Goal oriented requirements engineering in data warehouses: A comparative study

Ingeniería de requerimientos orientado a objetivos en almacenes de datos: un estudio comparativo

A. Cravero Leal1, S. Sepúlveda2, A. Mate3, J.-N. Mazón4 and J. Trujillo5

ABSTRACT

Data warehouses provide historical information about the organization that needs to be analyzed by the decision makers; therefore, it is essential to develop them in the context of a strategic business plan. In recent years, a number of engineering approaches for goal-oriented requirements have been proposed, which can obtain the information requirements of a data warehouse using traditional techniques and the objectives of the modeling. This paper provides an overview and a comparative study of the treatment of the requirements in the existing approaches to serve as a starting point for further research.

Keywords: Requirements Engineering, Goal Oriented, Data Warehouses.

RESUMEN

Los almacenes de datos proveen información histórica de la organización que requiere ser analizada por los tomadores de decisiones, por lo que es primordial desarrollarlos en el contexto del plan estratégicos del negocio. En los últimos años se han propuesto una serie de enfoques de ingeniería de requerimientos orientada a objetivos que permiten obtener los requisitos de información, a cubrir por el almacén de datos, mediante técnicas tradicionales y del modelado de objetivos. Este trabajo, ofrece una visión general y un estudio comparativo del tratamiento de los requisitos en los actuales enfoques con el fin de servir de punto de inicio a posteriores trabajos de investigación.

Palabras clave: Ingeniería de Requerimientos, Orientado a Objetivos, Almacenes de Datos.

Received: June 24th 2013
Accepted: February 28th 2014

Introduction

Data Warehouses (DW) are a collection of historical data of any type of organization. The historical data are analyzed by the decision-makers, converting the data into strategic information to support the decision-making process (Giorgini, et al., 2005). These DWs integrate a huge amount of data coming from heterogeneous data sources in a multidimensional (MD) model. This model enables the users to access the data in a more natural way, by means of its structure, composed of facts (measures of analysis) and dimensions (context of the fact analysis) (Kimball, & Ross, 2002; Mazón, et al. 2007).

Diverse studies have indicated that most of these MD models do not include the required information due to a lack of communication between the DW developers and the business analysts. The main reason is that DWs are traditionally designed by using the information available in the operational data sources that will integrate the MD model, without taking the main users, the business analysts and their needs into account in the design process (Avila, et al., 2008). Therefore, it is necessary to carry out a Requirements Engineering (RE) phase, which includes the modeling of the DW information requirements from the user needs, e.g.: the analysis of the strategic goals that must be met (Giorgini, et al., 2005; Mazón, et al., 2007).

In order to carry out the RE phase, there are a series of approaches that provide the information requirements in a systematic way. These approaches use goal modeling for elicitation, specification and validation of requirements. In this work, a general overview is shown, as well as a comparative study, of the eight Goal Oriented (GO) approaches for DW. This paper presents a detailed and updated review presented in the chronological study (Cravero, & Sepúlveda, 2012), focusing specifically on the techniques used in the requirements engineering stage and goal modeling. In this comparative study, we analyze how they represent actor goals, the techniques used at each stage of the RE phase and the goal models employed by each approach because there is no...
established standard to carry out the aforementioned phase. The main motivation of this work is to serve as a starting point for researchers in a new area, which is becoming increasingly important in the DW field and for decision support systems. Consequently, this comparative study can be useful for researchers in achieving a common understanding and providing a solid foundation for the research community.

The remainder of this paper is structured as follows. Firstly, in the basic concepts section, we describe what a DW is and what GO is about. Then, the main GO approaches for the proposed DW are described to the best of our knowledge. Then, a comparative study is detailed. Finally, in the last section, we describe the main conclusions and the future work.

Basic Concepts

DataWarehouses

The classic definition of a DW was proposed by Inmon (Inmon, 1996) as a subject oriented, non-volatile, integrated, and time variant collection of data in support of management’s decisions.

The main contribution of a DW is its capability to convert data into strategic information, supporting the decision-making process at higher levels of the organization. In order to achieve faster and more flexible queries, the data are structured in a multidimensional way (known as star schema) where the information is classified as facts and dimensions (Kimball, & Ross, 2002). The facts are the numeric data, which represent a specific industrial activity to be analyzed. The dimensions are the individual perspectives of the data, which define the granularity (level of detail) to be adopted for the representation of a fact. The units of the facts and their values are called measures (Kimball, & Ross, 2002).

Goal Oriented Requirements Engineering

Requirements Engineering: The IEEE definition of requirement is “a condition or capability that must be presented in a system or its components in order to satisfy a contract, standard, specification or other formal document” (IEEE, 1998). However, for the users of the software systems, the requirements are the necessary conditions or capabilities needed to solve a problem or achieve a goal (Kavakli, E., & Loucopoulos, 2008).

However, the RE is related to those activities as the elicitation, specification and validation of requirements, in order to understand exactly the user needs and translate those needs to a set of unambiguous, accurate sentences. The main techniques that have been traditionally used in each stage in RE according to various authors are (Sommerville, 2007):


2. Specification: Scenarios, Templates or patterns, Use Cases, Natural Language, Formal Languages.


For a long time, the RE was only focused on what the system should do; however, in later works, Yu and Mylopoulos showed the necessity of understanding the organization’s environment and the interaction that it should have with the system (Yu, E., & Mylopoulos, 1998). Recently, the systems are considered as a contribution to the business solutions, and the relationship between the systems and their environment is expressed in terms of relationships based on goals (Kavakli, E., & Loucopoulos, 2008). This is due to the more dynamic environments in which the businesses develop currently, where the systems are used to change the business process, instead of automation (Kavakli, E., 2002). This is the reason why modeling techniques, which include the business goals as an important element, are needed. In this sense, the introduction of goals offers a way to clarify the system requirements, thus creating a new RE approach, a Goal Oriented one, which Kavakli has named GORE (Goal Oriented Requirement Engineering) (Kavakli, E., 2002; Kavakli, E., & Loucopoulos, 2008).

Overall, GORE focuses on the activities that precede the formulation of software system requirements. The following main activities are normally presented in GORE approaches: goal elicitation, goal refinement and various types of goal analysis, and the assignment of responsibility for goals to agents (Lapouchnian, 2005).

While there is a consensus on how to develop a data-driven approach, there is no consensus on how to develop a GORE phase. Therefore, we present this comparative study in order to show the diverse RE techniques used by GORE for the different proposals.

According to the identified tasks, Kavakli classifies the different GORE approaches used in RE (see Table 1). A detailed description and a comparative analysis of each approach can be found in (Kavakli, E., 2002; Kavakli, E., & Loucopoulos, 2008).

### Table 1. Goal oriented analysis roles in relation to the activities in GORE. Based on (Kavakli, E., & Loucopoulos, 2008)

<table>
<thead>
<tr>
<th>RE Activity</th>
<th>Goal Analysis Contribution</th>
<th>Goal-Oriented Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specification</td>
<td>2. Relationship between the KAOS, GBRM, NFR, GSC business strategic goals and the functional and non-functional components of the system</td>
<td>Goal-Oriented Approach</td>
</tr>
<tr>
<td>Validation</td>
<td>3. Validation of the system GQM, GSM, URN specifications with the organization goals</td>
<td>Goal-Oriented Approach</td>
</tr>
</tbody>
</table>

Kavakli (Kavakli, E., 2002) argues that a distinction between the foreseen needs in early stages and the ones in later stages exists, which can lead to conceptions of the goals in different ways. In the early GORE stages, it is more important to model and analyze the actor needs (in our case the business analysts) and their interests and also analyze how they can be compromised by the decision of introducing a new system (Kavakli, E., 2002; Kavakli, E., & Loucopoulos, 2008). The later stages are centered on the future goals and how these can be implemented in terms of system components.

Goal Oriented analysis in Data Warehouses

The DW is an important part of the budget for information technology in most organizations. Successful projects have confirmed a high level of user satisfaction and return of investment. Despite the recognized potential, many projects fail to deliver the information expected to support the decision-making process (Weir, et al., 2003; Winter, & Strauch, 2003). There seems to be a consensus in the community that behind these failures there is a poor requirements definition stage (Giorgini, et al., 2005; Naveen Prakash, & Gosain, 2008).
In this sense, the primary objective of the RE phase for a DW is to model the users, their goals and the relationships between them, in order to achieve the strategic business goals (Frendi, & Salinesi, 2003; Giorgini, et al., 2005; Mazón, et al., 2007). Therefore, this phase is crucial to the development of DWS, allowing developers to locate the DW in the context of the organization, and align it with its strategic objectives.

The RE phase for DW should be based on the GORE framework proposed by Kavaklı for the following reasons: (i) the DW is intended to provide adequate information to support decision-making, thus helping to meet the strategic objectives of an organization, (ii) the requirements of a DW are difficult to obtain from scratch, as information analysts often only express the general expectations about the goals that the DW should support, and (iii) the DWs have many users with different types of interests, and, therefore, with different interrelated goals that must be modeled to obtain an MD model that satisfies them (Giorgini, et al., 2008). Currently, this type of approach of RE for DW is known as Goal Oriented.

Various GO approaches exist in the DW field and are summarized later. Note that because some authors do not describe in their studies all the software engineering techniques used for the RE phase, we obtained this information from other sources (Romero, & Abelló, 2009) and (Golfarelli, & Rizzi 2009). We have included an alias for each approach for an easier identification.

**Bo01: Bonifati et al. (2001)**

In this approach, developers obtain the needs of the DW users through interviews, expressing their expectations though the paradigm Goal / Question / Metric (GQM), which consists of a set of forms that are completed in four steps. (Bonifati, et al., 2001). Unfortunately, their studies do not mention other software engineering techniques used for the RE phase.

**SP03: Silva-Paim and Castro (2003)**

They create an approach named DWARF, which is divided into a series of well-defined stages. Each stage presented in a development cycle, applies different levels of abstraction that detail the application more deeply each time, with the goal of creating a baseline for requirements. To identify and validate the information requirements of the DW they use a variety of software engineering techniques such as interviews, brainstorming, checklist, prototypes, use case scenarios using the NFR framework for non-functional requirements, and traceability matrices to support change management (Silva-Paim, & Castro, F. B. 2003).

**GS06: Gam y Salinesi (2006)**

In this approach, named CADWA, the developers obtain the information requirements from: (1) the goals presented by the strategic business plan, (2) the goals of decision makers, (3) the structure of transactional systems, and (4) the structure of the existing DW models that can be reused (Gam, & Salinesi, 2006). With these sources, they create a model using the MAP goal model (Etien, & Salinesi, 2005), to represent the current and future interests of decision makers.

**Ma07: Mazón et al. (2007)**

This approach (Mazón, et al., 2007) uses the i* framework (Yu, E., 1995) to incorporate within a model, through interviews, the strategic, decisional and informational goals of each analyst. A set of information requirements (tasks and resources) is obtained from the informational goals, which is incorporated into a conceptual model for the DW actor using an adapted i* model, which will give rise to the design of a conceptual MD model using a UML profile (Mazón, et al., 2007). The proposed approach is supported by a CASE tool named DaWaRA (Glorio, et al., 2008).

**Gi08: Giorgini et al. (2005, 2008)**

In this approach, named GRaND, two perspectives are used: (1) shaping the organization (which consists of strategic analysis, analysis of each actor’s goals to obtain the facts and analysis of attributes), and (2) modeling of decision-making, which focuses on decision makers (Giorgini, et al., 2005). The requirements are obtained through interviews, which are documented in templates and tables. To model the goals of the organization, they use the Tropos model, containing a variant of i*, GRaND is supported by a CASE tool called DW-Tool (Giorgini, et al., 2008).

**PG08: Prakash y Gosain (2008)**

In this approach, the authors focus on the broader context of the organizational goals to design a DW. In this sense, the strategic goals enable to identify the set of decisions that are relevant, which helps with determining the content of the DW (Prakash, N., & Gosain, 2008). In order to obtain a technical point of view, they use the concept of scenario information, which is written for each decision that is going to be supported and is available in the GDI (Goal-Decision-Information model) organization scheme. (Prakash, N., & Gosain, 2003).

**DT12: Di Tria et al. (2012)**

The aim of this approach is the definition of a sequential hybrid method (Di-Tria, et al., 2012); it takes into account the advantages of Dimensional Fact Model (Golfarelli, & Rizzi, 2009), formalization of user requirements represented by UML multidimensional schemas (Mazón, et al., 2007) obtained from i* framework, to gather all the advantages of each schema. The authors present a framework to be used for the conceptual design of data warehouses. Such a framework starts from the analysis of the business goals defined by decision makers. Using these goals, a schema representing information requirements is first produced. Then, discovery of facts and dimensions from the information requirements help to suitably derive an initial conceptual schema.

**AC13: Cravero et al. (2013)**

This paper addresses data warehouse design from a business perspective by highlighting business strategy analysis, alignment between data warehouse objectives and a firm’s strategy, goal-oriented information requirements’ modeling and how an underlying multidimensional data warehouse model may be derived (Cravero, et al., 2013). This approach considered business strategy using vision, mission, objectives, strategies, tactics requirements analysis, a business motivation model (BMM) for aligning DW objectives and organizational strategy, modeling such objectives with i* and deriving the underlying multidimensional model of the DW by means of a unified modeling language (UML) profile (Mazón, et al., 2007).

**Comparative analysis of the approaches**

Although there are a variety of methodologies and approaches for the design of DW, the researchers believe that the RE field is still very poorly developed. In this sense, Rizzi and other authors indicate that “A very few comprehensive methods that have been devised so far” (Giorgini, et al., 2008). Overall, these authors believe...
that some specific issues in RE, have not been properly investigated yet. Generally, as yet, there is no common strategy for the development of data warehouses (Cravero, et al., 2013).

The analysis presented is based on two aspects that are described in the following subsections. (I) Presents the first perspective of analysis, comparing the RE techniques used by each approach at each RE stage and how they treat requirements. (II) presents an analysis of the goal models used by each approach and the focus on the RE stages involved.

**Techniques and activities covered**

To make this comparison, the techniques for requirement elicitation, specification and validation are presented in Table 2. For each approach, the techniques described or set explicitly for the different GO approaches are indicated.

By analyzing the results, and keeping in mind that the proposals are chronologically ordered, it is interesting to observe the evolution of requirements in the DW environment. The first proposals were more focused on obtaining the requirements by means of interviews and checklists. If we observe Table 2, we can see that the more recent proposals remark on the necessity of obtaining the requirements by means of prototypes and ethnography, to solve the aforementioned problem.

### Table 2 – Standard RE techniques used by various GO approaches

<table>
<thead>
<tr>
<th>Elicitation</th>
<th>Bo01</th>
<th>SP03</th>
<th>GS06</th>
<th>Ma07</th>
<th>Gi08</th>
<th>PG08</th>
<th>DT12</th>
<th>AC13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviews</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Brainstorming</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Questionnaires and checklist</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Prototypes</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ethnography</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specification</th>
<th>Bo01</th>
<th>SP03</th>
<th>GS06</th>
<th>Ma07</th>
<th>Gi08</th>
<th>PG08</th>
<th>DT12</th>
<th>AC13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural Language</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Templates or patterns</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Formal languages</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Validation</th>
<th>Bo01</th>
<th>SP03</th>
<th>GS06</th>
<th>Ma07</th>
<th>Gi08</th>
<th>PG08</th>
<th>DT12</th>
<th>AC13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviews or walk-throughs</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Traceability matrices</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Expert opinion</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Model validation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

By analyzing the results, and keeping in mind that the proposals are chronologically ordered, it is interesting to observe the evolution of requirements in the DW environment. The first proposals were more focused on obtaining the requirements by means of interviews and checklists. If we observe Table 2, we can see that the more recent proposals remark on the necessity of obtaining the requirements by means of prototypes and ethnography, to solve the aforementioned problem.

### Table 3 – Most common GORE techniques used by various GO approaches

<table>
<thead>
<tr>
<th>Elicitation</th>
<th>Bo01</th>
<th>SP03</th>
<th>GS06</th>
<th>Ma07</th>
<th>Gi08</th>
<th>PG08</th>
<th>DT12</th>
<th>AC13</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFR</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MAP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tropos</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>GDI</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Conclusions

In order to carry out the RE phase in the DW field, there are a number of approaches that are capable of obtaining the requirements in a systematic way, using the goal modeling for either the elicitation, specification or validation of these requirements. This study showed an overview and a comparative study of requirements treatment in eight GO approaches to DW because there is still no established standard to perform this phase.

The analysis presented is based on two aspects:

1. The first analyzes the RE techniques used by each approach in the stages of elicitation, specification and validation. It was noted that the most commonly used techniques are interviews (100%), formal languages (63%), checklists and templates (50%).

2. The second analyzes the use of goal models at each RE stage. At this point, we found that most of the approaches discussed use the goal model to elicit and specify requirements. Approaches are used to produce logical multidimensional schemas, but with time, most of them generate conceptual schemas. One reason for this situation could be that Kimball (Kimball, 1996) introduced multi-
dimensional modeling at a logical level as a specific relational implementation. Over the course of time, it has been argued that it is necessary to generate schemas at a platform independent level and that in fact (Mazón, et al., 2007), the multidimensional design should span the three abstraction levels (conceptual, logical and physical), similar to the method used in the relational databases field (Romero, & Abelló, 2009).

As seen in Table 3, different RE tasks, require reasoning about different types of goals. In particular, during requirements elicitation, one needs to reason about the current organizational goals and how these are realized in existing system components. In addition, during requirements elicitation, we need to understand the motivation for changing the current situation (i.e., we need to model the change goals). In contrast, in requirements specification the focus is on future business goals and how these can be operationalized into system components. Finally, during the validation of system requirements, the focus is on the stakeholders’ evaluation goals and how the derived specification conforms to these goals. Therefore, we can differentiate between four types of goals at the RE level namely: current goals, change goals, future goals and evaluation goals.

Thanks to this comparative study, we can develop an improved DW development methodology, with a better understanding of the RE stage, which includes the modelling of the business strategy and aligns each actor goals with it, providing a more complete and consistent view of the system.

Acknowledgements

This work was financed by the Universidad de La Frontera. DIU-FRO Project DI13-0047.

References


Inman, W. (1996), Building the data warehouse (2nd ed.).


Prakash, N., & Gosain, A. (2008). An approach to engineering the requirements of data warehouses (pp. 49-72).


