Case Study on Requirements Engineering in Information Mining Project: Metallurgical Enterprise

Pollo-Cattaneo, M.F.  

Pesado, P.  
School of Computer Science. National University of La Plata. Buenos Aires, Argentina. ppesado@lidi.info. unlp.edu.ar

Britos, P.  
Information Mining Research Group. National Lab ID + i in Visualization, Computer Graphics and Creative Code. University of Rio Negro at El Bolsón, Río Negro, Argentina. paobritos@gmail.com

García-Martínez, R.  
Information Systems Research Group. National University of Lanús. Buenos Aires, Argentina. rgarcia@unla.edu.ar

ABSTRACT
There can be no debate that Information Mining Projects cause processing tools to arise with the sole purpose of converting available organization data into useful knowledge on account of decision-making. Considering the aforementioned, this type of projects demands due diligence in the requirement specification process, as the latter needs to be thorough and traceable throughout the entire project, and therefore, process associated with requirements engineering are not to be reused in any future projects among the like. Similarly, latest methodologies within the field of Information Mining fail to take practices associated to stakeholders and costumers’ requirements management into account. With this aim in mind, a solution model to the Information Mining Project Management necessities is proposed.

CCS Concepts
• Computing methodologies→Modeling and simulation
  • Information systems→Information systems applications

Keywords
Information Mining, Requirement Management, Process, Methodology, Requirement Engineering.

1. INTRODUCTION
The basis of Information Mining is centered on data processing for nontrivial knowledge collection, an organizational task in which analysis and synthesis tools are indispensable [1]. The aforementioned knowledge being unknown, it can be further exploited by organizations on account of the decision-making process [2]. Information Systems specialists claim data relationships, fluctuations and dependencies to be the core of the process, rather than the data itself [3]; whether these relationships reflect the reality and are then regarded as valid for doing so, the fact that they revolutionize the criteria for decision-making cannot be dismissed [4]. Taking the CRISP-DM [5], P3TQ [6], and SEMMA [7] standout Information Mining Methodologies as examples, it can be seen that these all fail to take into account aspects of those related to both project management and the organizational context where the project is taken on, thus failing to engineer the key concepts in business knowledge [8]. All things considered, this paper employs a proposed model which contributes to the development of a thorough requirements management in the context of an Information Mining project. In order to do so, the detected problem is first described (Section 2), and then a solution proposal is suggested (Section 3). Afterwards, a case study in which the proposed process is implemented is then shown (Section 4), and lastly, conclusions and future work lines are presented (Section 5).

2. PROBLEM DESCRIPTION
What a proper Information Mining project needs is a due - consistent and traceable the project throughout - requirements specification, which should, at first, allow for an orderly project management, and also leave out any possible requirements misunderstandings [9]. Nonetheless, there is one glaring difference between this type of projects and traditional software projects, as an Information Mining Project demands no software product construction, but none other than the transformation of data into knowledge: a mere process. For this reason, it is clear that requirements for this type of projects do not abide by any definitions of restrictions and/or functionalities of those which the software product does have to fulfill in the field of Software Engineering [10].

At the origination of an Information Mining project, its objectives describing customer’s general necessities - what he wants to get as a final result of the project, generally linked to strategic and tactical business goals [11] - must be identified. This is because by applying this field’s algorithms to available data, the latter is transformed into knowledge in order to accomplish all sorts of objectives. What is more, guessing out the organization’s real expectations is key to obtaining the desired final project [12]. So as to attain a clearer overview of the project, all parties taking part in it need to manage the same vocabulary [13]. Once the project’s aims have been identified, an initial survey on available sources of information has to be carried out. Based upon an analysis of the project objective and information sources, the scope of the project can be defined, thereby obtaining a group of particular objectives. These may be achieved with the aid of Information Mining-based algorithmic processes [14]. Thus, the business problem - the mere motive of the entire project - can then be solved with methodologies other than those originated in Software Engineering, in that they overlook practical aspects of the characteristically Information Mining requirements specification [15].

Latest methodologies - CRISP-D, P3TQ and SEMMA - are centered on Knowledge Discovery in Databases Process (KDD) and emphasize available data detection, together with a simplified
The model applied in this paper elaborates on the Process Model for Information Mining projects from [18], which is based on the CRISP-DM methodology [19], considering Small and Medium Enterprises (SME) aspects. Despite this methodology including phase-spread activities of CRISP-DM, especially “Business Comprehension” and “Data Comprehension”, COMPETISOF's details none by only indicating what are the techniques to be applied in any activity [20]. Similarly, Requirements Engineering documenting templates defined in [13] are implemented in this paper.

3. PROPOSED SOLUTION MODEL

A proposed process model is split into five orderly phases: “Project Definition”, “Business Process Engineering”, “Business Conceptualization” and “Information Mining Process Specification”. “Project Definition” phase aims to define the project scope, stakeholders, and objectives to accomplish. The “Business Process Engineering” phase seeks to identify and survey the most relevant business processes in the project. The purpose of “Business Process Engineering” is to locate data repositories where information of the various business processes is stored and to survey their contents. The phase of “Business Conceptualization” attempts to define business in terms of concepts developed and vocabulary managed in order to understand the business jargon, therefore revealing business technical words' meanings coined in the business context. Finally, “Information Mining Process Specification” phase intends to identify Information Mining Processes available for use in solving business processes problems, prior to developing the planification of the remaining project activities. A description of the proposed model phases and activities is displayed at [21].

4. CASE STUDY

In this section, the results obtained applying the proposed process model to a case study, are offered. Such case study uncovers behavior patterns allowing the description of trailers used as resources in the production area of a metallurgical enterprise. These patterns will be taken under consideration for decision making when planning the assembly line of units. Following, the products obtained in the five phases of the process are shown (for clarity generated graphics and application models of each phase they are also represented in [21]).

4.1 Applying the First Phase of Project Definition

The following activities are described: “Identifying Project Objectives” (Figure 1), “Identifying Project Stakeholders” (Figure 2), and “Identifying the Project Scope” (Figure 3).

Activity: Identifying Project Objectives.
Input: Stakeholder meetings information.

<table>
<thead>
<tr>
<th>Process:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Documentation analysis to detect project objectives.</td>
</tr>
<tr>
<td>2. Generation of a functional decomposition tree (figure 4).</td>
</tr>
<tr>
<td>3. Correspondence between function decomposition tree and objective documentation templates (figure 5).</td>
</tr>
</tbody>
</table>

Output: Project and Requirement Objectives Templates (figure 6).

Activity: Identifying Project Stakeholders.
Input: Stakeholder meetings information.

<table>
<thead>
<tr>
<th>Process:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Documentation analysis for project stakeholders detection and registration in template.</td>
</tr>
<tr>
<td>2. Documentation analysis to detect success criteria and project expectations.</td>
</tr>
<tr>
<td>3. Success criteria “AND/OR” graph development (figure 7).</td>
</tr>
<tr>
<td>4. Expectation knowledge map development (figure 8).</td>
</tr>
<tr>
<td>5. and Success Criteria Documentation Template (figure 9) and correspondence between expectations knowledge map and Expectations Documentation Template (figure 10).</td>
</tr>
</tbody>
</table>

Output: Project Stakeholders, Project Success Criteria and Project Expectations Templates (figure 11).

Activity: Identifying the Project Scope.
Input: Stakeholder meetings information.

<table>
<thead>
<tr>
<th>Process:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Documentation analysis to define work boundaries in the project, detection of business problems and generation of the Project Scope Definition Template.</td>
</tr>
<tr>
<td>2. Documentation analysis to define project assumptions and information restrictions.</td>
</tr>
<tr>
<td>3. Development of a Project and Requirement assumptions conceptual map (figure 12) shows an example of conceptual map of project assumptions.</td>
</tr>
<tr>
<td>4. Develop to knowledge map of project restrictions and requirement restrictions (by means of example, figure 13 shows a knowledge map for project restrictions).</td>
</tr>
<tr>
<td>5. Correspondence establishment between the assumptions conceptual map and the assumptions documentation template (figure 14) and between a restrictions knowledge map and the restrictions documentation template (figure 15).</td>
</tr>
<tr>
<td>6. Documentation analysis to define project risks and contingency plans associated with each of these risks and generation of risks and contingency plans documentation templates.</td>
</tr>
</tbody>
</table>

Output: Project Scope Definition Template, Project Restrictions Template, Project Assumptions Template, Requirement Restrictions Template, Requirement Assumptions Template, Project Risks Template, Project Contingency Plans Template, Requirement Risks Template and Requirement Contingency Plans Template (figure 16 shows a project objective template example).

Activity: Identifying Project Stakeholders Activity
Input: Stakeholder meetings information.

<table>
<thead>
<tr>
<th>Process:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Documentation analysis for project stakeholders detection and registration in template.</td>
</tr>
<tr>
<td>2. Documentation analysis to detect success criteria and project expectations.</td>
</tr>
<tr>
<td>3. Success criteria “AND/OR” graph development (figure 7).</td>
</tr>
<tr>
<td>4. Expectation knowledge map development (figure 8).</td>
</tr>
<tr>
<td>5. and Success Criteria Documentation Template (figure 9) and correspondence between expectations knowledge map and Expectations Documentation Template (figure 10).</td>
</tr>
</tbody>
</table>

Output: Project Stakeholders, Project Success Criteria and Project Expectations Templates (figure 11).

Activity: Identifying Project Objectives Activity
Input: Stakeholder meetings information.

<table>
<thead>
<tr>
<th>Process:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Documentation analysis to detect project objectives.</td>
</tr>
<tr>
<td>2. Generation of a functional decomposition tree (figure 4).</td>
</tr>
<tr>
<td>3. Correspondence between function decomposition tree and objective documentation templates (figure 5).</td>
</tr>
</tbody>
</table>

Output: Project and Requirement Objectives Templates (figure 6).

Activity: Identifying Project Scope Activity
Input: Stakeholder meetings information.

<table>
<thead>
<tr>
<th>Process:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Documentation analysis to define work boundaries in the project, detection of business problems and generation of the Project Scope Definition Template.</td>
</tr>
<tr>
<td>2. Documentation analysis to define project assumptions and information restrictions.</td>
</tr>
<tr>
<td>3. Development of a Project and Requirement assumptions conceptual map (figure 12) shows an example of conceptual map of project assumptions.</td>
</tr>
<tr>
<td>4. Develop to knowledge map of project restrictions and requirement restrictions (by means of example, figure 13 shows a knowledge map for project restrictions).</td>
</tr>
<tr>
<td>5. Correspondence establishment between the assumptions conceptual map and the assumptions documentation template (figure 14) and between a restrictions knowledge map and the restrictions documentation template (figure 15).</td>
</tr>
<tr>
<td>6. Documentation analysis to define project risks and contingency plans associated with each of these risks and generation of risks and contingency plans documentation templates.</td>
</tr>
</tbody>
</table>

Output: Project Scope Definition Template, Project Restrictions Template, Project Assumptions Template, Requirement Restrictions Template, Requirement Assumptions Template, Project Risks Template, Project Contingency Plans Template, Requirement Risks Template and Requirement Contingency Plans Template (figure 16 shows a project objective template example).
Figure 5. Correspondence between a Functional Decomposition Tree and Objective Documentation Templates

Figure 6. Output Products of the “Identifying Project Objectives” Activity

Figure 7. Success Criteria “AND/OR” Graphs

Figure 8. Expectation Knowledge Map

Figure 9. Correspondence between Success Criteria AND/OR Graph and Success Criteria Documentation Template

Figure 10. Correspondence between an Expectations Knowledge Map and the Expectations Documentation Template

Figure 11. Output Products of the “Identifying Project Stakeholders” Activity

Figure 12. Conceptual Map of Project Assumptions
4.2 Applying the Second Phase of Business Process Engineering

The following activities are described: “Identifying Business Processes” (figure 17) and “Surveying Business Processes” (figure 18).

Activity: Identifying Business Processes.

Input: Project Scope Definition Template, Project Objective Template, Requirement Objective Template, Project Success Criteria Template, and Project Expectations Template.

Process:
1. Documentation analysis to define business activities related to the project objectives and generation of the business processes diagram template.
2. Generation of the column of business process, associated with each stakeholder in the project stakeholder template (defined in the first phase of the process).

Output: Business Process Diagram Template (figure 19).
Activity: Surveying Business Processes.
Input: Business Process Diagram Template.

Process:
1. Conduction of interviews with stakeholders to understand each business process and to generate a business process template per process.

Output: Business Process Templates. The business processes defined are two: "Units Production" and "Production Planning" (figure 20).

**Figure 18. Surveying Business Processes Activity**

---

**Figure 19. Output Products of “Identifying Business Processes” Activity**

---

**Figure 20. Output Product of “Surveying Business Processes” Activity**

4.3 Applying the Third Phase of Business Process Data Engineering

The following activities are described: “Identifying Data Repositories” (figure 21) and “Surveying Data Repositories” (figure 22).

Activity: Identifying Data Repositories.
Input: Business Process Diagram Template, templates corresponding to every business process and Stakeholders interviews information.

Process:
1. Documentation analysis to define the data repositories used or consulted by each business process and generation of the data repositories template.

Output: Data Repositories Template (figure 23).

**Figure 21. Identifying Data Repositories Activity**

---

**Figure 22. Surveying Data Repositories Activity**

---

**Figure 23. Output Product of “Surveying Data Repositories” Activity**
Activity: Developing a Business Dictionary.

Input: Business Process Diagram Template, templates corresponding to every business process, Data Repositories Template and Data Structure Template.

Process:
1. Documentation analysis to define the business keywords. Upon identification, words are validated by stakeholders and thus the business dictionary is built up.
2. Wording glossary build-up, detailing keywords. By means of example, in figure 27 two keywords of the utmost importance to the business domain.
3. Building of the Concept-Relation and Concept-Attributes-Value chart (figures 28 and 29, respectively) upon the wording glossary, and generation of the Concept relations graph from the two latter charts (figure 30).
4. Correspondence establishment between wording glossary and definitions, acronyms and abbreviations template (figure 31).
5. Correspondence establishment among concept-attribute-value chart, wording glossary, and requirement-related attributes template (figure 32).

Output: Business Dictionary Template, Template of Abbreviations, Acronyms, and Definitions and Template of Attributes related to Requirements. Figure 33 shows an extract of each template.

Activity: Developing a Business Model.

Input: Data Repositories Template, Data Structure Template and Business Dictionary Template.

Process:
1. Documentation analysis to establish the relationships among business processes, data repositories and business words and business model diagram build-up.

Output: Business Model Diagram (figure 34).
4.5 Applying the Fifth Phase of Information Mining Process Specification

The following activities are described: “Formalizing Business Problems” (figure 35), “Identifying Information Mining Processes” (figure 36) and “Developing a Project Plan” (figure 37).

**Activity: Standardizing Business Problems.**

**Input:** Project Scope Definition Template, Project Objective Template, Requirement Objective Template and Business Process Diagram.

**Process:**
1. Analysis of the project objectives, project scope definition, business processes and data repositories and development of a business problems list.
2. If necessary, project objectives and requirement templates are verified and updated in parallel to the former process.

**Output:** Business Problem Template (figure 38).

**Figure 35. Formalizing Business Problems Activity**

**Activity: Identifying Information Mining Processes.**

**Input:** Business Problem Template.

**Process:**
1. Identification of the Information Mining processes to be applied to solve each of the business problems.
2. If necessary, assumptions, restrictions and requirement-related attributes templates are validated and updated in parallel to the former process.

**Output:** Information Mining Process Template (figure 39).

**Figure 36. Identifying Information Mining Processes Activity**

**Activity: Developing the Project Plan.**

**Input:** Project Scope Definition Template, Project Objective Template, Requirement Objective Template, Project Stakeholder Template, Project Risks Template, Requirement Risks Template and Information Mining Process Template.

**Process:**
1. Development of the project plan upon analysis of all prefetched documentation.
2. Announcement of the specified plan to business stakeholders.

**Output:** Project Plan Template (figure 40). Under no circumstances does this template aim to replace or overlap project administration activities, but it does aim to enclose the pivotal information for ulterior stages concretion.

**Figure 37. Developing the Project Plan Activity**

**Figure 38. Output Products of “Standardizing Business Problems” Activity**
The research presented in this paper has been partially financed by the 11211 EIU TIBA, research project of the National University of Technology - UTN (for its acronym in Spanish) Buenos Aires.

6. ACKNOWLEDGMENTS

5. CONCLUSIONS

This paper's main contribution is to provide the community with tools so as to enable them to carry out a thorough management of requirements in the context of an Information Mining Project. Consequently, this paper seeks to improve current methodologies in which this approach has been severely overlooked, which is why the process model proposed was applied to controversial case in which this approach has been severely overlooked, which is why the statement of every document process can lead to a higher rate of automation in these processes' techniques. Moreover, it is considered that a software tool capable of assessing the condition of templates and minutes advised by the process - additionally distinguishing the various versions of a single document, thereby reporting each version's contents together with a change history - is highly desirable project to be substantiated.

6. REFERENCES


