# Overcoming Human Weaknesses in Validation of Knowledge-Based Systems

Rainer Knauf
rainer.knauf@tu-ilmenau.de

Setsuo Tsuruta
tsuruta@sie.dendai.ac.jp

Avelino J. Gonzalez
gonzalez@ucf.edu

**Abstract:** Human experts employed in validation exercises for knowledge-based systems often have limited time and availability. They often have different opinions from each other as well as from themselves over time. We address this situation by introducing the use of validation knowledge used in prior validation exercises for the same knowledge-based system. We present a Validation Knowledge Base (VKB) that is the collective best experience of several human expert validators. We also present the concept of Validation Expert Software Agents (VESA) that represent a particular expert's knowledge. VESA is a software agent corresponding to a specific human expert validator. It models the validation knowledge and behavior of its human counterpart by analyzing similarities with the responses of other experts. We also describe experiments with a small prototype system to evaluate the usefulness of these concepts.

## 1 Introduction

In spite of significant advances in recent years, validation of knowledge–based systems still requires significant involvement on the part of human validators. In contrast to verification, validation typically involves rigorous and often extensive testing of the system.

Our work is implemented in the context of our previously described Validation Framework [KGA02]. Briefly, the methodology incorporated in the Validation Framework provides for the generation of a practical test case set that meets certain user–specified and domain–specific validation criteria. It subsequently provides a format for expert validator review of test cases and results using a variation of a Turing Test for the validation of knowledge–based systems. In this step, humans play the role of expert validators as part of a validation panel. Their task is to (1) solve the test cases posed to the system under evaluation, (2) review and provide their judgment on the correctness (the ratings) of all anonymous solutions (the system's as well as the panel's own). Lastly, the Framework also provides an algebra to compute a validity statement (quantifiable determination of the system's overall validity) as well as a means to refine its rules. This Framework is specifically designed to work with rule-based systems, so we limit our discussion in this article to such systems. [Kn00] provides a complete and detailed description of this Validation Framework.

In the Validation Framework, the outcome of the validation exercise is heavily influenced by the quality of the human expert validators' contribution. However, expert validators may not always be available or even willing to cooperate with the validation exercise, thereby causing delays. Furthermore, they may not always agree among themselves. Expert validators are a scarce resource, have limited time, and are expensive to employ. This is in reference, of course, to Feigenbaum's original description of the knowledge acquisition bottleneck [Fe84].

Based on the insight that knowledge-based systems require special evaluation and maintenance (see [CB93]) several theories, approaches, and techniques for their verification and validation have been developed. These are found in [OO93, Me93, MP94, LO94, ZP94, Pr98, Pr99, MM00, OL01] and [KGA02]. Also, see [DDL02] for the most recent overview of these techniques. Interestingly, these publications reveal that many authors initially approached the problem through formal approaches. These mostly focused on verification, but also on validation to a lesser degree. Over time, they supplemented their approaches to include system refinement and addressed the human factor as part of this system refinement. Refer to [MV96, BKH00, DDL02, KK03] and [Kn02] for typical examples. The issue of human involvement is addressed for specific application domains in [AKT93, Br95, Ra96, OL98] and [MV96]. Nevertheless, a general approach addressing imperfect human knowledge within the validation process, both from the standpoint of improving the process as well as making the process less burdensome to the human expert validators, so far appears to be missing.

Our specific contributions in this article are two: (1) a **Validation Knowledge Base** (VKB) that contains collective validation knowledge from prior validation exercises, and (2) **Validation Expert Software Agents** (VESA) that can replace missing validators during validation exercises. The VKB can also help in reducing expert validator workload. These concepts, while seemingly simple, contain complex techniques to maintain currency in the validation knowledge (learn) and to compensate for gaps in the knowledge.

Both concepts are related to Case-Based Reasoning (CBR) approaches [Ko93, Alt98]. In fact, the VKB is a case base. A difference to CBR systems are (1) the use of the cases and the particular way to maintain the case base. Cases in the VKB are not used to derive solutions to new new problems by adopting solutions of former problems. Instead, it is very much intended to bring (non derived) former solutions into the rating process of the validation technology. The maintenance of a VKB aims at collecting those solutions to processed cases, which gained the maximum approval within the expert panel.

The VKB concept shares some ideas with the concept of so–called Experience Factories (EF) [BCR94a]. However, an essential objective of EF is supporting the process of project organization. Therefore, the strictly separate learning from project organization, which seems to be important for their intent [BCR94b]. The VKB concept, on the other hand, aims at learning domain knowledge, only. The intent of VKB is to support the refinement of systems by contributing topical knowledge.

The VESA concept is also a little different from CBR approaches. A basic difference is the way to look for similarities in the case base: The subject of looking for similarities are not the application domain problem cases, but individuals who solved or rated problem cases in the past. After finding a most similar problem solver or rater, we adapt his/her solution or rating towards a solution of the missing solver or rater.

Sections 2 and 3 describe these two concepts in earnest. Section 4 introduces the tests performed to evaluate the effectiveness of these concepts when applied to a simple but non-trivial knowledge-based system. Section 5 discusses the results of the tests, section 6 outlines the learning curve, and section 6 provides a summary and outlook.

## 2 Collective Validation Knowledge in the VKB

To improve the validation process, the validation knowledge used in prior exercises, namely the set of test cases (the test inputs and the best rated solutions) along with their authors, must persist from one validation exercise to the next. This is effectively accomplished by the Validation Knowledge Base. The VKB and its historical validation knowledge can also significantly reduce the involvement of expert validators by eliminating their need to solve old test cases whose solutions are already found in the VKB. The expert validator panel needs only to solve new test cases created by the Validation Framework that are not already part of the VKB. However, they still must rate all solutions. We continue by describing the internal structure of the VKB.

The task of the expert validators doesn't change at all by introducing the VKB. Since the Turing Test concept of the original framework [Kn00] is intended to be an anonymous rating process, the panel even doesn't realize that there are some solutions to test cases that are not provided by the system or the current panel.

The VKB's set of previous (historical) test cases and their best rated solutions, which can be described by 8-tuples $[t_j, E_{Kj}, E_{Ij}, sol_{Kj}^{opt}, r_{IjK}, c_{IjK}, \tau, D_j]$, where $t_j$ is a test case input, $sol_{Kj}^{opt}$ is a solution associated to $t_j$, which gained the maximum experts' approval in a prior validation exercise, $E_{Kj}$ is the list of experts who provided this particular solution, $E_{Ij}$ is a list of experts who rated this solution, $r_{IjK}$ is the rating of this solution, which is provided by the experts in $E_I$, $c_{IjK}$ is the certainty of this rating, $\tau$ is a time stamp associated with the validation session in which the rating was provided, and $D_j$ is an informal description of any aspects of the test case that could not be described formally by the other seven elements in the 8-tuple. Additionally, a list of supporters $E_S \subseteq E_{Ij}$ for each solution $sol_{Kj}^{opt}$ is kept in VKB. A supporter is a rating expert, who provided a positive rating for $sol_{Kj}^{opt}$. $E_S$ can easily be computed from $E_{Ij}$ and $r_{IjK}$ of each entry in the VKB.

The VKB functions in the second step, the test case experimentation. In the original approach, the test case generation procedure consist of two steps (a) generating a quasi exhaustive set of test cases $QuEST$ and (b) reducing it down to a reasonably sized set of test cases $ReST$ [KGA02]. Exactly between these two sub–steps is the 'entry–point' of the external validation knowledge stored in a VKB that has been constructed in prior validation sessions. Both $QuEST$ and the historical cases in VKB are subjected to the criteria-based reduction procedure that aims to build a subset of test cases in $QuEST$ or VKB. The cases in VKB are included in the reduction process to (1) ensure that they meet the requirements of the current application and (2) their number is small enough to be the subject of the time consuming and expensive test case experimentation.

## 3 Individual Validation Knowledge in a VESA

A VESA can model individual knowledge that is different from the accepted knowledge of the majority of experts. Thus, the VESA has the potential to maintain excellent and innovative individual human expertise. As with VKB, the task of the expert validators doesn't change at all by introducing the VESAs. The VESAs act simply as "colleagues" of the current expert panel, who also provide and rate solutions to test cases. Again, since the concept of the original framework [Kn00] is intended to be an anonymously, the panel even doesn't realize that there are some solutions to test cases that are not provided by the system or the current panel.

Formally, a *VESA*$_i$ acts as follows when requested to provide an assumed solution of expert $e_i$ for a test case input $t_j$:

1. In case $e_i$ solved $t_j$ in a former session (with a value other than $unknown$), his/her solution with the latest time stamp will be provided by VESA$_i$.
2. Otherwise,
   (a) All validators $e'$, who ever delivered a solution to $t_j$ form a set $Solver_i^0$, which is an initial dynamic agent for $e_i$: $Solver_i^0 := \{e' : [t_j, E_{Kj}, \ldots] \in VKB, e' \in E_{Kj}\}$
   (b) Select the most similar expert $e_{sim}$ with the largest set of cases that have been solved by both $e_i$ and $e_{sim}$ with the same solution and in the same session. $e_{sim}$ forms a refined dynamic agent $Solver_i^1$ for $e_i$: $Solver_i^1 := e_{sim} : e_{sim} \in Solver_i^0, |\{[t_j, E_{Kj}, \_, sol_{Kj}^{opt}, \_, \_, \tau, \_] : e_i \in E_{Kj}, e_{sim} \in E_{Kj}\}| \rightarrow max!$
   (c) Provide the latest solution of the expert $e_{sim}$ to the present test case input $t_j$, i.e. the solution with the latest time stamp $\tau$ by VESA$_i$.
3. If there is no such most similar expert, provide $sol := unknown$ by VESA$_i$.

If a VESA$_i$ is requested to provide a rating to a solution of a test case input $t_j$ on behalf of expert validator $e_i$, it models the rating behavior of $e_i$ as follows:

1. In case $e_i$ rated $t_j$ in a former session, VESA$_i$ adopts the rating with the latest time stamp $\tau$ and provides the same rating $r$ and the same certainty $c$.
2. Otherwise,
   (a) All validators $e'$, who ever delivered a rating to $t_j$ form a set $Rater_i^0$, which is an initial dynamic agent for $e_i$: $Rater_i^0 := \{e' : [t_j, \_, E_{Ij}, \ldots] \in VKB, e' \in E_{Ij}\}$
   (b) Select the most similar expert $e_{sim}$ with the largest set of cases that have been rated by both $e_i$ and $e_{sim}$ with the same rating $r$ and in the same session. $e_{sim}$ forms a refined dynamic agent $Rater_i^1$ for $e_i$: $Rater_i^1 := e_{sim} : e_{sim} \in Rater_i^0, |\{[t_j, \_, E_{Ij}, sol_{Kj}^{opt}, r_{IjK}, \_, \tau, \_] : e_i \in E_{Ij}, e_{sim} \in E_{Ij}, \}| \rightarrow max!$
   (c) VESA$_i$ provides the latest rating $r$ along with its certainty $c$ to the present test case input $t_j$ of $e_{sim}$.
3. If there is no such most similar expert $e_{sim}$, VESA$_i$ provides $r := norating$ along with a certainty $c := 0$.

## 4 Evaluation of VKB and VESA – A Test Prototype

To maximize the probability of enlisting adequate participation from expert validators for our experiment, we selected an amusing application problem: the selection of an appropriate wine for a given dinner. We built a simple system by consulting the topical literature and deriving some informal knowledge from it. One might argue, that this "application" is quite naiv and far away from similar to real world applications. Remember, the entire concept aims at the validation of intelligent (AI–) systems. Such systems are typically employed in application field with a very vague and disputatious domain knowledge.

The problem of selecting an appropriate wine depends on three inputs (a) the main course ($s_1$), (b) the kind of preparation ($s_2$), and (c) the style of its preparation ($s_3$). The input space of the considered classification problem is $I = \{[s_1, s_2, s_3]$: $s_1 \in \{pork, beef, fish, \ldots\}$, $s_2 \in \{boiled, grilled, \ldots\}$, $s_3 \in \{Asian, Western\}\}$. In fact, this is a quite simplistic classification. Remember, the experiment needs to be performed by voluntary humans, who do this job for free and must not bombarded with detailed information in bulk.

The output $O = \{o_1, \ldots, o_{24}\}$ contains 24 different kinds of wine [Kn04] [1]. Expressing the informal knowledge with these input and output specification as HORN clauses leads to a rule base $R$ consisting of 45 rules [Kn04]: $r_1$: $o_1 \leftarrow (s_1 = fowl)$, $r_2$: $o_1 \leftarrow (s_1 = veal)$, $r_3$: $o_2 \leftarrow (s_1 = pork) \wedge (s_2 = grilled)$

According to the test case generation technique as described in [KGA02], we computed a *Quasi Exhaustive Set of Test Cases* ($QuEST$) that contains 145 cases. To generate the *Reasonable Set of Test Cases* ($ReST$), we applied four criteria according to the semantic of the test cases and received 42 test inputs form the reasonable set of test cases $ReST$.

Available resources were three human experts $e_1$, $e_2$ , and $e_3$ and the computed reasonable set of test cases $ReST$ with 42 test inputs $\{t_1, \ldots, t_{42}\}$. The objective of our test program was to evaluate the feasibility of the $VKB$ and $VESA$ concepts. By feasibility, we mean whether or not the VKB can provide external knowledge from prior validation exercises to improve the validation process. The improvement was considered to be the introduction of test cases and solutions that would not have otherwise been considered by the panel. Furthermore, we mean whether or not VESAs can provide the same responses as their corresponding humans. In effect, we sought answers to the following questions: (1) Does the VKB increasingly contribute to the validation exercises in direct relation to the number of validation exercises? (2) Does the VKB increasingly contribute valid knowledge (best rated solutions) in direct relation to the number of validation exercises? (3) Does the VKB increasingly exhaust human expertise in direct relation to the number of validation exercises? (4) Do the VESAs' representations of their corresponding expert validators improve in direct relation to the number of validation exercises?

We refer to the resulting VKBs and VESAs[2] of an $i$–th session as VKB$^i$, VESA$_1^i$, VESA$_2^i$, and VESA$_3^i$. $ReST^i$, on the other hand, is the set of test cases generated for the current

[1] This is the initial output set. Of course, the human expertise might bring new outputs in the process.

[2] VESA$_1$, VESA$_2$, and VESA$_3$, which model the behavior of $e_1$, $e_2$, and $e_3$.

session, i.e. its top index is larger than that of the VESAs by one, because their indices refer to the current session whereas the VKB's and VESAs' indices refer to the result of the preceding session.

For a fair evaluation of the usefulness of VKB, the intersection of the test case inputs found in the VKB and those computed for the $ReST$ ($EK$ = external knowledge) needs to be considered in each validation session. $EK^i$ denotes the external knowledge hold by the VKB within the $i$–th experimentation session. Of course, $EK^1$ is empty, because there has no VKB been built before the $1^{st}$ session. All other $EK^i$ are formed by the subset of $ReST^i$, for which there are also solutions in the VKB$^i$ (the VKB which is available in the $i$–th session). The consideration of only this external knowledge is because this is the only knowledge that can be introduced into the rating process by the VKB from outside the current human expertise: $EK^1 = \emptyset$, $EK^2 = \Pi_1(\text{VKB}^1) \cap \Pi_1(ReST^2)$, $EK^3 = \Pi_1(\text{VKB}^2) \cap \Pi_1(ReST^3)$, and $EK^4 = \Pi_1(\text{VKB}^3) \cap \Pi_1(ReST^4)$.

We designed a set of metrics to address the four questions. After each session (session # $i$), beginning with the second session, we determine[3]:

- the number $rated_i$ of cases from VKB$^{i-1}$, which were the subject of the rating session and relate it to $|EK^i|$: $Rated_i := rated_i/|EK^i|$
- the number $best_i$ of cases from VKB$^{i-1}$, which provided the optimal (best rated) solution and relate it to $|EK^i|$: $BestRated_i := best_i/|EK^i|$
- the number $intro_i$ of cases from VKB$^{i-1}$, for which a new solution has been introduced into *VKB* and relate it to $|EK^i|$: $Intro_i := intro_i/|EK^i|$
- the number $ident_i$ of solutions and ratings, which are identical responses of $e_{i-1}$ and VESA$^{i-1}$ and relate it to the number of required solutions and ratings:
  $ModelRating_i := ident_i/|required\_responses|$

The above four questions can now be addressed as follows: (1) $Rated_4 > Rated_3 > Rated_2$, (2) $BestRated_4 > BestRated_3 > BestRated_2$, (3) $Intro_4 < Intro_3 < Intro_2$, and (4) $ModelRating_4 > ModelRating_3 > ModelRating_2$.

## 5 Results

1. $Rated_4 > Rated_3 > Rated_2$ ? With $Rated_4 \approx 0.85$, $Rated_3 \approx 0.071$, and $Rated_2 \approx 0.071$ this requirement was met at least in the step from the $3^{rd}$ to the $4^{th}$ session. The contribution effect could not really be expected as a result of the sessions before that. However, after the $3^{rd}$ session, a remarkable number (24 out of 28) possible cases of VKB$^3$ have been introduced in the rating process.
2. $BestRated_4 > BestRated_3 > BestRated_2$ ? With $BestRated_4 \approx 0.071$, $BestRated_3 = 0$, and $BestRated_2 = 0$ this requirement was also met when going from the $3^{rd}$ to the $4^{th}$ session. In the $4^{th}$ session VKB$^3$ contributed solutions for two cases, that had not been provided by the human experts, but won the 'rating contest'. Here, the VKB introduced new knowledge which turned out to be more valid than the knowledge provided by the human experts.

[3]In the first session the VKB is empty and thus, not able to contribute any external knowledge.

3. $Introduced_4 < Introduced_3 < Introduced_2$ ? With $Introduced_4 \approx 0.61$ , $Introduced_3 \approx 0.57$, and $Introduced_2 = 0.5$ this requirement was not met. The underlying assumption for this question a static domain knowledge, which needs to be explored systematically. This was not true for the considered domain. In interesting problem domains there is a change over time of the domain knowledge.
4. $ModelRating_4 > ModelRating_3 > ModelRating_2$ ? With $ModelRating_4 = 0.6$ , $ModelRating_3 \approx 0.63$, and $WellModeled_2 \approx 0.51$ we can at least claim that $ModelRating_4 \geq ModelRating_3 \geq ModelRating_2$ is almost met. That these numbers are not convincing is due to the human factor in the experiment and the approach itself: (1) All experts changed their opinion during the experiments for a remarkable number of cases. (2) In particular, the rating process of a VESA on the basis of a last consideration of this case in a solving (not rating) session is based on the assumption the domain is deterministic by nature, which is certainly not true for many interesting problem domains.

## 6 Lessons Learned

**Improvements to VKB** The VKB includes all aspects of "collective historical experience" that have been provided by former validation panels. However, some issues need to be addressed:

- Outdating Knowledge

  Because the number of solutions likely to be introduced in the rating process increases with the number of sessions, the probability of acquiring new external knowledge also increases over time. However, domain knowledge might become outdated because of new insights. A strong indication of this would be when a solution contributed by the VKB repeatedly receives poor ratings whenever it is introduced in a rating session. One approach to solve this problem is to analyze the prior ratings of each entry in the VKB and remove those entries that received poor ratings for an extended period.

- Completion of VKB towards other than (former) test cases

  The fact that a VKB can only provide external knowledge (solutions) to cases that have been test cases in former validation sessions, turned out to be a limitation on the practical value of the concept. According to the idea of Case–Based Reasoning (CBR), the so–called *locally–weighted Regression* and, as far as investigated, the way human experience works, we propose a derived version of the so–called *$k$ Nearest–Neighbor* ($k$–NN) data mining method to bring about a decision among the $k$ most similar cases in the case base. The developed concept is subject of an upcoming paper.

**Improvements to VESA** The VESA approach, on the other hand, needs a more general revision. The issues that need to be considered towards a next generation VESA are as follows:

- Computation of a most similar expert

  It turned out to be likely that the computation of a most similar expert results in several experts with the same degree of similarity with respect to their previous responses. It did not happen in the experiment when determining the reply of a VESA to a request for a solution or a rating, but it happened, that the similarities are very close to each other. In this case, we suggest using the expert with the most recent identical behavior to maximize the probability that the latest thinking is employed.

- Continuous validation of VESA

  The authors analyzed the experimentation results to validate VESA's validation knowledge. In fact, this continuous validation of the VESAs should be performed by employing the VESAs in the background at all times when its human counterparts are available. By (a) submitting VESA's solution to the rating process of its human counterpart and (b) comparing VESA's rating with the one of its human counterpart, a VESA can easily be validated and statements about its quality can be derived. A VESA learns from its human counterpart by collecting the solutions and ratings of him/her, even if it is not validated. So its learning capability can be validated by comparing the VESAs outputs with the reply of its human counterpart to the same request for a solution or a rating.

- Completion of VESA towards other than (former) test cases

  The fact that a VESA can only provide validation knowledge (solutions and ratings) to cases that have been used in prior (solving or rating) sessions turned out to be a limitation of the practical value of the concept. The test cases of a current exercise are often different from test cases that have been considered in prior sessions. Following the intention of representing the individual expertise of its human counterpart, the VESA approach needs to be refined by a concept of a "most likely" response of this human source in case there is no "most similar" expert who ever considered an actual case in the past. Again, the $k$–NN data mining method might bring a decision among the $k$ most similar experts.

## 7 Summary

Application fields of intelligent systems are often characterized by having no other source of domain knowledge than human expertise. This source of knowledge, however, is often uncertain, undependable, contradictory, unstable, it changes over time, and furthermore, it is quite expensive. To address this problem, a validation framework has been developed that utilizes the "collective expertise" of an expert panel [KGA02].

However, even this approach does not yet utilize all opportunities to acquire human knowledge. With the objective of also using 'historical knowledge' of previous validation sessions, a Validation Knowledge Base (VKB) has been introduced as a model of the 'collective experience' of expert panels. Primary benefits are more reliable validation results by incorporating external knowledge and and/or a reduced need for current human input, for example smaller expert panels to reach the same quality of validation results. Furthermore, Validation Expert Software Agents (VESA) are introduced as a model of a particular expert's knowledge. Whereas the VKB can be considered (centralized) collective human expertise, a VESA can be considered a (decentralized) autonomous expertise, which is likely to be similar to the expertise of the modeled human counterpart. The VKB is more reliable, but may miss minor, yet possibly excellent human expertise. A VESA, on the other hand, can maintain such minor but possibly excellent human expertise.

A TURING Test experiment with a small prototype system indicates the usefulness of these concepts to model the collective (VKB) and individual (VESA) validation expertise. Generally, the idea of VKB is certainly the appropriate way to establish new sources of knowledge for system validation towards more reliable systems.

## References


[AKT93] Agarwal, R.; Kannan, R.; Tanniru, M.: Formal Validation of a Knowledge–Based System Using a Variation of the Turing Test. Expert Systems with Applications, 6(1993), pp. 181-192.

[Alt98] Althoff, K.-D.; Birk, C.; v. Wangenheim, G.; Tautz, C.: Case–Based Reasoning for Experimental Software Engineering. In Lanz, Bartsch–Spörl, Burkhard, Wess (eds.): Case–Based Reasoning Technology – From Foundations to Applications. Springer Verlag, 1998, pp. 235-254.

[BCR94a] Basili, V.R.; Caldiera, G.; Rombach, H.D.: Experience Factory. In Marciniak (ed.): Encyklopedia of Software Engineering, 1(1994), John Wiley & Sons, pp. 469-476.

[BCR94b] Basili, V.R.; Caldiera, G.; Rombach, H.D.: Goal Question Metric Paradigm. In Marciniak (ed.): Encyklopedia of Software Engineering, 1(1994), John Wiley & Sons, pp. 528-532.

[BKH00] Bultman, A.; Kuipers, J.; van Harmelen, F.: Maintenance of KBS's by Domain Experts: The Holy Grail in Practice, Proc. of the 13th International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems 2000 (IEA/AIE'00), Springer Verlag, Lecture Notes in Artificial Intelligence 1821, 2000.

[Br95] Brown, C.E.; Nielson, N.L.; O'Leary, D.E.; Phillips, M.E.: Validating Heterogeneous and Competing Knowledge Bases Using a Black-box Approach. Expert Systems with Applications, 9(1995): 4, pp. 591-598.

[CB93] Coenen, F. and Bench-Capon, T.: Maintenance of Knowledge–Based Systems. Academic Press, 1993.

[DDL02] Djelouah, R.; Duval, B.; Loiseau, S.: Validation and Reparation on Knowledge Bases. Lecture Notes in Computer Science, Proc. of the 13th International Symposium on Foundations of Intelligent Systems, 2002, ISBN 3-540-43785-1, pp. 312-320.

[Fe84] Feigenbaum, E.A.: Knowledge engineering: The applied side of artificial intelligence. Annals of the New York Academy of Sciences, 1984, pp. 91-107.

[KK03] Kelbassa, H.-W.; Knauf, R.: The Rule Retranslation Problem and the Validation Interface. Proc. of 16th International Florida Artificial Intelligence Research Society Conference (FLAIRS-2003), Menlo Park, CA: AAAI Press, 2003, pp. 213-217.

[Kn00] Knauf, R.: Validating rule-based systems – A complete methodology. Aachen: Shaker, Berichte aus der Informatik, Zugl. Ilmenau, Technische Universität, Habilitationsschrift, ISBN 3-8265-8293-4, 2000.

[KGA02] Knauf, R., Gonzalez, A.J., and Abel T.: A framework for validation of rule-based systems. IEEE transactions of systems, man and cybernetics – Part B: Cybernetics, 32(2002): 3, pp. 281-295.

[Kn02] Knauf, R.; Philippow, I.; Gonzalez, A.J.; Jantke, K.P.; Salecker D.: System Refinement in Practice – Using a formal Method to Modify Real–Life Knowledge. Proc. of 15th International Florida Artificial Intelligence Research Society Conference Society (FLAIRS-2002), Menlo Park, CA: AAAI Press, 2002, pp. 216-230.

[Kn04] Knauf, R., Tsuruta, S., Uehara, K., Onoyama, T. and Kurbad, T.: The power of experience: On the usefulness of validation knowledge. Proc. of 17th International Florida Artificial Intelligence Research Society Conference 2004 (FLAIRS-2004), Menlo Park, CA: AAAI Press, 2004, pp. 337-342.

[Ko93] Kolodner, J.L.: Case-based Reasoning, Morgan Kaufman, CA, 1993.

[LO94] Lee, S.; O'Keefe, R.: Developing a strategy for expert system verification and validation, IEEE Transactions on Systems, Man, and Cybernetics, 24(1994): 4, pp. 643-655.

[Me93] Meseguer, P.: Conventional Software and Expert Systems: Some Comperative Aspects Regarding Validation, Knowledge Oriented Software Design, Elsevier Science Publishers, 1993, pp. 193-204.

[MP94] Meseguer, P.; Plaza, E.: The VALID project: Goals, Development and Results, International Journal of Intelligent Systems, 9(1994): 9, pp. 867-892.

[MV96] Meseguer, P.; Verdaguer, A.: Export System Validation through Knowledge Base refinement, International Journal of Intelligent Systems, 11(1996): 7, pp. 429-462, 1996.

[MM00] Mosqueira-Rey, E.; Moret-Bonillo, V.: Validation of intelligent systems: a critical study and a tool, Expert Systems with Applications, 18(2000), pp. 1-16.

[OO93] O'Keefe, R.M.; O'Leary, D.E.: Expert system verification and validation: A survey and tutorial, Artificial Intelligence Review, 7(1993), pp. 3-42.

[OL01] O'Leary, D.E.: Special issue on verification and validation issues in databases, knowledge-based systems, and ontologies. International Journal on Intelligent Systems, 16(2001): 3, pp. 263-264.

[OL98] O'Leary, D.: Knowledge Acquisition from Multiple Experts: An Empirical Study. Management Science, 44(1998): 8, pp. 1049-1058.

[Pr99] Preece, A.: COVERAGE: Verifying Multiple–Agent Knowledge–Based Systems, Knowledge-Based Systems, 12(1999), pp. 37-44.

[Pr98] Preece, A.; Grossner, C.; Chander, P.G.; Radhakkrishnen, T.: Structure–Based Validation of Rule–Based Systems, Data and Knowledge Engineering, 26(1998), pp. 161-189.

[Ra96] Ram, S.: Design and validation of a knowledge-based system for screening product innovations. IEEE Transactions on Systems, Man and Cybernetics, Part A, 26(1996): 2, ISSN 1083-4427, pp. 213-221.

[ZP94] Zlatareva, N.P.; Preece, A.: State of the Art in Automated Validation of Knowledge-Based Systems, Expert Systems with Applications, 7(1994): 1, pp. 151-168.