A FRAMEWORK FOR KNOWLEDGE-BASED CONFIGURATION OF PROJECT MANAGEMENT INFORMATION SYSTEMS

Solvita Berzisa, Janis Grabis

Riga Technical University, Faculty of Computer Science and Information Technology, 1 Kalku, Riga, Latvia, berzisa@gmail.com, grabis@iti.rtu.lv

Abstract. Project management is a complex process regulated by project management methodologies, standards and other requirements. Project management information systems are used to support project management activities. To ensure effective use of the project management information system, it is necessary that the system could be configured according to requirements of the chosen methodology. The configuration provides a structural framework for project management while accumulated project management knowledge provides basis for efficient project execution. Therefore, the project management information systems also should provide means for representing and reusing knowledge. The objective of this paper is to propose architecture of knowledge-based system that ensures project management information system configuration process with appropriate knowledge. This architecture is referred as the knowledge-based configuration system. Project management knowledge repository and case-based reasoning principles are used in development of the proposed architecture. This architecture allows storing project management and configuration information and organizing and reusing this information.

Keywords: PMIS configuration, knowledge-based system, case-based reasoning.

1 Introduction

Establishing project management (PM) procedures and setting up a PM environment are among the main tasks of project initialization [1][2]. PM methodologies are used to support this activity. However, the methodologies are fine-tuned for every particular project to better account for specific features. PM knowledge can be used to assist the fine-tuning. The objective of this paper is to develop an approach for using knowledge in configuring the PM environment. Given that the PM environment usually is implemented using the project management information system (PMIS), the knowledge utilization is proposed a part of the overall configuration of PMIS. The configuration defines data structures and processes supported by PMIS. Consequently, knowledge accumulated during previously completed projects can be used to identify what data structures and processes should be considered in the configuration for the current project.

Case-based reasoning (CBR) principles are used for development of the PM knowledge-based configuration system (KBCS). This architecture allows to store theoretical and practical knowledge in the repository and reuse it in PMIS configuration for similar projects. All used description of the PMIS configuration has been added to knowledge repository as a new case. The paper describes the PMIS configuration approach and the PM KBCS architecture.

The contribution of this research is development of the PM KBCS architecture. To our knowledge, knowledge processing has been applied to addressing individual project management activities while there are no applications in setting-up the project management environment and in configuration of PMIS in particular. Some PM knowledge management solutions have been developed for software planning [3][4][5] and size estimation [6], and constructing schedule development [7][8]. All above mentioned PM knowledge management solutions have been provided for some PM tasks and commonly have been associated with certain areas. The proposed KBCS architecture allows PM information and PMIS configuration information and knowledge reuse in preparing of description of the PMIS configuration. PM KBCS can be used for projects in different areas assuming that an appropriate methodology is available. The proposed architecture with some modifications also could be used for knowledge representation in PMIS.

This paper is divided in six sections. Section 2 is a description of the PMIS configuration approach. Main elements and technologies of the knowledge-based system are described in Section 3. This section also provides a brief comparison of the proposed system with some of existing systems. Structure of PM KBCS architecture is explained in Section 4. An example of knowledge use in PMIS configuration is described in Section 5. Section 6 concludes and discusses future work.

2 PMIS Configuration Approach

It is assumed that a project is performed using a project management methodology and PMIS is used to support project management activities. Before its productive usage, PMIS must be configured according to requirements set by the project management methodology used. The purpose of the PMIS configuration approach elaborated in [9] and shown in Figure 1 is to provide a systematic solution for transforming the informally specified requirements $R_j$ into an executable PMIS configuration $I^j$, where $j$ indicates the project and
$k$ refers to the particular PMIS. This transformation process consists of two steps. The first step is transformation of $R_j$ into the standardized form $C_j$, where $C_j$ is structured in a format defined by the template $S$:

$$C_j = T_1(R_j,S) \quad (1)$$

A configuration client is used to support this manual transformation. The second step is transformation of $C_j$ into an executable configuration of PMIS $I_{jk}$:

$$I_{jk} = T_2^k(C_j) \quad (2)$$

One configuration file $C_j$ can be transformed to the various PMIS configuration $I_{jk}$ by using different transformations $T_2^k$. The same transformation can be applied to different $C_j$ assuming the same target PMIS.

The template $S$ is defined using the XML schema for Configuration of Project Management Information System (XCPM) [10]. XCPM allows describing PM data, processes and knowledge. The schema is developed on the basis of the PM concept model that has been obtained after conceptual modelling of the project management domain [11]. The concept model describes the PM domain concepts and it relations. The standard template also includes abstract elements available for representing project features not supported by default. Given the schema provides a comprehensive definition of the project management domain, it can be used to represent various PM methodologies and standards. As the result of transformation $T_1$, the schema is populated with data characterizing particular project and $C_j$ is represented as an XML file.

The configuration process is supported by appropriate knowledge from the PM knowledge repository. The PM knowledge repository includes information about PM methodologies, previously used PMIS configurations $C_j$, $j=1,…,m$, and other relevant information. Two types of knowledge are required during the configuration process: configuration and operational. The configuration knowledge is necessary for $T_1$ and is used in configuration client. Operational knowledge is used for $T_2^k$ and during usage of PMIS. It includes diagrams, templates, documents, checklists and other.

### Figure 1. PMIS configuration approach

#### 3 Knowledge-based Systems

A knowledge-based system (KBS) is the computer system that contains stored knowledge and solves problems like humans would. KBS has been an active research theme for the last twenty years. Consequently, there are a lot of different KBS developed for different fields: medicine, transport, customer service management, logistic, enterprise, finance etc. Knowledge-based configuration is one of the most successful application areas of KBS hence several approaches have been developed to tackle configuration tasks [12]. It is used in a number of domains, such as, telecommunication, engineering, process control, software configuration management, computer design, electronics, chemical design and others.

All KBS systems consist of four components: a knowledge base, an inference engine, a knowledge engineering tool, and a specific user interface [13]. The knowledge base stores knowledge and can be implemented as a repository, a relational database or a case library. Knowledge modelling and engineering activities and tools deal with development of the knowledge base. The inference engine is mostly one of reasoning technologies: rule-based, case-based and hybrid.

Rule-based reasoning uses ‘if-then-else’ rule statements [13]. Knowledge is represented in rules. New problem is solved by finding appropriate rule in the knowledge database. Rule-based reasoning has been used in some configuration systems [12] and expert systems such as product planning, teaching, system development, knowledge representation and other [13]. Only one of reviewed PM KBS [6] used rule-based reasoning together with data clustering.

The most frequently considered knowledge processing method is CBR that uses the past experience [14]. CBR is used in KBS for project management [3][4][5][7] and also in the proposed PM KBCS architecture. Knowledge is represented with the cases. A new problem is solved by finding the similar past cases and re-using in the new problem situation. CBR consists of the case library and four-step process: retrieve, reuse, revise and retain [14]. During the retrieve step, the most similar cases are founded in the case library according the new
case description. Information and knowledge from similar cases are copied or adapted to solve the new problem in the reuse step. Proposed problem solution is tested, evaluated and repaired during the revise step. During the retain step, the new case and knowledge is added to the case library for future reuse. Following technologies is used to ensure CBR implementation: data warehouse and online analytical processing (OLAP) [15] [16] [17]; data mining [18][19]. CBR principles have been used in Google translator [20]. Some KBS already use both rule-based reasoning and CBR [21] [22] [23].

Proposed PM KBCS is similar with the recommender system for software project planning [3] and the knowledge-based logistics strategy system [15]. The recommender system for software projects planning uses project attributes and weight to find similar cases and ensures statistical information about cases similarity and performs multiple cases analysis. The proposed PM KBCS also uses the set of project attributes and similarity weights for the similar case search but weight values are binary. The set of project attributes is variable and similarity weights depend of the knowledge search area. PM KBCS ensures list of the knowledge search area related knowledge with statistical information about frequency. In the knowledge-based logistics strategy system data warehouse and OLAP is used for CBR implementation. Similar cases are searched in the data warehouse according to defined parameters and full case description is in the repository. This KBS ensures numeric data (logistic strategy) and symbolic data (logistic workflow) processing. In the proposed PM KBCS case attributes and classification are stored only in a relational database structure, but also full case description is in the repository. PM KBCS also ensures structured and unstructured knowledge store.

Also in KBS is used: ontology, a neuron network, a data mining and other artificial intelligence methods. Ontology is a shared description of the concepts and relation in domain knowledge. Ontology in KBS is used as storage system [24] and knowledge representation method [25] [26]. Data mining is a process of extracting patterns from data and is an important tool to transform data into information. Data mining techniques is used CBR case clustering/classification [18] and for knowledge mining [27].

### 4 Architecture of the PM Knowledge-Based Configuration System

In order to support knowledge utilization in the PMIS configuration, PM KBCS is elaborated. Principles of case-based reasoning are used in design of PM KBCS. The architecture of PM KBCS is shown in Figure 2. Cases in PM KBCS are a) empirical knowledge or configuration files $C_j$, $j=1,..., m$, that previously have been used for configuration of PMIS and b) PM methodologies, standards and best practices $H_i$, $i=1,..., p$.

The CBR process is managed using the configuration client, which has three main modules. The first module is used to describe a new case using set of the project attributes $A_j$, where $j+1$ is used to identify the new case. The retrieve step is performed by the information retrieval module that find the similar cases according to the information search knowledge area $M_k$ (e.g., risk management) and the defined similarities $L_k$, $k=1,..., n$. Sets of cases similar to the new case $j+1$ and search knowledge area $s$ are denoted by $H_{j+1,s}$ and $C_{j+1,s}$, for theoretical and empirical knowledge, respectively, is the result of information retrieval. The information processing and display module performs CBR reuse and revise steps. This module collects and processes gathered information and displays it to the user. CBR retain or learning is performed after the new configuration file $C_{j+1}$ has been used for configuration on PMIS (Figure 1, transformation $T^2_k$). $C_{j+1}$ is added to the PM knowledge repository as a new case.

---

**Figure 2. PM KBCS architecture**
4.1 PM Knowledge Repository

The PM knowledge repository is conditionally divided into two parts: the library and the case register (Figure 2). In the library, knowledge is stored in three ways: theoretical (\(H = \{H_i\}_{i=1,\ldots, p}\)), empirical (\(C = \{C_j\}_{j=1,\ldots, m}\)) and unstructured (\(D = \{D_k\}_{k=1,\ldots, s}\)). Unstructured knowledge is in various formats e.g. diagrams, forms templates or documents, but this information in some ways is related with data stored in \(H\) or \(C\).

The case register collects information that is needed for knowledge organization, search and retrieval. It stores information about the case descriptions (\(P^H = \{P^H_i\}_{i=1,\ldots, x}\) and \(P^C = \{P^C_i\}_{i=1,\ldots, y}\)), the project attributes \(A_i\), its values and the case similarity (\(L = \{L_k\}_{k=1,\ldots, z}\)).

Knowledge classification is ensured by the set of project attribute \(A_i= (a_1, \ldots, a_n)\). Each attribute \(a_i\) also has defined values. Attribute default values, mandatory and multi-choice are stored with the attribute configuration. XCPM schema \[10\] defines the set of typical attributes represented using the EnvironmentFactors element (Figure 3). This element also allows storing of the additional attributes with the OtherFactor element.

Each case description \(P^H_i\) or \(P^C_i\) includes a set of case describing attributes \(A_i\) and the case identifiers \(H_i\) or \(C_i\). The case description \(P^H_i= (A_i, H_i)\) defines the theoretical knowledge case \(H_i\), but \(P^C_i= (A_i, C_i)\) similarly describes the empirical cases \(C_i\). In situation when in the case description contains attributes with multi-choice values, the case has more than one case description according to a number of possible combinations of attribute values.

\[
\text{Area} \quad \text{Type} \quad \text{Methodology} \\
\text{Standards} \quad \text{ProjectBudget} \quad \text{TeamSize} \\
\text{PMProcessModel} \quad \text{ClientType} \quad \text{Rules} \\
\text{ProjectProduct} \quad \text{ProjectAction} \quad \text{OtherFactor}
\]

Figure 3. XCPM schema fragment: EnvironmentFactors

The case similarity \(L_k = (M_k, B_k, X_k)\) depends of the knowledge search area \(M_k\) and the knowledge type \(B_k, k= 1, \ldots, n\). The knowledge search area specifies what PM knowledge is searched, e.g., risk register, quality criteria, change request, documents, etc. The knowledge type defines what case search \(L_k\) is provided: theoretical \((H)\) or empirical \((C)\). Importance of each attribute in the search of similar cases is defined by the set of the attribute similarity measurement \(X_k=(x_1, \ldots, x_n)\). The count of element for the similarity measurement \(X_k\) is equal with the count of the project attributes. The value of \(x_i\) can be 0 or 1. If \(x_i=0\) then cases need to compare according to the \(i\)th attribute \(a_i\) during the similar case search, but if \(x_i=1\) then attribute \(a_i\) is ignored.

4.2 Information Retrieval

The information retrieval module ensures retrieving similar case identifiers from the case register. Input data is the set of the project attributes \(A_i\) and the search knowledge area \(M_i, s = 1, \ldots, v\). Output data are retrieved sets of the cases identifiers \(H_{i+1,s}\) and \(C_{i+1,s}\). The information retrieval process consists of two steps:

1. Similarity clarification. Appropriate similarity is searched in the set of similarities \(L\) according to the search knowledge area \(M_i\). The result is a subset of the similarities:

\[
L_{i+1,j} = \{L_k \mid L_k \in L \text { and } M_k = M_i \text { and } k = 1, \ldots, n\}
\]  \hspace{1cm} (3)

2. Case search. Using similarities from the subset \(L_{i+1,j}\), similar cases are searched in \(P^H\) and \(P^C\). The search in each of case sets is performed separately and search attribute similarity already can different. With type \(B_k\) in \(L_k\) is defined the case set with similarity can be used. Consequently, the case search is performed in two stages:

2.1. Search in set in the theoretical case \(P^H\). For each \(L_k\) where \(L_k \in L_{i+1,j}\) and \(B_k = H\). The set of case identifier \(H_{i+1,s}= (H_k|H_k \in P^H_i)\) is retrieved, which attribute values \(a_i\) are equal to new case attributes values \(a_{i+1}\) of those attributes that similarity measurement \(x_i=1\):
\[ \sum_{i=1}^{n}(x_i | x_i \in X_k \text{ and } X_k \in L_k) = \sum_{i=1}^{n}(x_i | x_i \in X_k \text{ and } X_k \in L_k) \times (1 \text{ if } \exists((a_{i\alpha}|a_{i\alpha} \in A_x \text{ and } A_x \in P_x^L) = (a_{ij+1}|a_{ij+1} \in A_{j+1})) \] (4)

2.2. Search in set in the empirical case \( P^E \). For each \( L_k \) where \( L_k \in L_{j+1,s} \) and \( B_k = C \). The set of case identifier \( C_{j+1,s} = \{C_j | C_j \in P_x^L \} \) is retrieved, which attribute values \( a_{j+1} \) are equal to new case attributes values \( a_{i\alpha} \) of those attributes that similarity measurement \( x_i = 1 \):

\[ \sum_{i=1}^{n}(x_i | x_i \in X_k \text{ and } X_k \in L_k) = \sum_{i=1}^{n}(x_i | x_i \in X_k \text{ and } X_k \in L_k) \times (1 \text{ if } \exists((a_{i\alpha}|a_{i\alpha} \in A_x \text{ and } A_x \in P_x^L) = (a_{ij+1}|a_{ij+1} \in A_{j+1})) \] (5)

4.3 Information Processing and Display

The information processing and display module processes the similar cases found and provides recommendation for configuration of PMIS. For example, if a user configures the risk list, then module shows data elements used in similar projects or chosen methodologies. During information processing all similar cases from the sets of case identifier \( H_{j+1,s} \) and \( C_{j+1,s} \) are found in the library and information about the search knowledge area \( M_t \) is extracted for further processing. To ensure that the user is not burdened with too much information and important information is shown at first, following activities are carried out during information processing process:

1. Grouping of information in order to ensure that every information element is displayed only once. Information element value analysis is not undertaken only grouping by the full-text.
2. Collection of statistic about the each information element. Statistic includes frequency estimates expressed in % between similar cases. This statistic is also supplemented with information element associated methodologies/standards listing.
3. Information element ranking in descending order to ensure that statistically most frequent used information is displayed first.

During describing PMIS configuration \( C_{j+1,s} \), knowledge is shown in a form of suggestions and users either use of ignore these suggestions.

5 PM KBCS Application Example

An example is used to demonstrate the knowledge assisted configuration process. This example shows configuration of the risk register as a part of the PM environment setup. The risk register is used to name all risk factors relevant to the project, and it is implemented as a list containing multiple data elements. The knowledge assisted configuration helps to determine what data elements should define in the risk register.

The risk register is prepared for the following project: IT company realized outsourcing maintenance project for government software. The project team includes 7 specialists and year budget is EUR 80000. The project is organized in an iterative manner and works according the PMBOK methodology and complies with ISO quality standards.

The first step of the knowledge assisted configuration process is definition of the project attributes \( A_z \) and their values (Figure 4).

![Figure 4. Definition of project attributes \( A_z \)](image-url)
The case similarity is clarified from the set of similarity \( L \) (Figure 5). In the example, the search knowledge area \( M \) is the risk register. The search result is the subset that includes two similarities: \( L_5 \) and \( L_6 \). These similarities further are used for case search. Similarity \( L_5 \) is used for search in the set \( P^H \) (\( B = H \)), and \( L_6 \) is used for search in the set \( P^C \) (\( B = C \)). Cases in the set \( P^H \) (Figure 6) are searched after methodology (\( \alpha_1 \)) because only the similarity measurement \( x_1 \) is equal to 1 for the similarity \( L_5 \) (Figure 5). Methodology ‘PMBOK’ is mentioned only in the case description \( P_9^H \) (\( H_9 = 1256 \)). Cases in the set \( P^C \) (Figure 7) are searched after 6 attributes: methodology, area, type, client type, project type and action. Two cases form the set \( P^C \) are the same abovementioned attributes values as new project: \( P_2^C \) (\( C_2 = A152 \)) and \( P_6^C \) (\( C_6 = A962 \)). Three cases are found as the result of the information retrieving process.

<table>
<thead>
<tr>
<th>( L )</th>
<th>Knowledge Area (( M ))</th>
<th>Methodology (( \alpha_1 ))</th>
<th>Standard (( \alpha_3 ))</th>
<th>Area (( \alpha_2 ))</th>
<th>Type (( \alpha_4 ))</th>
<th>Project (( \alpha_5 ))</th>
<th>Client (( \alpha_6 ))</th>
<th>Rules (( \alpha_7 ))</th>
<th>Product (( \alpha_8 ))</th>
<th>Action (( \alpha_{10} ))</th>
<th>Type (( \beta_2 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L_5 )</td>
<td>Risk register</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( L_6 )</td>
<td>Risk register</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Figure 5. Set of the similarity \( L \)**

<table>
<thead>
<tr>
<th>( P^H )</th>
<th>Methodology (( \alpha_1 ))</th>
<th>Standard (( \alpha_3 ))</th>
<th>Area (( \alpha_2 ))</th>
<th>Type (( \alpha_4 ))</th>
<th>Project (( \alpha_5 ))</th>
<th>Client (( \alpha_6 ))</th>
<th>Rules (( \alpha_7 ))</th>
<th>Product (( \alpha_8 ))</th>
<th>Action (( \alpha_{10} ))</th>
<th>Case ID (( \beta_1 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_2^H )</td>
<td>PMBOK</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1256</td>
</tr>
</tbody>
</table>

**Figure 6. Set of the theoretical knowledge case descriptions \( P^H \)**

<table>
<thead>
<tr>
<th>( P^C )</th>
<th>Methodology (( \alpha_1 ))</th>
<th>Standard (( \alpha_3 ))</th>
<th>Area (( \alpha_2 ))</th>
<th>Type (( \alpha_4 ))</th>
<th>Project (( \alpha_5 ))</th>
<th>Client (( \alpha_6 ))</th>
<th>Rules (( \alpha_7 ))</th>
<th>Product (( \alpha_8 ))</th>
<th>Action (( \alpha_{10} ))</th>
<th>Case ID (( \beta_1 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_2^C )</td>
<td>PMBOK</td>
<td>ISO</td>
<td>IT</td>
<td>Outsourcing</td>
<td>120000</td>
<td>10</td>
<td>Iterative</td>
<td>government</td>
<td>-</td>
<td>software</td>
</tr>
<tr>
<td>( P_6^C )</td>
<td>PMBOK</td>
<td>-</td>
<td>IT</td>
<td>Outsourcing</td>
<td>50000</td>
<td>5</td>
<td>Agile</td>
<td>government</td>
<td>-</td>
<td>software</td>
</tr>
</tbody>
</table>

**Figure 7. Set of the configuration knowledge case descriptions \( P^C \)**

During information processing, the risk register configuration information is extracted from the full case description (Figure 8). This configuration information includes information about the risk register data elements. Then all found risk register data elements are grouped and ordered. The result for a user is shown as in Figure 9. The user sees the list of data elements (i.e., data columns) that is ordered descending, statistic about frequency and methodology if this column is already mentioned in it. The column configuration information also can be analyzed and shown. Knowledge processing and display result (Figure 9) shows that such columns as ID, description, condition and owner are mentioned in all cases found in the knowledge repository.

If the other information, e.g., change requests, is prepared for the described project than the sets of similar cases (\( H_{k+1,s}^H \) and \( C_{k+1,s}^C \)) can be changed because a different knowledge search area (\( M\)) is considered and subsequently other similarity measurements (\( X_i \)) are used.

<table>
<thead>
<tr>
<th>Case: 1256</th>
<th>Case: A152</th>
<th>Case: A962</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>Description</td>
<td>Condition</td>
</tr>
<tr>
<td>Category</td>
<td>Priority</td>
<td>Response action</td>
</tr>
<tr>
<td>Probability</td>
<td>Affect</td>
<td>Owner</td>
</tr>
<tr>
<td>Cost</td>
<td>Time</td>
<td>Responsible</td>
</tr>
<tr>
<td>Affect</td>
<td>Owner</td>
<td>Response strategy</td>
</tr>
<tr>
<td>Owner</td>
<td>Probability</td>
<td>Process</td>
</tr>
</tbody>
</table>

**Figure 8. Risk register column lists of retrieved case**
6 Conclusion and Future Work

Having improvement of efficiency of PMIS configuration and implementation as an overall goal, the architecture of knowledge-based configuration system for PMIS has been described in this paper. This architecture is a part of the PMIS configuration process and provides users with appropriate configuration and PM knowledge. The KBCS architecture consists of two parts: the configuration client that provides case-based reasoning for knowledge search and processing, and the PM knowledge repository that include the case register and the library. The project attributes and environment factors have been used for the case classification.

During implementation and configuration of PMIS, CBS generates configuration suggestions on the basis of historical data and theoretical knowledge. That helps process designers and project managers to get configuration guidance. In the example explored, the user can see data elements frequently used in setting-up the risk register. Configuration of PMIS and accumulation and use of historical data are greatly aided by the relatively well-defined scope of project management. The XCPM schema is used to standardize definition of the project management domain.

The paper explores only data similarity. However, processes play important part in PMIS. Using knowledge about processes and analyzing process similarity are main future research directions.

7 Acknowledgement

This work has been supported by the European Social Fund within the project „Support for the implementation of doctoral studies at Riga Technical University“.

References


