In: Aamodt, A., Althoff, K.-D., Magaldi, R. & Milne, R. (1995). *Case-Based Reasoning: A New Force In Advanced Systems Development.* Tutorial, London, April 27, 1995, Unicom Seminars & AI Intelligence, UK (143 pages)

Evaluating Case-Based Reasoning Systems

Klaus-Dieter Althoff
Centre for Learning Systems and Applications (LSA)
Department of Computer Science
University of Kaiserslautern
67653 Kaiserslautern, Germany

Abstract

Evaluation is an important issue for every scientific field and a necessity for an emerging software technology like case-based reasoning. This paper is a supplementation to the review of industrial case-based reasoning tools by K.-D. Althoff, E. Auriol, R. Barletta and M. Manago which describes the most detailed evaluation of commercial case-based reasoning tools currently available. The author focuses on some important aspects that correspond to the evaluation of case-based reasoning systems and gives links to ongoing research.

1 Introduction

Evaluation is an important issue for every scientific field and a necessity for an emerging software technology like case-based reasoning. However, in different contexts the notion of evaluation is used differently. Firstly, evaluation can be seen as evaluating *one* case-based reasoning system, i.e. validating to which degree an application problem has been solved (cf. Grogono, Batareh et al., 1991; Hollnagel, 1989). Secondly, evaluation can be viewed as evaluating *different* case-based reasoning systems, i.e. comparing systems (cf. Rothenberg, 1989; Drenth & Morris, 1992). Thirdly, the notion of evaluation can also be used for evaluating different development *methodologies* for case-based reasoning systems, i.e. comparing system development methodologies (cf. Hilal & Soltan, 1991; Aamodt, 1993b, Aamodt, 1994b). Examples are Components of Expertise (Steels, 1990; Steels, 1993), CommonKads (Wielinga, Van de Velde et al., 1993), Generic Tasks (Chandrasekaran & Johnson, 1993; Allemang, 1994), or Kemras (Hilal & Soltan, 1991).

We would like suggest an integrative view on case-based reasoning system evaluation that integrates several important aspects from currently known views. Our experience is based on several successfully finished or still ongoing case studies (Althoff, 1992a; Althoff, 1992b; Althoff, Auriol et al., 1995a; Althoff, Auriol et al., 1995c; Aamodt & Althoff, 1993; Aamodt, 1994c, Althoff & Aamodt, 1995; Althoff, 1995). The main idea is to combine different kinds of evaluation criteria:

- Domain and application task oriented criteria (e.g., size, theory strength, openness).
  These criteria are used to describe the requirements that arise from a certain domain and application task and, thus, need to be satisfied if an application problem is to be solved (e.g., Althoff, 1992a; Althoff, 1992b; Aamodt 1993b; Benjamins, 1993; Bartsch-Spörl, 1995).
- Technically and ergonomically oriented criteria (e.g., case and knowledge representation, similarity assessment, user acceptance).
  These criteria are used to describe the capabilities of case-based reasoning systems and, thus, allow the comparison of different systems (e.g., Althoff, Auriol et al., 1995a, Rothenberg, 1989).

- Knowledge engineering criteria (e.g., ease of use of methodology, development phases, tools)
  Methodologies are needed in order to increase the efficiency, reliability, maintainability, and modifiability of systems. This, of course, depends to a high degree on the above mentioned criteria (e.g., Aamodt & Althoff, 1993; Althoff & Aamodt, 1995) and led us to our goal of developing case-based reasoning systems based on the relationship of domain/task criteria and technical/ergonomic criteria (cf. also fig. 2).

In the next section we will focus on the combination of such criteria within one framework and on a methodology how to use this framework. In chapter three we will describe two case studies on how to utilise domain criteria using the natural domain of *classification of marine sponges* (Manago, Conruyt & Le Renard, 1992) and the technical domain of *fault diagnosis for CNC[1] machine tools* (Pfeifer & Faupel, 1993; Althoff, Faupel et al., 1989) which both have been used in the review of industrial case-based reasoning tools (Althoff, Auriol et al., 1995a). In section four we shortly report on the feedback this review had on the INRECA Esprit project (cf. next section) which represented the context for it. In section five we summarise some interesting future activities.

2 Framework and Methodology

In the past there have been several events on case-based reasoning where both research prototypes and commercial case-based reasoning tools have been demonstrated. Such events allow the assessment of the underlying technology up to a certain degree of detail. There have also been a number introductory articles on case-based reasoning including those focusing on commercially available tools. However, for the INRECA Esprit project, which deals with the integration of induction and case-based reasoning for decision support and diagnosis, it turned out that the respective level of detail was not sufficient to be of any value for the project. Especially we were not able to achieve a concrete guidance for the design and development of the final integrated INRECA system. Therefore, we decided to carry out an evaluation of case-based reasoning technology that would allow us

- to achieve deeper "insights" for the final integrated INRECA system,
- to find a reasonable and realistic combination of research and application issues, and
- to have a more concrete estimation of what can be expected from case-based reasoning technology in the near and the far future.

While the first aspect is, of course, INRECA-specific the two other aspects are hopefully of more general interest. In addition, there is a fourth aspect that was not originally intended but soon turned out to be of high value:

- to get a guideline how to evaluate commercial case-based reasoning tools.

We started with the set of criteria defined in Traphöner, Manago et al. (1992) and identified the basic capabilities of current case-based reasoning systems (cf. fig. 1), the latter being to a high degree influenced by the work of Aamodt (1993a) and Aamodt and Plaza (1994). What came out is a set of criteria being applied to five industrial case-based reasoning tools (Althoff, Auriol et al., 1994), of which the results have been made available in Althoff, Auriol et al. (1995a). The same set of criteria is currently used to evaluate the final integrated INRECA system (Althoff, Auriol et al., 1995c).

For the evaluation, we selected criteria that produce quantitative or qualitative results, where the latter split into mainly domain-dependent ones as well as domain-independent ones. The quantitative criteria have used to evaluate aspects like noise, incompleteness, performance, correctness, and consistency. Domain-dependent qualitative criteria have been used to gain information on case representation and similarity assessment. We will present two examples in the next sec-

[1] Computerised numerical control

In: Aamodt, A., Althoff, K.-D., Magaldi, R. & Milne, R. (1995). *Case-Based Reasoning: A New Force In Advanced Systems Development.* Tutorial, London, April 27, 1995, Unicom Seminars & AI Intelligence, UK (143 pages)

tion. What we call domain-independent qualitative criteria were used to collect information about the case-based reasoning tools concerning technically and ergonomically oriented issues:

- Technically oriented criteria, e.g.:
  - Case and knowledge representation
  - Organisation of the case library
  - Similarity assessment
  - Handling of noisy and incomplete data
  - Performance
- Ergonomically oriented criteria, e.g.:
  - Control of application development
  - Validating and testing of the application system
  - Acquisition and maintenance of knowledge and data
  - Explain ability and modelling support
  - User acceptance

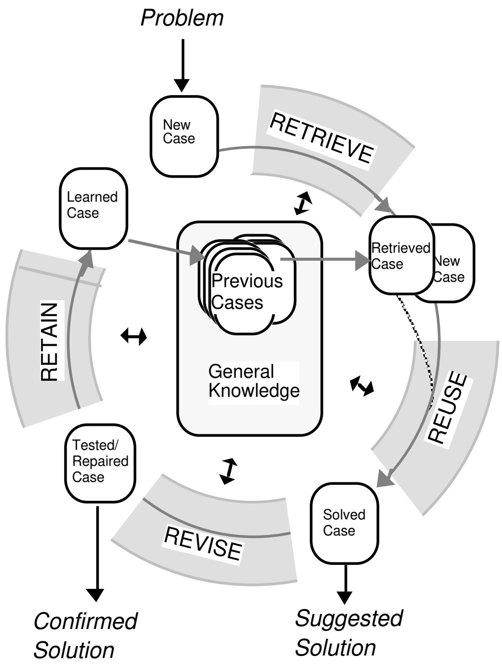

Fig. 1. The case-based reasoning cycle (due to Aamodt, 1993b; Aamodt & Plaza, 1994)

The qualitative criteria are thought to make it easier to interpret the results of the evaluation. While on the first look quantitative criteria appear to be much more precise and, thus, more useful for case-based reasoning system evaluation, the problem is that these criteria measure only isolated aspects of the systems. The guiding principle of Gaschnig, Klahr et al. (1983) for these criteria is that "anything can be measured experimentally as long as it is exactly defined how to take these measurements". However, normally a huge number of measurements have to taken to cover all interesting aspects of case-based reasoning systems. Therefore, Rothenberg rejected such criteria completely (Rothenberg, 1989). We decided in Althoff, Auriol et al. (1995a) to use eight quantitative criteria.

Our set of qualitative domain-independent criteria is motivated by two other evaluation principles of Gaschnig, Klahr et al. (1983):

- Complex objects or processes cannot be evaluated by a single criterion or number;
- The larger the number of distinct criteria evaluated or measurements taken, the more information is available on which to base an overall estimation.

Starting from the huge initial criteria set of Traphöner, Manago et al. (1992), in Althoff, Auriol et al. (1995a) we kept all criteria for which our evaluation team was able to select a reasonable —and in the team completely agreed— distribution of values. However, a small evaluation team cannot cover all interesting issues and aspects. This is also summarised in another evaluation principle of Gaschnig, Klahr et al. (1983):

- In general people will disagree about the relative significance of various criteria according to their respective interests.

We decided to introduce qualitative criteria that are also domain-dependent up to a certain degree, i.e. that reflect a domain and/or application task characteristic. Here the interpretation of results remains to a very high degree on the side of the user, i.e. the reader of the evaluation report. Interesting domain and application task criteria are:

- General criteria, e.g.:
  - Size: Number of cases, classes, attributes, and attribute values
  - Complexity: Number of different relationships
  - Openness: Environmental dependency
  - Theory strength: Certainty of relationships
  - Change over time: Static versus dynamic
- Concept structure, e.g.:
  - Neighbourhood notion available?
  - Target concept fixed?
  - Test selection necessary, i.e. acquisition of additional symptom values?
- Integration of reasoning strategies, e.g.:
  - one or the other
  - side by side
  - interleaved
  - seamless

We will give two examples of domain-dependent qualitative criteria in the next section, namely those that have been applied in Althoff, Auriol et al. (1995a).

Our current view on case-based reasoning system evaluation is that all kinds of criteria, namely technical, ergonomic, domain and application task criteria need to be covered. As shown in fig. 2, these criteria are not only useful for evaluation tasks but can also be utilised for application and tool development or integration tasks. Here especially the relation between domain and application task criteria on the one hand side and technical and ergonomic criteria on the other hand side is very interesting. In practise such knowledge is used for a long time. More recent work tries to directly use it within case-based reasoning system development which allows a system development methodology that heavily combines top-down and bottom-up approaches using both high-level structures of generic descriptive components and example instances of existing case-based reasoning systems. On the one hand side existing systems could be described by relating them to high-level characterisations, on the other hand side high-level characterisations could be made more concrete by relating them to existing systems (Aamodt & Althoff, 1993; Althoff & Aamodt, 1995).

We agree with Rothenberg (1989) that the importance of quantitative criteria is very restricted. However, in Althoff, Auriol et al. (1995a) we gained many insights from carrying out these measurements which gave us important information for the definition of the qualitative criteria as well as the decisions on the respective results. What we recognised can be summarised by contrasting foreground and background characteristics:

In: Aamodt, A., Althoff, K.-D., Magaldi, R. & Milne, R. (1995). *Case-Based Reasoning: A New Force In Advanced Systems Development.* Tutorial, London, April 27, 1995, Unicom Seminars & AI Intelligence, UK (143 pages)

- *Foreground characteristics* include all the aspects that are *directly* concerned with the test data, the experiments, the criteria list, as well as the technical information about the respective procedures of their application.
- *Background characteristics* encompass all the aspects that are only *indirectly* concerned with the data, the experiments, the criteria, and the application procedures, but had important influence on the results: building the evaluation team, selecting the test domains, defining the experiments, defining the set of criteria, as well as all content-related information about conducting the experiments and applying the criteria set.

Both kinds of evaluation characteristics have to be transparent and need to be consistent with the overall goal of the evaluation.

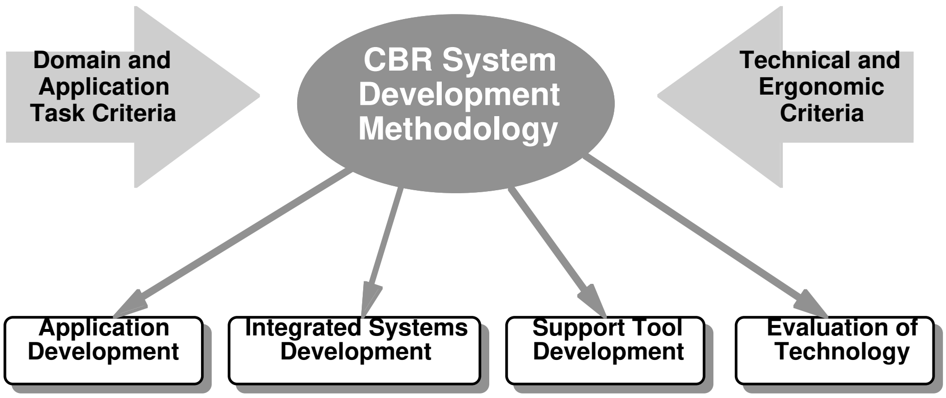

Fig. 2. CBR System Development

3 Case Studies

In this section we describe two different case studies —carried out in the context of Althoff, Auriol et al. (1995a)— that deal with the classification of marine sponges and fault diagnosis of CNC machine tools. The domain-dependent qualitative criteria that have been analysed are case representation and similarity assessment, respectively.

3.1 Analysing a Natural Domain: The Marine Sponge Domain

We tried out to which extent CBR tools are able to model the Marine Sponge domain. This task is not trivial because of the complexity of the sponge domain. Firstly, the sponges are not represented as flat structures but as nested objects. Secondly, this domain needs slots that can store one or several values for one attribute. Especially, most of these multi-valued slots contain complex objects, so that they cannot be represented as lists of primitive types.

In the context of Althoff, Auriol et al. (1995a), the goal of the Marine Sponge domain is improving the expert's view on the descriptive model for sponges in terms of accuracy and exhaustivity, i.e. finding a systematic way to think about describing sponges. The sponge expert needs a support tool for writing descriptions of sponges in a guided way. On course of time, the descriptive model of the expert is combined with the descriptions from several other non-expert people.

Developing a descriptive model of sponges arises all the problems of natural domains in contrast to technical domains like the CNC domain (see next sub section). Among others, there are the following problems:

- The problem of homology, namely that there are no absolute values. For example, the classification of a big sponge and a small sponge could be confused because the micramphidisque of a big sponge is of a similar size as the mesamphidisque of a small sponge.

In: Aamodt, A., Althoff, K.-D., Magaldi, R. & Milne, R. (1995). *Case-Based Reasoning: A New Force In Advanced Systems Development.* Tutorial, London, April 27, 1995, Unicom Seminars & AI Intelligence, UK (143 pages)

- The problem of measurement:
  The work done is pattern-driven interpretation of subjective data. Thus, identification in the first place depends to a high degree on the available apparatus. Looking at the distribution of length of amphidisque, one could, for instance, come up with three different classes. Afterwards improved measurement may allow the discovery of a fourth class. On the other side, it is often not clear whether a new class has to be introduced or a new object.
- Descriptions are always incomplete:
  This leads to the strategy to acquire redundant information to enforce consistency. This also helps to deal with the problem of variability of sponge characteristics.
- All parameters are related to one another:
  One strategy to handle this problem is to make all the available knowledge explicit.

For dealing with these problems, a descriptive model of sponges would be very helpful. From the above problem descriptions, it is also obvious that a descriptive model has to be developed incrementally. Interesting for the expert is to achieve a classification that is as precise as possible. Thus, all available information has to be collected. For achieving sponge database consistency, a good questionnaire has to be provided.

From these domain and application requirements the conclusion is that a fully structured, object-oriented language is necessary, allowing hierarchical descriptions involving complex sub objects and relationships between these sub objects.

3.2 ANALYSING AN ENGINEERING DOMAIN: THE CNC DOMAIN

The CNC domain is difficult to handle because of the huge amount of unknown values. In addition, the overall number of attributes and classes is very high. This domain also illustrates the necessity of having seamlessly integrated systems (Pfeifer & Richter, 1993; Althoff, 1992a+b), even if considered as a more closed domain. However, it is not possible to view it as a static domain (Althoff & Wess, 1991; Althoff, 1992a; de la Ossa, 1992). One reasonable approach here is to use/view cases as exceptions (Althoff & Wess, 1991; Althoff, 1992a).

The CNC domain is a good example for the need of having a well-suited and flexible test selection component that includes, among others, the following features:

- Asking for the next symptom value.
- Accepting any symptom value at any step in the user-system interaction.
- Accepting any number of symptom values.
- Accepting user hypotheses (cf. Stadler & Wess, 1989; Althoff, 1993).

The PATDEX system (Wess, 1993) was originally designed for dealing especially with the CNC domain. Though PATDEX's similarity measure is well-suited for it, in the similarity assessment scenario below it will retrieve the wrong case (CaseIncorrect). However, it is able to improve this situation by its test selection component (asking for symptom values to improve the given information) and/or by the application of causal and/or heuristic rules that allow the derivation of additional symptom values based on already known ones. In addition, PATDEX can use weights for the relative importance of the given attributes for the respective classes (diagnoses). These weights can be automatically updated to avoid misclassifications (Wess, 1993; Richter, 1992). The use of default values for symptoms and its special handling of pathologic/abnormal symptom values also allow an improved treatment of situations similar to that given in the similarity assessment scenario. Recent extensions of PATDEX also allow the use of a deep model of the underlying CNC machine's behaviour to improve the similarity assessment (Pews & Wess, 1993; Bergmann, Pews & Wilke, 1994).

Our experiences with the PATDEX system were the motivation for applying the following similarity assessment scenario in Althoff, Auriol et al.(1995a) to the evaluated case-based reasoning tools. Here we describe the role of PATDEX within this scenario.

There are two reference cases and one query case given from the CNC domain:

In: Aamodt, A., Althoff, K.-D., Magaldi, R. & Milne, R. (1995). *Case-Based Reasoning: A New Force In Advanced Systems Development.* Tutorial, London, April 27, 1995, Unicom Seminars & AI Intelligence, UK (143 pages)

CaseCorrect	**CaseIncorrect**	**Query**
ErrorCode = i59	ErrorCode = i59	ErrorCode = i59
I/OStateOut7 = logical-1	?	?
Valve5Y1 = switched	Valve5Y1 = not-switched	?
I/OStateOut24 = logical-0	?	?
Valve5Y2 = not-switched	?	Valve5Y2 = not-switched
PipesClampingReleaseDevice = ok	?	?
I/OStateIn32 = logical-1	I/OStateIn32 = logical-1	I/OStateIn32 = logical-1
DIAGNOSIS = IOCardIN32i59Defect	DIAGNOSIS = MagneticSwitch5Y1Defect	**?**
$\mathrm{SIM(Query,CaseCorrect)} = \frac{1*3}{1*3+0.5*4} = \underline{\underline{0.6}}$	$\mathrm{SIM(Query,CaseIncorrect)} = \frac{1*2}{1*2+0.5*2} = \underline{\underline{0.67}}$	

Table 1. CaseCorrect, CaseIncorrect, the Query case and the respective similarities

The underlying assumption is that IOCardIn32I59Defect is the correct diagnosis and that the symptom values of CaseCorrect reflect the current fault situation. We assume that the underlying similarity measure is the static similarity function underlying PATDEX-1 (cf. Stadler & Wess, 1989; Althoff, 1993):

$$\mathrm{SIM(Case1,Case2)} = \frac{a*card(E)}{a*card(E)+b*card(D)+c*card(U1)+d*card(U2)}$$

- E set of symptoms with the *same* values for Case1 and Case2
- D set of symptoms with *different* values for Case1 and Case2
- U1 set of symptoms with *known values for Case1* but not for Case2
- U2 set of symptoms with *known values for Case2* but not for Case1
- a=1, b=2, c=1/2, d=1/2 (the default weights used by PATDEX-1)
- card cardinality of sets

Given the above defined Query case, PATDEX comes up with the wrong diagnosis. The reason for this is that PATDEX uses its well-informed similarity measure based on Tversky's contrast model (Tversky, 1977) and explicitly considers unknown symptom values. However, the first wrong judgement (cf. table 1) can be improved (cf. table 2) by the use of general domain knowledge, namely the application of causal or heuristic rules like the following one:

Causal Rule R1:

IF ErrorCode=i59 & Valve5Y2=not-switched THEN I/OStateOut24=logical-0.

In: Aamodt, A., Althoff, K.-D., Magaldi, R. & Milne, R. (1995). *Case-Based Reasoning: A New Force In Advanced Systems Development.* Tutorial, London, April 27, 1995, Unicom Seminars & AI Intelligence, UK (143 pages)

CaseCorrect	**CaseIncorrect**	**Query-2**
ErrorCode = i59	ErrorCode = i59	ErrorCode = i59
I/OStateOut7 = logical-1	?	?
Valve5Y1 = switched	Valve5Y1 = not-switched	?
I/OStateOut24 = logical-0	?	I/OStateOut24 = logical-0
Valve5Y2 = not-switched	?	Valve5Y2 = not-switched
PipesClampingReleaseDevice = ok	?	?
I/OStateIn32 = logical-1	I/OStateIn32 = logical-1	I/OStateIn32 = logical-1
DIAGNOSIS = IOCardIN32i59Defect	DIAGNOSIS = MagneticSwitch5Y1Defect	**?**
$SIM(Query\text{-}2,CaseCorrect) = \frac{1*4}{1*4+0.5*3} = \underline{\underline{0.73}}$	$SIM(Query\text{-}2,CaseIncorrect) = \frac{1*2}{1*2+0.5*3} = \underline{\underline{0.57}}$	

Table 2. Situation after applying rule R1

An alternative improvement strategy (results in table 3) is the acquisition of further information using a complete test selection sub component:

Applying test selection to extend and thereby improve the available information:
Test 5Y1: "What is the status of Valve5Y1?" <— Valve5Y1=switched.

CaseCorrect	**CaseIncorrect**	**Query-3**
ErrorCode = i59	ErrorCode = i59	ErrorCode = i59
I/OStateOut7 = logical-1	?	?
Valve5Y1 = switched	Valve5Y1 = not-switched	Valve5Y1 = switched
I/OStateOut24 = logical-0	?	?
Valve5Y2 = not-switched	?	Valve5Y2 = not-switched
PipesClampingReleaseDevice = ok	?	?
I/OStateIn32 = logical-1	I/OStateIn32 = logical-1	I/OStateIn32 = logical-1
DIAGNOSIS = IOCardIN32i59Defect	DIAGNOSIS = MagneticSwitch5Y1Defect	**?**
$SIM(Query\text{-}3,CaseCorrect) = \frac{1*4}{1*4+0.5*3} = \underline{\underline{0.73}}$	$SIM(Query\text{-}3,CaseIncorrect) = \frac{1*2}{1*2+2*1+0.5*1} = \underline{\underline{0.44}}$	

Table 3. Situation after ascertaining status of Valve5Y2

Important problems that need to be handled in the CNC domain are system-guided and/or user-guided completion of available information, the handling of a huge number of attributes, classes, cases and unknowns. An attribute-value based case representation using symbolic value ranges only is sufficient.

From these domain and application requirements the conclusion is that value-based indexing is possible and necessary to allow sufficiently efficient case retrieval. An explicit dealing with the high number of unknowns requires a Tversky-like similarity measure. In addition, general domain knowledge is necessary for supporting both the test selection and the similarity assessment component.

4 Results

The evaluation of industrial case-based reasoning tools (Althoff, Auriol et al., 1995a) had a positive feedback for the INRECA project. In several aspects the INRECA system could be improved. It also showed where the INRECA system contributes to the current state of case-based reasoning technology (Althoff, Auriol et al., 1994). One important issue here is the identification of four levels of integration between induction and case-based reasoning within one system architecture (cf. fig. 3), especially the seamless integration (cf. fig. 4) of these two techniques (Auriol, Manago et al., 1994; Wess, 1995; Althoff, Bergmann et al., 1995).

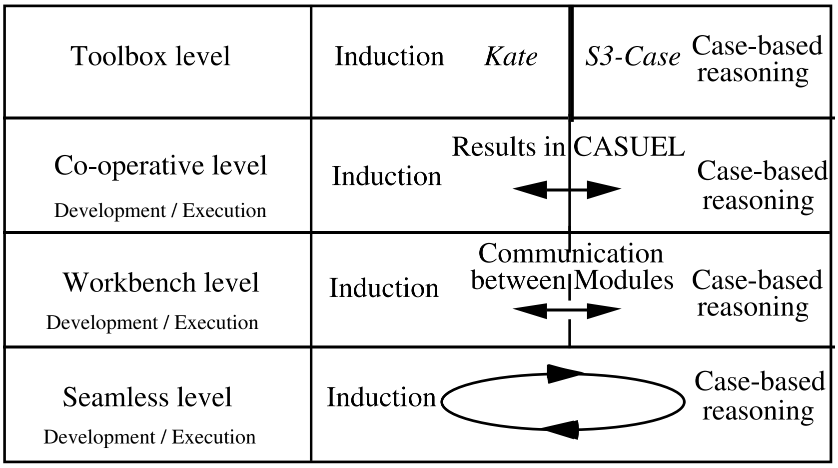

Fig. 3. Levels of Integration in INRECA

The INRECA tree (an extension and modification of a *k*-d tree), with its associated algorithms, can be configured to be a pure decision tree or a pure indexing tree. A learning mechanism automatically extracts knowledge about the similarity of cases from a decision-tree-like INRECA tree and embeds such knowledge into the underlying similarity measure. By this learning method (that is comparable to PATDEX's adaptation method for its similarity measure, cf. Wess, 1993; Richter, 1992; Althoff, Wess et al., 1994), INRECA's k-dimensional indexing tree gradually evolves into a decision tree. Thus, on the seamless level INRECA offers a system evolution over time from pure case-based reasoning to pure induction based on a concept learning strategy (cf. Fig. 4). The theoretical background for this process is described in Wess & Globig (1994a), Wess & Globig (1994b), and Wess (1995).

In: Aamodt, A., Althoff, K.-D., Magaldi, R. & Milne, R. (1995). *Case-Based Reasoning: A New Force In Advanced Systems Development.* Tutorial, London, April 27, 1995, Unicom Seminars & AI Intelligence, UK (143 pages)

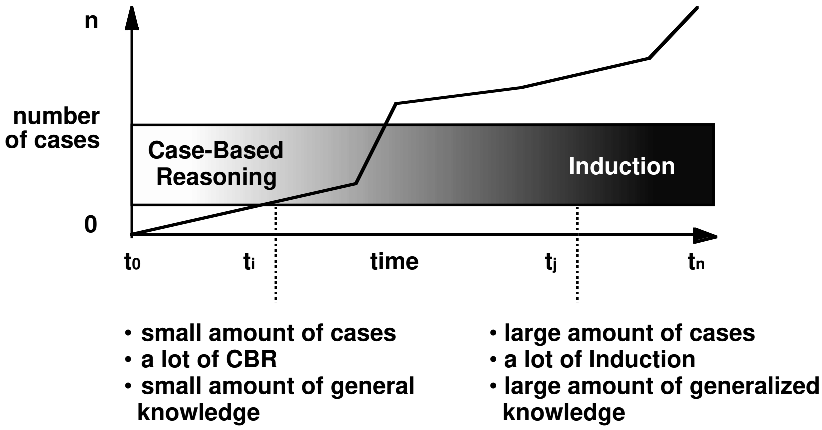

Fig. 4. Seamless Integration in INRECA

For the CAR domain (cf. Althoff, Auriol et al., 1995a) Wess (1995) also shows that the seamlessly integrated INRECA system has significantly improved its performance (cf. fig. 5). Auriol, Wess et al. (1995) also show this for an additional domain.

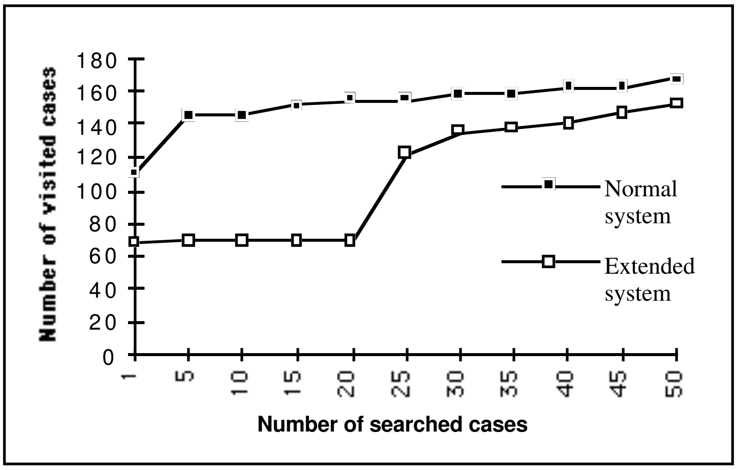

Fig. 5. Improving similarity assessment through learning (Wess, 1995)

5 Future Activities

We currently apply the set of criteria used in Althoff, Auriol et al. (1995a) to the final integrated INRECA system (Althoff, Auriol et al., 1995c) and hope to achieve valuable feedback on the usefulness of these criteria. In addition, we work on the further development of the set of domain and task criteria (Althoff & Aamodt, 1995; Weis, 1995; Althoff, 1995). This includes a thorough analysis of a set of existing case-based reasoning methods based on a defined evaluation framework, a methodology for analysing case-based reasoning methods with respect to their assumptions about application tasks and domain knowledge characteristics, and the development of a set of generic integrated case-based problem solving and learning architectures based on the above analysis (Althoff & Aamodt, 1995).

In: Aamodt, A., Althoff, K.-D., Magaldi, R. & Milne, R. (1995). *Case-Based Reasoning: A New Force In Advanced Systems Development.* Tutorial, London, April 27, 1995, Unicom Seminars & AI Intelligence, UK (143 pages)

6 Acknowledgement

The presented work is based on a cooperation with *Agnar Aamodt* whom I would like to thank for many of the underlying ideas that have been developed in intensive discussions with him or that have their roots in Agnar's scientific work. Thanks go also to *Stefan Wess* who contributed to my work with his competent and practical view on case-based reasoning.

The author is indebted to *Harald Holz*, *Carsten Priebisch*, *Alexandre Meissonnier*, and *Wolfgang Wilke* who heavily supported me in putting the evaluation criteria to use as reported in Althoff, Auriol et al. (1995a), and again Carsten and Alexandre who are still doing so. Thanks also to *Karl-Heinz Weis jun.* for his untiring work on domain and task criteria and all related aspects.

Funding for this work has been provided within the INRECA project by the Commission of the European Communities (Esprit Contract 6322) to which the author is greatly indebted. The partners of INRECA are AcknoSoft (Paris, prime contractor), tecInno (Kaiserslautern), IMS (Dublin), and the University of Kaiserslautern.

7 References

Aamodt A (1993a). *An Overview on CBR*. Tutorial at the First European Workshop on Case-Based Reasoning (EWCBR-93), Kaiserslautern

Aamodt A (1993b). *Knowledge Level Modelling and the Symbol Level Gap*. Slide copies of an invited presentation during the Knowledge Level Seminar at the University of Trondheim

Aamodt A (1994a). Explanation-Based Case-Based Rasoning in *Topics in Case-Based Reasoning* (Eds Wess, S, and Althoff, K-D, and Richter, M M) pp. 274-288, Springer Verlag

Aamodt A (1994b). Knowledge Acquisition and Learning by Experience – The Role of Case-Specific Knowledge in *Integration of Knowledge Acquisition and Machine Learning* (Eds Kodratoff, Y, and Tecuci, G), Kluwer, forthcoming

Aamodt A (1994c). Case-Based Reasoning in *Proc. MLnet Summer School on Machine Learning and Knowledge Acquisition* (Ed Rouveirol, C) pp. 1-58, Dourdan

Aamodt A and Althoff K-D (1993). Problem Solving and Sustained Learning from Experience: Analyzing Methods with respect to Domain Characteristics in *Proc. MLnet Workshop on "Learning and Problem Solving"* (Ed van Someren, M), Blanes

Aamodt A and Plaza E (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications, March 1994*

Aha D W (1994). Preface. *Proc. AAAI Workshop on "Case-Based Reasoning"* (Eds Aha, D W) pp. iv, Seattle

Allemang D (1994). Combining Case-based Reasoning and Task Specific Architectures. *IEEE Expert*

Althoff K-D (1992a). *Eine fallbasierte Lernkomponente als integrierter Bestandteil der MOLTKE-Werkbank zur Diagnose technischer Systeme*. Doctoral Dissertation, University of Kaiserslautern; also: Sankt Augustin, Germany: infix Verlag

Althoff K-D (1992b). Machine Learning and Knowledge Acquisition in a Computational Architecture for Fault Diagnosis in Engineering Systems. *Proc. ML92 Workshop on "Computational Architectures for Supporting Machine Learning and Knowledge Acquisition"* (Ed Weintraub, M), Aberdeen

Althoff K-D (1993). Lernverfahren in MOLTKE in *Diagnose von technischen Systemen– Grundlagen, Methoden und Perspektiven der Fehlerdiagnose–* (Eds Pfeifer, T, and Richter, M M) pp. 173-200, Deutscher Universitäts-Verlag

In: Aamodt, A., Althoff, K.-D., Magaldi, R. & Milne, R. (1995). *Case-Based Reasoning: A New Force In Advanced Systems Development.* Tutorial, London, April 27, 1995, Unicom Seminars & AI Intelligence, UK (143 pages)

Althoff K-D (1995). *Description, Analysis, and Development of Integrated Problem Solving and Learning Architectures for Decision Support and Diagnostic Reasoning*. forthcoming

Althoff K-D and Aamodt A (1995). Relating Problem Solving and Learning Methods to Task and Domain Characteristics: An Analytical Framework. *forthcoming*

Althoff K-D and Auriol E and Holz H and Manago M and Meissonnier A and Priebisch C and Wess S and Wike W (1994). *An Evaluation of Commercial Case-Based Reasoning Tools.* INRECA (Esprit P6322), Deliverable D5

Althoff K-D and Auriol E and Barletta R and Manago M (1995a). *A Review of Industrial Case-Based Reasoning Tools*. AI Intelligence, Oxford, UK

Althoff K-D and Auriol E and Bergmann R and Breen S and Dittrich S and Johnston R and Manago M and Traphöner R and Wess S (1995b). Case-Based Reasoning for Decision Support and Diagnostic Problem Solving: The INRECA Approach in *Proc. 3rd German Workshop on Case-Based Reasoning* (Eds Bartsch-Spörl, B, and Janetzko, D, and Wess, S) pp. 63-72, LSA Report LSA-95-02, Centre for Learning Systems and Applications, University of Kaiserslautern

Althoff K-D and Auriol E and Bergmann R and Breen S and Manago M and Meissonnier A and Priebisch C and Traphöner R and Weis K-H and Wess S and Wilke W (1995c). *An Evaluation of the Final Integrated INRECA System*. INRECA (Esprit P6322), Deliverable D6, forthcoming

Althoff K-D and Bergmann R and Globig C and Wess S and Auriol E and Manago M (1995). On the Seamless Integration of Induction and Case-Based Reasoning for Decision Support and Diagnosis. *Submitted*

Althoff K-D and Faupel B and Kockskämper S and Traphöner R and Wernicke W (1989). Knowledge Acquisition in the Domain of CNC Machining Centers: the MOLTKE Approach in *Proc. EKAW-1989* (Eds Boose, J, and Gaines, B, and Ganascia, J-G), (1989) pp. 180-195

Althoff K-D and Wess S (1991). Case-Based Knowledge Acquisition, Learning, and Problem Solving in Diagnostic Real World Tasks in *Proc. EKAW-1991* (Ed Smeed, D), Glasgow and Crieff

Althoff K-D and Wess S and Bergmann R and Maurer F and Manago M and Auriol E and Conruyt N and Traphöner R and Bräuer M and Dittrich S (1994). Induction and Case-Based Reasoning for Classification Tasks in *Information Systems and Data Analysis — Prospects-Foundations-Applications—* (Eds Bock, H H, and Lenski, W, and Richter, M M) pp. 3-16, Springer Verlag

Auriol E and Manago M and Althoff K-D and Wess S and Dittrich S (1994). Integrating Induction and Case-Based Reasoning: Methodological Approach and First Evaluations in *Proc. Second European Workshop on Case-Based Reasoning (EWCBR-94*) (Eds Keane, M, and Haton , J P, and Manago, M) pp. 145-156, AcknoSoft Press, Paris

Bartsch-Spörl B (1995). KI-Methoden für innovative Design-Domänen in *Expertensysteme 95* (Eds Richter, M M, and Maurer, F) pp. 137-151, infix Verlag

Benjamins R (1993). *Problem Solving Methods for Diagnosis*. Ph.D. thesis, University of Amsterdam

Bergmann R and Pews G and Wilke W (1994). Explanation-Based Similarity: A Unifying Approach for Integrating Domain Knowledge into Case-Based Reasoning for Diagnosis and Planning Tasks in *Topics in Case-Based Reasoning* (Eds Wess, S, and Althoff, K-D, and Richter, M M) pp. 182-196, Springer Verlag

Chandrasekaran B and Johnson T (1993). Generic Tasks and Task Structures: History, Critique and New Directions in *Second Generation Expert Systems* (Eds David, J-M, and Krivine, J-P, and Simmons, R) pp. 232-272, Springer Verlag

De la Ossa A (1992). *Knowledge Adaptation: Analogical Transfer of Empirical Knowledge Across Domains for Case-Based Problem Solving*. Doctoral Dissertation, University of Kaiserslautern

Drenth H and Morris A (1992). Prototyping expert solutions: an evaluation of Crystal, Leonardo, GURU and ART-IM. *Expert Systems* 9, 1, 35-45

Gaschnig J and Klahr P and Pople H and Shortliffe E and Terry A (1983). Evaluation of Expert Systems: Issues and Case Studies in *Building Expert Systems* (Eds Hayes-Roth, F, Waterman, D A, and Lenat, D B) pp. 241-282, Addison-Wesley

Graner N and Sleeman D (1993). A Multistrategy Knowledge Refinement and Acquisition Toolbox in *ECML-93 workshop on "Integrated Learning Architectures"* (Ed Plaza, E), 14 pages, Vienna

Grogono P and Batarekh A and Preece A and Shinghal R and Suen C (1991). Expert system evaluation techniques: a selected bibliography. *Expert Systems 8*, 4, 227-239

Hilal D K and Soltan H (1991). A suggested descriptive framework for the comparison of knowledge-based systems methodologies. *Expert Systems 8*, 2, 107-114

Hollnagel E (1989). Evaluation of expert systems in *Topics in Expert System Design* (Eds Guida, G, and Tasso, C) pp. 377-418, North Holland

Manago M and Althoff K-D and Auriol E and Traphöner R and Wess S and Conruyt N and Maurer F (1993). Induction and reasoning from cases in *Proc. First European Workshop on Case-Based Reasoning (EWCBR-93)* (Eds Richter, M M, and Wess, S, and Althoff, K-D, and Maurer, F) pp. 313-318, SEKI-REPORT SR-93-12, University of Kaiserslautern

Manago M and Conruyt N and Le Renard J (1992). Acquiring Descriptive Knowledge for Classification and Identification in *Current Developments in Knowledge Acquisition - EKAW ´92* (Eds Wetter, Th, and Althoff, K-D, and Boose, J, and Gaines, B, and Linster, M, and Schmalhofer, F) pp. 392-405, Springer Verlag

Pews G and Wess S (1993). Combining Case-Based and Model-Based Approaches for Diagnostic Applications in Technical Domains in *Proc. First European Workshop on Case-Based Reasoning (EWCBR-93)*(Eds Richter, M M, and Wess, S, and Althoff, K-D, and Maurer, F) pp. 325-328, SEKI-REPORT SR-93-12, University of Kaiserslautern

Pfeifer T and Faupel B (1993). Die Anwendung von MOLTKE: Diagnose von CNC-Bearbeitungszentren in *Diagnose von technischen Systemen–Grundlagen, Methoden und Perspektiven der Fehlerdiagnose–* (Eds Pfeifer, T, and Richter, M M) pp. 42-67, Deutscher Universitäts-Verlag

Pfeifer T and Richter M M (1993). Einleitung in *Diagnose von technischen Systemen–Grundlagen, Methoden und Perspektiven der Fehlerdiagnose–* (Eds Pfeifer, T, and Richter, M M) pp. 1-7, Deutscher Universitäts-Verlag

Richter M M (1992). Classification and Learning of Similarity Measures in *Proc. of the 16th Annual Meeting of the German Society for Classification*, Springer Verlag

Rothenberg J (1989). Expert system tool evaluation in *Topics in Expert System Design* (Eds Guida, G, and Tasso, C) pp. 205-229, North Holland

Stadler M and Wess S (1989). *Konzept und Implementierung eines fallbasierten, analogieori-entierten Inferenzmechanismus und dessen Integration in ein regelbasiertes Expertensystem zur Diagnose eines CNC-Bearbeitungszentrums*. Project Thesis, University of Kaiserslautern

Steels L (1990). Components of expertise. *AI Magazine 11*, 2, 29-49

Steels L (1993). The componential framework and its role in reusability in *Second Generation Expert Systems* (Eds David, J-M, and Krivine, J-P, and Simmons, R) pp. 273-298, Springer Verlag

In: Aamodt, A., Althoff, K.-D., Magaldi, R. & Milne, R. (1995). *Case-Based Reasoning: A New Force In Advanced Systems Development.* Tutorial, London, April 27, 1995, Unicom Seminars & AI Intelligence, UK (143 pages)

Traphöner R and Manago M and Conruyt N and Dittrich S (1992). *Industrial Criteria for Comparing the Technologies of INRECA*. INRECA (Esprit P6322), Deliverable D4

Tversky A (1977). Features of Similarity. *Psychological Review 84*, 327-352

Weis K-H jun (1995). *An integrated approach for case-based reasoning*. Diploma Thesis, University of Kaiserslautern, forthcoming

Wess S (1993). PATDEX - ein Ansatz zur wissensbasierten und inkrementellen Verbesserung von Ähnlichkeitsbewertungen in der fallbasierten Diagnostik in *Expertensysteme 93, Proc. of the 2nd German Conference on Expert Systems* (Eds Puppe, F, and Günter, A) pp. 42-55, Springer Verlag

Wess S (1995). *Fallbasiertes Schließen in wissensbasierten Systemen zur Entscheidungsunterstützung und Diagnose*. Doctoral Dissertation, University of Kaiserslautern, *submitted*

Wess S and Globig C (1994a). A Case Study on Case-Based and Symbolic Learning. *Proc. AAAI Workshop on "Case-Based Reasoning"* (Eds Aha, D W) pp. 43-47, Seattle

Wess S and Globig C (1994b). Case-Based and Symbolic Classification Algorithms - A Case Study Using Version Space in *Topics in Case-Based Reasoning* (Eds Wess, S, and Althoff, K-D, and Richter, M M) pp. 77-91, Springer Verlag

Wielinga B J and Van de Velde W and Schreiber G and Akkermans H (1993). Towards a unification of knowledge modeling approaches in *Second Generation Expert Systems* (Eds David, J-M, and Krivine, J-P, and Simmons, R) pp. 299-335, Springer Verlag