

Protocols from periodic inspection of buildings in text mining analysis - advantages and problems of analysis

*Marcin Gajzler*¹

Poznan University of Technology, Institute of Structural Engineering, Piotrowo 5 Poznań, Poland

Abstract. The article presents the problem of the analysis of post-inspection protocols of buildings in order to obtain the necessary data and information for the construction of a system supporting the technical management of residential buildings. The attention was primarily focused on the document itself - a protocol of control and automated technique of obtaining data and information from this document - text mining technique. Problems appearing in the analysis and the potential for their removal were indicated. The conclusions formulated relate to improvement of the process of conducting the review and formulating the post-control protocol in the aspect of its use in automated data collection and information transfer involving state of buildings.

1. Introduction

The subject of the article results from a broader issue that is currently being implemented - the construction of a system supporting the technical management of buildings with the use of available knowledge resources. The problem of building management has been developed for a long time as an element combining science and practice [1, 6, 12]. The technical conditions of the building in the maintenance policy and technical building management is a matter repeatedly undertaken in scientific works [4, 5, 8, 12]. The system being under construction is responsible for supporting decisions related to shaping the policy of maintenance, renovation and modernization of residential buildings. It is dedicated to entities dealing with real estate management - property managers. In the course of the conducted research, it

¹ Corresponding author: marcin.gajzler@put.poznan.pl

turned out that an important management problem is the planning of renovation policy with a limited availability of funds. Long-term planning of activities is necessary - in the perspective of several years, in order to prepare for them, and in particular to provide the necessary financial resources. It is helpful in this respect to forecast both the technical condition, the necessary expenses as well as the consequences of choosing a particular method of action in the maintenance policy. One of the knowledge resources that are used in the construction of the technical management support system are documents that occur during the operation of the building and are associated with the implementation of compulsory periodic inspections of technical condition of buildings in Poland. In the course of control of construction works, the following factors are assessed: essential elements of the construction of the facility, installations and selected technical equipment. As a result of the inspections carried out, a protocol is created describing the current technical condition and formulating recommendations in relation to the irregularities found. This protocol is an element building object book and is in the form of a text document, therefore, for the purpose of its analysis, the proposed text mining analysis is justified.

2. Basics for data and text mining analysis.

Mining analyzes compared to classical statistical methods are relatively young techniques having roots in the techniques of knowledge discovery and machine learning. Interestingly, in the course of using the mining approach, it is also possible to use classical statistics techniques. Mining techniques rely on the processing of potentially used data in order to find new interesting and often accidental dependencies. The mining approach, taking into account the results of the analysis, can be divided into targeted and non-targeted - accidental. Text mining is a younger version of the data mining date based on the analysis of text documents (data). It involves the development of a formal / quantitative representation of text data, which are subsequently analyzed using various available techniques (including statistical methods). The general text mining algorithm is presented on Figure 1.

A wider look at the methods and possibilities of conducting analyzes was presented in [11]. On the other hand in 2010[2] the application of text mining methods in the analysis of issues related to technological decision support in construction were presented. It can be stated that the suitability of the method has been demonstrated, but also some important limitations have been pointed out, in particular the sensitivity of quantitative data to the analysis and a complete loss of their context in the case of text mining analysis.

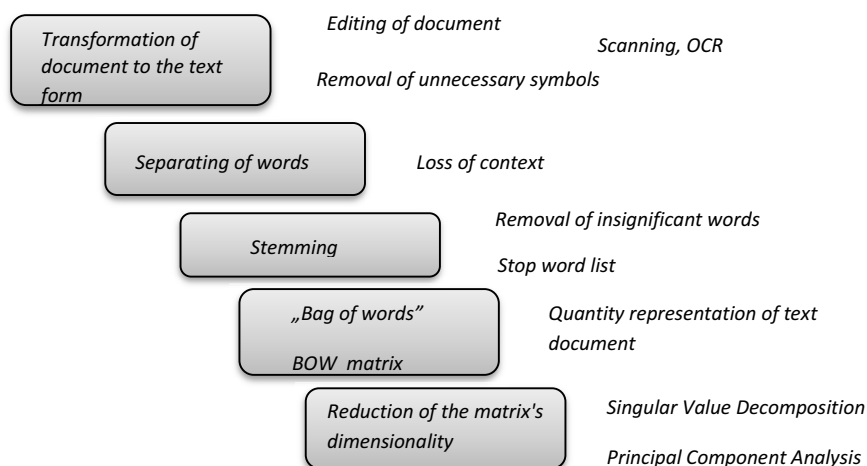


Fig. 1. The basic algorithm of the text mining process\

3. Control of building objects

The control of construction works is primarily due to the legal provisions in force in Poland and applies in particular to multi-family residential buildings. This obligation rests with the owners or established managers of the property. The inspection is entrusted to a person with appropriate qualifications to perform independent technical functions in construction. Checks are carried out annually, and expanded inspections are carried out every 5 years. The scope of control of construction facilities is described, among others in [1, 3]. Apart from the fact that the inspection obligation is imposed by the law in force in Poland, the inspection of the technical condition of the building itself is important from the point of view of conducting a proper maintenance policy of the building and guaranteeing the proper technical condition of the building. The results of the inspection allow for monitoring the technical condition of the building and in the event of the first signals of any irregularities to plan the taking of appropriate actions. At the same time, it should be noted that in fact, the necessary steps to maintain the proper technical condition are limited by the availability of funds (primarily financial), and therefore the maintenance activities are correlated with the prior acquisition of funds for necessary actions.

The post-inspection report is the result of the conducted technical inspections. It is a study containing general characteristics of the building undergoing inspection, taking into account basic surface-cubic parameters, construction and technological parameters as well as the age of the building. The basic element of the post-inspection protocol is the description of the technical condition of individual building elements starting from the foundations to the roof covering. It is not possible to precisely specify elements of the structure, because it should be borne in mind that not all elements are directly visible. In addition, individual elements related to a specific building are subject to the assessment of the technical condition.

Description of the technical condition is carried out by verbal description, where the symptoms of possible degradation occurring on the assessed element are mentioned, such as cracks, scratches, deflections, corrosion phenomena. You can also find a description of the possible causes of these degradation symptoms. In addition, a qualitative assessment of the element is made using often any dictionary, eg good, bad, satisfactory, and also recommendations for the assessed element are made. During analysis of post-inspection reports, the quantitative assessment of the technical condition by defining the degree of technical wear expressed can be also met. As a rule, in the protocols approach for estimation of the degree of technical wear, is not set, although the most common approach is based on a symptomatic visual assessment. In addition to the above-mentioned elements, the post-inspection report may contain a number of additional descriptive information, eg concerning the maintenance and repair works carried out in the last period and their impact on the current technical condition. The photo service complements the post-inspection protocol.

The analyzes of post-inspection protocols of 40 buildings from the last 10 years carried out in the course of the research indicate that these protocols are not standardized in terms of both the form and the dictionary used to describe the technical condition.

4. Text mining analysis – possibilities and barriers

The use of the text mining method in the analysis of periodic inspection reports of buildings in the context of the construction of a system supporting the technical management of residential buildings is justified and necessary at the same time. Post-inspection reports are the basic source of information on the technical condition of the building. It would be unreasonable to omit such vital information about the building. The problem that requires the use of the text mining method is that the description of the technical condition of the building in the post-inspection report is represented by a text description. In addition to it, there are also often quantitative information regarding the technical condition (eg technical wear of the facility), however, due to the lack of standardization protocols do not occur each time, and also do not allow for in-depth knowledge of factors conditioning the value presented in the protocol. The text description of the technical condition allows for a broader understanding of both the factors affecting the building, symptoms and possible directions of action.

Of course, the significance of quantitative information should not be marginalized, since these, combined with other available information, allow to learn about interesting relationships. An example of such are the dependencies of expenditures in subsequent years and the technical condition characterized by the degree of technical wear. Analyzing the mentioned dependencies of expenditures and technical wear of the building, some recurring models are noticed. In most of them, a typical relationship is already observed, namely the increase in maintenance costs causes in the perspective of one to three years the improvement of the technical condition and the reduction of the determined degree of technical wear. However,

this does not apply to all objects, and there are also situations in which the occurrence of outlays does not cause a noticeable change in the technical condition of the building. This is the case in which, either the inputs take inadequate quantities or occur too late. Based on these dependencies, one can infer about proper or incorrect maintenance policy, which is an element of technical building management.

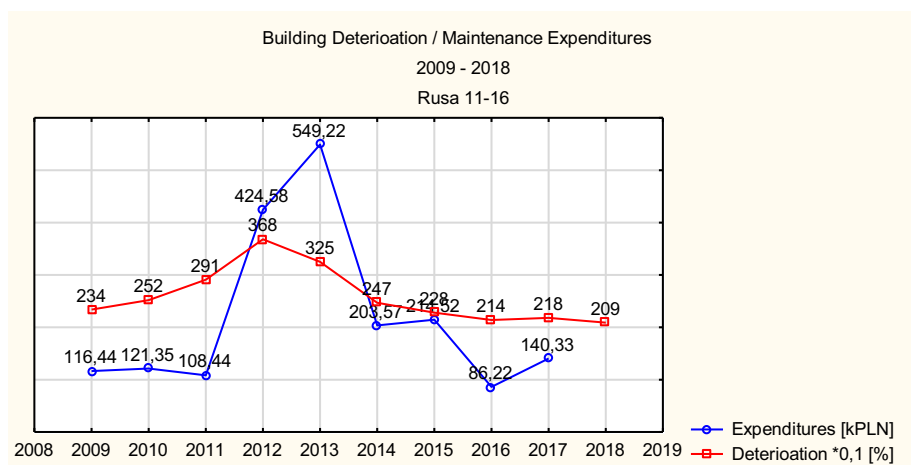


Fig. 2. Example of dependence of maintenance expenditures and the degree of technical wear for a residential building in precast concrete technology from 1979

However, focusing on the analysis of text mining, it is worth commenting on the possibilities offered by the analysis and the barriers and limitations associated with it. The author's first experiments with text mining analyzes were presented in []. Then they were used to support technological decisions. Text documents analyzed at that time constituted technical cards for various building materials. A characteristic feature of these cards was the occurrence of various quantitative data in the text description (eg strength values in MPa, application and bonding times in minutes, application temperatures in degrees of Celsius). The post-inspection reports analyzed today are much poorer in quantitative values. They occur sporadically or not at all. The dominant element of the post-control protocol is a pure text description. Even if in the technical condition description there is a need to characterize some measurable value, eg deflections, damage - scratches, cracks, due to the fact that the description of the technical condition of the building is a derivative of the inspections performed, measurable values are usually characterized by qualitative variables - fuzzy (eg permissible, over-standard deflection, deep scratches, etc.). Potentially such a qualitative description - fuzzy may seem like a weak point in technical issues, in the case of the text mining method allows for a more complete consideration of the information contained in the analyzed description. The above conclusion will be presented on the example of a text description fragment. The first fragment was taken from the post-control protocol in

which the quantitative data occurred, while the second fragment from the more common plain-text description.

Fragment 1:

„Cracks in concrete approx. 3 mm between the side wall of balcony and the wall of the building. Contamination of supports by algae. Measured deflection of the 7 mm bracket. Cracks on the joints of the cover plates”.

Fragment2:

„On the basis of the lack of cracks in the bearing walls, it can be assumed that the condition of the foundations is correct. Negligible scratching of selected load-bearing walls in the common part. It has been assumed that the condition of load-bearing walls (internal) and their joints are good”.

Classical analysis of both fragments allows to obtain the right information, and most importantly, the correct interpretation of the phenomena described. It turns out that in the case of text mining analysis, both fragments can be included with different results in terms of information. The decisive element here is the occurrence in the first fragment of numerical values which in the course of the text mining process lose their context and informative meaning. This is due to the fact that in the course of the text mining process, individual parts of the text are subject to various processes of machining, in particular separation, reduction and steaming. Often numerical values are reduced as a symbol (not text), and if they even go through the text mining process, they are an abstract element that does not carry any information load. Separated from units and names of values are devoid of any context and, at the most, they introduce unnecessary noise. As a result of the text mining process in the basic approach, a frequency matrix is obtained, which counts the frequency of the extracted and processed words. The frequency matrix is also difficult to interpret in terms of the information originally conveyed by the text, but the awareness of individual components of the text mining process allows categorically to state that numerical values lose their context and information load.

The frequency matrix is also the first basic quantitative representation of the text. Taking into account the length of the text description, it can be expected that the frequency matrix will be characterized as oversized. To reduce it, the decomposition of singular values is used. The representation thus obtained is already a complete abstract representation of the text. However, awareness should be paid that despite the abstract form, the information load is contained in this form. Taking into account the building management system for building management, the text mining analysis allowed obtaining the BOW frequency matrix with the dimension 400x273 (the number of rows results from the number of buildings in the database and the adopted 10-year analysis period, while the number of columns concerns words included in the analysis resulting from the text mining text documents). After applying the decomposition, a matrix of 400x5 can be obtained, where each of the 400 rows is responsible for representing the technical condition of a specific building in a given year according to the established nomenclature:

019: [2,3413 1,4193 1,5616 2,4809 3,1036]

018: [1,9803 1,4232 1,9425 2,0377 2,1807]

.....

.....

040: [2,5501 1,7052 3,1428 2,3882 1,3209]

where the first two digits of the row number correspond to the number of the building, and the third number is denoted to the time in years back from the year of analysis, eg the number 9 meant 2009, while the correct matrix was peculiar values representing the technical condition for a given building and a given year of its operation.

Taking into account the technical building management system being built, the technical condition is one of 11 variables (building year, building technology, usable floor space, number of floors, technical condition, maintenance costs, ownership structure, number of owners, ownership share of the city, conservation supervision) included in the analysis. Due to the use of quantitative and intelligent methods available in the group of data mining methods, it was necessary to bring a description representing the technical condition to an acceptable form in the presented analyzes. The set of cases characterized by the 11 variables creates a base used for CBR [7, 13] for new cases. The case selection mechanism from the database includes several variables, including similarity of technical condition. For this purpose, agglomeration methods have been used, in particular the Ward method [3]. However, the use of this method is justified only after prior preselection of cases from the database and is the last element of selection. Empirical methods have shown that the building selection is a selective variable. It is so important that, as shown by a cursory analysis of the case database, it is strongly correlated with the year of construction of the building. This is confirmed by the development of housing construction in Poland, which can be divided into brightly visible stages: traditional construction, prefabricated construction and a return to traditional and mixed technologies.

5. Conclusions

Conducted in connection with the construction of a system supporting the technical management of residential buildings the analysis of test documents - post-inspection protocols pointed to numerous problems and barriers to the application of the analysis. The first group are technical problems regarding the analysis itself. The software on which the Statistica Data Miner is run does not have any dictionaries implemented to allow the use of the steaming process for documents in Polish. This makes it necessary to conduct this part of the analysis in a manual way. There are of course other software solutions (PS Clementine PRO and Applica MPJ module), but these are not so extensive in the field of analysis methods. Another problem in text-based analysis is the occurrence of quantitative data in text documents that lose context and information in the course of the process. A very important problem regarding the text documents themselves is the lack of their standardization in terms of the information contained therein as well as the form itself and the dictionary used. The proposal solving most of the problems mentioned above is the construction of a universal form allowing for the standardization of both the description of the technical condition of the building and the available description dictionary. It seems even legitimate to build a mobile application available for

devices such as smartphone and tablet, which using a fixed form and an accessible dictionary will allow you to generate a post-control form and simultaneously obtain a text document - standardized in electronic form, ready to conduct analyzes. Such approach would also allow dynamic and significant development of the available case database of the management support system.

References

1. Bucóń R., M. decyzyjny wyboru wariantów remontu lub przebudowy budynków mieszkalnych. Politechnika Lubelska, 2017
2. Gajzler M., TEDE, **16** (2010), 219-232
3. Gajzler M., *Proc.Eng.* **208** (2017), 20-26
4. Halim, T., Muthusamy, K., Chia, S.Y., Lam, S.W.: Facilities, **29** (2011) , 114-132
5. Horner, R.M., El-Haram, M.A., Munns, A., *Int. J. of Quality in Maint., Eng.* **3** (1997) , 273-280
6. Kapliński O., *Eng. Struct. Techn.*, **10**, (2018), 27-33
7. Kończak A., Paślawski J., *Proc. Eng.* **57** (2013), 583-588
8. Moghaddam, K.S., Usher, J.S., *J. of Quality in Maint. Eng.*, **16** (2010) 3, 271-287.
9. Nowotarski P., Paślawski J., Lean and agile management synergy in construction of high-rise office building. *Archives of Civil Engineering*, **62(4)**, (2016), 133-148,
10. Plebankiewicz E., Zima K., Wieczorek D., *Archiv. of Civ. Eng.*, **62** (2016), 149-166
11. Shu-Hsien Liao, Pei-Hui Chu, Pei-Yuan Hsiao, *Exp. Sys. with Applic.*, **39** (2012) 11303-11311
12. Tsang, A. H. C.: *J. of Quality in Maint. Eng.*, **1** (1995) 3-17
13. Zima K.. *Proc. Eng.*, **122**(2015), 57-64