International Journal of Contemporary Management
Volume 14 (2015) Number 3, s. 99–113
doi:10.4467/24498939IJCM.15.006.4308
www.ejournals.eu/ijcm

THE CONCEPT OF A KNOWLEDGE BASE TO AID IN COST ESTIMATING OF SPORTS FACILITIES

*Aleksandra Radziejowska**
*Krzysztof Zima***

Abstract

Background. Reasoning based on cases is one of the heuristic techniques which is used in making right decisions in complex situations. Inference systems typically use previously acquired knowledge which is the basis for creating tools to represent and process information. So far, there has been a wide variety of solutions developed to make inference in conditions of incomplete knowledge, but this process appears in very few studies related to civil engineering.

Research aims. This article will provide a concept of the model of knowledge base and a method of describing sports facilities. The cost estimation model will be based on the prices of the integrated works which, being based on the indicated system, will allow to create the module supporting cost and tender calculations.

Methodology. The authors present the main assumptions of the concept of estimating the costs of sports facilities with the use of the system of inference on the basis of cases (CBR – Case-Based Reasoning).

Key findings. The indicated method is designed to improve the accuracy of estimating the costs of planned sports investments and improve the decision-making process at the planning stage.

Keywords: CBR, reasoning based on cases, costing

INTRODUCTION

Searching for historical knowledge and decision support based on CBR (Case-Based Reasoning) as opposed to relying on individual experiences can accelerate the estimation of costs. Models based on CBR,

* AGH University of Science and Technology in Cracow.
** Cracow University of Technology.

are better than other models for long periods of use by maintaining quality and the ability to solve problems despite the lack of clarification of certain information (Kim, An & Kang, 2004, pp. 1235–1242; Duverlie & Castelain, 1999). CBR may be based on both qualitative and quantitative data, reflecting the types of data sets that exist in the real world (Mendes, Mosley & Counsell, 2002).

CBR can be used to estimate construction costs based on finding similar projects already completed. This provides a simple way of measuring the cost of construction, whereas in most studies there are non-linear relationships between the cost and the influencing factors (Chou *et al.*, 2006; Emsley *et al.*, 2002; Lowe, Emsley & Harding, 2006, pp. 22–30). An example of using the CBR reasoning of cases in the estimation of costs may be a model CBR using AHP method for establishing the weights of criteria proposed in (An *et al.*, 2007, pp. 2573–2579), or CBR model using genetic algorithms to estimate the cost of construction (Sae, 2011, pp. 570–581) or the unit cost of residential construction projects (Dogan, Arditi & Gunaydin, 2006, pp. 1092–1098).

On the other hand Sae-Hyun Ji *et al.* (Sae *et al.*, 2012, pp. 43–52) presented a CBR method that uses existing cases to solve new problems designed to calculate the costs of adapting military barracks in Korea based on 129 historical data. CBR models have also been used in the calculation of the cost of drilling wells (Mianaei, Iranmanesh, Akbari, 2012, pp. 186–193) or the pumping station (Marzouk & Ahmed, 2011). Created models are also forecasting both, time and cost of construction at an early stage of a construction project (ChoongWan *et al.*, 2010, pp. 739–752).

The paper will present the concept of support of estimating construction costs in the initial phase of investment based on reasoning of the cases. The process of information management and utilization of historical data is different from the previously applied model approach. Previous models feature a global assessment of construction projects or works by assigning certain criteria resemblance to the entire investment. In the proposed model a more detailed approach has been adopted using the CBR in estimating the costs of individual elements of a building's circuits. It should be added that in support systems for estimating costs of sports facilities such as football pitches, running tracks, or skateparks there is a lack to fill, for which the authors present the proposed model.

The Concept Of Model Support For Cost Estimates Based On The Reasoning

An important feature which distinguishes CBR from other inference methods is the mechanism of a continuous learning system in a way independent of the expert. The learning process is carried out by collecting cases (solutions) in a database and makes them available to solve new problems in the future. Resolved cases, located in the database typically contain knowledge specific to a given problem. It is worth noting that the Knowledge Base does not necessarily contain a complete knowledge of the field (Figure 1) (Zadora & Wolny, 2002).

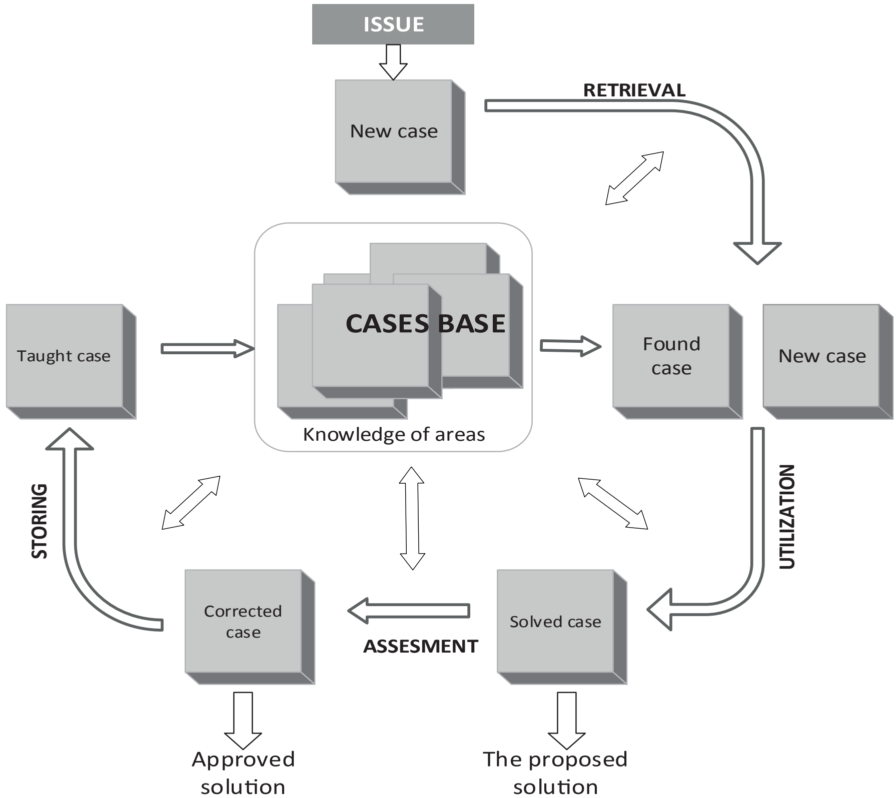

Figure 1. The operating cycle of CBR

Source: based on Zadora & Wolny, 2002.

CBR imitates, in more faithful way, human intelligence and learning process because the applicant system modifies its behavior based on accumulated experience. CBR is a method of solving problems consisting of searching for analogies between the existing situation and previous cases where components are adequately described in the databases.

Below the methods of CBR systems are presented:

1) The first step is to analyze the problems considered and extract the input data needed to determine the mutual similarity between cases.
2) The next step is to find (Retrieval) cases most similar to a new problem from the set of cases accumulated in the Case Base. After finding the appropriate cases, they are scheduled and the most similar case is returned to the user for viewing. Very often an old case does not fit neatly into the new one, which requires modification or making changes in the old solution, so as to adapt them to the new problematic situation.
3) In the next stage of the cycle of the algorithm the closest problematic case is sent for Reuse.
4) The next step is to assess and adapt new solutions of the collected and analyzed situations (Revision). The proposed solution is tested and evaluated, and the necessary corrections/adjustments are made. Reaction is obtained and analyzed. If it does not proceed as expected, its result has to be explained. The procedures following the determination of the course include an explanation of defects and attempts to fix them. All information is stored so as to be able to predict and prevent future, possible defects.
5) The last stage of this cycle is the process of learning (Retainment), i.e. memorizing new solution in the database (Base Case). The new case is presented in a form that the system may retrieve relatively to CBR passed case. The aim is to recover the useful cases, i.e. those which have the potential to provide a solution to a new problem immediately. The new case updates the library for future cases be used. By adding new situations to the library, the system performs incremental learning process (Figure 2).

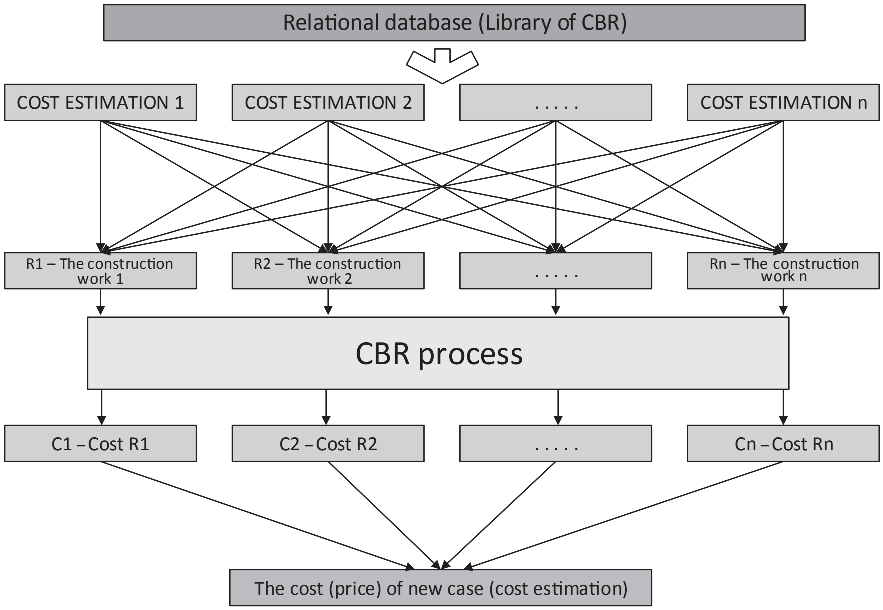

Figure 2. The process of CBR

Source: own work.

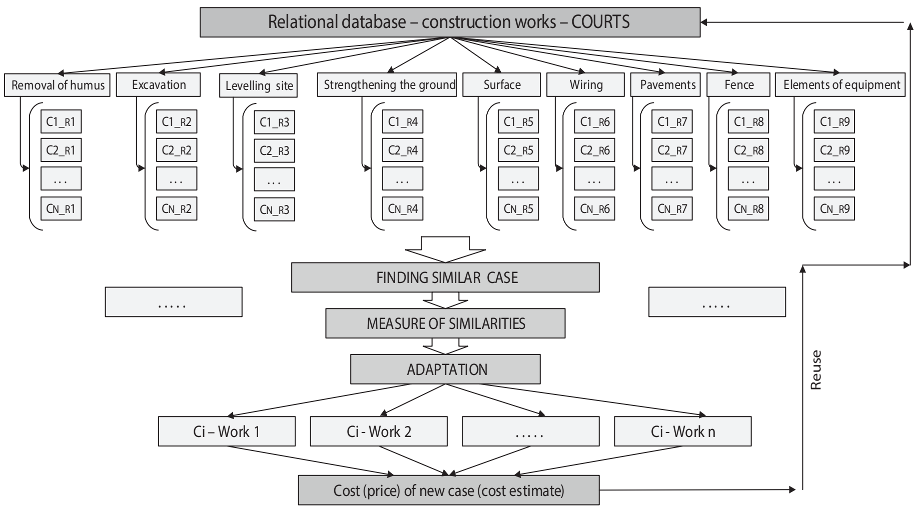

Figure 3. Scheme for the presented example

Source: own work.

Computational example for sports fields

In order to formalize the description of CBR functioning a job description (C_case) should be formulated, in which the user will determine the actual situation of the proposed project:

$$C\_case = \{ W_{A1}, W_{A2}, \dots , W_{AJ}\} \quad (1)$$

where:

W_{AJ} – *the value of j-th attribute describing the situation which occurred in the past.*

The CBR database is a collection of *Cases* in which the structure member functions are included defined for these specific solutions, which were carried out in the past:

$$Cases = \{case^{1} [SP_1, OP_1, GRP_1, OK_1], \dots , case^{i} [SP_i, OP_i, GRP_i, OK_i] \} \quad (2)$$

where:

$case^{i} [SP_i, OP_i, GRP_i, OK_i]$ – i-th case of estimating,

SP_1 – estimates situation for the i-th case (timing, location),

OP_1 – a description of the i-th estimating case,

GRP_i – graphical representation of the i-th estimating case,

OK_1 – constructional description of the i-th case.

Basing on the value of each criterion, described in C_case, a properly chosen calculation mechanism is introduced, which will determine the degree of similarity between the description of the individual tasks in the database and case currently being analyzed.

Thanks to applied algorithms the user can obtain both: information and solutions on the level of the component functions and solutions for the general issues (Ociepka, 2011, pp. 287–291).

In the next step, having regarded the founded limits for the degree of similarity a set of cases is generated which is the base of solutions for the particular case. Table 1 shows an example of sports fields construction cost estimating problem.

The database are other works or elements R1, R2, ..., Rn indexed according to the classification of works OmniClass, Table 1. Costs of works R1, R2, ..., Rn were collected from the offer cost estimate (Figure 3) selected in the tender for the execution of sports fields.

Table 1. Fragment of base cases for works involving the execution of sports fields

Localizatiof of sports fielsds	**21-07 20 30 10 Pedestrian Pavement**	**21-07 20 30 20 Pedestrian Pavement Curbs and Gutters**	**21-06 10 60 30 Arena Football Boards**	**21-05 10 70 50 Athletic Equipment**				
	[PLN/m²]	**[PLN/m]**	**[PLN/m²]**	**volleyball [PLN/u]**	**Tenis [PLN/u]**	**football [PLN/u]**	**handball [PLN/u]**	**basketball [PLN/u]**
Dąbrówka Nowa	76.68	50.28	nd	1347.93	320.04	nd	1431.80	nd
Humniska	41.40	32.36	111.15	nd	3561.20	nd	1293.40	2034
Wolanów	75.99	15.77	20.62	nd	nd	9398.25	nd	nd
Kraczkowa	56.03	17.67	135.00	3323.80	nd	nd	1651.92	2897.30
Seredzice	58.55	49.75	100.00	2936.66	nd	nd	1625.00	1889.76
Biskupie	nd	nd	16.71	nd	nd	nd	nd	nd
Komorniki	93.38	15.94	80.02	6092.67	2193.74	nd	4253.80	3927.56
Bytyń	35.39	24.62	95.40	2353.44	nd	4838.44	nd	2279.82
Wąsocze	nd	nd	16.88	nd	nd	1715.45	nd	nd
Bytom	36.86	32.54	109.00	1199.99	nd	nd	1584.44	2010.84
Smulsko	60.24	36.62	197.25	3776.43	nd	nd	2202.50	4207.5
Stalowa Wola	46.90	31.46	113.00	2719.48	3126.72	nd	1820.85	6361.36

Source: own work.

Then the system sorts the selected solution in accordance with the index consistent with the classification code OmniClass (eg. 21-07203010 pedestrian pavement), thus generating an ordered set of cases Case_u (Figure 3). So that the present listing is a set of possible solutions that can be applied or adapted by the user for the task considered cost estimation (Ociepka, 2011, pp. 287–291).

Table 2 shows a fragment of a database for the criterion grasslands or surfaces for sports fields. For the selected criterion in the table are indicated the sub-criteria of construction of the case (in this case the surface type).

The costs of execution of sports fields pavements have been adjusted due to valorization factor taking into account the change in prices of construction works between the period (quarter, year) at which it was publicly contracted and the period in which it is considered a new case (today). The correction factor was adopted in accordance with the publication of "Aggregated valorization-prognostic indicators ZWW Publishing Promotion limited liability company". The adjustment also concerned the price difference between the location of a new investment (new case), and the location of investments selected from the database – a regional coefficient. After taking into account those factors, new revised unit costs of the works has been received.

By examining an example $R_k \in \{R_1, R_2, \ldots, R_n\}$ consisting of execution of the playing surface for the new case, in which a field of natural grass surface was to be built in the province Wielkopolska with an area of 1450 m^2 field calculations were carried out of similarities between the new Nc and subsequent cases from the database Cn (Table 3).

The calculation of similarities in Table 3 were made using the following formulas.

For the criteria described by the linguistic (type of surface, purpose):

$$SIM\left(w^{Nc}, w^{Cn}\right) = 1 - \frac{\left|\eta\left(w^{Nc}\right) - \eta\left(w^{Cn}\right)\right|}{M-1} \tag{3}$$

where:

w^{Nc} – value of the sub-criterion for the new case,

w^{Cn} – value of the sub-criterion for previous cases,

$\eta(w^{Nc})$, $\eta(w^{Cn})$ – value in an ordered array $\eta(w) = 1, 2, \ldots, n$,

M – the number of values in an ordered array $\eta(w)$.

For the criteria described by the numerical values (the size of the field):

$$SIM\left(w^{Nc}, w^{Cn}\right) = 1 - \frac{\left|w^{Nc} - w^{Cn}\right|}{w_{\max} - w_{\min}} \tag{4}$$

where:

w_{max}, w_{min} – maximum and minimum values for the sub-criterion of previous cases.

Table 2. Fragment of a database for the criterion: Grasslands or surfacing sports fields

Year / quarter	13.06. 2014	06.08. 2014	06.03. 2014	28.04. 2014	04.03. 2014	27.05. 2014	07.08. 2014	04.03. 2014	24.02. 2014	11.09. 2014	24.07. 2014	28.02. 2014
Indexation factor	0,967	0,966	0,964	0,967	0,964	0,967	0,966	0,964	0,964	0,966	0,966	0,964
	Investment name – City (province)											
Ssurface type [zł/m2]	Wolanów (mazowieckie)	Humniska (podkarpackie)	Seredzice (mazowieckie)	Kraczkowa (podkarpackie)	Biskupie (wielkopolskie)	Komorniki (łódzkie)	Bytyń (wielkopolskie)	Wąsocze (wielkopolskie)	Bytom (śląskie)	Smulsko (łódzkie)	Stalowa Wola (podkarpackie)	Dąbrówka Nowa (kujawsko-pomorskie)
Regional coefficient	0.99	1.00	0.99	1.00	0.93	0.94	0.93	0.93	0.96	0.94	1.00	0.98
grassy surface	20.62*	–	–	–	16.71	–	–	16.88	–	4.88	–	–
	19.78**	–	–	–	14.95	–	–	15.10	–	4.45	–	–
artificial grass	–	111.20	–	–	–	–	95.40	–	–	–	–	–
	–	107.42	–	–	–	–	85.52	–	–	–	–	–
polyurethane surface	–	–	135.00	100.00	–	80.02	–	–	109.00	–	113.00	–
	–	–	129.09	96.69	–	72.98	–	–	101.36	–	109.15	–

* the cost of construction of the case

** the job cost after taking into account indexation and regional factors

Source: own work.

Table 3. Similarity measurement for criterion: Grasslands or surfacing sports fields

Assessment criteria	Wolanów (mazowieckie)	Humniska (podkarpackie)	Seredzice (mazowieckie)	Kraczkowa (podkarpackie)	Biskupie (wielkopolskie)	Komorniki (łódzkie)	Bytyń (wielkopolskie)	Wąsocze (wielkopolskie)	Bytom (śląskie)	Smulsko (łódzkie)	Stalowa Wola (podkarpackie)	Dąbrówka Nowa (kujawsko--pomorskie)
Surface type	natural grass	synthetic grass	polyurethane	polyurethane	natural grass	polyurethane	synthetic grass	natural grass	polyurethane	polyurethane	polyurethane	no data
Destination	football	multifunctional	multifunctional	multifunctional	football	multifunctional	multifunctional	football	multifunctional	multifunctional	multifunctional	multifunctional
The size of the pitch	7 000	1320	613,11	1234	1056	1383,48	800	2970	1196	1012	968	bd
The new case – calculation of similarity												
Surface type	1	0	0	0	1	0	0	1	0	0	0	0
Destination	1	0	0	0	1	0	0	1	0	0	0	0
The size of the pitch	0.13	0.98	0.87	0.97	0.94	0.99	0.90	0.76	0.96	0.93	0.92	0
SIM(Nc, Cn)	0.13	0	0	0	0.94	0	0	0.76	0	0	0	0

Source: own work.

And so, for example, comparing according to the first formula for sub-criterion for surface SIM (Nc – natural grass, C1 (Wolanów) – natural grass) = 1, and comparing SIM (N_c – trawa natural, C2 (Humniska) – synthetic grass) = 0. In turn comparing the second formula for sub-criteria for field size SIM (N_c – 1450 m^2, C_1 (Wolanów) – 7000 m^2) = 0.13. The global similarity SIM card (N_c, C_n) for grasslands or pavement criterion was calculated by multiplying the sub-criteria. Only 3 solutions have similarity greater than 0, which was a prerequisite for their consideration. The highest score achieved similarity for solution C5 (Biskupie) = 0.94 and it was adopted for the calculation of the cost. The estimated cost of laying a natural grass surface is thus equal to 1450 m^2 * 14.95 (from Table 2), ie 21,677.50 PLN.

Similar steps should be followed for the entire scope of work in the relevant new case. Estimated value of the costs of the investment will be equal to the sum of all costs C_i.

Verification of the model

Obtaining information that will be used in the knowledge base CBR, is based on the analysis of similar case estimates in the past. As a result, due to the fact that the analysis of events is performed by different people who define and describe the events according to their own perceptions and use different wording, (often cost estimate their own positions*) a kind of semantic problem is formed. To illusttrate different ways of determining the same events, an ambiguous description of a selected topic follows. An interesting fact is that in describing the event, only 1/5 of respondents use the same vocabulary, even if we have to deal with experts in their chosen field (Mirończuk & Maciak, 2013, pp. 95–106). However, there are many methods to verify the result, one of which is shown below.

Practical verification of the results obtained from the proposed model can be made by comparing the estimated total cost of the new case to the actual costs for the highlighted set of testing, in order to calculate the mean absolute error of estimation MAEE is:

$$MAEE = \frac{\sum \frac{|C_{CBR} - C_{ACT}|}{C_{ü}}}{n} \times 100\% \quad (5)$$

where:
MAEE –mean absolute error of estimation,
C_{CBR} – the estimated cost of the new case,
C_{AKT} – the current cost of the highlighted event of the test set
n – the number of test cases.

Table 4. The steps in the process of CBR

1	**Input**	Isolating the input data needed to determine the mutual similarity between cases
2	**Retrieve**	Finding the most similar case or set of cases
3	**Reuse**	Utilizing the knowledge contained in similar case to solve current issues
4	**Revise**	Evaluation of the usefulness of the proposed solution
5	**Retain**	"Learning" – adding a new case to the database for later use in future issues

Source: own work.

CONCLUSIONS

In conclusion, we can determine the cycle of the system implementing the method of reasoning by cases as five major processes that follow each other in sequence shown in Table 4.

The analyzed issue located at the entrance of the system is compared with the cases that have been previously stored in the database of cases. The similarity criterion allows for the isolation of cases meeting the conditions of the test situation. After selecting the most favorable solution for a new case it is added to the database.

CBR systems show high effectiveness in supporting issues related to decision-making in many fields of science. CBR is not a popular system in issues related to construction, but as shown by the above considerations, it should be considered for application in this area of economy. The Database determined for sports facilities shows that the system allows selection of the optimum solution for a new case in an efficient manner (Kempa, 2007, pp. 283–290).

It should be remembered that CBR requires equipment with adequate mechanisms for raising and processing of knowledge and experience and effective mechanism for searching databases, in order to determine the similarity between the current situation and that analyzed on the basis of the stored cases (Mirończuk & Maciak, 2013, pp. 95–106).

The article showed the concept of support estimation costs CBR model. Using the mechanism of CBR for estimating the cost of individual works can increase the accuracy of calculations by adjusting the criteria of the past cases. In the article an example of estimating the costs for one criterion was illustrated – the execution of playing fields turf and pavement.

The method presented is a good example of the possibilities the learning systems might have. Basing on the material presented we can see advantages offered by CBR system. This kind of method may successfully be used to promote the formation of estimates, project planning, or the entire investment.

REFERENCES

An, S.-H., Kim, G., & Kang, K. (2007). A case-based reasoning cost estimating model using experience by analytic hierarchy process. *Building and Environment*, *42(7)*, 2573–2579.

Choong Wan, K., Tae Hoon, H., Chang Taek, H., & Kyo Jin K. (2010). CBR-based hybrid model for predicting a construction duration and cost based on project characteristics in multi-family housing projects. *Canadian Journal of Civil Engineering*, *37*, 739–752.

Chou, J.-S., Peng, M., Persad, K.M., & O'Connor, J.T. (2006). Quantity-based approach to preliminary cost estimates for highway projects. *Transportation Research Record*, *1946*, 22–30.

Dogan, S.Z., Arditi, D., & Günaydın, H.M. (2006). Determining attribute weights in a CBR model for early cost prediction of structural system. *Journal of Construction Engineering and Management*, *132(10)*, 1092–1098.

Duverlie, P., & Castelain, J.M. (1999). Cost estimation during design step: Parametric method versus case based reasoning method. *International Journal of Advanced Manufacturing Technology*, *15(12)*, 895–906.

Emsley, M.W., Lowe, D.J., Duff, R., Harding, A., & Hickson, A. (2002). Data modelling and the application of a neural network approach to the prediction of total construction costs. *Construction Management and Economics*, *20(6)*, 465.

Kempa, A. (2007). Integracja systemów ekspertowych i systemów wnioskujących na podstawie przypadków. *Technologia przetwarzania wiedzy w SWO*, 283–290.

Kim, G.-H., An, S.-H., & Kang, K.-I. (2004). Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, *39(10)*, 1235–1242.

Lowe, D.J., Emsley, M.W., & Harding, A. (2006). Predicting construction cost using multiple regression techniques. *Journal of Construction Engineering and Management*, *132(7)*, 750–758.

Marzouk, M.M., & Ahmed, R.M. (2011). A case-based reasoning approach for estimating the costs of pump station projects. *Journal of Advanced Research*, *2(4)*, 289–295.

Mendes, E., Mosley, N., & Counsell, S. (2002). The Application of Casebased Reasoning to Early Web Project Cost Estimation. In: *Proceedings – IEEE Computer Society's International Computer Software and Applications Conference*. Oxford, United Kingdom: Institute of Electrical and Electronics Engineers Computer Society.

Mianaei H.S., Iranmanesh S.H., & Akbari, H. (2012). *Application of Case-Based Reasoning in Cost Estimation of Drilling Wells*. 2012 International Conference on Industrial and Intelligent Information (ICIII 2012), IPCSIT, 31. Singapore: IACSIT Press, 186–193.

Mirończuk, M., & Maciak, T. (2013). Propozycja mieszanego przetwarzania półstrukturalnego modelu opisu zdarzeń z akcji ratowniczo-gaśniczych państwowej straży pożarnej PSP. *BiTP*, *29(1)*, 95–106.

Ociepka, P. (2011). Zastosowanie metody CBR do wspomagania procesu projektowo-konstrukcyjnego. *Wybrane Problemy Inżynierskie*, *2*, 287–291.

Sae-Hyun, J., Moonseo, P., & Hyun-Soo, L. (2012). Case adaptation method of case-based reasoning for construction cost estimation in Korea. *Journal of Construction Engineering and Management*, *137(1)*, 43–52.

Sae-Hyun, J., Moonseo, P., & Hyun-Soo L. (2011). Cost estimation model for building projects using case-based reasoning. *Canadian Journal of Civil Engineering*, *38(5)*, 570–581.

Wrona, S., & Zadora, P. (1997). Przetwarzanie wiedzy pochodzącej z przeszłości w celu rozwiązania problemów klasyfikacyjnych. In: H. Sroka, S. Stanek (red.), *Inteligentne systemy wspomagania decyzji w zarządzaniu. Transformacje systemów*. Katowice: AE.

Zadora, P., & Wolny, W. (2002). *Systemy wnioskujące na podstawie przypadków. Systemy wspomagania organizacji*. Katowice: AE.

KONCEPCJA BAZY WIEDZY DLA WSPOMAGANIA KOSZTORYSOWANIA OBIEKTÓW SPORTOWYCH

Abstrakt

Tło badań. Wnioskowanie na podstawie przypadków jest jedną z technik heurystycznych, która służy przy podejmowaniu właściwych decyzji w skomplikowanych sytuacjach. Systemy wnioskowania wykorzystują zazwyczaj zdobytą wcześniej wiedzę, na podstawie której tworzone są narzędzia służące do reprezentacji i przetwarzania informacji. Dotychczas opracowano wiele różnorodnych rozwiązań pozwalających na wnioskowanie w warunkach niepełnej wiedzy, jednak proces ten pojawia się w nielicznych opracowaniach związanych z wiedzą inżynierską.

Cel badań. W artykule przedstawiona zostanie koncepcja modelu bazy wiedzy wraz ze sposobem opisu obiektów sportowych. Model szacowania kosztów oparty będzie na cenach robót scalonych, które na podstawie wskazanego systemu pozwolą na utworzenie modułu wspomagającego kalkulacje kosztorysowe inwestorskie oraz ofertowe.

Metodyka. Autorzy przedstawią główne założenia koncepcji szacowania kosztów obiektów sportowych przy użyciu systemu wnioskowania na podstawie przypadków (CBR – Case Based Reasoning).

Kluczowe wnioski. Wskazana metoda ma na celu poprawę dokładności szacowania planowanych kosztów inwestycji sportowych i ulepszenie procesu decyzyjnego na etapie planowania.

Słowa kluczowe: CBR, wnioskowanie na podstawie przypadków, kosztorysowanie