DE GRUYTER OPEN
DOI: 10.1515/ace-2015-0056
ARCHIVES OF CIVIL ENGINEERING
Vol. LXII ISSUE 1 2016
© 2016 by M. Gajzler. This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivs license (http://creativecommons.org/licenses/by-nc-nd/3.0/)

USEFULNESS OF MINING METHODS IN KNOWLEDGE SOURCE ANALYSIS IN THE CONSTRUCTION INDUSTRY

M. GAJZLER[1]

The mining methods are classified as the methods of data analysis and the knowledge acquisition and they are derived from the methods of "Knowledge Discovery". Within the scope of these methods, there are two main variants associated with a form of data, i.e.: "*data*" and "*text mining*". The author of the paper tries to find an answer to a question about helpfulness and usefulness of these methods for the purpose of knowledge acquisition in the construction industry. The very process of knowledge acquisition is essential in terms of the systems and tools operating based on knowledge. Nowadays, they are the basis for the tools which support the decision-making processes. The paper presents three cases studies. The mining methods have been applied to practical problems – the selection of an adhesive mortar coupled with alternative solutions, analysis of residential real estate locations under construction by a developer company as well as support of technical management of a building facility with a large floor area.

Keywords: data text mining, knowledge in construction, DSS, technological decisions, localization problem

1. INTRODUCTION

The engineering activities are based on knowledge, skills and – to a significant extent – on experience [4, 7, 10]. Taking into account the high dynamics of conditions and development of technology and building materials the construction engineer is required to continuously increase their competences by acquiring knowledge, improving skills and gaining experience. These elements, after transformation, and especially generalization, will be a valuable and possibly universal resource and a reliable reference point in case of new problems.

[1] PhD., Eng., Poznan University of Technology, Institute of Structural Engineering, Ul. Piotrowo 5, 60-965 Poznan, Poland, e-mail: marcin.gajzler@put.poznan.pl

Unauthenticated
Download Date | 7/14/16 11:36 AM

On the other hand, time pressure and the considerable number of problems to be solved promote the search of methods, which enable rapid analysis of resources available to an engineer. The number of available resources is significant, but also their form varies sometimes. The mining methods are one of the possible methods of analysis, which enable – in an automated way – to analyze varied in form resources and based on this information they allow to form an image of knowledge. The image of knowledge, created in this way, may be useful for engineering activity. The application of the above mentioned methods can greatly improve solving of tasks and decision-making problems that are associated with the investment process or even with the full life cycle of a building facility. The usefulness of the mining methods is based essentially on the following idea – from the data (which are very often already exploited) to the new knowledge [2, 6, 9, 15]. The attempts of application of the mining methods are presented in three practical problems: the support of material and decisions on technology (selection of an adhesive mortar and some alternatives solutions), the problem of residential real estate locations (determining the attractiveness of locations) and the problems of technical management of a building facility (collection of information about events in the life cycle of a building).

2. METHODS OF DATA ANALYSIS

In order to systematize the concepts of the discussed problems the author presents basic definitions [11]. These concepts are often prioritized in accordance with the so-called "DIKW pyramid" (i.e. Data, Information, Knowledge, Wisdom) (Fig. 1). The analysis will skip the highest prioritized concept – "the wisdom", because this concept already belongs to another area (the philosophical sciences).

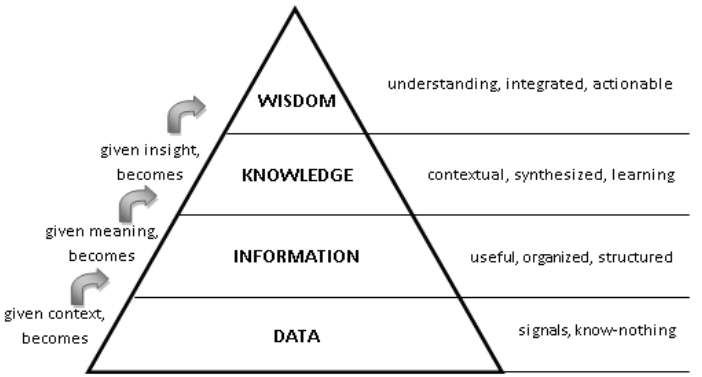

Fig. 1. Pyramid: Data – Information – Knowledge – Wisdom (DIKW)

Unauthenticated
Download Date | 7/14/16 11:36 AM

Based on available data, information and knowledge applies universality to concepts in many fields, including the technical science. Taking into account the hierarchical position of these concepts the data is the first one. The data is often referred to as a raw material or an information carrier. The data are itself a context-free element that only properly interpreted and often enriched with essential descriptions may prove its usefulness. The data in itself are often not useful precisely due to the lack of context. However, there are some methods which enable, even from the context-free data, to obtain some information and knowledge. Thanks to the development of the computer science and the computer technology the exploration processes of dependencies among the data – including the context – proceed relatively smoothly and quickly. The information is the second concept. Together with achieving a higher level in the DIKW pyramid, the recipient's entropy decreases. The entropy, in the information theory, is a measure of the amount of information and it is defined as the average-weighted information carried by a single message, where the weights are the probabilities of sending of particular messages. Referring to the DIKW pyramid it can be concluded that the information is the properly interpreted data (located the level below in the presented pyramid). Their interpretation should be especially useful for the recipient. It's this process which allows data gain appropriate context, which has previously been neglected. The last element in the analysis, and the penultimate in the DIKW pyramid, is the knowledge. Davenport and Prusak presented an interesting definition of knowledge [5] "knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds of knowers. In organizations, it often becomes embedded not only in documents, repositories but also in organizational routines, processes, practices, and norms.". By the definition, it shows the complexity of knowledge which combines many elements – including the contextual information. The second aspect is utilitarianism of knowledge, because it allows leading the processes in conclusion to the new experiences and information. This last value is crucial in the decision making process. Among the methods of data analysis, which (in accordance with the above presented definition) enable the information to create a specific area of knowledge, it can be classified with taking into account the degree of involvement of a person conducting the analysis. Therefore, one can distinguish between a group of "manual" methods – classical, and a group of "automated" methods [10]. The manual methods require human involvement to the extent that a human is the interpreter of data and therefore a human is responsible for transformation of the data into the form of knowledge. In the last stage the image of knowledge firstly supplies the mental model of a person doing the analysis and further down – in the method of knowledge acquisition – it can be acquired

Unauthenticated
Download Date | 7/14/16 11:36 AM

and saved to the formal form. The second group of methods of data analysis – defined as “the automated ones” – allows easing a human in and it focuses on the methods of search and detection of the data dependencies. The special attention should be paid to the methods of data “drilling”– the so-called “the data mining methods”, which derive from the group of methods of "the Knowledge Discovery". These methods rapidly enable, an analysis of a variety of resources, even those potentially used. At the same time, one should be aware of the fact that these methods (in the author’s opinion) only allow fora rough image of knowledge. A broader look at the methods of data analysis is presented in the following works [6, 10].

3. SOURCES OF KNOWLEDGE

Referring to the earlier quoted Davenport and Prusak’s definition of knowledge, the authors state that the knowledge “… it often becomes embedded not only in documents, repositories but also in organizational routines, processes, practices, and norms.". In the author's opinion the sources (included in the definition) do not exhaust all available catalogued sources. The mental model of an expert is one of the most valuable sources of knowledge. The problem might lie in obtaining the knowledge due to the abstractness of "the storage" – the memory in a human brain. In relation to the construction industry there are large groups of different sources of knowledge. At the same time it can be noted that the concept of “knowledge” concerns the already interpreted data and contextual information. The available sources of knowledge were classified into 4 groups. In the first group there are numerous text studies. In the analysis it the following sources of knowledge had not been taken into account: textbooks and books, because they are commonly occurring for each problem area. In the second group there are the numerical statements that define certain measurable values of different factors, such as: financial, factors of construction production, distances, time. In the next group there are observations. The construction industry, as a branch related to the production sphere, allows to observe, the production processes, which in spite of diversified projects, are characterized by a certain repetitiveness, which cannot be said about the entire project. Moreover, the observations may be subject to different events occurring outside the production and investment activity. Certain events might occur during the life cycle of a building facility resulting in events such as: repairs, modernizations, etc. In order to minimize the negative consequences some experiments and simulations are conducted for the real system. The observations of these activities results might also be a valuable source of knowledge. In another group there are the so called "*abstract warehouse*” defined in the analysis, which generally typically require manual approaches.

Unauthenticated
Download Date | 7/14/16 11:36 AM

In the last group there are the virtual models, and among them the tools that are used to collect and process of data/information/knowledge. The BIM models and the universal smart tools (such as the trained artificial neural networks or the knowledge bases in various forms) dominate this group. Table 1 is the overview of the available sources of knowledge in the construction industry, with some comments concerning the form and the potential possibilities and the limitations of the acquisition process and the further formalization.

Table 1. Classification of groups of knowledge sources in construction (own work)

Group of sources	Sources (examples)	Comments
Text data – descriptive	Technical standards	- large and numerous resources, - included expertise knowledge in quality description, which also includes clarifications and detailing, - relate to variety of phenomena and problems, - occur with numeric values giving context.
	Technical specifications of execution and acceptance of construction work, expertise and opinions	
Numerical data	Calculations	- possibility of using quantitative analysis methods (statistics, data mining), - significant relationships between values and units (context of numerical values), - large number of documents.
	Financial analysis	
	Quantitative statements of building materials and labor intensity	
Observations	Observations of executives processes	- need to involve an observer or possibility of recording, - usually a long lasting process of knowledge acquisition, - need to identify repeatable processes in order to observe regularities.
	Experience, experiments, simulations	
	Observations of events	
Abstract warehouses	Experts, experienced engineers	- application of automated methods seems to be limited, - in case of experts – possibility of detailed identify and obtain explanations, - necessary dedicated approach and post-processing.
	Design documentation – drawing part	
Virtual models	Building Information Models	- advanced models, - based on intelligent techniques and tools of computer-aided decision making process.
	Artificial Neural Networks, Knowledge bases	

4. SUPPORTING OF TECHNOLOGICAL DECISIONS

The selection of material and technological solutions in the realization of a construction process is one of the many problems with decision-making character [12, 13, 14]. A specific example of analysis considered a selection of an adhesive mortar for ceramic tiles with high technical properties. Due to construction of the external surface (which will be subjected to frost, rainfall and other external factors) the selection of an adhesive mortar was limited. Despite this, the available number of building materials (from different manufacturers) meeting these requirements is significant. In selection of building materials, contractors often attach too much importance to the

Unauthenticated
Download Date | 7/14/16 11:36 AM

material prices. This results in situations that selection of building material, in terms of quality, turns out to be ineffective and as a result contractors incur significant losses related to the necessity of warranty repairs. The losses are greater due to the loss of building material's expenditures – the adhesive mortar (with lower technical parameters), the loss also includes the additional labour expenditures and other building material's expenditures (i.e. the ceramic tiles). One of author's first hand experiences related to participation in court cases, was when a contractor (purchasing a building material – the adhesive mortar for tiles with a value of more than 70 000 PLN) incurred losses equivalent to more than 3 000 000 PLN. It proves that the decisions related to the selection of appropriate building material and technological solutions are very important. To assist the selection process the text mining analysis was used. The text mining analysis consisted of analysis of the text documents – the technical cards and available materials. The author analysed only fragments of technical cards – the section of technical properties and the material application range. The similar analysis is presented in (6) except that it concerns the repair building materials for concrete structures. Based on the analysis of 18 building materials a tree diagram was created, on the basis of which it is possible to quickly define some alternatives for the selected building materials solutions. The comparative verifying analysis shows that the building materials, shown in the diagram, have very similar properties and range of applications. The meaning of this analysis shows that using of the text mining techniques (by creating a quantitative representation of text documents) maes it possible to use the statistical analysis. This enables, knowledge acquisition about the subject of research, and in the considered example about the examined building materials, without the prior in-depth analysis of the text documents relating to the subject.

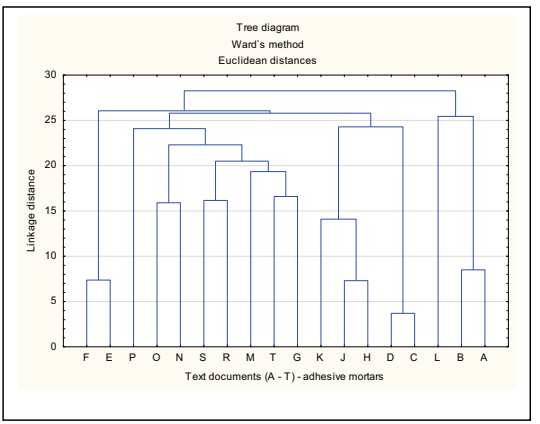

Fig. 2. Diagram of tree elaborated according to Euclidean distance method for 18 analysed technical data cards of adhesives mortars (Data Miner Statistica)

Unauthenticated
Download Date | 7/14/16 11:36 AM

5. SUPPORT OF LOCATION DECISIONS

The support of location decisions is another problem, in which the mining analysis was applied. The comprehensive example of such analysis, together with the theoretical basis, is presented in the following work [1, 9]. The supplement of previously conducted analysis is taking into account the sales results of apartment units for some selected locations in the city of Poznan (as a factor determining the generally understood attractiveness). These data were obtained from two large developer companies operating in the residential market in Poznan. Such analysis enables, on one hand, defining certain selection criteria by potential customers and to analyze the relationships between them, and, on the other hand, it enables to determine the selection location accuracy based on the real sales results of apartment units. It should also be noted that the demand for a selected group of real estates is not only caused by the attractiveness of locations, but it depends on additional factors (e.g. price, apartment standard, size of apartments, preferential conditions) in addition to the impact of individual factors (weight) which is not known.

The analysis includes some selected location factors, described as the distance from the characteristics places, i.e.: distance from the means of public transport (PT), distance from the airport (AP), distance from the kindergarten (KG), distance from the green areas (GA), distance from the noisy streets (NS), distance from the places of entertainment (EC).

These distances have been expressed by the average travel time by car or by the average time to reach them on foot based on the Google Maps in a uniform manner by the factor.

The attractiveness of real estate was classified for the selected 17 locations constructed by the developer companies in the last 7 years for which the data are available. The data concern the following parameters:

- the prices (Price) reduced to the comparative period,
- the standards of apartment units (for all units wih the similar standard – developers' standard),
- the medium apartment floor areas (MS)
- the sales results of apartments (assuming realization of the sales plans during the defined by the company period) (SP).

The attractiveness was classified based on the last factor. One should be fully aware of the fact, that despite the location analysis, such defined attractiveness is a complex function of many factors unrelated to the location.

Unauthenticated
Download Date | 7/14/16 11:36 AM

The data for the analysis are summarized in Figure 3.

Przepisy | Drzewa | Uczenie maszyn | Segmentacja i grupowanie | Text Mining

	1 P_T	2 A_P	3 K_G	4 G_A	5 N_S	6 E_C	7 M_S	8 Price	9 S_P	10 Attractiv.
Loc. 1	2	29	2	2	3	6	66,7	5992	149	H
Loc. 2	2	31	5	2	4	10	73,9	5579	104	H
Loc. 3	3	12	1	2	0	4	54,9	6347	89	M
Loc. 4	4	24	3	1	10	13	79,2	5389	91	M
Loc. 5	1	14	4	10	0	10	53,7	6121	77	L
Loc. 6	10	47	1	1	10	16	77,4	5189	88	M
Loc. 7	12	9	3	1	0	8	74,1	6430	71	L
Loc. 8	4	28	5	0	9	12	68,7	4894	96	H
Loc. 9	5	14	4	3	2	4	71,8	5220	88	L
Loc. 10	2	29	2	2	3	6	66,7	5880	149	H
Loc. 11	2	31	5	2	4	10	73,9	5610	117	H
Loc. 12	3	12	1	2	0	4	54,9	6421	89	M
Loc. 13	4	24	3	1	10	13	79,2	5120	91	M
Loc. 14	1	14	4	8	2	11	55,9	6219	70	L
Loc. 15	8	40	1	1	10	16	76,6	5320	88	M

Fig. 3. Data set used in analysis (part) – the markings: P_T – public transport, A_P – airport, K_G – kindergarten, G_A – green areas, N_S – noisy streets, M_S – main (average) size, Price – price, S_P – sales plan (realization), Attrractiv.- attractiveness

Based on the above table, it is difficult to directly formulate the general conclusions. In order to help, and also to detect invisible at this moment dependences, the mining techniques were used. In the first step the statistics concerning determination of validity of the dependent variables on an output variable – attractiveness were used.

The analysis of this graph shows an interesting relationship. It can be concluded that in the case of a purchase of the apartment unit, the price is not the only factor determining the selection. Among the major factors, there are two, i.e. (N_S and A_P) that can be identified with generated excessive noise. Following this reasoning, it appears that during purchasing, the customers look for peaceful and quiet places, where are relatively close to a kindergarten or other similar institutions. Among the major factors, there is also a price, which seems to be natural.

Apart from such analysis, the mining methods enable creation of other, more advanced, models such as the classification trees, or (in a case of larger number of population) they enable to introduce the predictive models based on the artificial neural networks. Based on them (the location characteristics) some conclusions may be drawn regarding the sales forecasts. Such model can be naturally extended with a range of other factors. The data from the recent years show the behaviour of the banking sector and its stimulation of the housing sector.

Unauthenticated
Download Date | 7/14/16 11:36 AM

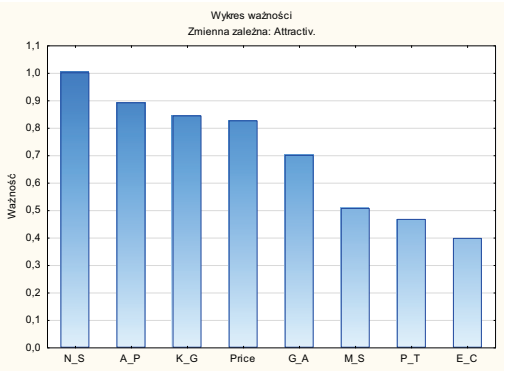

Fig. 4. Graph of validity for an output variable – "Attractiveness" (Data Miner Statistica)

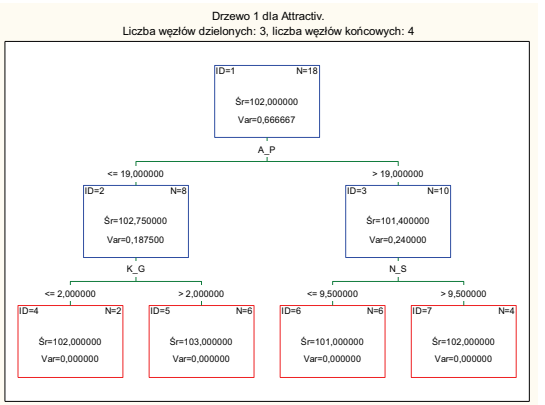

Fig. 5. Example of classification tree for variable – "Attractiveness" (Data Miner Statistica)

6. SUPPORT OF TECHNICAL MANAGEMENT

It is assumed that the life cycle of a building facility is definitely longer than the investment process. During the life cycle of a building facility, maintenance, renovation and repair work are some of the managerial duties which ensure the reliability of a building, as well as the possibility of its efficient use. As a rule, these works consist in conducting regular, required by the legislation, inspections and controls of a building and realization of the recommendations resulting from them. In addition, the problem related to the technical management is appropriate reaction in some incidental situations, i.e. carrying out activities/work consisting of removal of current damages and failures.

Unauthenticated
Download Date | 7/14/16 11:36 AM

During these activities (work) it is valuable to collect information about these events, consequences and the steps taken together with their effectiveness. Because of such gathered information, it is possible to create a tool supporting the management process. This will be another knowledge based tool.

The advisory system HASIFR is an example of such tool. It supports technical management in the selection of building materials and technological solutions for the purpose of conducting the processes of repairs of concrete industrial floor finishing [7, 8]. In case of creating of this system, the knowledge, which is the basis of the system, was gained by the acquisition method (interview, questionnaire) from a domain expert. The current research involves the knowledge acquisition from observations of the real cases. Among others the mining methods are applied in this knowledge acquisition.

The research is done in the warehouse of the automotive wholesale network. These warehouses are the high bay stores. The specificity of these warehouses is done using them as storage for, in addition to spare parts, a large number of chemicals (i.e. oils, fluids, etc.). During the service life of these warehouses some minor damages of building elements and its equipment associated with the movement of forklifts and during unloading of supplies can be observed. In addition, the failure of automotive chemicals packaging is a common phenomenon, which usually leads to further oiling of the floor, penetration of liquids with varying degrees of chemical aggression into the floor, especially in the place of the earlier damage. As a result, in extreme cases, the repairs must be done causing the need for exclusion of some part of a building..

In order to obtain knowledge about: incidental events and their types, their frequency, time of occurrence and consecutively: response methods, time of repair of damage/failure, it a model form was developed – a card of event notification. The layout of the card is modular and prepared in a special way so that there was no lapses of important information related to the different character descriptions. This research uses the text mining techniques. The scheme of the research run is shown in Fig. 6. The research uses the alternative technique OCR (Optical Character Recognition) in order to quickly digitize the paper text documents and further the possibility of applying the text mining techniques.

Currently, the sample of cards is not large enough to conduct a meaningful analysis. In the author's opinion the selection of the problem in order to illustrate the method is not accurate. It turns out that the frequency of incidental events is low (Table 2).

Unauthenticated
Download Date | 7/14/16 11:36 AM

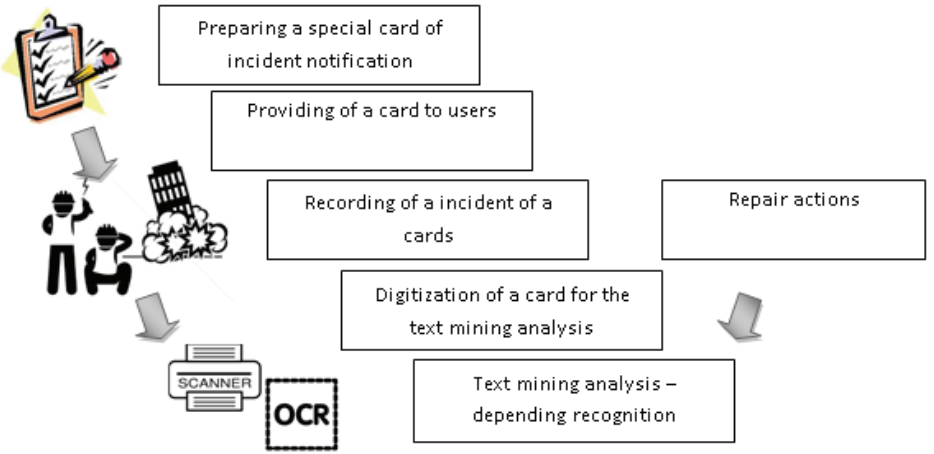

Fig. 6. Scheme of research recording incidental events

Table 2. Recording incidents in a building

Incident	Number of incidents in examined buildings during research period (7 buildings /16 months)
Damage of rack post cover	17
Oiling of floor, non-neutral substance	15
Damage of landing ramp	9
Damage of gate	4
Others	6

In terms of the effectiveness of the approach/method some conclusions may be formulated in order verify in practice such approach. In the author's opinion, the purposefulness of collecting of information about a building and events is not subject to questioning. In this case, you can rely on the BIM technology, as early as at the design stage of a building, the gathering of information about the building is done. This process is gradually developed and continued at the stage of building facility service life, and the model itself is actively used. In the author's opinion, in case of buildings without the BIM model, it is appropriate to build such models as the secondary ones, even at the stage of their service life. There are some advanced methods and tools (e.g. a laser scanning) allowing a quick way to get a digital image of a building structure, which is important in improving of this process [3].

Unauthenticated
Download Date | 7/14/16 11:36 AM

7. CONCLUSIONS

Based on the analysed cases, it is problematic to clearly determine the helpfulness and usefulness of the mining tools in the analysis of available sources in the construction industry. On the one hand, these techniques have managed relatively well and quickly with a large number of text documents, but on the other hand, there is a reasonable concern that the non-text data are blurred, deprived of context and finally removed from the analysis. In this situation, the analysis of text documents (that contain a significant share of the relevant numerical values) becomes shallow and ineffective.

The prior deliberate elaboration/preparation of cards is a certain remedy for this problem. The cards should be elaborated in such a way as to either minimize the share of the non-text values or cause the separation of these values from the text data so that it is possible to conduct separate analysis. The earlier elaboration/preparation of cards is not always possible, and it is usually time-consuming and laborious, besides it generates some additional processes associated with the data processing. In relation to the data mining techniques, it might be indicative of the helpfulness and usefulness of this technique, especially in the analysis of a large number of data and numerical summaries. According to the idea of the data mining even the "worn-out" data give a chance to find out some new relationships and knowledge. The analysed examples can prove that.

In summary, providing an answer to the thesis about the helpfulness of the mining methods however we should be inclined to confirm this thesis. These methods are useful and worth of applying, however they have some limitations. The above mentioned preparation of some forms-models and cards can be used as remedies to the limitations as well as the applications of a hybrid approach in the knowledge acquisition, where other knowledge acquisition approaches can bu used in parallel to the mining methods, even the classical ones.

Unauthenticated
Download Date | 7/14/16 11:36 AM

REFERENCES

1. Bagočius, V Zavadskas, K. E., Turskis, Z., Selecting a location for a liquefied natural gas terminal in the Eastern Baltic Sea, Transport, Vol. 29, Issue 1, p. 69-74, 2014
2. Berry M., Linoff G., Data mining techniques for marketing, sales and customer support. Wiley, New York, 1997
3. Braun, A. Tuttas, S. Borrmann, A. Stilla, U., A concept for automated construction progress monitoring using BIM-based geometric constraints and photogrammetric point clouds. Journal of Information Technology in Construction, Vol. 20, p. 68-79, 2015
4. Brown, C.B.Elms, D.G. Engineering decisions**:** Information, knowledge and understanding. Structural safety, Vol. 52, Issue PA, p. 66-77, 2015
5. Davenport T. H., Prusak L., Working Knowledge, Harvard Business School Press, Boston MA, 1998
6. Gajzler, M. Text and data mining techniques in aspect of knowledge acquisition for decision support system in construction industry. Technological and Economic Development of Economy (2/2010), p. 219 – 232, 2010
7. Gajzler, M. The idea of knowledge supplementation and explanation using neural networks to support decisions in construction engineering. Procedia Engineering, Vol. 57, p. 302-309, 2013
8. Gajzler, M. The support of building management in the aspect of technical maintenance. Procedia Engineering, Vol. 54, p. 615-624, 2013
9. Gajzler, M. Zagadnienie wyboru lokalizacji z wykorzystaniem metodyk data mining. Budownictwo i Inżynieria Środowiska, Vol. 2, No. 3, p. 253 – 261, 2011
10. Gajzler, M. Analysis of knowledge sources and processing in the construction area. Technical Transactions, Y. 111, Issue 1B, p.137-144, 2014
11. Grabowski, M., Zając, A., Dane, informacja, wiedza – próba definicji. Zeszyty Naukowe/Uniwersytet Ekonomiczny w Krakowie, no. 798, p. 99-116, 2009
12. Huo, Z.-G. Zhou, Z.-G. Approaches to multiple attribute decision making with hesitant fuzzy uncertain linguistic information. Journal of Intelligent and Fuzzy Systems, Vol. 28, Issue 3, p. 991-998, 2015
13. Książek, M.V., Nowak, P.O., Kivrak, S., Rosłon, J.H., Ustinovichius, L., Computer-aided decision-making in construction project development. Journal of civil engineering and management, Vol. 21, Issue 2, p. 248-259, 2015
14. Monghasemi, S., Nikoo, M.R., Khaksar Fasaee, M.A., Adamowski, J. A., Novel multi criteria decision making model for optimizing time-cost-quality trade-off problems in construction projects. Expert systems with applications, Vol. 42, Issue 6, p. 3089-3104, 2015
15. Shu-Hsien Liao, Pei-Hui Chu, Pei-Yuan Hsiao, Data mining techniques and applications – A decade review from 2000 to 2011. Expert systems with applications, Vol. 39, Issue 12, p. 11303–11311, 2012

Received 22. 06. 2015

Revised 14. 07. 2015

Unauthenticated
Download Date | 7/14/16 11:36 AM

LIST OF FIGURES AND TABLES:

Unauthenticated
Download Date | 7/14/16 11:36 AM

UŻYTECZNOŚĆ METOD MININGOWYCH W ANALIZIE ŹRÓDEŁ WIEDZY W BUDOWNICTWIE

Słowa kluczowe: data i text mining, wiedza w budownictwie, wspomaganie podejmowania decyzji, decyzje technologiczne, zagadnienie lokalizacji

STRESZCZENIE:

Metody miningowe są klasyfikowane jako metody akwizycji wiedzy wywodzące się z metod „knowledge discovery". W zakresie tych metod występują odmiany: data oraz text mining. Artykuł staje przed próbą odpowiedzi na pytanie o ich użyteczność na potrzeby akwizycji wiedzy w budownictwie. Proces akwizycji wiedzy jest nieodzowny w aspekcie operowania systemami i narzędziami bazującymi na wiedzy. Stanowią one aktualnie podstawę narzędzi wspomagających podejmowanie decyzji. Sformułowane w oparciu o analizę przypadków wnioski wskazują na przydatność tych technik, jednocześnie definiując pewne ograniczenia związane z ich stosowaniem. Elementem wniosków są metody redukcji ograniczeń, m.in. poprzez stosowanie podejścia hybrydowego w procesie akwizycji wiedzy. Przydatność metod analizy miningowej scharakteryzowano na trzech zagadnieniach: pierwsze dotyczy wyboru materiału budowlanego – zaprawy klejowej o wysokich właściwościach technicznych, a także znalezienia materiału alternatywnego. Drugie z zagadnień dotyczy problematyki wyboru lokalizacji kolejnych inwestycji mieszkaniowych przez przedsiębiorstwo deweloperskie na podstawie danych wynikających z wcześniejszych doświadczeń – lokalizacji, realizacji i sprzedaży nieruchomości mieszkaniowych. Trzecie zagadnienie dotyczy pozyskiwania wiedzy o zdarzeniach incydentalnych obiektach magazynowych aspekcie zarządzania technicznego.

W przypadku pierwszego zagadnienia – wyboru materiału oraz rozwiązania alternatywnego wykorzystano metodykę text mining. W analizie oparto się na częściach opisowych kart technicznych materiałów różnych producentów o różnym poziomie wartości parametrów technicznych. Poprzez budowę reprezentacji ilościowej tekstu (macierze BOW – *Bag Of Words* i kolejno jej symplifikacje) dotyczącego opisu materiałów, zastosowano metody grupowania (met. odległości euklidesowej), w celu wykazania podobieństw między materiałami w zakresie zastosowań i właściwości technicznych. Uzyskano w rezultacie wykres drzewa, w oparciu o który możliwy był wybór rozwiązań alternatywnych. Proces wsparty był poprzez oprogramowanie Data Miner Statistica. W oparciu o przykład sformułowano wnioski o przydatności metody, ale i jednoczesnym ograniczeniu zastosowania ze względu na wysokie prawdopodobieństwo utraty części danych i zanikowi ich kontekstu (dane ilościowe). Okazało się, że w trakcie obróbki tekstu występuje proces podziału słów i sprowadzenie ich do tzw. rdzenia. W procesie tym często dochodzi do rozdzielenia słów, w tym i rozdzielenia wartości liczbowych od kolejnych słów czy symboli. Prowadzi to do sytuacji, w której istotne dla budownictwa wartości parametrów ulegają zagubieniu czy tracą zupełnie kontekst i tym samym wartość dla problemu.

Kolejny przypadek analizy miningowej został zaprezentowany na bazie danych uzyskanych z przedsiębiorstw deweloperskich. Istotnym elementem warunkującym późniejszy efekt finansowy dla przedsiębiorstwa jest staranny wybór przyszłej lokalizacji inwestycji mieszkaniowej. W tym przypadku wykorzystano dane dotyczące wcześniej realizowanych inwestycji, gdzie czynnikami wejściowymi były określone odległości, natomiast poprzez realizację zakładanego planu sprzedażowego wskazywano na atrakcyjność lokalizacji. Do analizy zastosowano metodyki data mining. Metody te są bardzo bogate jeśli uwzględnić możliwości modelowania zjawiska - od prostych statystyk po złożone modele inteligentne. Wybór dostępnych modeli podyktowany jest z reguły charakterem problemu, a także dostępną próbką danych. W analizowanym przypadku nie dysponowano liczebną próbką, aby zastosować bardziej zaawansowane metody.

Unauthenticated
Download Date | 7/14/16 11:36 AM

Wykorzystano analizę istotności cech wejściowych i ich wpływu na zmienną wyjściową „atrakcyjność" oraz dalej zbudowano przykładowe drzewo klasyfikacyjne. Wnioski płynące z analizy pokazują interesującą zależność, że to wcale nie cena jest czynnikiem decydującym. Bardzo istotą wagę przykłada się do czynnika związanego z oddziaływaniem akustycznym – hałasem, odległością od źródeł hałasu (ruchliwe arterie komunikacyjne, lotnisko). W aspekcie aktualnej sytuacji w Poznaniu, gdzie z rozwojem lotniska cywilnego (Ławica), a wcześniej wojskowego (Krzesiny) wprowadzono obszary ograniczonego użytkowania, wnioski te wykazują się aktualnością. Czynnik oddziaływania akustycznego był i jest często podnoszony w mediach w aspekcie roszczeń odszkodowawczych, co powoduje wzmożoną „czujność" potencjalnych klientów. W zakresie metodyki data mining wskazano na jej dostępność i łatwość, a także istotną efektywność w zakresie pozyskiwania wiedzy o zjawisku.

Trzecie zagadnienie związane z zarządzaniem technicznym obiektami magazynowymi dotyczyło pozyskiwania wiedzy o zdarzeniach incydentalnych w obiekcie. Zdarzeniom tym zwykle towarzyszy adekwatne działanie naprawcze. Sytuacja taka stanowi przykład pozyskiwania wiedzy z obserwacji. Zastosowanie metod miningowych miało na celu zautomatyzowanie tego procesu. Uwzględniając wnioski z wcześniejszych przypadków, a zwłaszcza utraty części danych i ich kontekstu przy obróbce text mining, opracowano wzór formularza – karty zgłoszenia zdarzenia. Kolejno wykorzystano ogólnodostępne techniki OCR (*Optical Charakter Recognition*) w celu digitalizacji dokumentów. Tak przygotowane dokumenty w wersji elektronicznej zostały poddane analizie miningowej.

Na podstawie przeanalizowanych przypadków trudne jest jednoznaczne określenie przydatności i użyteczności narzędzi miningowych w analizie dostępnych źródeł w budownictwie. Z jednej strony techniki te stosunkowo dobrze i szybko poradziły sobie z dużą ilością dokumentów tekstowych, z drugiej strony istnieje uzasadniona obawa, że dane pozatekstowe ulegają rozmyciu, pozbawienia kontekstu i finalnie usunięciu z analizy. Pewnym środkiem zaradczym jest uprzednie umyślne opracowanie formularzy w taki sposób, aby albo minimalizować udział wartości pozatekstowych lub spowodować rozdział tych wartości od danych tekstowych, tak aby możliwe było przeprowadzenie odrębnych analiz. W odniesieniu do technik data mining można wskazać na przydatność wysoką użyteczność tej techniki, a zwłaszcza w analizie dużej liczby danych i zestawień liczbowych. Zgodnie z ideą data mining nawet „wyeksploatowane" dane dają szansę poznania nowych zależności i wiedzy. Przeanalizowane przykłady są tego dowodem.

Reasumując i odpowiadając tezie o przydatności metod miningowych należy się jednak skłonić ku jej potwierdzeniu. Metody te są przydatne i warte stosowania z tym, że należy się liczyć z pewnymi ograniczeniami. Środkiem zaradczym jest wspomniane przygotowywanie pewnych szablonów i formularzy lub też stosowanie podejścia hybrydowego w procesach akwizycji wiedzy, gdzie równolegle z metodami miningowymi stosowane będą inne podejścia akwizycyjne, nawet choćby klasyczne.

Unauthenticated
Download Date | 7/14/16 11:36 AM