2007".

Cysneiros L. & Sampaio, J. C. (2004), Nonfuncional Requirements: from elicitation to conceptual models, IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 30, NO. 5.

Dale M. (2004), "Defining user requirements for a large corporate data warehouse: an experiential case study", 9th Australian Workshop on Requirements Engineering, pp. 5.1-5.11.

Davyt N. (2001), "Ingeniería de requerimientos: una guía para extraer, analizar, especificar y validar los requerimientos de un proyecto", Facultad de Ingeniería, Universidad ORT del Uruguay.

DiLauro L. (2000), "What´s next in monitoring technology? Data Mining Finds a Calling in Call Centers".

Eisenfeld B. (2003), et al., "42 percent of CMR software goes unused", http://www.gartner.com, February 2003.

Firesmith D. (2003), "Engineering Security Requirements", in *Journal of Object Technology*, vol. 2, no. 1, January-February, pp. 53-68

Gacitúa R. (2001), Identification of requirements: a focus based on a verb taxonomy, Theoria, Vol. 10: 67-78, ISSN 0717-196X.

Gerhard Armin P. (1997), "Requirements Acquisition and Specification for Telecommunication Services", tesis doctoral, University of Wales, Swansea, UK.

Glinz M. (2000), "Problems and Deficiencies of UML as a Requirements Specification Language", Proc. of the 10th Int. Workshop on Software Specification and Design (IWSSD'00).

Gordijn, J., et al. (2000), "Value Based Requirements Creation for Electronic Comerce Applications", Proceedings of the Hawaii Internacional Conf. On System Sciences, January 4-7, Hawaii.

Gordijn, J. (2003), "Value-based Requirements Engineering Exploring Innovative e-Commerce Ideas", VRIJE UNIVERSITEIT.

Gordijn J., & Akkermans. H. (2007), Business Models for Distributed Energy Resources In a Liberalized Market Environment. In *The Electric Power Systems Research Journal*, Vol. 77(9):1178–1188, Elsevier.

Gorschek T. & Claes W. (2006), "Requirements Abstraction Model", Requirements Eng, 11: 79–101, DOI 10.1007/s00766-005-0020-7.

Han J., & Kamber M. (2001), "Data Mining: Concepts and Techniques", Academic Press.

Hayes J. & Finnegan P. (2005), "Assessing the of potential of e-business models: towards a framework for assisting decision-makers", European Journal of Operational Research 160(2): 365-379.

Heitmeyer, C. & Bharadwaj, R. (2000), "Applying the SCR Requirements Method to the Light Control Case Study." *Journal of Universal Computer Science 6*, 7: 650-678.

Heninger, K., et al. (1978), "Software Requirements for the A-7E Aircraft." *Technical Report 3876*. Washington, D.C.: Naval Research Laboratory.

Heninger, K. L. (1980) "Specifying Software Requirements for Complex Systems: New Techniques and their Application." *IEEE Trans. on Software Engineering SE-6*, 1 (January): 2-13.

Hermiz, K. (1999), "Critical Success Factors for Data Mining Projects", DM Review Magazine, February, 1999,

http://www.dmreview.com/issues/19990201/164-1.html.

Johnston, S. (2004), "Rational UML Profile for Business Modeling", Julio, 2004, Disponible en línea

   http://www-128.ibm.com/developerworks/rational/library/5167.html#author1

Kantardzic M., & Zurada J. (2005), "Trends in Data Mining Applications:From Research Labs to Fortune 500 Companies", Next Generation of Data Mining Applications", IEEE, Wiley

Kelley Ch., & Adelman, S. (2003), "¿Where can I find sources about failed data mining projects and the reason for their failure?", DM Review Online Published in April 2003. DMReview.com.

Kdnuggets. (2007),

   http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm.

Knorr, K. & Rohrig, S. (2001), "Security Requirements of E-Business Processes," 73-86. Towards the E-Society: E-Commerce, E-Business, and E-Government. First IFIP Conference on E-Commerce, E-Business, E-Government, Zurich, Switzerland, Oct. 4-5, 2001. Norwell, MA: Kluwer Academic Publishers, (ISBN 0792375297).

Kotonya G., & Sommerville I. (1998), *Requirements Engineering. Processes and techniques*", USA. J. Wiley.

Koubarakis M., & Plexousakis D., Business process, modelling and design: a formal model and methodology, BT Technol J Vol 17 No 4 October 1999.

Lagha B., et al. (2004), "An ontology for e-business models" In Value Creation from E-Business Models. W. Currie, Butterworth- Heinemann.

Larman C. (2003), "UML y Patrones, una introducción al análisis y diseño orientado a objetos y al proceso unificado", 2ª. Edición, Ed. Prentice Hall.

Laudon K., & Laudon J. (2004), *Sistemas de Información Gerencial"*, Ed. Prentice Hall.

Leiwo, J. Et al. (1999), "Organizational Modeling for Efficient Specification of Information Security Requirements," 247-260. *Advances in Databases and Information Systems: Third East European Conference, ADBIS'99*. Maribor, Slovenia, Sept. 13-16, 1999.

Maciaszek L. (2005), *Requirements Analysis and System Design"*, 2a Edition, Ed. Addison Wesley.

Marbán Ó., et al. (2008), "Towards Data Mining Engineering: a Software Engineering Approach", preprint submitted to Elsevier Science.

Martinez A., et al. (2002), "From Early Requirements to User Interface Prototyping: A methodological approach", 17th IEEE International Conference on Automated Software Engineering 2002, September 23-27, Edinburgh, UK

McDonald P.P., et al. (2006), "Growing its contribution: The 2006 CIO Agenda", Gartner group, http://www.gartner.com.

Medina, J. C. (2004), "Análisis comparativo de técnicas, metodologías y herramientas de Ingeniería de Requerimientos", Tesis Doctoral, CINVESTAV, Junio de 2004, México, D.F. México.

Meta Group Research. (2003), The Top 5 Global 3000 Data Mining Trends for 2003/04 Enterprise Analytics Strategies, Application Delivery Strategies, META Group Research-Delta Summary, 2061.

Mylopoulos J. et al. (2002), "Towards Requirements-Driven Information Systems Engineering: The Tropos Project.To appear in Information Systems", Elsevier, Amsterdam, The Netherlands.

Ochoa, A. (2006), "Uso de Técnicas de Educción para el Entendimiento del Negocio", Tesis de Magister en Ingeniería del Software. Escuela de Postgrado. Instituto Tecn. de Buenos Aires.

Ortín M. J. et al. (2000), "De los procesos de negocio a los casos de uso" , JISBD 2000, Valladolid, España   Fecha: Noviembre 2000.

Osterwalder A. et al. (2005), "Clarifying business models: origins, present, and future of the concept", Communications of AIS, Volume 15, Article, May.

Martyn A. (1995), Business Processes - Modelling and Analysis for Re-engineering and Improvement. John Wiley & Sons, Chichester, England.

Pérez C. & Santín D. (2007), "Minería de Datos Técnicas y Herramientas", Ed. Thomson.

Piatetsky-Shaphiro G. & Frawley W. (1991), "Knowledge Discovery in Databases", AAAI/MIT Press, MA.

G. Piatetsky-Shaphiro. (2000), "Knowledge Discovery in Databases: 10 Years After", SIGKDD Explor. Newsl., 1(2):59–61.

Rilston F., Paim S., & Castro J. (2003), "DWARF: An approach for requirements definition and management of data warehouse systems", 11th IEEE International Requirements Engineering Conference (RE'03), p 75.

Robertson S. & Robertson J. (1999), "Mastering the Requirement Process", Ed. Addison – Wesley.

Russell, N., et al. (2006), "On the Suitability of UML 2.0 Activity Diagrams for Business Process Modelling", Third Asia-Pacific Conference on Conceptual Modelling (APCCM2006), Australia. Conferences in Research and Practice in Information Technology, Vol. 53.

Sánchez M. A. (2006), "Una recomendación para el desarrollo de software en un contexto de negocio bajo demanda de acuerdo a la especificación MDA y la arquitectura SOA", tesis doctoral, Universidad Pontificia de Salamanca.

Santander V. & Castro J. (2002), "Deriving Use Cases from Organizational Modeling", Proceedings of the IEEE Joint International Conference on Requirements Engineering (RE'02), IEEE.

Schewe K.D. (2000), "UML: A modern dinosaur? – A critical analysis of the Unified Modelling Language", 10th European – Japanese Conference on Information Modelling and Knowledge Bases, Saariselk, Finlandia.

Slagell, M., et al. (2002), "A Software Fault Tree Approach to Requirements Analysis of an Intrusion Detection System." Requirements Engineering 7, 4 (December): 207-220.

Sommerville I. (2002), "Ingeniería de Software", 6ta. Edición, Ed. Addison Wesley.

Sommerville I. (2005), "Ingeniería de Software", 7ma. Edición, Ed. Addison Wesley.

Sparx Systems (2008) (portal), "UML 2.1 Tutorial", [en línea], disponible en: http://www.sparxsystems.com.au, [Consulta: 13 de febrero de 2008].

Vérosle J., et al. (2003) A generic model for WLAN, hostpots- A roaming, business case in The Netherlands, WMASH'03, Septembre 19, San Diego, California, USA.

Weber M. & Weisbrod J. (2003), "Requirements engineering in automotive development: Experiences and challenges". IEEE Software, pages 16 –24, Enero/Febrero.

White S., IBM Corporation, "Introduction to BPMN", Stephen A. White. All Rights Reserved, www.bptrends.com, BPTrends July, 2004.

Wilcox, P. & Gurau C. (2003), "Business modelling with UML: the implementation of CRM systems for online Retailing", Journal of Retailing and Consumer Services 10, 2003, available online at www.sciencedirect.com.

Wohed, Petia, et al. (2006), "On the Suitability of BPMN for Business Process Modelling", In proceedings, 4th International Conference on Business Process Management 4102/2006, pages pp. 161-176, Vienna, Austria.

Yu E. (1995), *Modelling Strategic Relationships for Process Reengineering*, Ph.D. thesis, Department of Computer Science, University of Toronto.

Yu Eric S.K. & Mylopoulos J. (1996). "AI Models for Business Process Reengineering.", in IEEE Expert Intelligent Systems and Their Applications. IEEE Computer Society. Volume 11, Number 4. pp. 16-23.

Yu, E. & Mylopoulos J. (1997), "Enterprise Modelling for Business Redesign: the i* Framework", Special Issue: Enterprise Modelling Papers, SIGGROUO, Vol.18, No. 1, April.
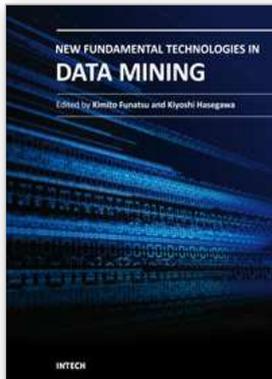
New Fundamental Technologies in Data Mining

Edited by Prof. Kimito Funatsu

The progress of data mining technology and large public popularity establish a need for a comprehensive text on the subject. The series of books entitled by "Data Mining" address the need by presenting in-depth description of novel mining algorithms and many useful applications. In addition to understanding each section deeply, the two books present useful hints and strategies to solving problems in the following chapters. The contributing authors have highlighted many future research directions that will foster multi-disciplinary collaborations and hence will lead to significant development in the field of data mining.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Oscar Marban, José Gallardo, Gonzalo Mariscal and Javier Segovia (2011). From the Business Decision Modeling to the Use Case Modeling in Data Mining Projects, New Fundamental Technologies in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-547-1, InTech, Available from: http://www.intechopen.com/books/new-fundamental-technologies-in-data-mining/from-the-business-decision-modeling-to-the-use-case-modeling-in-data-mining-projects

INTECH
open science | open minds