Verification and validation of a knowledge-based system

Jean-Marc Desharnais, Alain Abran, André Mayers, Julien Vilz, François Gruselin

In industry, a knowledge-based system (KBS) must often accommodate the considerable number of references required to support a particular knowledge domain [1]. The size of such a knowledge repository makes its detailed verification challenging and its subsequent maintenance onerous. New technologies can help improve the verification and the validation (V&V) of these knowledge repositories. To investigate the effectiveness of new technologies for V&V, we developed two subsequent versions of a KBS designed to improve the consistency of software measurement using ISO 19761. This paper also illustrates the quantitative benefits and lessons learned for the KBS V&V processes.

1 Introduction

There are problems related to the considerable number of supporting references to be included in the construction of a KBS:

- The knowledge of the expert (even if considerable) is not necessarily well defined. The user requirements for a KBS are then ill-defined [1];
- A text approach is sometimes the only way to express the knowledge of the expert; it is then difficult to verify the consistency of each text, when there are many text within the KBS;
- There are many links between the various parts of the knowledge system to consider.

Verification and validation (V&V) of a KBS is fairly new as a research topic. In 2001, Preece [2] noted that it is still difficult to draw conclusions about the efficiency of different V&V techniques because of a lack of available data. Information on knowledge-based systems is often textual and of a semantic nature, and, as such, it can be successfully inspected manually for anomalies related to the quality of knowledge, while formal methods have not yet proved to be convenient for most V&V projects. Under a specific set of conditions, automated support tools for detecting anomalies should be useful, provided that the data are structured enough to allow for some level of automation. Hayes et al. have worked on Case-Based Reasoning (CBR) using XML, but their approach is not directly related to the V&V process. Instead, they were looking “to extend the incremental CBR approach to network applications, to examine the distributed architecture to such a system and to situate the first two strands as part of a process of creating open standards for case-based network computing…” [3] The V&V terminology used by these authors is somewhat ambiguous, in that the same terms are used for different V&V techniques, but often in different contexts. All authors agree, however, on Barry Boehm’s [4] definition of verification (doing the system right) and validation (doing the right system), even though the techniques they use are different.

ISO 19761 [5] standardizes the measurement rules of the COSMIC-FPP measurement method. The main objective of COSMIC-FPP is to provide the size of the software based on the Functional User Requirements (FUR). The measurer must identify the different functional processes of the software. For each functional process, the measurer must also identify the data movements, which are Entry, Exit, Read, and Write. The number of data movements corresponds, by convention, to the functional size of the software.

2 Overview of our KBS prototypes

The COSMIC-Xpert KBS [6-8] was developed as a diagnostic tool to help measurers derive the correct measurement results when using ISO 19761 to obtain the functional size of software. The design of this diagnostic tool consists of a knowledge system built using both case-based and rule-based approaches. The first prototype included 105 case problems and almost 800 files (hyperlinks) for the references required to support the KBS rules repository. Verification and validation of such a large number of rules (files) is, of course, very challenging. This led us to the development of a second Web-based KBS prototype (XML, XSL, Java Server Page), the design of which was much easier to verify and validate.

The functionality and the design of both prototypes were similar from the point of view of the COSMIC-FPP measurer, but used different technologies: Microsoft Visual Basic 6 (language) and Microsoft Access (database) for the first prototype [7], and a Web approach with an XML structure for the second prototype [9]. The first prototype included two interfaces:

- An interface for the measurement expert who must put into the diagnostic tool the knowledge required for the establishment of diagnostics, and who must maintain it;
- An interface to support the measurer in his measurement tasks.
- In the second prototype, the measurer interface is Web-based, and the measurement expert interface has been replaced by a single input for both the administrator and the expert.

3 The V&V process

More than a hundred case problems were used to populate the first KBS prototype, producing nearly 800 files, each between half a page and 3 pages in length (for a total of more than 1200 pages). In the first KBS prototype, the knowledge was stored in multiple types of documents, in which, depending on the type of document, both the structure and the content may vary.

Verification process

Verification must be carried out on the whole KBS, including the 779 files. A key verification challenge is to ensure that each concept (definition, principles, context, rules, etc.) appearing in one specific document be used in exactly the same way in documents of all types. We also need to verify that all the links are used.

Table 1 presents an overview of our verification process. A detailed verification process was designed for each verification criterion.

<table>
<thead>
<tr>
<th>Verification Criteria and Techniques</th>
<th>Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion: Coherence</td>
<td>Creating XML files matching determined XML schemas (XSD)</td>
</tr>
<tr>
<td>Technique: Inspection</td>
<td>XML schema (with XSD) and XSL (reusing part of several XML files in an output)</td>
</tr>
<tr>
<td>Criterion: Redundancy, reusability</td>
<td>Automation with a tool to ensure that all links are present</td>
</tr>
<tr>
<td>Technique: Static verification</td>
<td></td>
</tr>
<tr>
<td>Criterion: Completeness</td>
<td></td>
</tr>
<tr>
<td>Technique: Production of links between</td>
<td></td>
</tr>
<tr>
<td>- topological concepts and case problems</td>
<td></td>
</tr>
<tr>
<td>- keywords and topological concepts</td>
<td></td>
</tr>
<tr>
<td>- case problems and themes</td>
<td></td>
</tr>
<tr>
<td>- themes and recommendations</td>
<td></td>
</tr>
<tr>
<td>- overall links</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Verification process

There were a number of outcomes for each type of verification criterion.
- Conformity with the structural standard of five types of XML document (keywords, topological concepts, themes, and recommendations);
- Conformity with the ontological definitions verified through keywords;
- Syntax and grammar for all user documents.

Applying this structural standard using our verification criterion significantly reduced the number of documents (from 779 to 150).

Validation process

The validation process for this KBS consisted of analyzing whether or not the recommendations were appropriate. We applied the validation process to the recommendations, since validation is related to user results. This implies that the validation process is not as sensitive to the technology used as the verification process. However, there is a relation between verification and validation: if there is an effective verification process, it is easier to execute the validation process. For example, the KBS is not working well because there are many missing links, it will not be possible to validate all the recommendations. Table 2 shows the validation process which was applied independently for each of the two prototypes.

<table>
<thead>
<tr>
<th>Validation Criteria</th>
<th>Validation Techniques</th>
<th>Validation Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- Correctness (to ensure that all recommendations are correct based on the various possibilities)</td>
<td>Decision tree Inspection</td>
<td>Generate all the recommendations from all the case problems</td>
</tr>
<tr>
<td>2- Correctness (to ensure that the % linking criteria to the various concepts are correct)</td>
<td>Decision tree Inspection</td>
<td>Generate a report of the various concepts with all associated percentages</td>
</tr>
<tr>
<td>3- Reliability (to ensure that the context of each case problem described is related to the case study)</td>
<td>Manual inspection by an expert Automation not yet possible.</td>
<td>Generate HTML files (through XSL) which use information on the context</td>
</tr>
</tbody>
</table>

Table 2: Validation process

Two different validation approaches were used to execute the validation plan, one for each of the two prototypes. For prototype 1, the validation was carried out by four COSMIC-FPP experts, while a single COSMIC-FPP expert executed the validation for prototype 2.

The validation outcomes of the first prototype were analyzed based on the following criterion ratings: agreement, disagreement and don’t know. On the basis of the experts’ comments, the disagreement and don’t know responses were related to a misunderstanding of some case problems, mainly because they were either ambiguous or ill-defined. For the majority of the case problems (80%), the measurement experts agreed with the recommendations. The range of variation between the experts is low for the agreements (between 79% and 85%), and higher for the disagreements and don’t knows.

The validation for prototype 2 was executed by a single measurement expert. In general, the measurement expert agreed with the recommendations for the 104 case problems, but proposed some changes to the content of many of them (for example, he proposed adding definitions to 120 recommendations). He also established the maximum number of recommendations (1480) and suggested using only a subset (545) of these, because many of the recommendations were very similar. He also checked that the percentages (%) associated with the various concepts were correct, which he confirmed. However, because there was only one measurement expert for this validation, the recommendations should eventually be cross-validated by other experts.

4 Benefits and lessons learned

The main benefits are:
- the number of anomalies in our KBS was reduced considerably;
the KBS will be easier to maintain in the future, not only because the number of files is reduced, but also because there is a structural link between the various parts of the KBS; the information in the KBS is consistent because there is only one source of information; the information in the KBS is non-redundant for the same reason; we know that the expert agrees with the recommendations.

Few authors have demonstrated the efficiency of the various verification and validation approaches for KBS because of the lack of available data, as noted by Preece [2]. To investigate the effectiveness of new technologies for V&V and maintenance for KBS, we have developed two subsequent versions of a KBS.

New technology can help improve both the verification and the maintenance of knowledge repositories. The KBS used in this study was implemented using two different prototypes. For each prototype, we designed and executed a V&V process.

In summary, the V&V of a single KBS, but one developed using two different techniques, made it possible to demonstrate the efficiency of one technique over the other. Moreover, our approach is in accordance with the emerging use of XML in the CBRS domain.

Finally, we think that our methodology could be integrated into software engineering practices in such a way that software designers could manipulate a large amount of knowledge within their KBS with few errors.

References


Contact

Jean-Marc Desharnais, Département de génie électrique École de Technologie Supérieure (ETS) – Université du Québec jmdeshar@ele.etsmtl.ca
Alain Abran: aabran@ele.etsmtl.ca
André Mayers: andre.mayers@dmi.usherbrooke.ca
Julien Vilz: jvi@info.fundp.ac.be
François Gruselin: fgruselin@xpectis.com

Dr. Jean-Marc Desharnais is a Professor at the École de Technologie Supérieure (ETS) – Université du Québec. For many years he was involved in the field of software measurement (productivity, estimation, quality) as consultant and searcher. In the framework of his Ph.D., his main interest was the construction of a KBS for beginner in the measurement field. The main subjects of his publications in the last 15 years were: productivity measure, software estimation, verification and validation, functional measure and knowledge base system.

Dr. Alain Abran is a Professor and the Director of the Software Engineering Research Laboratory at the École de Technologie Supérieure (ETS) – Université du Québec. He is currently Co-executive editor of the Guide to the Software Engineering Body of Knowledge project. He was also actively involved in software engineering standards as the international secretary for ISO/IEC JTC1 SC7 - Software and System Engineering; he is also Co-chair of the Common Software Measurement International Consortium (COSMIC).

André Mayers holds a master degree in psychology and a Ph. D. in computer science. He is currently a professor of computer science at the University of Sherbrooke. Its research interests relate to artificial intelligence and cognitive sciences. He is interested more particularly in cognitive process modeling, the design of user models and their integration in systems which need for a fine tuned model of their user in order to act intelligently with them.

Julien Vilz got a master in Information Technology at the University of Namur (Belgium) in 2003 by presenting jointly with François Gruselin a statement about the verification and validation of hybrid expert system, realized mainly at the University of Quebec in Montreal at the end of 2002. At present time Julien Vilz is researcher in the database application engineering laboratory at the University of Namur. He takes part at the development of a E-business dedicated case-tool for little and middle enterprise.

François Gruselin obtained a mastery in Information Technology in the University Faculties Our Lady of the Peace to Namur (Belgium) in 2003 by presenting collectively with Julien Vilz a statement about the verification and the validation of hybrid expert systems. A big part of this work was realized at the UQAM and at ETS at the end of 2002. At present, François Gruselin works for a private society in Luxembourg, XPECTIS, in a project of exchange of transactional information for MasterCard (‘clearing’).