TECHNOLOGICAL AND ECONOMIC DEVELOPMENT OF ECONOMY
Baltic Journal on Sustainability
2010
16(2): 219–232

TEXT AND DATA MINING TECHNIQUES IN ASPECT OF KNOWLEDGE ACQUISITION FOR DECISION SUPPORT SYSTEM IN CONSTRUCTION INDUSTRY

Marcin Gajzler

Poznan University of Technology, Piotrowo 5, 60-965 Poznan, Poland
E-mail: marcin.gajzler@put.poznan.pl

Received 12 October 2009; accepted 27 April 2010

Abstract. This article presents the possibilities of using mining techniques in building Decision Support Systems. One of the biggest problems is the issue of gaining data and knowledge, their mutual representation and reciprocal usage. Data and knowledge make up the resources of the system and are its key link. It has been estimated that 70% to 80% of the sources available for general use are text documents. The text mining technique is defined as a process aiming to extract previously unknown information from text resources (e.g. technological cards). The fundamental feature of text mining is the ability to converse text documents in formal form, which opens up great possibilities of conducting further analysis. This article presents chosen IT tools using text mining technique, along with the elements of the text mining analysis. The main objectives are the simplification of the process of knowledge acquisition, its automation and shortening as well as the creation of ready-made models containing knowledge. Previous tests with knowledge acquisition (surveys, questionnaires) were time-consuming and exacting for experts.

Keywords: decision support systems, knowledge acquisition, text mining, AI models, advisory system.

Reference to this paper should be made as follows: Gajzler, M. 2010. Text and data mining techniques in aspect of knowledge acquisition for decision support system in construction industry, *Technological and Economic Development of Economy* 16(2): 219–232.

1. Introduction

Present trends which can be observed in construction industry indicate an increased share of automation and information technology in many branches of economy (Boddy *et al.* 2007; Chau, Albermani 2003; Chen 2008; Hanna, Lotfallah 1999; Hola, Schabowicz 2007; Schabowicz and Hola 2008; Jang, Skibniewski 2008; Kaplinski 2007, 2009). The time of response to a stimuli is unquestionably essential. The desire is that the trend is as short as possible, the

ISSN 1392-8619 print/ISSN 1822-3613 online
doi: 10.3846/tede.2010.14
http://www.tede.vgtu.lt

response appropriate and desired, and at the same time the load on human as low as possible. From this point of view one can observe an effective development of systems that assist human actions. These include systems assisting decision making with rich – today – internal classification (expert systems, assisting, management, diagnosing etc). Different aspects connected with building systems assisting the decision-making are discussed in this article. Addressing the issue related to the DSS, the problem of data and knowledge acquisition often arises (Anderson 1996; Kaklauskas *et al.* 2007; McCovan, Mohamed 2007; Naimavičienė *et al.* 2007; Shaked, Warszawski 1995; Ping Tserng, Lin 2004; Ustinovichius *et al.* 2007; Zavadskas *et al.* 1995). Due to the fact that DSS operate within the decision problems, which is often described by linguistic variables thus having the features of fuzziness, data and knowledge acquisition is difficult and time-consuming. From this point of view it is sensible to apply automatic knowledge acquisition techniques. These techniques can be text and data mining (Berry, Linoff 2000; Cohen *et al.* 2003; Creese 2004; Fayyad *et al.* 1996; Feldman 2006; Haerst 1999). A large number of text sources (technical papers, specifications, etc.) creates an excellent proving ground for the above-mentioned techniques.

2. DSS resources – data and knowledge bases

Data and knowledge are among the basic resources of the assisting systems. Analyzing the structure of any expert system (here it will be understood as a model of DSS system) we can differentiate two blocks – data base and knowledge base. Both of them can be updated in a certain way. It is unquestionably a condition of the whole system being up-to-date.

Modern techniques allow, among other things, to constantly monitor given data structures and directly monitor the fluctuations of phenomena (sensors working on the basis of wire and wireless communication, constant search for resources) (Cheng *et al.* 2008; Jang, Skibniewski 2008; Maas, Vos 2008; Paslawski, Karlowski 2008). In case of expert systems in which the knowledge came directly for the field expert, the updating of knowledge was problematic. An example of such problem is a Hybrid Advisory System (Fig. 1), in which the knowledge base is a derivative of the mental model of a field expert (Gajzler 2008a, 2008b). By knowledge acquisition (methods of direct intelligence supported by paper form), this knowledge was recorded in form of rules. It is worth to mention here that only the verbal model was subject to acquisition which was a part of the mental model. In the analysis of this case the problem of knowledge updating was discussed in two ways.

The first was to guarantee the most universal profile of the knowledge contained in the knowledge base. Due to this, among other things, the fuzzy sets were used which took into account the linguistic variables that had a meaning range. The fuzzy variables were less sensitive to not being up-to-date than the quantitative (sharp) ones. A second solution which had been proposed in this thesis was to perform the acquisition session again followed by a repeated processing of the gathered data. For obvious reasons the second solution was more problematic and unquestionably more energy- and time-consuming. Due to this the presented problem concerns methods of data and knowledge updating contained in the resources of DSS system and the structure itself of the primary bases of the system.

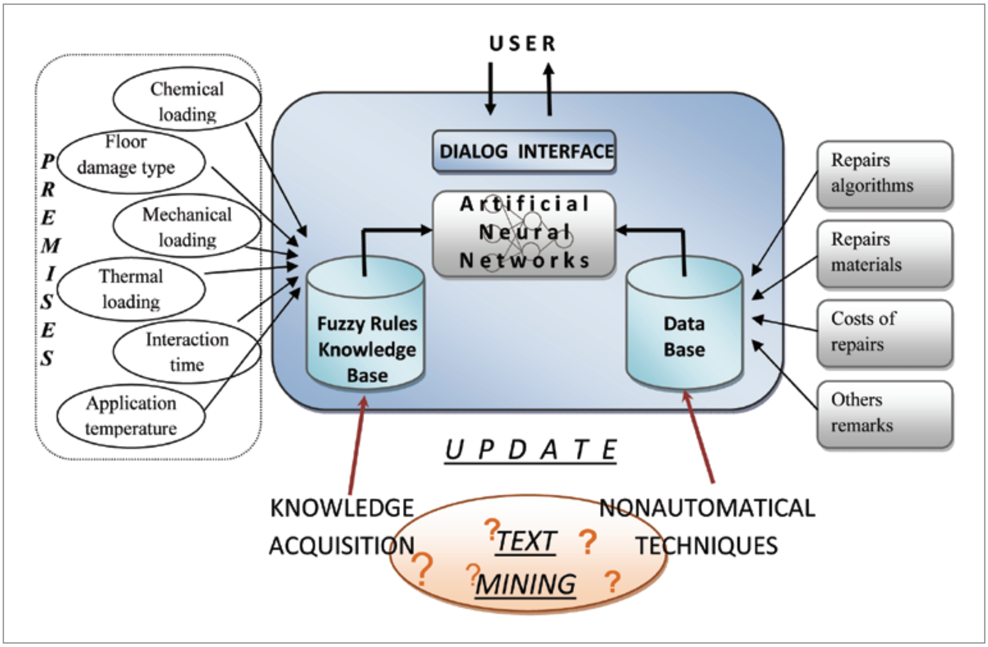

Fig. 1. Elements of the Hybrid Advisory System (HAS) for industrial floor repairs

While addressing the problem, it is worth mentioning the differences in representation of data and knowledge. We have two main methods of knowledge representation. The first one is symbolical representation which is further composed of procedural and declarative representation. The procedural representation lets on the definition of the procedures set representing the knowledge domain (e.g. procedures, ascertainments). The declarative representation consists in the definition of the set specific for the analyzed domain of facts or rules (e.g. semantic networks, rules, frames). The second one is non-symbolical representation with AI methods. As we can see, there are many possibilities. When it comes to knowledge, one of the most frequently and widely used representation is the rule representation. It is a natural representation of the expert knowledge. It is possible to make a direct recording based on the expert declaration. Obviously, it is not that simple and, in reality, the process of knowledge base building is quite complex (Mulawka 1997; Zavadskas *et al.* 1995; Hajdasz 2008 a, b).

3. Knowledge acquisition

One of the stages along the way of knowledge database building is – amongst other things – the stage of knowledge acquisition. It is preceded by the preparation activities i.e. the ones aimed at recognizing the problem or choosinge the representation. The Acquisition stage is associated with the selection of the knowledge source. The acquisition itself can be performed in various ways (Fig. 2).

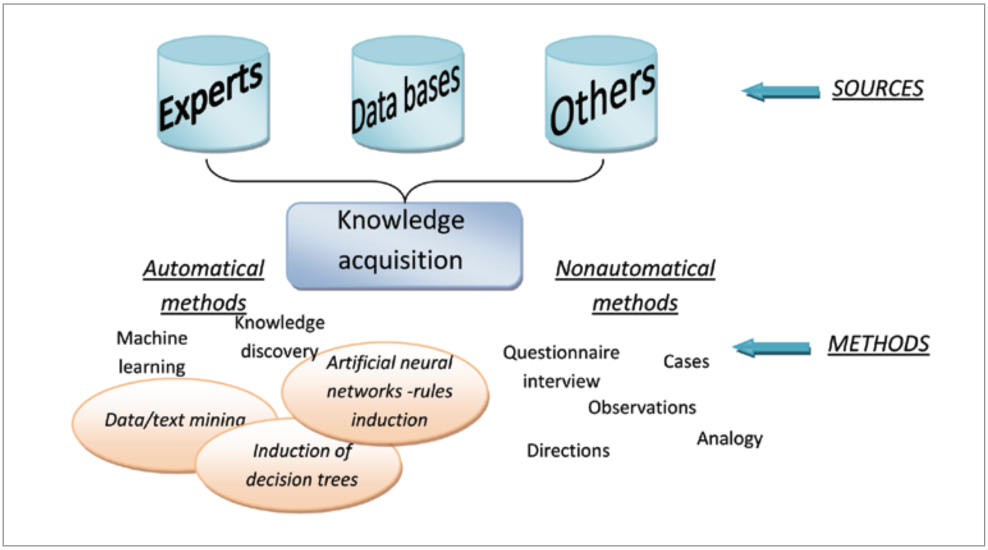

Fig. 2. Methods of knowledge acquisition

Current experience of the author is associated in this matter with the subject of the field expert. In the course of building the Hybrid Advisory System (HAS), the main source of knowledge was the expert – engineer having many years of experience in the analyzed problem (repair of concrete industry floors). The particular type of acquisition of expert knowledge indicated some weak points of such approach. The expert himself had wide knowledge of the problem. This knowledge had been contained in the so-called mental expert model. Unfortunately, at the stage of acquisition the knowledge was collected in the verbal model. The verbal model due to its language limitations and stiff terms is poorer than the mental model. In addition, the recording itself had its limitations which, again, made the knowledge poorer.

During the process of knowledge acquisition, the text sources have provided some support. They have been some sort of fill-up of the verbal model and have also been a sort of supplements of the expert's knowledge. In the further process of the database building these sources played a role of a form of verification of the recorded knowledge. It was then when the idea of text data was born, without an expert, in order to build the knowledge database. Still, one problem remained: in what way this could be done in order to obtain a relatively good effect with low cost and labor? The text mining techniques brought the solution to the problem.

4. Text documents and mining techniques

It has been estimated that about 70–80% of the information and knowledge resources are text sources. They are very rich and numerous, they can be stored or archived and, first of all, they are comprehensible to a human who can easily process them. Besides these advantages, the text documents have a number of disadvantages. The most important ones include: the increasing level of their number, multilinguisticness, noisiness and difficulties in the assessment of the quality of the information contained in the text.

In the field of building large numbers of technical papers function in relation to given products, specifications and directions of the building works, technical instructions for machines and devices. A kind of advantage of these documents, as it will be seen in case of text mining, is often their structured and ordered profile. This allows for efficient savings in time and labor in the analysis of these documents.

What, in the light of that, are the aforementioned text mining techniques? These techniques are quite young because the first information about them dates back to the literature of the 1990s. The term for data mining appeared much earlier. In this context the text mining technique is treated as a variant of data mining technique in relation to the text sources. According to author's opinion the next step in the text mining techniques evolution can for example be photo/view mining, which is a technique related to the sources of information and knowledge in the form of graphic documents (photographs, pictures).

Marti Haerst is one of the inventors of the text mining technique who defined it as a process aimed at extracting from the text resources the previously unknown information. Relating this definition to data mining one can easily notice a difference in the resources from which the information is gained. In case of data mining these are the resources with a defined data structure with values expressed with classic measuring scales, whereas for text mining they are text resources often without a defined structure and, above all, expressed in linguistic variables. The idea, however, is common – the exploration of data and knowledge.

So how is the text mining process built? This process consists of several stages, starting from defining the goal and the analysis scope by conversion of text documents and performing calculations after the interpretation of the results. The main kernel of the text mining analysis is brought to the conversion of text documents to a form convenient for analysis and to performing of adequate calculations (analysis). The method of the conversion of text documents is shown in Fig. 3.

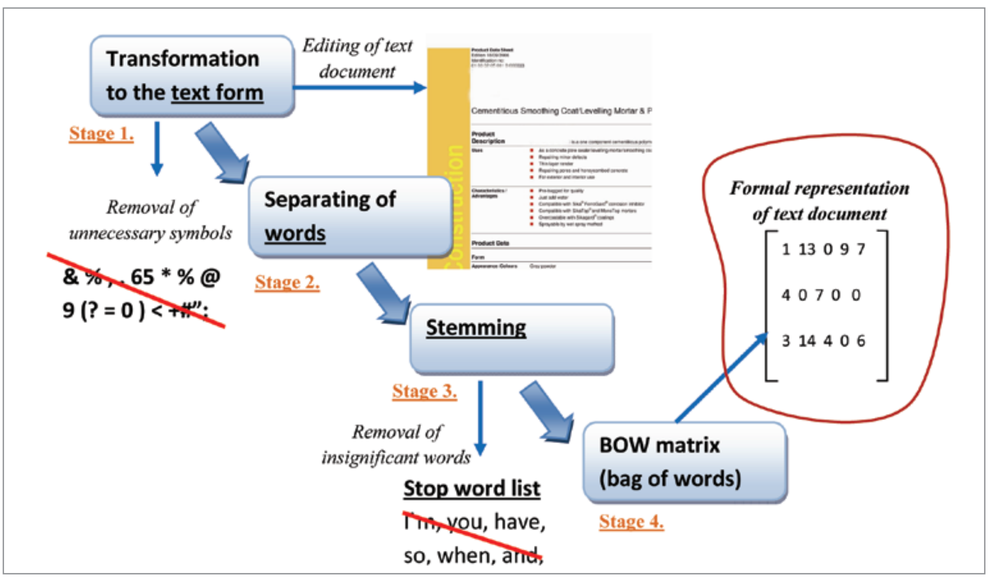

Fig. 3. Main stages of the text mining process

As we can see in Fig. 3 the text mining process is a multistage process aimed at creating a formal representation of the text document in the form of frequency matrix (BOW), which most often becomes a database in the further stage of analysis. Besides the frequency matrix, there can also be more complex representations e.g. considering complex terms or the order of term occurrence.

The first stage of the method is a transformation of the document to the text form i.e. removal/substitution of all unnecessary symbols, removal of formatting signs. The second stage is the division of the documents into words. The next one is reduction to core (stemming) i.e. bringing the words to their basic form. Often the reduction to core is accompanied by the stage of elimination of insignificant words thanks to the application of the stop-list i.e. the list containing words insignificant from the matter point of view. Obviously, on the basis of the reverse rule in relation to the stop-list the document can be analyzed in order to limit it only to the significant words. In each case it is necessary to build such a list. After this stage it is possible to count the presence of a particular word in the given document and as result to create the matrix of the frequency (BOW), which after eventual conversions can be subject to further analysis leading towards the analysis conclusions.

As we can conclude from Fig. 4, the BOW matrix contains in its columns the frequencies of occurrence of particular words in a document and the number of columns considers the general amount of words in a document. This gives for a set of average text documents an extremely large matrix. In order to avoid it, the reduction is applied. First such example has already been given – it was a stop list and reduction to core (stemming). Another extremely valuable and interesting possibility is the analysis of main contents and the decomposition in characteristic values – SVD (Singular Value Decomposition) analysis (Fig. 4). A new coordinate system is built, new components are selected and as a result we obtain a new matrix with significantly reduced dimensions. One disadvantage for the matrix operating with the characteristic values is the difficulty and actually impossibility of interpreting them. The frequency matrices and their derivatives (binary logarithmic) are easy to interpret, whereas the matrix with characteristic value does not give such opportunity.

In this way we obtain the formal form of the text document, which is subject to analysis by means of available statistic or intelligent methods in order to conclude some regularities.

The next part of the article presents an example of analysis of text mining for a set of several text documents (technical specification of building materials – materials for building and repairs of industrial floors and concrete constructions). After the frequency matrix is

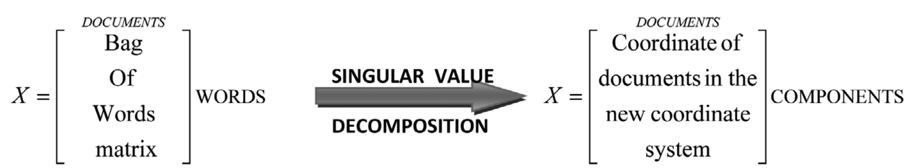

Fig. 4. BOW matrix and the method of SVD decomposition

obtained, techniques are presented, which allows for extraction of knowledge so that it can be used in the building of DSS system. These techniques are:

– classification trees based on which it is possible to generate the rules;

– taxonomic model allowing for learning the structure of the documents set.

Obviously, they are not the only possibilities. Another example can be artificial neuron networks. They, however, in order to operate correctly require a larger number of cases (number of text documents).

The analyzed example is limited only to 19 documents and, from this point of view, the technique was not presented because on this basis an insufficient number of teaching and checking cases would have been obtained.

5. Example

During the building of the knowledge base with a hybrid system for repairs of industrial floors, the author faced several problems. One of them was the aspect of accruing the knowledge and data. Primarily, in order to acquire the knowledge, a number of sessions with an expert had been carried out. It was in a way tiresome for both subjects.

Faced with large amounts of information available on the market and some structuring of that information, an attempt was made with the use of text mining technique. In order to do it a set of 19 technical papers for materials for repairs of concrete construction and for production of industrial floors was used. All calculations were performed in the Statistica StatSoft environment. The profile of the analyzed problem can be determined as the problem of non-pattern classification – for the taxonometric problem (analysis of concentrations) and as pattern in case of classification trees.

Stage 1 – building of formal representation of text documents

In case of later process of pattern classification as well as non-pattern one, it is necessary to build formal representation (matrix) for text documents. This process has already been discussed in chapter 4. It is the essence of the mining technique, which is the building of formal representation of the text document.

The analysis was performed for 19 text documents – technical papers. These documents came from producer and were highly organized and similar in their structure (Fig. 5).

Due to practical reasons (faster analysis and lower requirement of computer memory), only parts of the documents were analyzed. The chosen parts were: material name, description and application. The remaining parts of the technical papers were not considered. The software used for the analysis had the abilities to analyze the parts of texts beginning with and ending with particular phrases. In addition, the software could also read the information form of pdf documents. However, it needs to be said that the mechanism is not perfect yet. Having this in mind the created documents with txt extension were used. As a result, a sheet was created the rows of which listed the consecutive text documents (technical papers ad material associated to them), and the columns reflected the information contained in these documents (name, description, application). This sheet was the basis for the proper text mining analysis according to the description in chapter 4.

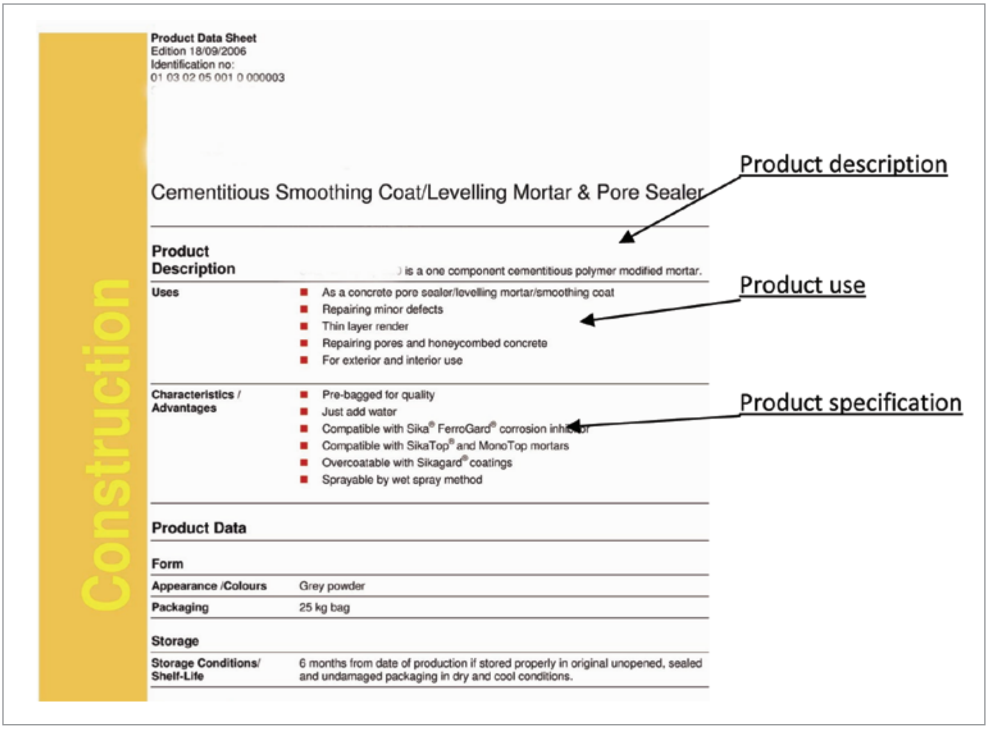

Fig. 5. Frame of text document – technologic sheet of repair-material

For a set of 19 text documents (in Polish language) with the text mining method, a frequency matrix was built as well as the derivative matrixes (logarithmic). These matrices had gigantic dimensions (19*338), which could result in slowing down further analysis. Therefore, a SVD decomposition with LSA (Latent Semantic Analysis) method was made, which resulted in significant reduction of the dimensions of the matrix. During the LSA there is a possibility to individually determine the number of new variables on the basis of a graph. This graph presents the rate of significance/importance of the new variables in the analysis (Fig. 6). As we can see, the graph slows decreases on the right-hand side, which suggests a decrease of significance of new contents. At some stage it is possible to cut off the less significant contents and, thus, decrease the dimensions of the space.

As a result of stage 1 a representation of text documents in formal form was obtained. These representations are:

– frequency matrix (Fig. 7) (and the derived logarithmic matrix);

– matrix for peculiar values (Fig. 8).

Stage 2 – non-pattern classification

The main aim of this analysis is to learn the structure of the set of 19 documents without learning the contents. In order to do that, a method of cluster analysis was used. Its effect will be a classification of the documents into groups of the most similar ones. The basis of the method is to perform successive fusions of n units into particular groups. The starting point

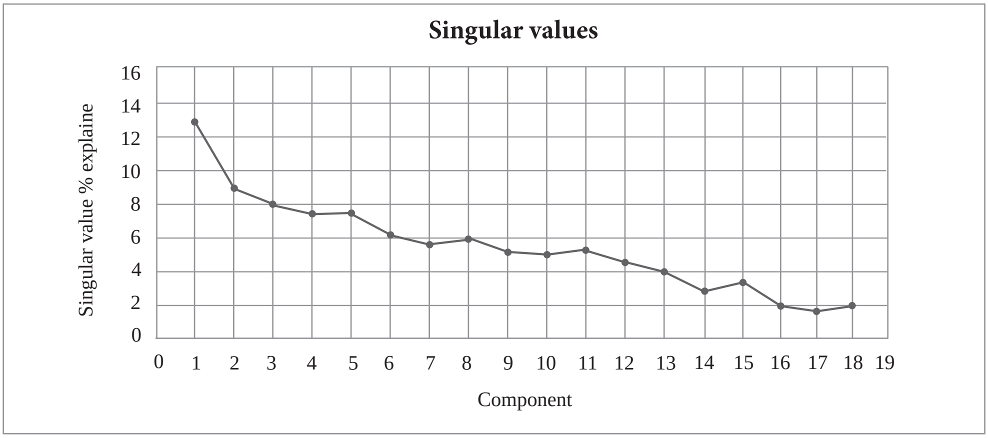

Fig. 6. Importance of following components in SVD analyze

	Word occurrences in files (SikaPL)		
	agression	acrylic	alkaline
1	1	0	0
2	1	0	1
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	1	0	0
9	1	0	0

Fig. 7. BOW matrix (fragment)

	SVD document scores (SikaPL)		
	Component 1	Component 2	Component 3
1	0.177506	0.069398	–0.358954
2	0.187124	0.068406	–0.340186
3	0.423287	0.134466	–0.243274
4	0.423287	0.134466	–0.243274
5	0.211350	–0.119085	0.388097
6	0.195781	–0.130385	0.396237
7	0.314950	0.062797	–0.013913
8	0.333534	0.003830	0.193024
9	0.331495	–0.015100	0.250495

Fig. 8. SVD matrix (fragment)

is the resemblance matrix of the units which make the tested population, which is determined on the basis of the approved resemblance rate. The Euclid's distance was the resemblance measure in the analyzed case (1).

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}.$$

The grouping itself was made on the basis of Ward's methods. Other known methods for clustering are: method of closest proximity, method of furthest proximity and centroidal method.

The applied Ward's method relies on grouping of objects (documents) on the basis of minimizing the sum of squares of variations of any two clusters, which can exist at every stage of the analysis. This method is very effective but often unable to identify groups with large range of variations of particular features and, due to that, creates clusters with low magnitude.

The result of such analysis of clusters is the hierarchical tree (Fig. 9), which indicates resemblances in particular documents. For the grouping itself we can be aided by another graph (Fig. 10). It allows for determining the place of cut of the hierarchical tree in order to separate particular groups. The cut places can be effectively chosen in the place on the graph where the greatest jump is visible.

In relation to the results obtained in the analysis, there is a visible resemblance in the documents labeled as "5" and "6", further "8" and "9", consecutively "1" and "2". The documents

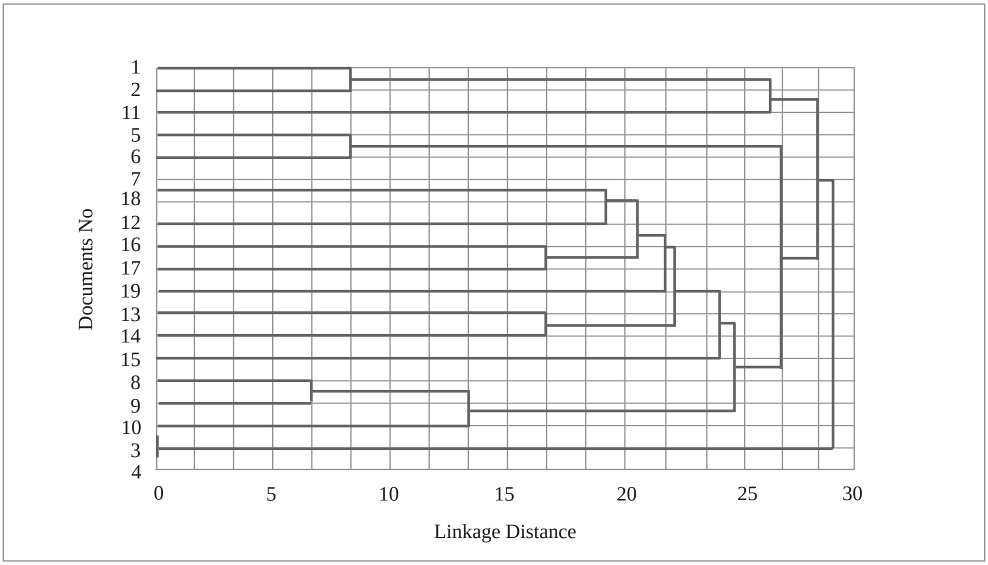

Fig. 9. Hierarchical tree diagram

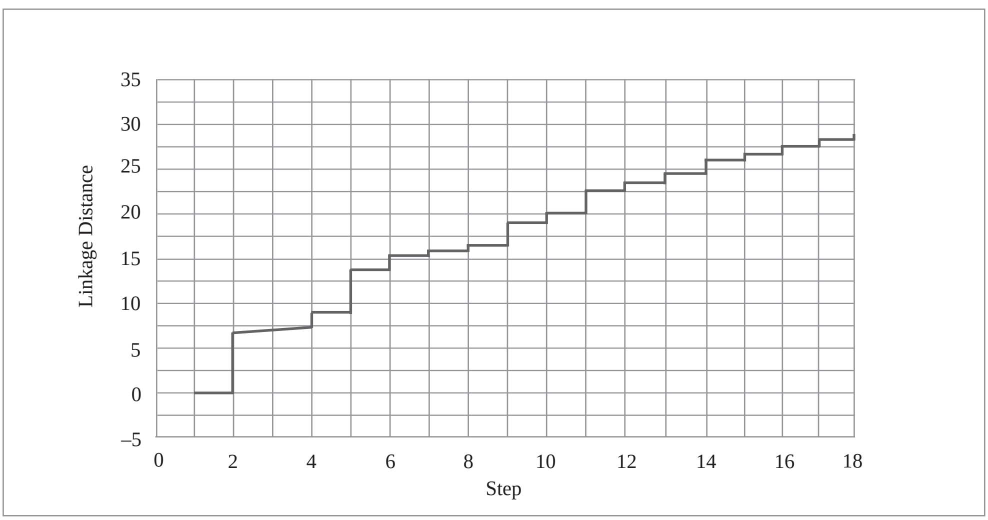

Fig. 10. Plot of linkage distances across steps

labeled as "3" and "4" have been classified as identical. These results have been confirmed by "natural" analysis of the contents of these documents. Obviously, by controlling the level of cut-off of the hierarchical tree, it is possible to obtain a defined number of classes.

Stage 3 – pattern classification

The aim of the pattern classification is to assign the documents with determined parameters to previously known and determined states. In order to accomplish this task, a minor modification of the documents used in the analysis was made. As they were related to building materials of different application, each of the documents was assigned a class corresponding to the level of usefulness in repairs of the concrete construction (repair of damage of the concrete construction which required rebuilding of 40 mm, the construction is subject to operation of aggressive factors). Therefore, each of the documents was assigned to one of the classes:

– H – high usefulness; – M – medium usefulness; – L – low usefulness

The essence of the presented problem lies in developing a method/way of generating conclusions for new cases, that is in assigning new documents to one of the presented above classes. The method used in this analysis is the decision tree (classification), for which it is possible to generate general rules which can be used in building the knowledge base. The built of the decision tree has been based on a representation of text documents in the form of SVD matrix. In each node the considered features are the chosen (and reduced with consideration of the significance) contents of the SVD analysis. The created decision tree is a relatively simple tree and consists of two branches and four leaves (Fig. 11).

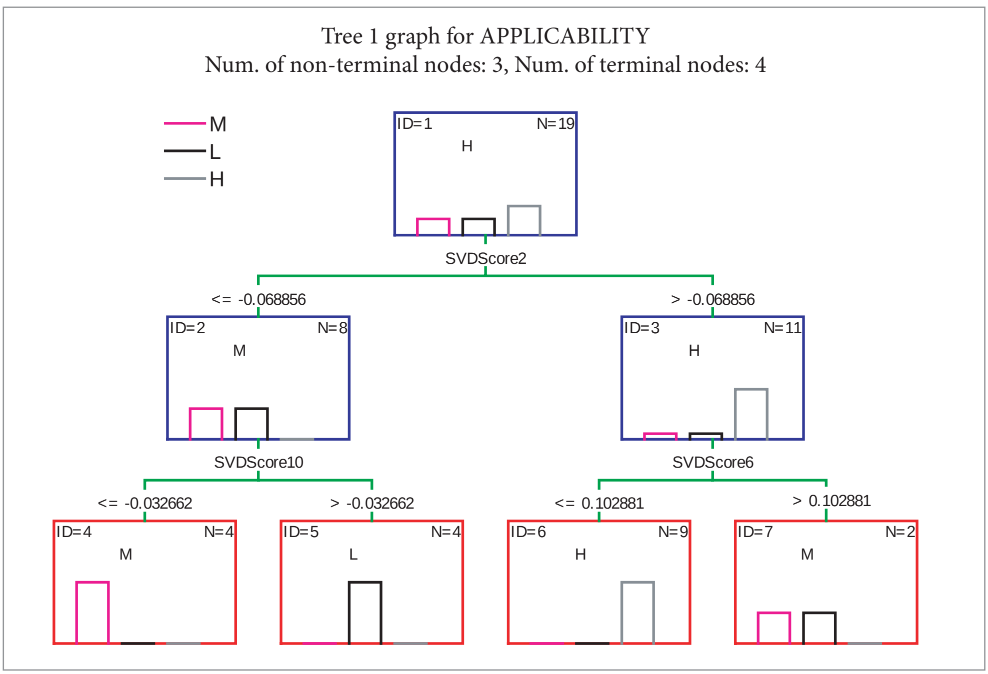

Fig. 11. Decision (classification) tree

For such built tree the rules can be generated automatically thanks to opportunities which are created in the software analysis. The rules are generated in the form of SQL code and on their basis the database can be built, which will be the resource for DSS.

The generated rules are universal i.e. they do not refer to particular materials/cases, but they represent their character, features and the values related to them. On their basis and with a new set of cases it is possible to carry on the classification of these cases into previously determined groups. A similar result cab be obtained with the use of artificial neuron networks. Unfortunately, in this case it is necessary to have a larger population of teaching cases.

6. Summary

The presented above text mining methods, being a class of data mining methods, are a promising tool in the analysis of text data. As it was presented in the examples, such analysis may be useful at the stage of building knowledge bases and data for DSS. Taking into account the existence of a significant number of text documents, text mining analysis may serve as a certain form of automation of knowledge acquisition task as well as building the database itself. Two cases of non-pattern and pattern classification provide the proof. From a scientific viewpoint, text mining analysis opens up a wide range of possibilities for using text documents e.g. in statistical device or in processing in intelligent models. It is based on the possibility of creating formal representation for the analyzed text documents and potentially facilitates the development of DSS as, instead of classical and long-term knowledge acquisition (acquisition sessions with an expert, manual analysis of databases and searching other sources), we can obtain a ready model containing knowledge, e.g. a trained artificial neural network. In practice, a certain weak point of text mining analysis consists in an insufficient number of software solutions. Since currently operating solutions are still being developed, their range of application is limited. Another inconvenience of text mining analysis lies in difficulties with interpreting numerical values. In most cases they are connected with a certain descriptive phrase (e.g. units of measurement). In this respect, it seems necessary to transform numerical values into corresponding verbal descriptions. Otherwise, the numerical values may become lost in the analysis. Text mining analysis is perfectly useful in the case of long verbal descriptions, such as technical instructions, technological guidelines and in such cases may be acknowledged as a valuable and useful method for building knowledge representation.

References

Anderson, T. 1996. Knowledge types: Practical approach to guide knowledge engineering in domains of building design, *Expert Systems* 13(2): 143–149. doi:10.1111/j.1468-0394.1996.tb00186.x

Berry, M. J.; Linoff, G. S. 2000. *Mastering Data Mining*. Willey & Sons, New York.

Boddy, S.; Rezgui, Y.; Wetherill, M.; Cooper, G. 2007. Knowledge informed decision making in the building lifecycle: An application to the design of a water drainage system, *Automation in Construction* 16(5): 596–606. doi:10.1016/j.autcon.2006.10.001

Chau, K. W.; Albermani, F. 2003. A coupled knowledge-based expert system for design of liquid-retaining structures, *Automation in Construction* 12(5): 589–602. doi:10.1016/S0926-5805(03)00041-4

Chen, J.-H. 2008. KNN based knowledge-sharing model for severe change order disputes in construction, *Automation in Construction* 17(6): 773–779. doi:10.1016/j.autcon.2008.02.005

Cheng, M.-Y.; Tsai, H.-C.; Lien, L.-C.; Kuo, C.-H. 2008. GIS-based restoration system for historic timber buildings using RFID technology, *Journal of Civil Engineering and Management* 14(4): 227–234. doi:10.3846/1392-3730.2008.14.21

Cohen, W. W.; Wang, R.; Murphy, R. F. 2003. Understanding captions in biomedical publications, in *Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 499–504.

Creese, G. 2004. Duo-Mining: combining data and text mining, *Dm Review Magazine*. Available from Internet: <http://www.dmreview.com/article_sob.cfm?articleID=1010449>.

Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. 1996. From data mining to knowledge discovery in databases, *AI Magazine* 17(3): 37–54.

Feldman, R. 2006. *Text mining handbook*. Cambridge University Press. doi:10.1017/CBO9780511546914

Gajzler, M. 2008a. Hybrid advisory systems and the possibilities of it usage in the process of industrial flooring repairs, in *Proceedings of the 25th International Symposium on Automation and Robotics in Construction. ISARC-2008*, Vilnius: Technika, 459–464. doi:10.3846/isarc.20080626.459

Gajzler, M. 2008b. *Hybrid advisory system for industrial concrete floors repairs*. PhD Thesis. Poznan University of Technology, Poznan (in Polish).

Haerst, M. A. 1999. Untangling text data mining, in *Proceedings of ACL `99, University of Maryland, June 20–26, 1999.* Available from Internet: <http://www.sims.berkeley.edu/~haerst/papers/ac199/ac199-tdm.html>.

Haidasz, M. 2008a. Visualizing simulated monolithic construction processes, *Journal of Civil Engineering and Management* 14(4): 295–306.

Haidasz, M. 2008b. Modelling and simulation on monolithic construction processes, *Technological and Economic Development of Economy* 14(4): 478–491.

Hanna, A. S.; Lotfallah, W. B. 1999. A fuzzy logic approach to the selection of cranes, *Automation in Construction* 8(5): 597–608. doi:10.1016/S0926-5805(99)00009-6

Hola, B.; Schabowicz, K. 2007. Mathematical-neural model for assessing productivity of earthmoving machinery, *Journal of Civil Engineering and Management* 13(1): 47–54.

Jang, W.; Skibniewski, M. 2008. Wireless network-based tracking and monitoring on project sites of construction materials, *Journal of Civil Engineering and Management* 14(1): 11–19. doi:10.3846/1392-3730.2008.14.11-19

Kaklauskas, A.; Zavadskas, E. K. and Trinkūnas, V. 2007. A multiple criteria decision support on-line system for construction, *Engineering Applications of Artificial Intelligence* 20(2): 163–175. doi:10.1016/j.engappai.2006.06.009

Kaplinski, O. 2007. *Methods and models of research in construction project engineering*. Polish Science Academy, Committee of Civil Engineering and Hydroengineering. Institute of Fundamental Technological Research, Warsaw (in Polish).

Kaplinski, O. 2009. Information technology in the development of the Polish construction industry, *Technological and Economic Development of Economy* 15(3): 437–452. doi:10.3846/1392-8619.2009.15.437-452

Maas, G.; Vos, A. 2008. Data collection to meet a contract requirements, in *Proc. of the 25th International Symposium on Automation and Robotics in Construction ISARC-2008*. Vilnius: Technika, 365–372. doi:10.3846/isarc.20080626.365

McCowan, A.; Mohamed, S. 2007. Decision support system to evaluate and compare concessions options, *Journal of Engineering and Management* 133(2): 114–123.

Mulawka, J. J. 1997. *Expert systems.* WNT. Warsaw (in Polish).

Naimavičienė, J.; Kaklauskas, A.; Gulbinas, A. 2007. Multi-variant decision support e-system for device and knowledge based intelligent residential environment, *Technological and Economic Development of Economy* 13(4): 303–313.

Paslawski, J.; Karlowski, A. 2008. Monitoring of construction processes in the variable environment, *Technological and Economic Development of Economy* 14(4): 503–517. doi:10.3846/1392-8619.2008.14.503-517

Ping Tserng, H.; Lin, Y.-Ch. 2004. Developing an activity-based knowledge management system for contractors, *Automation in Construction* 13(6): 781–802. doi:10.1016/j.autcon.2004.05.003

Schabowicz, K.; Hola, B. 2008. Application of artificial neural network in predicting earthmoving machinery effectiveness ratios, *Archives of Civil and Mechanical Engineering* 8(4): 73–84.

Shaked, O.; Warszawski, A. 1995. Knowledge-based system for construction planning of high-rise building, *Journal of Construction Engineering and Management – ASCE* 121(2): 172–182. doi:10.1061/(ASCE)0733-9364(1995)121:2(172)

Ustinovichius, L.; Zavadskas, E. K.; Podvezko, V. 2007. Application of a quantitative multiple criteria decision-making (MCDM-1) approach to the analysis of investments in construction, *Control and Cybernetics* 36(1): 251–268.

Zavadskas, E. K.; Kapliński, O.; Kaklauskas, A.; Brzeziński, J. 1995. *Expert systems in construction industry. Trends, potential & applications.* Vilnius: Technika.

DUOMENŲ RINKIMO METODAI STATYBOS SPRENDIMŲ PARAMOS SISTEMAI

M. Gajzler

Santrauka

Straipsnyje pateikiamos informacijos rinkimo metodų pritaikymo galimybės sprendimų paramos sistemoms statyboje. Daugiausia problemų sukelia informacijos gavimas, tinkamas jos atvaizdavimas ir naudojimas. Duomenys yra pagrindinis sistemos išteklius. Nustatyta, kad nuo 70 iki 80 % visų turimų bendrojo naudojimo informacijos šaltinių yra tekstiniai dokumentai. Tekstinės informacijos rinkimo technika yra suprantama kaip procesas, kuriuo siekiama išgauti anksčiau nežinomą informaciją iš tekstinių dokumentų (pavyzdžiui, technologinių kortelių). Pagrindinė šios technikos savybė – galimybė tekstinių dokumentų informaciją pateikti formalizuota forma, tai atveria plačių galimybių tolesnei analizei. Šiame straipsnyje pateikiamos pasirinktos IT priemonės, naudojamos tekstinei informacijai rinkti. Autoriaus tikslas – supaprastinti informacijos rinkimą, jį automatizuoti ir sutrumpinti, sukurti informaciją apimančius modelius. Ankstesni informacijos kaupimo metodai (apklausos, anketos) reikalavo daug ekspertų darbo ir laiko.

Reikšminiai žodžiai: sprendimų paramos sistemos, informacijos rinkimas, tekstų analizė, AI modeliai, konsultavimo sistema.

Marcin GAJZLER. PhD. C. E., Assistant Professor, Poznan University of Technology, Institute of Structural Engineering, Division of Construction Engineering and Management. Analyst at the stage of solving decision problems in the area of construction engineering. Research interests: decision-making theory, AI methods in construction industry (fuzzy logic, artificial neural networks, advisory systems, mining methods), investment process, economic and management in construction.