

# COMPARATIVE ANALYSIS OF CONTEMPORARY CASE BASED REASONING SOFTWARE FRAMEWORKS

Assoc. Prof. Atanasov A.

Department of Computer Science – University of Chemical Technology and Metallurgy, Bulgaria  
naso@uctm.edu

**Abstract** The paper presents the comparison of two Case-Based Reasoning (CBR) oriented software frameworks myCBR3 Workbench and CBR-Works ver. 4.3.0 for the development of predictive diagnosis and maintenance systems. Those frameworks were selected after detailed preliminary comparisons of previous versions of myCBR presented in [3], as well of the investigations of the capabilities of other popular CBR software systems [2]. The evaluation of myCBR and CBR-Works includes the capacity to support the:  $R^4$  CBR circle; clusterization of cases, variety of used similarity functions, etc. Specific abilities to provide GUI, database support, required knowledge to work with the systems were also considered.

**Keywords:** CASE-BASED REASONING, PREDICTIVE DIAGNOSIS, SOFTWARE COMPARATIVE ANALYSIS

## 1. Introduction

The methodology of Case-Based Reasoning (CBR) for solving problems or tasks, based on already known past decisions of same problems, is predicated on the hypothesis that similar problems have similar solutions [1]. CBR is a part of artificial intelligence, connected to automation of reasoning based on cases, current problem formulation, searching of one or more solved in the past similar cases and adaptation of old solutions to the current problem.

In this paper a deep comparative analysis of software frameworks myCBR 3 and CBR Works is made. The main purpose of this analysis is to determine whether these software frameworks are applicable for the development of CBR applications in the area of predictive diagnostics and maintenance.

The frameworks myCBR 3 and CBR Works are chosen on base of detailed previous research, published in [2,3] of most distributed and used in practice commercial and non-commercial software tools for CBR, such as CBR Works [6,8], myCBR [7], COLIBRI Studio [11], FreeCBR, CAT-CBR [5] etc.

### 1.1 Description of CBR reasoning mechanism

The stages of reasoning in CBR systems, based on cases, are known as classical R4 cycle. Cases are the main objects in CBR systems. Each case represents expert knowledge of a solved problem in a specific domain. The case can be represented either as a free text, containing a list of question and answers, or as a structural type when the case is represented as a record or object in the reasoning data base (Case Base).

All structural cases are described as a pair of problem-solution [1]. The problem  $\mathbf{p}_i = (\mathbf{a}_i, \mathbf{v}_i)$  is organized as a structure of attributes and values, described by the attribute vector  $\mathbf{a}_i = (\mathbf{a}_{i1}, \mathbf{a}_{i2}, \dots, \mathbf{a}_{in})$  and the value vector  $\mathbf{v}_i = (\mathbf{v}_{i1}, \mathbf{v}_{i2}, \dots, \mathbf{v}_{in})$ .

The solution  $\mathbf{s}_i$  can be represented in the same way by number of solution attributes and their values. In some cases the solution can be presented as vectors, defined by the specific tasks. For example in multidimensional supervised control tasks, the decision includes two vectors  $\mathbf{s}_i = (\mathbf{sp}_i, \mathbf{pr}_i)$ , where the first vector  $\mathbf{sp}_i = (\mathbf{sp}_{i1}, \mathbf{sp}_{i2}, \dots, \mathbf{sp}_{iq})$  consists of controllers sets on first hierarchical level, and the second  $\mathbf{pr}_i = (\mathbf{pr}_{i1}, \mathbf{pr}_{i2}, \dots, \mathbf{pr}_{im})$  – values of the target parameters, corresponding to the sets.

For solving an actual problem, the following 4 main tasks of CBR R4 cycle are iteratively performed (Fig. 1) [1, 2, 3]:

**Retrieve** – process of extraction of one (nearest neighbor) or a group of cases (k-nearest neighbors) having closest definition to the current problem. The global similarity between the problems of these cases (the new  $\mathbf{p}_{new}$  and the one in the case base  $\mathbf{p}_j$ ) is presented by following expression:

$$\text{sim}(\mathbf{p}_{new}, \mathbf{p}_j) = \sum_{i=1}^n w_i \text{sim}_i(\mathbf{p}_{new_i}, \mathbf{p}_{ji}), \text{ and } \sum_{i=1}^n w_i = 1$$

Where  $w_i$  is the weight of i-th attribute  $0 \leq w_i \leq 1$  and  $\text{sim}(\mathbf{p}_{new_i}, \mathbf{p}_{ji})$  is the local similarity between i-th attributes of the cases.

For global similarity measure the following metrics are most used: weighted Euclidian distance, Manhattan's metric, Humming's metric, Tversky's metric, Tchebishev's metric, and minimum or maximum metrics etc.

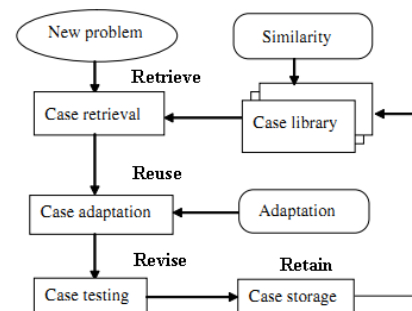


Fig.1 Classical Case-Based Reasoning R4 circle

**Reuse** – reuse of the solutions of chosen in the first step one or k-nearest neighbors.

- When only one nearest case is chosen, the solution of the new problem  $\mathbf{s}_{new}$  will be the solution of the chosen case  $\mathbf{s}_{NN}$ .

- When k-nearest neighbors are chosen, the solution of the new case is calculated on the base of adaptation of k-nearest neighbors' solutions or a part of them.

**Revise** – validate new solution made on the Reuse stage. This validation and test is mostly done by an expert or it is made based on simulation researches if there is a mathematical model available.

**Retain** – saving (retaining) the new solution (learned case) in the Case Base for future use, if it is successful

## 2. Comparative analysis of myCBR 3 Workbench and CBR Works platforms

The analysis of mentioned software frameworks is aimed to compare their capabilities for implementation of real-time and diagnostic CBR systems and takes into account following features:

- Type and level and availability of software documentation;
- Required knowledge to use the software;
- Software support, updates and new versions;
- Presence of graphical User Interface (GUI);
- Internal description of the cases (problems and solutions);
- Type of similarity functions (SF);
- Support of all parts of R4 of the CBR circle;
- Support of different case-base DB (data bases);
- Clusterization and indexation of the cases;
- Visualization of the cases;
- Application of software in specific scientific areas
- Support of interfaces to diagnostic or real-time systems;
- Support of fat and thin clients

**MyCBR 3 Workbench** [7] is one of the most popular CBR software applications. It is an open source platform written in Java, accessible to all users as SDK. MyCBR 3 Workbench is a

standalone application with certain capabilities and limitations. It is a successor of myCBR 2.6.6 plug-in of open source ontology editor Protégé [9]. MyCBR 3 is developed by the German Research Center for Artificial Intelligence DFKI [7]. Using SDK (Software Development Kit) to support thin (Web-based) or fat (standalone) CBR applications. Its programming code is well documented. The purpose of myCBR 3 is to minimize the efforts to create CBR applications. For its normal use, without modifying the source code, no programming skills are required, but expertise in a specific CBR-domain. Work with the platform and its features are described in the tutorial, available at [7]. At present time (May 2015) new versions 3.1 and 3 of myCBR are available for most popular operating systems.

The framework myCBR 3 supports description of cases with various attributes: numeric, symbol and string, logical, taxonomy type, etc. The templates of the cases are generated as concepts with number of attributes (Fig. 2.). It is possible one attribute itself to be a new concept with its own attributes. This way an object-oriented structure of the case can be implemented.

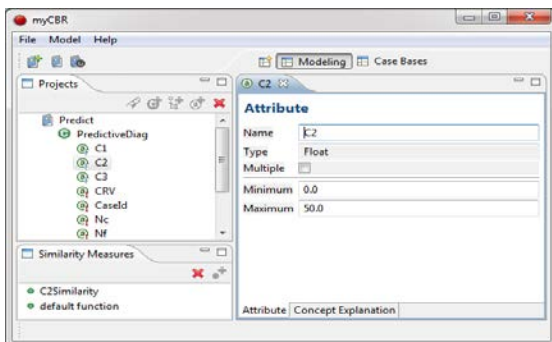


Fig. 2 Definition of concept and its attributes

In my CBR 3 the concept is a part of the CBR model which contains case attributes, global and local similarity functions and a textual concept description. It contains as well the weight of each attribute and information whether it is discriminant or not (part of the problem or part of the solution). The weight determines the significance of the attribute in relation to others. Attributes with a weight of zero (0) or discriminant value of false are not considered when searching the case-base DB.

The CBR cases are objects of the concept-class described by its attributes' values.

In myCBR 3 are given the opportunities to select the similarity functions (SF) on concept level (global SF) and to edit it on attribute level (local SF). At the concept level the SF are: Weighted sum, Euclidean difference, Maximum or Minimum distance. On attribute level the SF can be modified through the GUI, as shown below in the middle of Fig. 3 and they can be symmetrical, asymmetrical, step-type or smooth step-type, linear or polynomial. Special symbol (Fig 4) and symbol taxonomy (Fig 5) SF are available for attributes of non-numeric types.

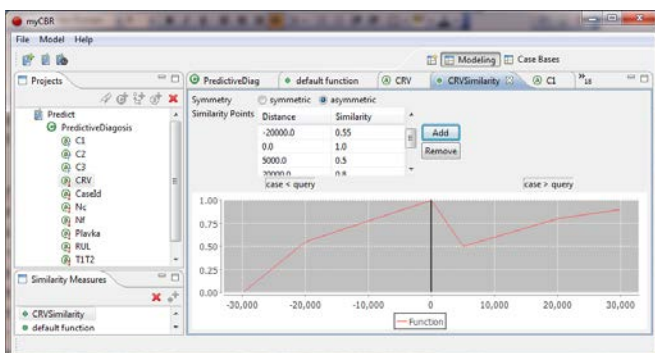


Fig. 3. Editing the local similarity function via GUI

In myCBR 3 the cases and their attributes can be created manually or automatically. The automatic generation of attributes is done during the import procedure from the Comma Separated Value /CSV/ file. Then to each column name from the CSV file is

assigned an attribute with the same name (See Table 1). For each row of the CSV file a new object (instance of the concept) is created in the Case-Base DB. Unfortunately, when cases are imported from CSV file the maximum-minimum range of variance of each attribute is calculated automatically on the basis of the maximum and minimum values for this attribute in all available cases, so the retrieval of solution for a new case with attribute out of this range is impossible. In this case manual modifications of mentioned ranges are required.

	Blue	LightGreen	Red	Black	Green
Blue	1.0	0.0	0.0	0.0	0.3
LightGreen	0.0	1.0	0.0	0.0	0.7
Red	0.0	0.0	1.0	0.0	0.0
Black	0.0	0.0	0.0	1.0	0.0
Green	0.3	0.7	0.0	0.0	1.0

Fig. 4. Applying Symbol attribute similarity function via GUI

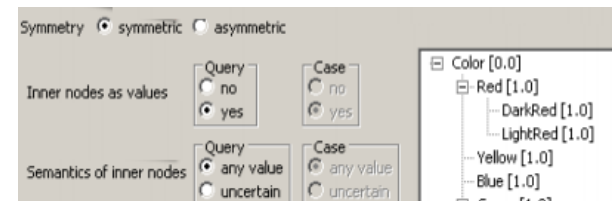


Fig. 5. Applying taxonomy similarity function via GUI

When the instance of the case is created manually several steps have to be undertaken. First a Case Base have to be created by selection of the tab Case Bases in the upper right part of the myCBR 3 application

. Then to select Case Base **Case Bases** tab on the bottom left window; to press icon Add Case Base and to provide name of Case Base DB for case-instances. Next step is to add (create) an instance by selecting tab Instances and pressing icon (a little bit confusing) Add Case Base. A new instance with name corresponding to the concept is created in left bottom window. With left mouse click on the created instance a form for adding case values is opened. In contrast to the forms in myCBR 2.6.6., here the attributes are sorted alphabetically which makes problems, if you have to take them, for example, from excel table.

With regard to maintenance of CBR 4R cycle phases, myCBR 3 supports only Retrieve and Retain. During the Retrieve phase all cases are extracted. They are presented sorted by degree of similarity based on the chosen global SF. The Query to the case-base DB could be done on the basis of all or part of the attributes, describing the new case.

Again, in contrast to myCBR 2.6.6 here on Retain phase it is impossible to save the Query as a new case or to use an old case as a basis for new Query.

MyCBR 3 does not work with external DB. It stores the cases in project file with unknown format. Authors of myCBR 3 claim that the concept template with attributes and local SF can be exported to other CBR platforms (COLIBRI Studio) but information how to do that is not available in the tutorial slides. MyCBR 3 can not support the case indexation and clusterization. The cases can not be graphically presented in the GUI. No interfaces to external systems and DB are available in myCBR3. It is valid regarding the interfaces to real-time or diagnostic systems.

**CBR-Works 4.3.0** [12] is a standalone application - shell intended to various CBR applications. There is no information about the programming language used for its development, but because it is using a kind of virtual machine it can be concluded that it is mostly Java-based. It is a product of Empolis Knowledge Management GmbH., Kaiserslautern, Germany, supported by ESPRIT project INRECA and Humboldt and Kaiserslautern Universities. The framework is built to support the full R4 CBR cycle and all related, processes as modeling the cases, attributes, SF and rules related to case attributes; support of its own Case Base, as well ODBC connections to databases from which it can import collected knowledge (cases from DB-tables); retrieval and filtering

the cases; providing server capabilities to thin Web-based client, etc. CBR-Works is GUI based application providing four managers to support all CBR knowledge modeling activities-Concept Manager, Type Manager, Case Explorer and Case Navigator.

The Concept Manager can be used to create concept model of the cases. It supports standard attributes (right part of Fig. 6) and object-oriented structures as hierarchical sub-concept trees (left part of Fig. 6). Each concept has its own description and annotation provided by tab Properties, global similarity (Average, Max, Min, Euclidian or Custom defined) and related rules (Fig. 7 - used for adaptation of the case in revised and reuse stages), as well questions regarding every attribute.

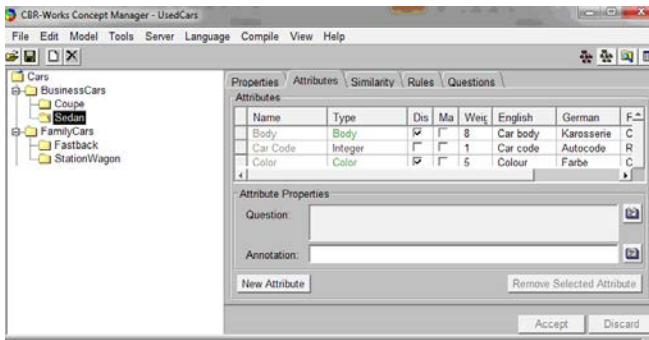


Fig. 6. Concept Manager View

Each concept attribute is described by name, type, weight discriminant and mandatory fields, as well its annotation and questions related to it. The mandatory checkbox is used to define whether the attribute can be used in confirmed cases. In CBR-Works each case can be in one of the four states: - unconfirmed (new case - not used for retrieval before testing); confirmed (used for retrieval); protected (its attributes cannot be modified) and obsolete (old case- not used for retrieval).

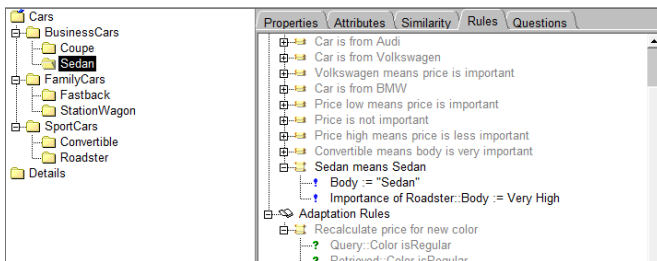


Fig. 7. Rules specified for attributes of concept Business Car and Sedan

The Type Manager is responsible to define the types of the user-defined attributes' and their similarities. Standard numerical and symbol and string types are supported, together with used defined interval or subset types. User defined taxonomy types, ranges and SF are easily edited via Type manager ( Fig 8).

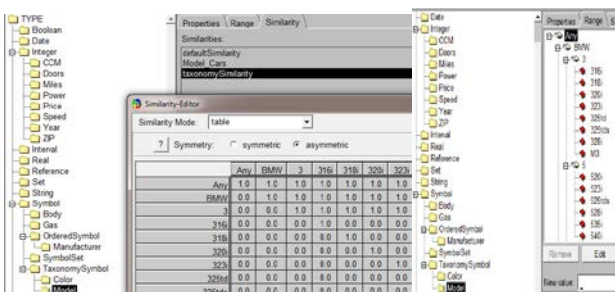


Fig. 8. Editing taxonomy similarity via Type Manager

The Explorer Manager can be used to create instances (Fig. 9) of the cases and to add them to the case base or to edit by attribute value edit or. From this manager the state of the cases can be modified or some case to be selected as a query case for retrieval.

The Case Navigator can be used for retrieval either using some available in the case base case or using retrieval wizard and typing some or all attributes' values. The results are sorted by highest similarity. Some results are presented in the next chapter on Fig 15.

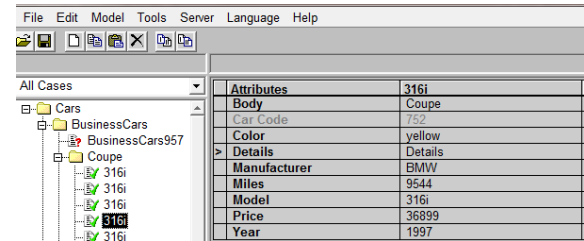


Fig. 9. Explorer Manager Input data for an instance (case)

As it is mentioned above the CBR-Works supports full R4 CBR circle. It provides completion and adaptation rules. The completion rule can be used to complete an existing case attributes or a query. The adaptation one is applied to derive a new result case from the query and retrieved cases (Fig. 10). Both rules are based on precondition and conclusion part related to the IF-THEN clause in the programming languages. The rules are related to specific attribute and how its value can be set in when certain condition or conditions are fulfilled or not. For consistent rules definition a specific Knowledge Engineer experience is required in the domain of CBR application. Except that no special programming skills are needed to model and develop a CBR system.

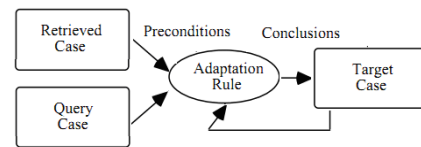


Fig. 10 Adaptation of the cases based on rules

CBR-Works does not support clusterization and indexation of the cases in the Case Base but it supports import via ODBC from extern databases, as well export to XML files of concept model, Case Base and the Rules. The product can be configured as server that can be accessed via Web-based thin clients. It means that third party diagnostics or real-time control systems can be connected to CBR-Works via provided interface.

After last version 4.3.0 CBR-Works is no longer supported or further developed but new CBR solutions and knowledge management systems are available on the company site. Nevertheless the product is well documented and some example CBR applications for training purposes are provided together with the product files. Free license code of the product can be obtained after registration on Empolis site.

3. Example with results

There are two CBR applications developed with myCBR and CBR Works for solving the problems related to the diagnostic of a Pierce-Smith converter. It is used for copper converting in metallurgy. The converter, periodically, (after 300-400 melts) can be repair. Decision to stop it for repairing, depends on the uneven thinning of its fireproof walls, number of working tuyeres and other factors. All data is processed using the methodology described in [10]. All cases concerned to the converter are given in table 1. Column A is a case Id, columns from B to J describe the problem attributes of the cases and column K—the decision attribute RUL is-Remaining Useful Life or after how many melts to stop the converter.

Table 1 Case-base data set

CaseId	A	B	C	D	E	F	G	H	I	J	K
1	Planka	NF	C1	C2	C3	DahaC	T1T2	CRV	Nc	RUL	
2	1	270	8	17.58	21.22	24.25	6.67	330	3780	299	29
3	2	250	10	14.00	19.4	26.70	12.70	321	10206	268	18
4	3	280	9	12.91	17.55	24.50	11.59	305	10206	286	6
5	4	300	12	24.84	27.88	30.83	5.99	265	7560	317	17
6	5	260	13	21.50	23.78	27.67	6.17	332	7560	301	41

First application is implemented on myCBR. On Fig. 11 the Predictive Diagnosis concept (on left upper part) and its attributes are presented, as well the definition of global (upper right part - Euclidian) and local similarity of these attributes (right most column SMF). The discriminant and weight columns describe whether certain attribute can be used in the retrieval (true) or not (false) and its importance of query (weight).

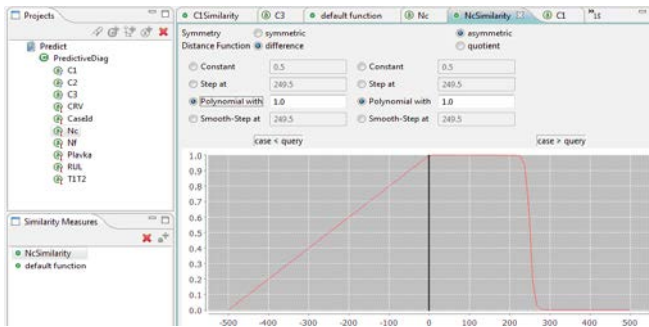


Fig. 11 Definition of the Predictive Diagnostic class properties

Figure 12 shows the results of the query to the case-base DB when we have a new case-problem. Upper left part on the figure contains the query values of some attributes describing a new problem. Upper right part contains all cases in the case base database sorted on the base of their proximity to the queried case.. In the estimations of the proximity the local and global SF are taken into account. Bottom part presents first four cases with highest proximity.

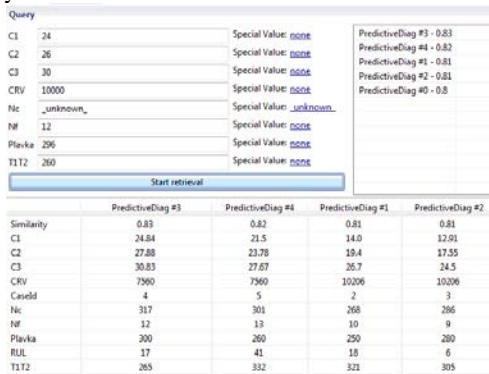


Fig.12 The results after the query to the case base DB

Figure 13 shows attribute definition in analogical CBR application, made with CBR Works.

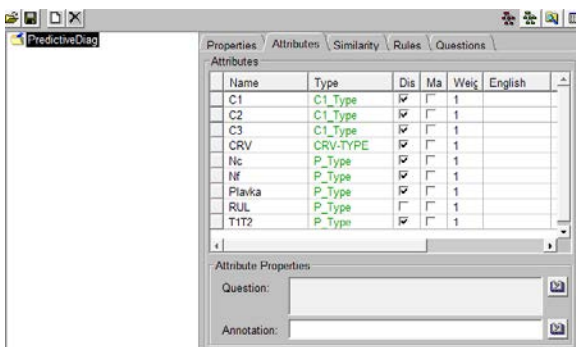


Fig.13 Definition of the attributes in CBR Works

On Fig. 14 setting the local similarity is presented and on Fig 15 the retrieval of results of similar query as shown on fig. 12.

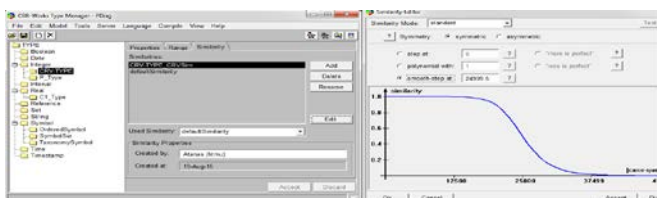


Fig. 14. Editing the similarity function for CRV attribute

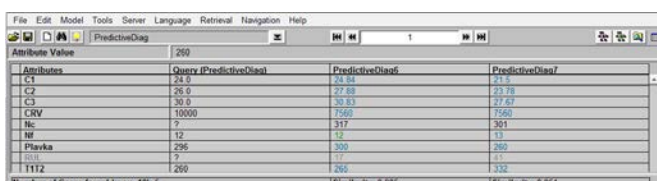


Fig. 15. Retrieval of results after the query to the case base DB

### 4. Conclusions

MyCBR 3 platform can be used for non-complex CBR applications development with partial CBR R4 cycle and with small number of cases. For development of CBR application with myCBR 3 no programming skills are required, only expert knowledge in the professional or scientific domain in which the CBR system is intended. Some design knowledge for case modeling is also desirable. MyCBR is not suitable for application with many attributes, as well with solution attributes containing large amount of text, especially when they must be visually presented in one window. Usage the GUI of myCBR 3, especially for Case Base modeling and case-instance generation is confusing in contrast with previous myCBR versions. The retrieval procedure does not work correctly with partial number of problem attributes if Special Value of the attributes whose value is not set explicitly. In this case Special Value must be set to \_undefined but not to \_unknown (which is default). In myCBR 3 is not possible to update the Case Base with the current query (new case) automatically. It must be done manually. The same is valid if you need to use old case for retrieval.

CBR-Works 4.3.0I can be applied for development for both non-complex CBR applications and complex supporting the full CBR R4 cycle. Development of complex applications however requires Knowledge Engineer experience, especially for synthesis of adaptation rules and their usage. The GUI of CBR-Works provides larger and better functionality for concept (case) and Case base structure modeling and retrieval, than this one in MyCBR. Filtering, query wizard, using a case from Case Base as a query and navigation over retrieved cases is also better than in myCBR. The product support work with external databases, as well import or export of the case base, concept and rules models data. As myCBR the CBR-Works also is not suitable for case modeling with attributes containing large text fields.

Based on the examples, given above, it is obvious that CBR-Works GUI and the support of external interfaces overmatches myCBR 3 and gives more options for weights and SF type modification of attributes and cases. This is of great importance for query adjustment and refining the case base. An advantage of myCBR is that is an open-source SDK and its code can be easily modified to the specific customer needs, but last one requires good programming skills. However, both frameworks can be used for development of CBR systems intended to diagnostics.

### 5. References

1. Richter M., Weber, R., Case-Based Reasoning A Textbook, Springer, 2013.
2. Boshnakov, K., L. Antonov, A. Atanassov, Comparative analysis of software implemented method of precedents in the tasks of diagnosis, Anniversary Scientific Conference "40 Years Department of Industrial Automation", UCTM-Sofia, 2011.
3. Atanassov, A., L. Antonov, Comparative Analysis Of Case Based Reasoning Software Frameworks jCOLIBRI and myCBR, Journal of University of Chemical Technology and Metallurgy, 47, 1, 2012
4. CBR\*Tools, [www-sop.inria.fr/ajxis/cbrtools/usermanual-eng/Intro-duction.html#OOFramework](http://www-sop.inria.fr/ajxis/cbrtools/usermanual-eng/Intro-duction.html#OOFramework)
5. Component Architecture Technology for developing CBR systems, CAT-CBR, [www.iiaa.csic.es/Projects/cbr/cat-cbr/](http://www.iiaa.csic.es/Projects/cbr/cat-cbr/)
6. Stefan Schulz. CBR(Works: A state-of-the-art shell for case-based application building. Proceedings of the 7th German Workshop on Case-Based Reasoning (GWCBR'99), 1999.
7. myCBR, <http://www.mycbr-project.net/>
8. CBR-Works, Reference Manual, TECINO GmbH, 1999.
9. Protege, <http://protege.stanford.edu/>
10. Boshnakov, K., A. Atanasov. Case-based Forecasting of the Campaign Length of Pierce-Smith Converters. Automation and Informatics No. 5-6,2012
11. <https://gaia.fdi.ucm.es/research/colibri/colibrstudio/installation>
12. <http://www.empolis.com/cbr/>