Domain ontology for analytical requirements elicitation

Fahmi Bargui, Hanene Ben-Abdallah, Jamel Feki

Mir@el Laboratory
FSEG, University of Sfax Tunisia, Po Box 1088
{fahmi.bargui,hanene.benabdallah,jamel.feki}@fsegs.rnu.tn

Abstract. Despite their potential in analytical requirements elicitation, goal-oriented approaches in Data warehouse (DW) projects have not been well exploited. The main difficulties in their adoption is the lack of assistance in goal elicitation and the lack of support in generating suitable information for decision-making, from the defined goals. To address these limitations, in this paper we introduce a decision-making ontology that formalizes the semantic relationships among the decision makers’ goals and explicitly relates the decision-making knowledge to the goals. In addition, this ontology is enriched with a set of queries and inference rules that provide for an automatic generation of suitable analytical requirements from the goals fulfilling any business process that the decision maker cares to analyze.

1 Introduction

A data warehouse (DW) is a special type of database dedicated for decision-making support. It organizes data into facts and dimensions based on Multidimensional (MD) modeling. Goal-oriented MD modeling (e.g. [1], [2]) gained an increasing interest in eliciting analytical requirements in DW projects thanks to their graphical languages and tool support. However, their widespread usage remains hindered by their lack of assistance in goal elicitation conducted through the decomposition of high level goals into concrete sub-goals. This goal decomposition requires domain knowledge mastered only by domain experts. Furthermore, our literature review highlighted that goal-oriented approaches provide little or no assistance in identifying appropriate performance indicators to measure the fulfillment degree of the elicited goals. The same shortage is also present in identifying information (i.e. data to be stored in the DW) that could be analyzed when the elicited goals are not met. In most cases, such information is informally collected and without an explicit link with the goals. Consequently, when decision maker’s goals change, it is difficult to trace the parts of the MD model that should be modified.
In this paper, we present a decision-making ontology that provides for the automation of analytical requirements elicitation. The next section, first, describes the ontology and illustrates it for the commercial domain. Section 3 shows how the ontology allows the analyst to overcome the lack of domain knowledge required during requirements elicitation. Finally, Section 4 synthesizes our proposal and presents our future work.

2 The decision-making ontology

2.1 Thesaurus part

Through an interview with decision makers at different hierarchical levels and belonging to different domains, we identified the terminology used in the decision-making process. This terminology includes nine concepts (Figure 1).

![Decision-making Ontology metamodel in Protégé (thesaurus part)](image)

A DecisionMaker is a person in the enterprise, who evaluates and controls the performance of a BusinessProcess composed of a set of related activities that produce a specific service/product for a particular customer. Each business process must fulfill a set of Goals during a period of time. A Goal is quantitative and calculated through a Formula to evaluate its achievement degree by comparing it to an estimated Target Value. The achievement degree of a goal is given by an Indicator. This later can be analyzed by the decision maker through a set of AnalysisAxes by aggregating (SUM, MAX, AVG…) the various values produced of the indicator over time.

In addition, an analysis axis is composed of several AnalysisLevels each of which represents a granularity echelon to aggregate the indicator values. An analysis level may have additional descriptions (AnalysisAttribute) and multiple analysis levels can be organized into an AnalysisHierarchy.
Furthermore, the above decision making concepts can be semantically related through the following nine relationships (Figure 1):

- **Control** \((D, P)\): The decision maker \(D\) controls the performance of the business process \(P\).
- **Fulfill** \((P, G)\): The business process \(P\) is defined to fulfill the goal \(G\).
- **Is Measured By** \((G, I)\): The fulfillment degree of the goal \(G\) is measured by the indicator \(I\).
- **Is Calculated Through** \((I, F)\): The value produced by the indicator \(I\) is calculated through the formula \(F\).
- **Is Analyzed Through** \((I, A)\): Indicator \(I\) is analyzed through the analysis axis \(A\).
- **Is Composed Of** \((A, H)\): The analysis axis \(A\) is composed of a hierarchy \(H\).
- **Has Level** \((H, L)\): Hierarchy \(H\) has level \(L\).
- **Is Described By** \((L, At)\): The level \(L\) is described by the analysis attribute \(At\).
- **Require** \((G, G')\): Fulfillment of the goal \(G\) requires the achievement of \(G'\).

Figure 2 illustrates an instance of our ontology meta-model for the commercial domain. In this instance, the concept SalesManager is defined to control the performance of the three business processes: Order, AuctionOrder and Delivery; each of which must fulfill some goals. For instance, the process Order is defined to fulfill the goal IncreaseSales. The realization of this later requires the achievement of three goals: IncreaseCustomers, ProductAvailable and IncreaseShops. Note that relationships among goals are represented in the ontology by means of predicates. For example, the realization of the goal IncreaseSales requires the availability of the products in the stock, and increasing the number of either customers or shops. This knowledge is represented in the ontology through the predicate: \((\text{Require}(\text{IncreaseSales}, \text{IncreaseCustomers}) \lor \text{Require}(\text{IncreaseSales}, \text{IncreaseShops})) \land \text{Require}(\text{IncreaseSales}, \text{ProductAvailable}))\).

![An extract of the commercial domain ontology (the thesaurus part)](image-url)
Figure 3, shows various levels organized by analysis hierarchies for the analysis axis Product. For example, for the hierarchy family, the indicator Turnover could be analyzed according to the level Category of a product. This gives the decision maker detailed values of the Turnover by category of product. The comparison of each realized value with an estimated target value allows the decision maker to judge what category of product is not sold as expected.

![An extract of the analysis axis concept (product)](image)

**Fig.3.** An extract of the analysis axis concept (product)

### 2.2 Inference rules and queries

The inference rules are defined by the domain experts and are identified based on the semantic relationships among the domain concepts. The inference rule’s premises express constraints on concepts and restrictions on their properties’ values and cardinalities. These premises are expressions considered as always true by the experts in the domain; that is, they represent knowledge (stored in the ontology) that cannot be contested. When an inference rule is applied, the reasoning engine matches the premises’ formulas with the knowledge stored in the ontology to infer a set of new knowledge, as conclusions, not explicitly stored. The deduced information is either retained by the user or can be part of the premises of other rules.

In our context, we have defined a set of inference rules and queries in order to assist the elicitation of the analytical requirement elements and to complete the filling of our natural language based analytical requirements template [3]. The components of this template are determined through an empirical study covering samples of decision-making processes (cf. [6]). Figure 4 shows a sample of eight queries and two inference rules. They are formalized respectively in the SQWRL and SWRL languages.

---

1 [http://www.w3.org/Submission/SWRL/](http://www.w3.org/Submission/SWRL/) and [http://protege.cim3.net/cgi-bin/wiki.pl?SQWRL](http://protege.cim3.net/cgi-bin/wiki.pl?SQWRL)
3 Analytical requirements elicitation

Our requirement elicitation process uses a set of queries and inference rules to extract information from the decision-making ontology (Figure 4). It is conducted through the following four steps by interaction with the decision maker and the requirements analyst:

- Business process identification: Query 1 (Figure 4) extracts from the ontology all business processes controlled by the given Decision Maker; this latter selects those business processes of interest.
- Goal identification: Query 2 extracts from the ontology all Goals that must be fulfilled by a business process selected by the Decision Maker. In addition, for complex Goals (i.e. might be difficult to fulfill at first); Rule 1 automatically finds any sub-goal indirectly fulfilled by the selected business process. Finally, to ensure the completeness of the set of goals, Rule 2 automatically finds all goals that are required by the found goals.
- Indicator retrieval: for each found goal, Query 3 extracts from the ontology the Indicator that measures it and Query 4 extracts its Formula.
- Analytical query generation: for each Indicator, Query 5 through Query 8 extract from the ontology, respectively: its analysis axis, their hierarchies, their analysis levels and analysis attributes. The identified elements are then combined to generate a query in natural language according to our query format [5]: 

\[
\text{<Analyze> + <the> <Indicator> + <by> + <Analysis axis> + <'s> + <Analysis level> + [and] + <Analysis Attribute>}. 
\]
Note that the stakeholders (i.e. decision maker and requirements analyst) are involved in each elicitation step in order to select from/confirm the found results.

4 Conclusion and future work

This paper presented an ontology for the decision-making domain. This latter can be used by a requirement analyst to identify automatically all: 1) pertinent business processes to analyze, 2) goals and sub-goals that must be fulfilled by the business processes, and 3) the indicators and their associated formulas measuring the achievement degree of the goals. This automatically identified information is then used to formulate automatically a set of analytical queries in natural language. The formulated analytical queries can be used by the decision maker to carry out different analyses when his/her goals are not met. In addition, the ontology helps the analyst to acquire domain knowledge necessary during requirements elicitation. Indeed, our method performs automatic decomposition of the decision makers’ goals by referring to the semantic relationships and the inference rules stored in the ontology.

We are currently working on finalizing the development of a toolset supporting our analytical requirements elicitation approach which is based on the presented decision-making ontology. In addition, we plan to examine how the evolution of the decision-making ontology can be handled.

References