KNOWLEDGE DISCOVERY IN DATA
IN CONSTRUCTION PROJECTS

J. SZELKA[1], Z. WRONA[2]

Decision-making processes, including the ones related to ill-structured problems, are of considerable significance in the area of construction projects. Computer-aided inference under such conditions requires the employment of specific methods and tools (non-algorithmic ones), the best recognized and successfully used in practice represented by expert systems. The knowledge indispensable for such systems to perform inference is most frequently acquired directly from experts (through a dialogue: a domain expert – a knowledge engineer) and from various source documents. Little is known, however, about the possibility of automating knowledge acquisition in this area and as a result, in practice it is scarcely ever used. It has to be noted that in numerous areas of management more and more attention is paid to the issue of acquiring knowledge from available data. What is known and successfully employed in the practice of aiding the decision-making is the different methods and tools. The paper attempts to select methods for knowledge discovery in data and presents possible ways of representing the acquired knowledge as well as sample tools (including programming ones), allowing for the use of this knowledge in the area under consideration.

Keywords: inference in engineering projects, discovering knowledge in data, rule-based knowledge representation

1. INTRODUCTION

It is characteristic of decision-making problems in civil engineering, in particular in construction projects, that on numerous occasions they are ill-structured. This being the case, in the processes of inference heuristic methods applied, among which considerable significance is attached to expert

[1] Assoc. Prof., Military Academy of Land Forces in Wrocław, Czajkowskiego 109, 51-150 Wrocław, Poland, University of Zielona Góra, e-mail: j.szelka@wso.wroc.pl
[2] PhD., Eng., College of Management "Edukacja" in Wrocław, Krakowska 56-62, 50-425 Wrocław, Poland, e-mail: z_wrona@wp.pl

inference. The computer-based aiding of such projects is performed with the use of expert systems. One of the major factors conditioning inference effectiveness is the quality of knowledge employed to make a decision. The nature of knowledge in civil engineering is mainly factual, which means that it can be obtained from standards, catalogue data or sets of parameters. To a large extent, however, the knowledge in this area comes from an expert's domain (or a team of experts) who through an interview between an expert and a knowledge engineer determines the form of occurring regularities or validates parameters. The acquisition of knowledge from the expert is regarded as one of the most labour-intensive and expensive stages in knowledge base development. It should be noted, though, that the engineer (decision-maker) generally has extremely powerful databases, which are a commonly used tools for the performance of information and analytical processes. They are hardly ever employed, however, in inference processes. In numerous areas of management successful attempts of knowledge acquisition from available data resources are made, with the use of such methods as artificial neural networks or case-based reasoning. The automation of the knowledge acquisition process seems reasonable also with regard to construction projects, provided that, however, the selection of appropriate methods and tools takes account of the specificity of the problem situations in this area.

The essence of the problem under consideration is shown in Fig. 1.
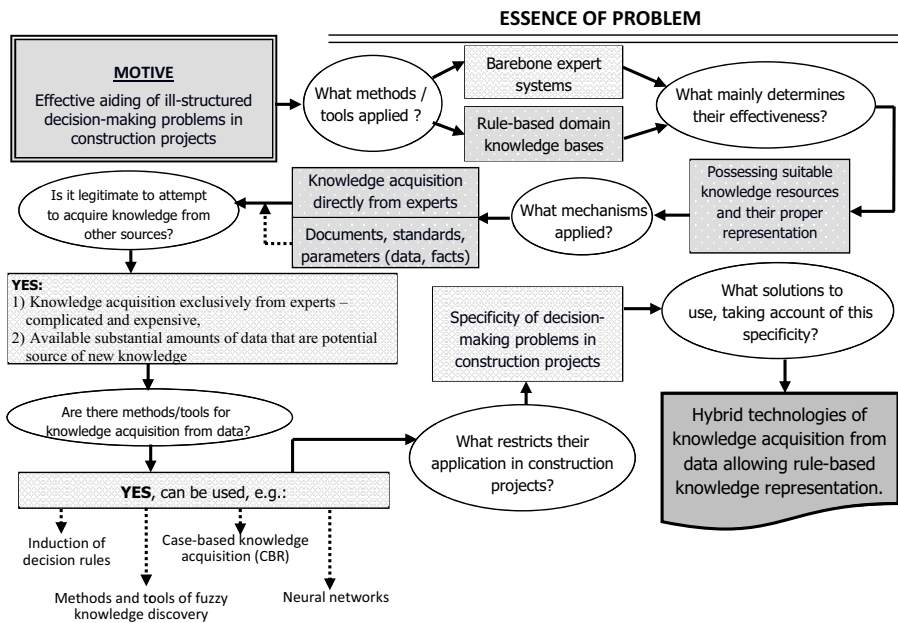

Fig. 1. The essence of the problem.

2. KNOWLEDGE IN THE PROCESS OF INFERENCE IN CONSTRUCTION PROJECTS

The problem of computer-aided decision-making becomes complicated for problems with poorly recognized structure. A qualitative or mixed method of problem description or no possibility to clearly identify relationships, make any attempt to reflect the situation under consideration with a precise mathematical model increasingly difficult or frequently even impossible. In such cases it is required to substitute heuristic actions for formal models.

Heuristics are interpreted as intuitive knowledge, rules and regularities formulated on the basis of experience. Not only do they enable one to reduce the complexity of the problem in question, they are also used in inference processes. Inference is a method that gives the possibility of deriving conclusions (new batches of knowledge) on the basis of facts and rules with the use of heuristics or other search methods [3]. Among the diverse solutions allowing for the practical use of inference in the decision-making practice of the engineer, a major role should be assigned to expert systems technologies.

Expert system knowledge is the result of the special mapping of the description of a segment of reality made by experts, its facts, phenomena, states of objects and relationships between them. As these resources usually include a variety of structures (descriptions in natural language, classifications, validations, numerical data sets, etc.), the process of acquiring knowledge and the representation of its portion in a homogeneous style is a task of substantial complicatedness [5].

On the whole, it can be stated that the knowledge acquired by the knowledge engineer in construction projects is in a form of facts. These are portions of information (statements, observations) depicting events, states, symptoms and situations from the area represented in the knowledge base. When used in the inference process, they allow for the activation of rules, i.e. the course of action oriented at conceptual descriptions and problem situations related to the recognition of objects, relationships between them or their behaviour. The activation of subsequent rules in the inference process can in turn expand the available knowledge of the problem under consideration.

The specification of knowledge for the expert system is the result of cooperation between the expert and the knowledge engineer (Fig. 2).

If in the knowledge obtained a content attribute and a form attribute should appear, then the content is the responsibility of the expert and the form, i.e. its notation, of the knowledge engineer The creation of a domain knowledge base in construction projects comes down to the performance of the following [1]:

1.   problem identification;

2.   knowledge acquisition;

3.   knowledge formalization;

4.   knowledge implementation;

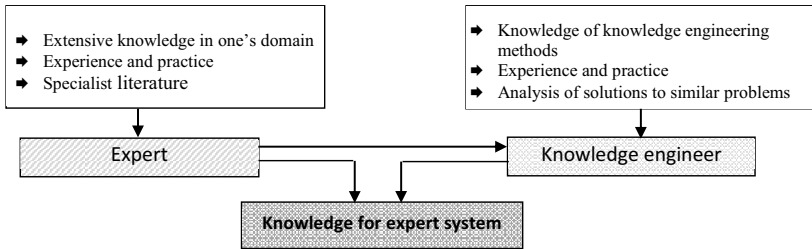5.   testing and initial knowledge verification.

Fig. 2. Participants in the knowledge acquisition process in accordance with [1]

All of the tasks related to knowledge base creation can generally be divided into two phases. The first one involves knowledge identification and acquisition, whereas the other one its formalization, implementation and testing (both phases are of considerable significance also in the process of automated knowledge acquisition).

Phase One focuses on the tasks related to the determination of the scope and context of a decision-making task, the definition of the sources of knowledge and its acquisition.

Phase Two includes model and implementation of activities with regard to the acquired knowledge and its testing.

The objective of problem identification is to recognize the problem situation in such a way that it will enable the specification of its semantic and the logical scope. The principal areas of activities aiming at problem identification usually include: the identification of participants in the knowledge base creation process, the identification of the functional aims of the proposed system and the identification of the context of the project.

Knowledge acquisition is the process of extracting knowledge from available sources. The way it is performed is primarily dependent on the information sources used. For instance, the sources of knowledge supporting problem-solving in bridge building are summarized in Table 1.

The accuracy of the decision (expert report) made by the expert system largely depends on the quality of knowledge accumulated in the knowledge base of the system. This is why it has to encompass the largest possible set of data, facts and rules concerning the construction of bridge facilities.

Table 1. Sources of domain knowledge for bridge building.

KNOWLEDGE SOURCE	SPECIFICATION
Experts' knowledge	Knowledge and experience of engineers specializing in construction and operation of bridge facilities
Documentation	Archival data on bridge facilities: design and operating documentation, rules and regulations, standards
Results of simulation analyses	Mainly for statistical and strength analyses of structures
Data resources	Records of structures, results of surveys, tests and expert reports, test loads

A substantial complexity of the processes related to the acquisition of expert knowledge, a large amount of time it requires and its high costs have become the reason for searching for other options to acquire knowledge. Given the essence of knowledge, defined as a body of reliable information in conjunction with the ability to use it (to formulate conclusions on the basis of the information), attention has been paid to IT storage mechanisms for structured data. The vast majority of state-of-the-art IT solutions, employed to collect and process information resources (data, knowledge), are currently created with the use of database technologies. In the area of construction projects, the greatest practical significance should be attached to two categories of databases: transactional and analytical.

The first one includes databases used for current (operating) recording and reporting activity, whereas the multidimensional structures of analytical databases can be used to support engineering analyses. In this case information resources can take the form of both operating (current and historical) data and various aggregates, composed from detailed data. It can be stated that in numerous cases the engineer/decision-maker holds a huge volume of data (and information resulting from data processing), accumulated in the databases at his disposal.

3. POSSIBILITIES OF KNOWLEDGE ACQUISITION FROM DATA

There are much fewer possibilities of understanding and analysing large data sets than of collecting and storing them. Possessing substantial volumes of data does not mean, therefore, that at the same time one possesses the knowledge contained in these data. The process of acquiring new, unknown knowledge from the information resources at one's disposal is referred to as knowledge discovery. Implicit knowledge can include previously unnoticed dependencies, patterns or relations [4].

Knowledge discovery is a complex process (Fig. 3), at the start of which data are collected, possibly "cleaned" (e.g. incorrect or missing data are handled), and if necessary also integrated (data from

different sources are merged). The data collected and pre-processed are then subject to further transformations, which most frequently include the following [2]:

1. data selection (choosing data that will undergo exploration);

2. data transformation (converting attribute types, defining inferred attributes);

3. exploration (extracting knowledge from data: generating rules, decision trees, neural networks, etc.);

4. results interpretation (choosing useful knowledge, visualising results).
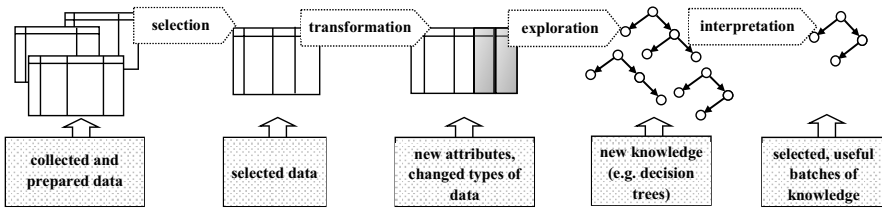

Fig. 3. Stages in knowledge discovery in accordance with [4].

In management practice, knowledge discovery tasks are performed by specialized systems, such as Intelligent Miner (IBM) and MineSet (Silicon Graphics). However, it is worth noting that for this purpose one can also use IT tools that are much more available. These include, for instance, programming languages such as MineSQL and DMQL (Data Mining Query Language), designed to acquire knowledge from relational databases [6].

4. KNOWLEDGE DISCOVERY FROM DATA IN CONSTRUCTION PROJECTS

The problem of the use of specific methods and tools for knowledge acquisition from data in engineering inference processes should be considered in the context of the specific nature of engineering projects, in particular construction ones. The complexity of the problem situations that occur in the case of these projects, the fact that they are heterogeneous and are often divided into stages (the conclusions from the previous stage are input information resources for the next one), lead to numerous complications when it comes to a clear selection of methods and tools supporting the solution of analytical and decision-making problems in this area. Additionally, in some cases, it is not possible to ensure completeness of the input data, and of those available, some may have a fuzzy (imprecise) nature. A large variety of problem situations, accompanied by the heterogeneity of information resources used in inference processes, cannot result in the heterogeneity of the

representation of available and acquired knowledge. The knowledge discovery systems used commercially frequently employ decision trees for representation purposes. However, in the case of complex engineering projects, they can become sizeable, which limits their effective application. It seems that for various reasons logical rules are a rational solution. They are the technique that is most frequently used for knowledge representation when knowledge is obtained in a traditional manner from experts and, in addition, they can also be used in case-based knowledge acquisition systems or in fuzzy systems. This results in the possibility of a unified knowledge representation for hybrid systems. These observations focus mainly on the selected methods and tools for knowledge discovery from data, along with their characteristic way of knowledge representation. Since a large portion of the data used in the construction projects is organized in relational database structures, special attention must be paid to the methods of the induction of decision rules, in particular the induction of classification and association. If the knowledge discovered with such methods is to be expressed as rules, then two relations that may exist between the data and the rules are determined: confirmation and violation. A record (row) in the database table confirms the rule if for the values of its attributes, both the premises and the conclusion of the rule take a logical truth value. If this principle is not satisfied, it is said that the record violates the rule. Furthermore, each rule can be assigned certain indicators (measures) of its assessment. Support is an example of such an indicator [6]. The support of the rule (S) indicates the number of records of relations that confirm the rule. Depending on the method of knowledge discovery, the rules obtained are referred to as classification, association or discrimination. A sample database table and the classification rules for the attribute ASSEMBLY METHOD found in it are shown in Table 2.

Table 2. Discovering classification rules for the case of the assembly of the DMS-65 span structure.

ASSEMBLY	NUMBER OF CRANES	ASSEMBLY METHOD
sections	2	sections
blocks	0	no assembly
blocks	1	blocks
sections	1	no assembly
individual components	1	individual components
Assembly = "sections" **&** Number_cranes >= 2 **➔** Assembly_method="sections" **(S=20%)**		
Assembly = "sections" **&** Number_cranes < 2 **➔** Assembly_method="no_assembly" **(S=20%)**		

The knowledge acquired in the form of classification or association rules can be supplemented with expert knowledge or can also supplement expert knowledge, acquired in the traditional manner.

A large part of the data in relational or analytical databases is assigned clear, precise values (usually numbers). The specificity of the construction projects, however, makes it necessary to define certain groups descriptively, e.g. owing to their special characteristics (type of the bottom = "gritty"), or because there is no access to accurate data (e.g. damage to the span = "minor"). What is more, the imprecision of data necessitates determining what set particular data belong to (e.g. determining the extent to which damage to the span can be considered "minor"). In such cases, one deals with the phenomenon of data or knowledge fuzzification.

Knowledge fuzzification may include fuzzification of data, facts as well as inference rules.

For instance, for the selection of a span reconstruction method, a fuzzy rule discovered based on the previously fuzzy attribute WIDTH.W.OBSTACLE (the width of the water obstacle) could be expressed as follows:

IF width.w.obstacle *IS* mean *AND* … *AND* span_type *IS* single, *THEN* bridge construction method *IS* span BIBER_with_additional_abutments.

Fuzzy knowledge discovery systems (with rule-based knowledge representation) can be useful both in the process of classifying the condition of a facility (e.g. a bridge) and the prediction of this condition (or the time of safe operation).

Except for the imprecision of information resources, available in construction projects, their incompleteness is yet another frequent phenomenon. The incompleteness of information resources (data, facts, rules) may consequently prevent the implementation of the inference process. One possible way of dealing with this problem is to use the CBR method. It allows one to analyze the similarities of the problem situation that has occurred with the cases recorded in the so-called database of cases and, after finding a sufficiently similar case, saving the adapted solution as a new batch of knowledge in the knowledge base [3]. This solution can be used in engineering facilities monitoring systems and facility condition classification systems based on the quantitative and qualitative description of its condition. Most importantly, in the case of the CBR method, a ruled-based knowledge representation generated on the basis of the database of cases can be used.

The specificity of construction projects (especially their heterogeneity) leads to considering the possibility of using hybrid solutions in the automated processes of knowledge acquisition [7], including the methods specific to precise and complete data and such that enable knowledge discovery for fuzzy data or the acquisition of knowledge required for the inference process, in the case of the incompleteness of knowledge (the CBR method). The idea of such a system, using both

the method of the induction of decision rules ("new" knowledge stored in the database of rules) as well as the CBR method, in the case of planning a crossing operation (NEW CASE) for which one wants to specify the type of crossing and select the right design, can be written schematically (Fig. 4):
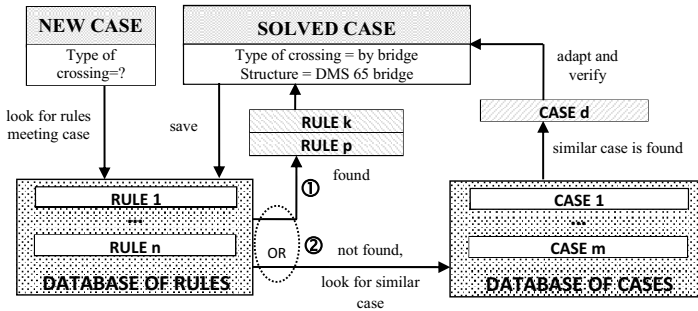

Fig. 4. The essence of using a hybrid knowledge discovery system in construction projects.

IF a new case meets at least one rule of the SET OF RULES *THEN* adopt the solution determined by the rule (rules) *ELSE* adapt the most similar solution from the DATABASE OF CASES

5. CONCLUSIONS

The most frequently used solutions in the field of computer-assisted inference in construction projects include expert systems, with the case-based representation of expert knowledge. Knowledge for these systems is almost exclusively acquired from experts (as a result of the dialogue: an expert – a knowledge engineer). High costs of these tasks, the vast amount of time it requires and significant complicatedness make one search for the alternative ways of knowledge acquisition, especially that such solutions have been used in various areas of management for years. The complexity of construction projects, the uniqueness of their parameters or the diverse nature of data make it difficult to pinpoint a single knowledge discovery solution (a method and tools) as rational in this area. It seems reasonable to attempt to combine several methods and tools specific to them (including query languages used for knowledge discovery), forming a hybrid system. The nature of the available data and the specificity of problem situations tend to indicate the methods of the induction of decision rules as the most appropriate ones. They should be, however, supported with the solutions allowing for knowledge discovery from fuzzy data and enabling case-based knowledge discovery. All of these methods allow for the rule-based representation of discovered knowledge and the uniformity of

representation seems to be an important condition for its use, without the need for complex conversion processes. A proposal for the architecture of a hybrid system for discovering knowledge from data, with the use of the discussed methods, seems to constitute a separate conceptual and methodological challenge.

REFERENCES

1. M. Bazewicz, Metody i techniki reprezentacji wiedzy w projektowaniu systemów, Wydaw. Politechniki Wrocławskiej, Wrocław 1994
2. D. T. Larose: Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych, PWN, Warszawa 2006.
3. A. Kempa: Modelowanie procesów biznesowych z wykorzystaniem metody case-based reasoning, [w:] Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą, red. J. Kacprzyk, L. Drelichowski, PSZW, Bydgoszcz 2005.
4. M. Nycz: Pozyskiwanie wiedzy menedżerskiej. Podejście technologiczne, Wyd. AE we Wrocławiu, Wrocław 2008.
5. H. Sroka, W. Wolny (red.), Inteligentne systemy wspomagania decyzji, Wydaw. AE w Katowicach, Katowice 2009.
6. J. Stefanowski, Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy. Rozprawa habilitacyjna, Wydaw. Politechniki Poznańskiej, Poznań 2001.
7. J. Bień, Modelowanie obiektów mostowych w procesie ich eksploatacji, Oficyna Wydaw. Politechniki Wrocławskiej, Wrocław 2002

ODKRYWANIE WIEDZY Z DANYCH W PRZEDSIĘWZIĘCIACH BUDOWLANYCH

*Słowa kluczowe***:** wnioskowanie w przedsięwzięciach budowlanych, odkrywanie wiedzy z danych, data mining, regułowa reprezentacja wiedzy

STRESZCZENIE:

Problemy decyzyjne w obszarze przedsięwzięć budowlanych charakteryzują się w wielu przypadkach słabym poziomem ustrukturyzowania. W procesie wnioskowania stosuje się wówczas często metody heurystyczne. Wśród nich szczególne znaczenie należy przypisać wnioskowaniu eksperckiemu. Informatyczne wspomaganie tych przedsięwzięć jest realizowane przy użyciu systemów ekspertowych. Jednym z ważniejszych czynników warunkujących skuteczność wnioskowania jest jakość wiedzy, wykorzystywanej do wypracowania decyzji. Przedsięwzięcie pozyskiwania wiedzy od ekspertów dziedzinowych zaliczane jest do najbardziej pracochłonnych i kosztownych etapów tworzenia bazy wiedzy.

Znaczny poziom złożoności procesów związanych z akwizycją wiedzy eksperckiej w obszarze przedsięwzięć budowlanych, jej czasochłonność i duże koszty stały się przesłanką do poszukiwania innych możliwości w zakresie pozyskiwania wiedzy (w tym – automatyzacji pozyskiwania wiedzy).

Można stwierdzić, że w wielu przypadkach, w dyspozycji inżyniera-decydenta jest olbrzymia ilość danych (i informacji, powstałych w wyniku przetwarzania danych), zgromadzonych w posiadanych bazach danych. Oprócz oczywistej możliwości ich wykorzystania w procesach informacyjnych i analitycznych, warto podjąć próbę potraktowania ich także, jako potencjalne źródło wiedzy.

Posiadanie znacznych ilości danych nie oznacza, że dysponujemy jednocześnie wiedzą, która jest zawarta w tych danych. Proces pozyskiwania nowej, nieznanej wiedzy z posiadanych zasobów informacyjnych określany jest mianem *odkrywania wiedzy* (ang. *Knowledge Discovery*). Nieznana dotychczas wiedza może obejmować niedostrzegane wcześniej zależności, wzorce, czy relacje [1].

Do metod o największym znaczeniu praktycznym można zaliczyć metody indukcji reguł decyzyjnych (w tym, m.in.: klasyfikacyjne, asocjacyjne), wnioskowanie na podstawie przypadków (CBR) oraz sztuczne sieci neuronowe.

Każda z zastosowanych metod odkrywania wiedzy charakteryzuje się specyficznymi własnościami, które warunkują potencjalne (racjonalne) obszary jej wykorzystania [2].

Ponieważ większość danych, wykorzystywanych w przedsięwzięciach budowlanych jest zorganizowana w strukturach relacyjnych baz danych, ze szczególną uwagą należy podejść do metod indukcji reguł decyzyjnych, a w szczególności indukcji reguł klasyfikacyjnych oraz asocjacyjnych. Pozyskana w postaci reguł klasyfikacyjnych lub asocjacyjnych wiedza może być uzupełniana poprzez wiedzę ekspercką, bądź też może stanowić uzupełnienie wiedzy eksperckiej, pozyskanej w tradycyjny sposób.

Znacznej części danych zawartych w relacyjnych, czy analitycznych bazach danych przypisane są jednoznaczne, precyzyjne wartości (zazwyczaj liczbowe). Specyfika przedsięwzięć budowlanych sprawia jednak, że pewne grupy danych są określane w sposób opisowy, np. z powodu ich specyfiki (rodzaj gruntu dna="piaszczysty", lub też z powodu braku dostępu do danych precyzyjnych (np. uszkodzenie przęsła="nieznaczne").

Brak precyzji danych rodzi też często potrzebę określania stopnia przynależności danych do określonego zbioru (np. określenia, w jakim stopniu uszkodzenie przęsła można uznać za „nieznaczne"). Rozmyte systemy odkrywania wiedzy (z regułową reprezentacją wiedzy) mogą być użyteczne zarówno w procesach klasyfikowania stanu obiektu (np. mostu), jak i predykcji takiego stanu (lub czasu bezpiecznej eksploatacji).

Oprócz nieprecyzyjności zasobów informacyjnych, dostępnych w przedsięwzięciach budowlanych, za dość częste zjawisko można uznać ich niepełność. Jednym z możliwych sposobów radzenia sobie z tym problemem jest wykorzystanie metody CBR (ang. *Case Based Reasoning*). Pozwala ona na analizę podobieństwa zaistniałej sytuacji problemowej do przypadków zapisanych w tzw. bazie przypadków i, po znalezieniu przypadku wystarczająco podobnego, zapamiętanie w bazie wiedzy zaadaptowanego rozwiązania, jako nowej porcji wiedzy.

Specyfika przedsięwzięć budowlanych (zwłaszcza ich niejednorodność) skłania do rozważenia możliwości wykorzystania w procesach zautomatyzowanego pozyskiwania wiedzy rozwiązań hybrydowych [3], obejmujących metody specyficzne dla danych precyzyjnych i kompletnych oraz takich, które umożliwiają odkrywanie wiedzy dla danych rozmytych lub pozyskiwanie wiedzy niezbędnej do realizacji procesu wnioskowania, w przypadku niepełności wiedzy (metoda CBR) – rys. 2.
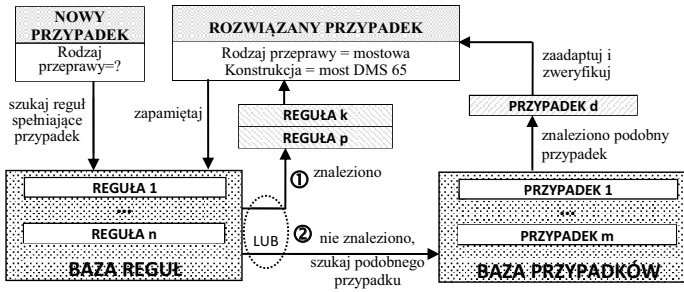
Rys. 2. Istota wykorzystania hybrydowego systemu odkrywania wiedzy w obszarze przedsięwzięć budowlanych

Złożoność przedsięwzięć budowlanych, niepowtarzalność w zakresie parametrów, czy zróżnicowany charakter danych sprawia, że niełatwo jest wskazać pojedyncze rozwiązanie (metodę i narzędzia) odkrywania wiedzy, jako racjonalne w tym obszarze. Zasadną wydaje się próba łącznego wykorzystania kilku metod i właściwych dla nich narzędzi (w tym - języków zapytań stosowanych do odkrywania wiedzy), tworzących system hybrydowy. Powinny one uwzględniać rozwiązania pozwalające na odkrywanie wiedzy z danych rozmytych oraz umożliwiać odkrywanie wiedzy w oparciu o przypadki.

LITERATURA

1. D. T. Larose: Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych, PWN, Warszawa 2006.
2. M. Nycz: Pozyskiwanie wiedzy menedżerskiej. Podejście technologiczne, Wyd. AE we Wrocławiu, Wrocław 2008.
3. J. Bień, Modelowanie obiektów mostowych w procesie ich eksploatacji, Oficyna Wydaw. Politechniki Wrocławskiej, Wrocław 2002