KBS LIFE-CYCLE AND VALIDATION

CAHIER N° 133
janvier 1996

C. HAOUCHÉ ¹
C. ROSENTHAL-SABROUX ¹


¹ LAMSade, Université Paris-Dauphine, Place du Maréchal De Lattre de Tassigny, 75775 Paris Cedex 16.
## CONTENTS

<table>
<thead>
<tr>
<th>Sections</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Résumé</td>
<td>1</td>
</tr>
<tr>
<td>Abstract</td>
<td>1</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>2</td>
</tr>
<tr>
<td>2. Classical main steps in a KBS life-cycle</td>
<td>3</td>
</tr>
<tr>
<td>2.1 Acquiring requirements</td>
<td>3</td>
</tr>
<tr>
<td>2.2 The analysis phase</td>
<td>3</td>
</tr>
<tr>
<td>2.3 Building the KBS</td>
<td>4</td>
</tr>
<tr>
<td>2.4 Validating the KBS</td>
<td>4</td>
</tr>
<tr>
<td>2.5 Installation and maintenance</td>
<td>5</td>
</tr>
<tr>
<td>3. A proposal for validation steps in a KBS life-cycle</td>
<td>5</td>
</tr>
<tr>
<td>3.1 CM validation</td>
<td>7</td>
</tr>
<tr>
<td>3.2 The validation of the CM representation</td>
<td>7</td>
</tr>
<tr>
<td>3.3 Inference paths validation</td>
<td>8</td>
</tr>
<tr>
<td>3.4 Validating the links between the KBS and the CM</td>
<td>9</td>
</tr>
<tr>
<td>4. KBS validation</td>
<td>9</td>
</tr>
<tr>
<td>5. Conclusion</td>
<td>13</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>14</td>
</tr>
<tr>
<td>References</td>
<td>14</td>
</tr>
</tbody>
</table>
CYCLE DE VIE ET VALIDATION D'UN SBC

Résumé

Nous nous intéressons aux étapes de validation dans le cycle de vie d'un système à base de connaissances (SBC). Nous nous sommes focalisées sur l'utilisation d'un modèle conceptuel, élaboré pendant la phase d'acquisition des connaissances, pour valider un SBC. Cela implique de définir de nouvelles étapes dans le cycle de vie du SBC et de définir les processus de validation inhérents à ces étapes. Plus pratiquement, nous proposons une méthode de test utilisant en partie un modèle conceptuel à la KADS, plus précisément la structure d'inférence, pour tester le SBC. Nous nous efforçons de préciser, pour chaque étape de validation considérée, quand cette validation doit avoir lieu, qui sont les acteurs de la validation et, enfin, comment cette validation doit être menée.

KBS LIFE-CYCLE AND VALIDATION

Abstract

This research deals with validation steps within the KBS life-cycle. Our paper is centered on the use of the conceptual model, elaborated during the Knowledge Acquisition phase, for the KBS validation. We will, on the one hand, show the implications of such a use, that is, the need to specify new steps in the life-cycle as well as validating their outputs. On the other hand, we will propose a procedure to use one part of a KADS conceptual model - the inference structure - to test the KBS. We will, as far as possible, and for each considered step of the life-cycle, precise the following points: When the system will be validated, that is at which step of the life-cycle, who will be validating the system and how this validation will be achieved.
1 Introduction

"The primary functions of a software process model are to determine the order of the stages involved in software development," as Boehm said in (Boehm, 1988). For the KBSs, there are essentially two kinds of development methodologies (Dieng, 1990). The first one is the incremental development technique. This methodology is also called rapid prototyping (Hayes-Roth et al., 1983). The major steps of this methodology are: identification, conceptualization, construction, implementation, and evaluation.

This methodology involves the following steps:
1. Identification: This step involves the identification of the problem or the software product that needs to be developed.
2. Conceptualization: In this step, the problem is broken down into smaller, manageable parts, and a conceptual model is created.
3. Construction: The conceptual model is then converted into a working prototype.
4. Implementation: The prototype is implemented using programming languages and tools.
5. Evaluation: The implemented software is evaluated to check its functionality and performance.

This methodology is particularly useful for developing complex software systems where the requirements are not well-defined at the beginning.
2 Classical main steps in a KBS life-cycle

Some studies have already dealt with KBS life-cycle (Halker & Welz, 1989; Chabaud et al., 1991; Wieginga et al., 1992; Dieng, 1990). The main steps that are identified are as follows (see Figure 1):

2.1 Acquiring requirements

This step is an informal one. It is conducted by means of rough interviews to identify the user’s needs. During this step, the expert, the knowledge engineer and the final user have to define these needs. Then, a decision is taken about the KBS appropriateness to solve the problem.

2.2 The analysis phase

The analysis phase aims at building a conceptual model (CM), that is a high level - implementation independent - description of the KBS. This phase is a cyclical one; it consists, on the one hand, of gathering the data that are relative to the problem. These data concern both the domain concepts and the different ways to use these concepts to solve the problem at hand. This gathering is conducted by the knowledge engineer who can use different elicitation techniques when interviewing the expert and the user (Dieng, 1990). We notice that these interviews are much more detailed than the previous ones and that they are relative to the domain knowledge. On the other hand, the knowledge engineer can use different kinds of methodologies
(KADS (Wielinga et al., 1992), COMMET\(^1\) (Canamero & Geldof, 1993), MACAO (Aussenac, 1989), etc.) in order to structure the knowledge that has previously been acquired and to build a conceptual model (Krivine & David, 1991; Wielinga et al., 1992). Let us notice that the CM building can lead to acquire new data and vice-versa, and this, until the CM is valid.

The KADS modeling layers

KADS\(^2\) offers structures for building a conceptual model to guide the knowledge acquisition process. A KADS conceptual model is four-level organized\(^3\), from the least abstract, which describes the domain concepts, to the most abstract, which represents the behavior of the system.

* The **domain layer** is a description of the domain concepts and of the relations between these concepts.

* The **inference layer**. In this layer, roles are attributed to the domain concepts. In addition, an inference structure describes all the links between the roles and the inferences that are involved in the problem solving method.

* The **task layer** specifies the goals assigned to the system and the way inferences are organized to achieve these goals.

2.3 Building the KBS

Despite the availability of a conceptual model, there is no systematic process to build a KBS from a conceptual model. However, some studies identify this phase as an operationalization one and propose tools, as the LISA language (Delouis, 1993) and the shell AIDE (Greboval & Kassel, 1992).

Two sub-phases are generally involved to build a KBS. One of them aims at choosing a representation formalism in order to build the knowledge base. The other one deals with the choice of the inference engine. Notice that these two sub-phases are strongly dependent.

2.4 Validating the KBS

Validating a KBS involves verifying the Knowledge Base and testing the KBS results and behavior.

Verification approaches

Various verifications tools have been developed. Most of them deal with detecting incoherencies (SACCO (Ayel, 1987), CHECK (Nguyen et al., 1985)) or with proving the KB coherency (COVADIS (Rousselet, 1988)). They also deal with detecting redundancies, potential circularities, subsumed rules (CHECK (Nguyen et al., 1985)).

\(^1\) KREST is the new name for COMMET.

\(^2\) See (Hickman et al., 1989; Wielinga et al., 1992) for a detailed presentation of the KADS methodology.

\(^3\) We used the KADS I methodology but we did not use the strategy layer. Notice that this layer is suppressed in KADS II.
Testing approaches

Two main approaches have been used to test KBSs: the black box approach and the white box approach.

a - The black box approach

A test case is composed of a set of data and of the results that are expected from these data. These data are used to run the system. The results obtained are then compared to the expected ones. Notice that this approach deals only with the inputs of the system and with the outputs it provides. It does not consider the structure of the system.

b - The white box approach

The white box approach focuses on the structural aspects of the KBS. It handles a KB structure to study if all the parts of this KB are exercised (Preece et al., 1993). For instance, if the KBS is rule-based, this approach allows to detect the rules which are not used on a given set of test cases, the ones that are often used, etc. Notice that some tools aim at generating test cases automatically; this is the case for SYCOJET (Vignollet, 1991).

2.5 Installation and maintenance

Two possibilities are to be considered. (i) The KBS is completely independent. In this case, it is used with no further effort. (ii) The KBS is to be integrated into a more complex information system. Then, some problems concerning interactions with other modules could occur. The maintenance phase may lead to review the conceptual model.

3 A proposal for validation steps in a KBS life-cycle
Figure 2: Additional Steps for a KBS Life-Cycle

the KADS conceptual model (the domain layer, the inference layer and the task layer) is represented. This step is performed by the knowledge engineer and/or by a programmer.

(B) Building the inference paths

We use the inference structure representation provided within the conceptual model representation, to derive all the inference paths. An inference path is a sequence of inferences between an initial inference and a final inference. An initial inference has only initial roles as inputs, and a final inference has a final role as output.

(C) Linking the implemented form of the CM and the KBS.

These links are used to build a high level trace during the testing phase. Up to now, there is no, to our knowledge, a method to establish such links. In our work, it is realised by the knowledge engineer who can use his knowledge of both the CM and the KBS.

Each step $s$ of the KBS life-cycle is expected to produce outputs $o(s)$ from inputs $i(s)$. Some of these inputs can be seen as a kind of specifications. The outputs $o(s)$ are represented using a given representation. In our opinion, two kinds of validations are thus to be performed on $o(s)$. (i) A validation, which we call internal validation, which aims at verifying that the output $o(s)$ is consistent with the specification $i(s)$ and its validation.

...
(ii) an *external validation* which aims at assessing that the outputs \( o(s) \) verify the specifications expressed in \( i(s) \).

We will focus on the validation of the conceptual model (2), the validation of the CM representation (A), the validation of the inference paths (B), the validation of the links (C). The KBS validation will be studied in the next section. We will specify the inputs and outputs for each of these steps as well as the kinds of validation which are necessary.

### 3.1 CM validation

For this phase, the inputs correspond to the CM and to the requirements which have been identified in phase (1). The outputs are a valid conceptual model. The representation formalism on which we focus is the one given by the KADS methodology.

The **CM internal validation** is performed by the knowledge engineer. As we said above, a KADS conceptual model is structured in 3 layers, namely: the domain layer, the inference layer and the task layer. Consequently, we identify two kinds of validation:
{\text{initial-specification}
  \text{role:}
  \text{input-for: !create \&assemble1 \&assemble2 \&assemble3}
  \text{output-for: !complete2}
  \text{linked-to-domain: \&lexeme-instance}
  \text{linked-to-kbs: }
}
{\text{!complete}
  \text{inference:}
  \text{input-roles: initial-specification1 \&domain-model0}
  \text{output-role: initial-specification}
  \text{linked-to-kbs: !regle-instanciation-composantes }
}

Figure 3:

\text{linked-to-domain, linked-to-kbs}. Each role is also described by a set of attributes (input-for,
3.4 Validating the links between the KBS and the CM

We focus now on the links\(^4\) between the KBS and the inference structure. The inputs for this step are the implemented representation of the inference structure, the KBS, and the links between them. The outputs are the valid links, that is associations established between the two structures. These links are represented by the attribute "linked-to-KBS" on each inference and role, and by the attribute "linked-to-CM" on each rule as exemplified on Figure 4.

```
{ !complete  
!inference:  
!input-roles: !initial-specification1 !domain-model0  
!output-role: !initial-specification  
!linked-to-kbs: !regle-instantiation-composantes } 
{!regle-instantiation-composantes  
!linked-to-cm: !complete  
!condition:  
(and ![?instance !occurrence-de ?mot]  
 ![?lexeme !entrec ?mot]  
 (meta-frame ![?lexeme ?m])  
 (fork  
 (then-do (copy-frame ?m (meta-frame ![?instance]))  
 (and ![?lexeme !s-applique-sur ?arg]  
 (then-do ![?instance !s-applique-sur ![?arg]])  
 (and ![?lexeme !case-frame ![?arg]]  
 (then-do ![?instance !case-frame ![?arg]]))) )} 
```

Figure 4: The links linked-to-kbs and linked-to-cm between the inferences and the rules

The links internal validation consists of verifying that these links are symmetrical. It is performed by the knowledge engineer and/or the programmer.

The links external validation aims at verifying that each inference of the inference structure representation is linked to the KB and reciprocally. This is performed by the knowledge engineer and/or the programmer.

4 KBS validation

The inputs for this step are the valid inference paths, the KBS and the links between the CM and this KBS. The output is the validated KBS. There are many representation formalisms. The most common ones are the rule-based formalisms and the object-oriented formalisms.

\(^4\)We recall that these links are established empirically by the knowledge engineer, relying on his knowledge and understanding of both the KBS and the CM.
The KBS internal validation consists of verifying whether the KB is represented in accordance with the formalism. Many tools have been developed to detect redundancies, subsumptions as well as incoherencies in such knowledge bases. This validation is mainly performed by the knowledge engineer and/or the programmer.

The KBS external validation

We focus here on the KBS test using the IP and we propose the following procedure:

1) The KBS is run on a data case and a code level trace is built.

2) The current inference path (CIP) is built from this trace using the links that are established between the CM and the KBS. This CIP constitutes a "high level trace".

3) The CIP is compared to the set of VIP. Two cases are to be considered:
   a) The CIP ∈ VIP: the testing procedure is iterated on a new data case.
   b) The CIP ∉ VIP: The CIP is shown to the expert. If the latter identifies a valid inference path, this path is added to the set of VIP; else, the links between the CM and the KBS are used to localize the code that has to be revised.

An example

In this section, we present an illustration of the previous validation procedure. This procedure is currently being tested on a real world KBS, AMD (Cavazza et al., 1992), for which we developed a CM (Hacouche, 1993). This CM describes a system developed to understand medical summaries. Here, "understanding" means building a representation of the summary, called a situation model. In this situation model, all the implicit knowledge carried in the summary is made explicit by using common-sense and medical knowledge. A preliminary task consists of building l-descriptions, that is structures that gather the semantic components through syntactic links between lexemes. For instance, the l-description l-descr25 (Figure 7) gathers the information which relates to the same concept tumeur and which is carried by the lexemes nodule-13 and gauche-15 (Figure 6).

To illustrate our testing procedure, we will focus on this subtask. The part of the inference structure concerning this subtask is thus presented (see Figure 8). On this inference structure, we can see the initial roles\(^5\) domain model 0 and initial specification and the final roles\(^6\) intermediary specification and complete specification. Inferences which have initial roles as inputs are called initial inferences (complete) and the ones which have a final role as an output are called final inferences (create, assemble1, assemble2, assemble3).

The initial set of inference paths The following set is built using a procedure which yields all the paths that link an initial inference to a final inference. These paths specify all the intermediary inferences between an initial inference and a final inference.

---

\(^5\) An initial role is an input to the inference structure.
\(^6\) A final role is either an output of the inference structure or a role that is specified as final by the expert as for instance intermediary specification.
Figure 5: The lexical inputs *nodule* and *gauche*

Figure 6: The lexeme instances *nodule-13* and *gauche-15*

Figure 7: The l-description *l-descr25* created for the concept *tumeur*
Figure 8: An inference structure for the sub-task Building l-descriptions

1 - (complete complete2 create)
2 - (complete complete2 create assemble1)
3 - (complete complete2 create assemble2)
4 - (complete complete2 create assemble2)
5 - (complete complete2 assemble1)
6 - (complete complete2 assemble2)
7 - (complete complete2 assemble3)

The set of Valid Inference Paths VIP The previous set of inferences is syntactically correct. However, some paths are not recognized as semantically correct by the expert. As an example, the expert tells that the path 5 is not semantically correct. In fact, among the lexeme instances, some will lead to the creation of l-descriptions, like *nodule* and some will not (*gauche*). On the other hand, the assembling process deals with transferring semantic components on l-descriptions. So the inference paths that do not create l-descriptions are not valid. For the same reason, the paths 6 and 7 are not valid.

Furthermore, all the semantically correct paths are not provided by the procedure, especially the ones that include “parallel inferences”. For example, the three inferences *assemble1 assemble2 assemble3* may occur separately or simultaneously. Thus, the expert also provides the additional paths given below:

8 - (complete complete2 create assemble1 assemble2 assemble3)
9 - (complete complete2 create assemble1 assemble2)
10 - (complete complete2 create assemble1 assemble3)
11 - (complete complete2 create assemble2 assemble3)
Finally, the set of VIP is as follows: 1 2 3 4 8 9 10 11.

Using the VIP during the testing phase Let us consider the following example: "The thyroid palpation reveals a left lobar nodule". The KBS tracer yields the following set of fired rules:

- lregle-instanciation-composantes (x 7 firings)
- lcreate-L-description (x 5)
- ltransfer-comp-on-l-descr (x 1)
- lspeify-links-on-l-descr-*with-polarity* (x 5)

The rule "lregle-instanciation-composantes" has thus been fired seven times in order to complete the description of the initial lexemes. The rule "lcreate-L-description" has been fired 5 times to create five l-descriptions, etc.

Without giving the whole rules, we give below the links that have been established between these rules and the inferences.

- The rule "lregle-instanciation-composantes" is linked to the inference !complete.
- The rule "lcreate-L-description" is linked to the inference !create.
- The rule "ltransfer-comp-on-l-descr" is linked to the inference !assemble1
- The rule "lspeify-links-on-l-descr-*with-polarity*" is linked to the inference !assemble3

These links permit to build the Current Inference Path which is given below:

(complete create assemble1 assemble3).

This path is not a member of the VIP. Generally speaking, if a CIP is not a member of the VIP, the expert has to decide if this CIP is a valid path or not. In the first case (i.e. the CIP is valid), as the CIP does not match the inference structure, the latter has to be modified to take it into account. In the second case, the code level has to be corrected. In our case, the expert considers that the CIP is valid, thus leading to modify the inference structure (Figure 9). Notice that this may lead to add new paths which will have to be validated.

5 Conclusion
Figure 9: The new inference structure for the sub-task Building l-descriptions

Acknowledgments

We would like to thank Jean Charlet and André Molia for their helpful advices.

References


Contents

1 Introduction ....................... 2

2 Classical main steps in a KBS life-cycle ......... 3
   2.1 Acquiring requirements .................. 3
   2.2 The analysis phase ...................... 3
   2.3 Building the KBS ...................... 4
   2.4 Validating the KBS .................... 4
   2.5 Installation and maintenance .......... 5

3 A proposal for validation steps in a KBS life-cycle ... 5
   3.1 CM validation ........................... 7
   3.2 The validation of the CM representation .... 7
   3.3 Inference paths validation ............... 8
   3.4 Validating the links between the KBS and the CM 9

4 KBS validation ...................... 9

5 Conclusion .......................... 13