Journal of Machine Construction and Maintenance
PROBLEMY EKSPLOATACJI
QUARTERLY 1/2017 (104)
ISSN 1232-9312
p. 53–57

Ewelina KOSICKA*, Małgorzata ŚLIWA**
* Lublin University of Technology, Lublin, Poland
** University of Zielona Góra, Zielona Góra, Poland
e.kosicka@pollub.pl

MACHINE FLEET FAILURE FREQUENCY CONTROL SUPPORT BY TEXT MINING METHODS

Key words: text mining, data mining, autonomous maintenance.

Abstract: The development of both control systems for machine fleets and computer-controlled production systems have provided companies with a wide spectrum of tools for collecting data about the operation of their machinery stocks. A vast number of companies only store historic data; however, they do not use these data to extract information with respect to improving the efficiency of their technical infrastructure. The paper discusses the application of data mining to control machine fleet failure frequency resulting from non-technical causes, i.e., due to human factors.

Wsparcie nadzoru nad awaryjnością parku maszynowego metodami text mining

Słowa kluczowe: text mining, data mining, eksploracja danych, autonomiczne utrzymanie ruchu.

Streszczenie: Rozwój systemów monitorowania stanu parku maszynowego, a także powszechna komputeryzacja obszaru produkcji, dały przedsiębiorstwom szeroki wachlarz narzędzi pozwalających rejestrować dane dotyczące funkcjonowania posiadanych maszyn. Znaczna część przedsiębiorstw jedynie gromadzi obszerne zbiory danych historycznych, nie podejmując działań zmierzających ku pozyskaniu z nich informacji mogących wpłynąć na poprawę efektywności posiadanej infrastruktury technicznej. W artykule przedstawiono możliwość zastosowania eksploracji danych tekstowych do prowadzenia nadzoru nad awaryjnością parku maszynowego powodowaną przyczynami nietechnicznymi w postaci czynnika ludzkiego.

Introduction

Although the first computers were created only 70 years ago, their dynamic development continues to this day, leading to the expansion of computers in nearly every single area of life. The computerization of the operation of companies resulted in a significant rationalization of many processes, while automation eliminated the problem of wasting time and production materials due to the elimination of the human factor from some production processes. The research conducted in 2015 demonstrates that only 33% of the investigated Polish companies are not computerized (Fig. 1).

The degree of the computerization of a company is affected by its size; nearly half of the investigated companies with the annual income of up to 300 million Polish zloty described themselves as computerized companies [1].

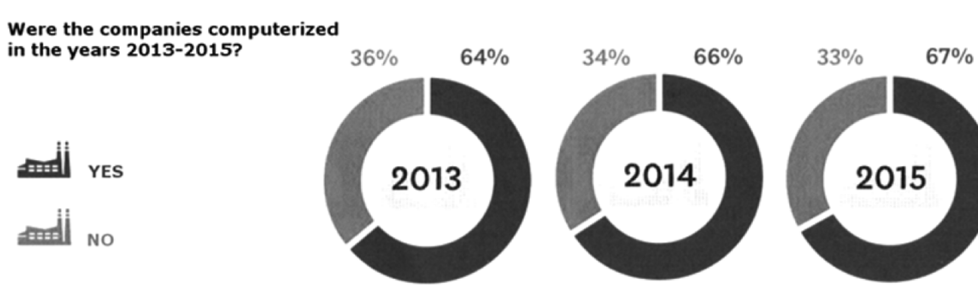


Fig. 1. Computerization of companies in the years 2013–2015 [1]

1. Data mining methods

Besides computer systems and production automation, companies can also implement advanced systems for machinery fleet control, thereby being able to monitor the condition of their technical infrastructure in real time. Unfortunately, many of the companies that decide to retrofit their machinery stocks or purchase machines provided with technical condition control systems take no actions with respect to the analysis of databases with registered parameters. As a result, they collect valuable historic data on machinery operation, but they do not extract any information thereof.

The rationalization of an approach to databases led to the development of data mining technologies [1, 6, 11] for discovering similarities, dependencies, or trends in large data sets. This term is also sometimes used as a synonym of knowledge discovery. In the literature on data mining [e.g., 1, 2–8], the authors make it specific that the term "knowledge discovery" should refer to the whole process of discovering knowledge consisting of stages that transform a given set of data into a different set that can then be used to support decision taking.

The use of tools based on the afore-mentioned *data mining* enables us to obtain information from structured databases described by values expressed on well-known measuring scales, whereas the possibility of automation of text data analysis developed in a natural language is ensured by *text mining* techniques [9, 10, 14, 15, 20, 22]. Hearst Marti A., the creator of *text mining,* describes them as a process for extracting previously unknown information from text resources [10]. The similarities between *text mining* and *data mining* include the following [16]:

– An exploratory approach to the analysis process,
– An emphasis on the usefulness of results, and
– The considerable use of the same methods and tools.

Speaking of text processing techniques, it is worth mentioning two methods for text processing: *deep text processing* and *shallow text processing*. Deep text processing is a computer linguistic analysis of interpretations and grammatical relationships in a text. This process does not take into account statistical relationships, and it uses processing based on in-built patterns [18, 19, 21]. Shallow text processing relates to actions that focus on recognizing non-recurrent text structures. This type of processing is mainly used to find, for instance, proper names or verb groups; however, it does not take into account the structure or function they play in the sentence [21].

The fields that make use of text data mining include medicine, economics, astronomy, and engineering [7, 17].

2. Operator duties under autonomous maintenance

The implementation of *Total Productive Maintenance* (TPM), which is a set of techniques and solutions for increasing the effectiveness of a machinery fleet [4], led to delegating some of the responsibility for technical conditions to operators. The scope of their responsibility was defined under *Autonomous Maintenance*, where the role of the operator is maintaining or restoring the efficiency of machines and devices [13]. Therefore, operators are responsible for running daily inspections, lubrication, simple repairs, machinery observation with respect to potential operational abnormalities, and precision control.

The implementation of Autonomous Maintenance has eliminated the approach where all actions regarding maintenance and repair are only taken by maintenance service staff. The participation of operators in some actions previously taken only by maintenance services enables one to decrease the number of occurring failures, and, as a result, leads to increased operational efficiency of the machinery fleet, reduced the number of unscheduled production shutdowns and product defects.

3. Analysis of machinery fleet failure frequency by text mining

Many production companies decide to rationalize their operation by the application of computer tools and implemented specialist Computer Maintenance Management Systems (CMMS).

The research conducted by Gartner Group and ARC Advisory [24] demonstrates that the implementation of CMMS effectively contributes to the following:

– Reduced stocks on hand (by approx. 15–35%),
– Reduced costs of logistics (by approx. 10–20%),
– Increased use of the company's assets (by approx. 3–10%), and
– Increased use of human resources (by approx. 10–20%).

CMMSs are a vast source of knowledge about the condition of a company's production infrastructure. They often contain data that can serve as a basis for rationalization or actions to improve operational efficiency of the machinery fleet by following the application of suitable tools.

An example of such data is records made by maintenance service about machinery failures. In most cases, the entries to the system are not made in a uniform way, and their form and content depend on a staff member. Therefore, it is recommended to systematize and provide guidelines with respect to report writing standards. Primarily, this would require the following [11, 25]:

- Impose a file format for writing reports generated on maintenance service intervention, e.g., Extensible Markup Language (XML), which is a common format of information exchange between systems;
- Sensitize the staff to the use of a uniform system of the codification of national characters in the text;
- The use of dictionaries built in text editing software to support correct spelling (elimination of spelling mistakes); and,
- Verify and eliminate errors in files.

In order to enable faster verification of documentation by text mining methods and to correctly diagnose the cause of a shutdown or a failure, it seems reasonable to design a report template divided into appropriate sections. In such a report, the part dedicated to the cause of failure would require a particular emphasis. Here, it would be necessary to note down in the form of keywords the causes of failure and maintenance service calls (Fig. 2). The machine in question can also be specified by adding a photograph or scanned copy of its data plate imported to the report from the collapsible menu.

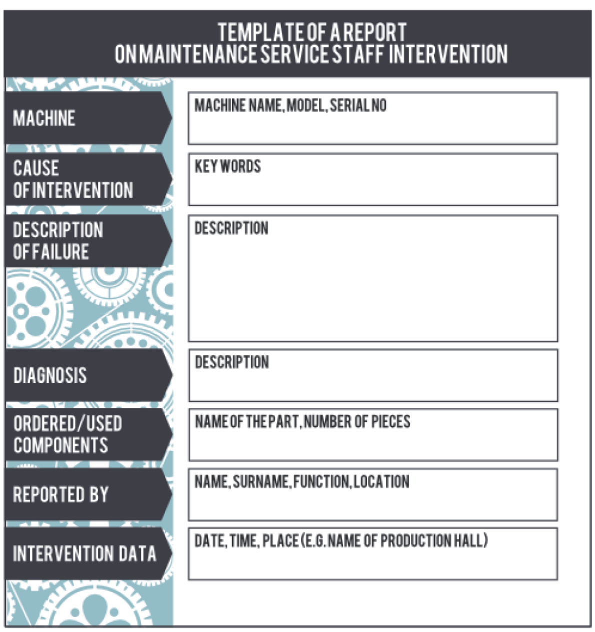


Fig. 2. Proposed template of a report on maintenance service intervention

Unfortunately, a majority of such entries are not analysed, and so companies do not take any steps to eliminate factors leading to potential failures. Some companies decide to analyse such databases, delegating this task to employees who do such analysis by hand.

The factors that contribute to failure occurrence are both technical and non-technical, i.e. caused by human factors. Many companies notice the relationship between the operator's knowledge of the machine and the occurrence of certain failures; therefore, they have developed their own systems of training with respect to machinery fleet operation. If the operator's fault leads to the failure of the machine, it seems reasonable to send this individual on extra training. Detecting such situations can be made convenient by the use of *text mining* techniques for the analysis of entries made in a computer system by maintenance service. Generating reports concerning the entire production line and then subjecting them to quick analysis to detect machine shutdowns or unnecessary calls of maintenance staff can eliminate such occurrences in the future, because an operator can be taught to take the appropriate actions to avoid such occurrences.

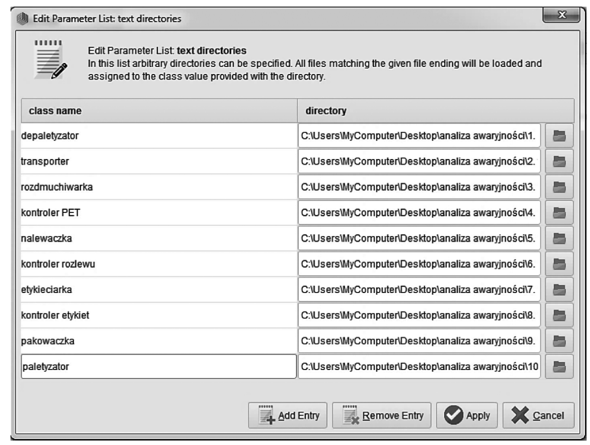
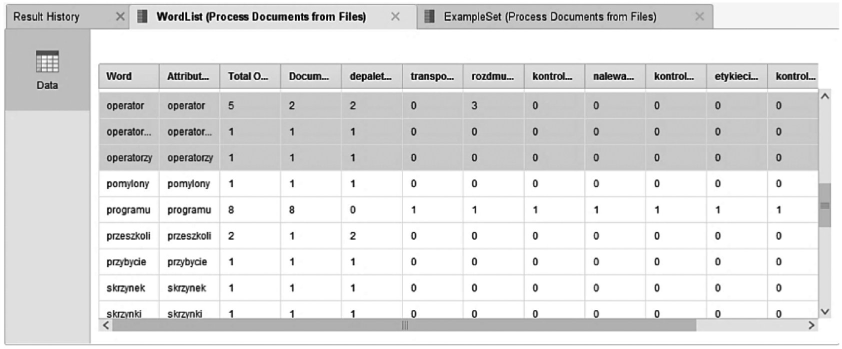


Fig. 3. Import of generated report into RapidMiner Studio programme

Figure 3 shows the import procedure of source files constituting a report with monthly records keyed in by the maintenance service staff in the form of a description of a registered shutdown or failure into the RapidMiner Studio ver. 7.1. programme for data mining. The records are not structured due to a lack of a specified procedure for making entries.

The reports concern 10 machines installed in a beer bottling plant: a depalletizer, a conveyor, a blow-moulding machine, a PET controller, a pouring machine, a discharge controller, a labelling machine, a label controller, a packaging machine, and a palletizer.

Following the application of suitable operators and data mining parameters, we obtained a list of words occurring in individual imported failure frequency reports. As a result, it was possible to indicate to the machine with reference to operator in the failure description (Fig. 4). In the reports concerning the depalletizer, there were altogether 5 such cases, while the report about the blow-moulding machine contained 3 such attributes. One can easily observe that along with the attribute "operator," the programme also detected attributes resulting from the declension of this word. The application of the tool equipped with a stemmer for reports analysis would enable changing all words into their basic grammatical form. However, the data-mining tool used in this case does not have the function of stemming in the Polish language, so the programme showed the inflected forms of the word "operator."



Word	Attribut...	Total O...	Docum...	depalet...	transpo...	rozdmu...	kontrol...	nalewa...	kontrol...	etykieci...	kontrol...
operator	operator	5	2	2	0	3	0	0	0	0	0
operator...	operator...	1	1	1	0	0	0	0	0	0	0
operatorzy	operatorzy	1	1	1	0	0	0	0	0	0	0
pomylony	pomylony	1	1	1	0	0	0	0	0	0	0
programu	programu	8	8	0	1	1	1	1	1	1	1
przeszkoli	przeszkoli	2	1	2	0	0	0	0	0	0	0
przybycie	przybycie	1	1	1	0	0	0	0	0	0	0
skrzynek	skrzynek	1	1	1	0	0	0	0	0	0	0
skrzynki	skrzynki	1	1	1	0	0	0	0	0	0	0

Fig. 4. List of report-generated attributes

The analysis of the production plant data with respect to the presence of words related to machinery operators in the plant's failure reports enabled us to determine the machines with the highest frequency of failures caused by operators that affected the maintenance reliability of these machines. Such operators may be sent to additional training in order to eliminate stoppage resulting from their mistakes.

Conclusions

Companies take numerous actions to improve the operational efficiency of their technical infrastructure. They are more and more aware of the significance of collected historical data in the form of either failure records or the values of the residual processes of the systems monitoring machinery operation.

In order to eliminate recurrent failures or ungrounded calls of maintenance service, it is recommended to examine the registers of machine shutdowns collected in IT systems. The very process of analysis can be considerably improved using specialist tools for data mining.

The objective of applying data mining tools to failure frequency records and thereby obtaining information about undesired occurrences caused by human factors is to support failure frequency control of a machinery fleet by *text mining* methods.

Treating the register of a company's technical infrastructure as a source of information and a basis for actions to be taken can help eliminate waste, rationalize performed operations, help design trainings for operators, and, most of all, improve the operational efficiency of a machinery fleet.

References

1. ASTOR: Inwestujemy w automatykę, ale dane nadal przetwarzamy ręcznie. Inżynieria&Utrzymanie Ruchu, 1(112)/2016, 14–15.
2. Chen, M.S., Han, J., Yu, P.S.: Data mining: an overview from a database perspective, IEEE Trans. Knowledge and Data Engineering. 1996, 8, 866–883.
3. Demski T.: Maszyna do czytania, czyli text mining w wykrywaniu nadużyt ubezpieczeniowych, StatSoft Polska Sp. z o.o., 2011, s. 98, [online].
4. Elliot B. R., Hill G.: Total Productive Maintenance. Is it time to move on? Logistics Solutions 1999;1.
5. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R.: Advances in knowledge discovery and data mining, MIT Press, 1996.
6. Fronczak E., Michalcewicz M.: Zastosowanie narzędzi eksploracji danych data mining do tworzenia modeli zarządzania wiedzą. Polskie Stowarzyszenie Zarządzania Wiedzą. Seria: Studia i Materiały. 2010, 27, 126–139.
7. Gibiec M.: Data mining w systemie monitorowania pracy kombajnów górniczych. Wiadomości Górnicze. 2011, 62/7–8, 398–405.
8. Han, J., Kamber, M.: Data mining: concepts and techniques. Morgan Kaufmann Pub.. 2006.
9. Hearst M., A.: Untangling Text Data Mining, Proceedings of ACL, 37th Annual Meeting of the ACL, New Jersey 1999, 3–10.
10. Hearst M.: Untangling Text Data Mining. In the Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland, June 20–26, 1999.
11. http://www.statsoft.pl/Portals/0/Downloads/Text_mining_w_wykrywaniu_naduzyc_ubezpieczeniowych.pdf, dostęp: 25 lipiec 2016.
12. Oxford University Press, 2002.
13. Kruczek M., Żebrucki Z.: Doskonalenie procesów utrzymania ruchu w przedsiębiorstwie branży hutniczej. Logistyka, 2012, 2, 787–797.
14. Kuligowska K., Lasek M.: Eksploracja danych tekstowych (text mining) w przedsiębiorstwie. Polskie Stowarzyszenie Zarządzania Wiedzą. Seria: Studia i Materiały, nr 13, 2008.
15. Liewa W.T., Adhityab A., Srinivasana R.: Sustainability trends in the process industries: A text mining-based analysis. Computer in Industry, 65(3), 2014, 393–400.
16. Lula P.: Text mining jako narzędzie pozyskiwania informacji z dokumentów tekstowych. StatSoft, 2005.

17. Marzec M., Uhl T.: Wykorzystanie metod eksploracji danych przy ustalaniu strategii eksploatacji taboru autobusowego. Problemy Eksploatacji. Maintenance Problems, 2013, 1, 45–58.
18. Mirończuk M.: Przegląd metod i technik eksploracji danych tekstowych. Studia i Materiały Informatyki Stosowanej, Tom 4, Nr 6, 2012, 25–42.
19. Neumann G., Piskorski J.: A Shallow Text Processing Core Engine. Computational Intelligence, No 18, 2002, 451–476.
20. Solka J. L. Text Data Mining: Theory and Methods. Statistic Survey 2/2008, 94–112.
21. Sołdacki P.: Zastosowanie metod płytkiej analizy tekstu do przetwarzania dokumentów w języku polskim. Rozprawa doktorska, Politechnika Warszawska 2006.
22. Ur-Rahman N., Harding J.A.: Textual data mining for industrial knowledge management and text classification: A business oriented approach. Expert Systems with Applications, 39(5), 2012, 4729––4739.
23. Zakrzewicz M.: Data Mining i odkrywanie wiedzy w bazach danych. Materialy konf. PLOUG'97, Zakopane, 1997.
24. Żabicki D.: Współpraca CMMS z systemami ERP i monitorowaniem produkcji. Inżynieria i Utrzymanie Ruchu Zakładów Przemysłowych. Nr 7/2010, 62–65.