

STUDY ON EFFECTIVENESS OF USING COLUMN-ORIENTED DATABASES IN THE PROCESSING OF MEASUREMENT CHARACTERISTICS OF AN ELECTRIC VEHICLE

Andrzej CZEREPICKI¹

¹ Warsaw University of Technology, Faculty of Transport, Warsaw, Poland

Abstract:

Electric vehicles are increasingly popular means of transport. One of the most important problems of their operation is to optimize the use of a battery pack. It requires to analyze the operational characteristics of a vehicle in motion, which are stored in a database. If the measurement data are collected from many vehicles, the efficiency of their analysis is important. The objective of this article is to study the possibilities of using modern column-oriented databases in order to increase the efficiency of the analysis of selected operational characteristics of an electric vehicle. The research problem is a comparative analysis of the processing efficiency of selected measurement characteristics of an electric vehicle in relational and column-oriented data structures. Important analytical functions were formulated and recorded in the form of database queries. An experiment consisting in multiple execution of functions packages on various database structures, including a column-oriented one, was carried out. The execution time of packages and the IT system load were collected and analyzed. The analysis of the experiment results allows to conclude that the use of the column-oriented data structures made it possible to shorten the time of executing the functions analyzing the energy consumption by the electric vehicle's drive system. Depending on the type of the analyzed characteristics of the vehicle and its method of representation in the database, a significant reduction of the analysis time compared to the relational structure was obtained. Also, a decrease in the load on the computer system during data processing on the column-oriented structures was noted. The use of the column-oriented databases in the processing and analysis of measurement operational characteristics of electric vehicles is justified and it can bring measurable effects. It should be considered that the effectiveness of solving depends on the number of the analyzed characteristics and the format of their representation in the computer.

Keywords: IT transport systems, electric vehicles, NoSQL databases.

To cite this article:

Czerepicki, A., 2019. Study on effectiveness of using column-oriented databases in the processing of measurement characteristics of an electric vehicle. Archives of Transport, 51(3), 77-84. DOI: <https://doi.org/10.5604/01.3001.0013.6164>



Contact:

1) a.czerepicki@wt.pw.edu.pl [<https://orcid.org/0000-0002-8659-5695>]

1. Introduction

Electric vehicles are increasingly popular means of transport. It applies both to public transport - electric buses, PRT vehicles (Choromański et. al., 2014), etc. - and means of personal transport, such as an electric car or a bicycle. One of the most important issues when travelling by electric means of transport is to optimize the use of a battery pack to cover the selected route for the purpose of possible minimization of energy consumption. It requires analyzing the operational characteristics of a vehicle, such as speed, acceleration, current voltage and battery discharge current in relation to a profile of the covered route. These data are collected by the vehicle measurement systems and can be stored in built-in mass memory devices or transmitted on-line to a remote server. One example of the telemetry system that meets the above assumptions is a system (Tomczuk et. al., 2017).

The analysis of the collected data allows to develop methods and algorithms supporting the estimation of electricity consumption on the chosen route and energy consumption optimization through the selection of the appropriate driving mode. In the situation when the measurement data enters the IT system from many vehicles at the same time, the data processing efficiency is an important aspect of the computer system functioning (Kemp et. al., 2016). In view of a large size of the processed data sets, they are most often stored on durable media, access to which is implemented with the use of a database. Database design is a complex multi-stage process and requires a systematic approach (Jachimowski et. al., 2017). The effectiveness of the transport IT system depends to a large extent on a properly selected data model.

In recent years, the intense development of alternative NoSQL (*Not only SQL*) database systems, which represent different data models than the relational one (Jing, 2018; Abramowa et. al., 2014), has been observed. NoSQL databases can be used to solve specific transport tasks. Effective representation of the transport network structure (Żochowska et. al., 2018) is possible with the use of a graph database (Czerepicki, 2016). Issues related to the use of the document-oriented NoSQL database to process information on urban transport accessibility for passengers are discussed in (Vela et. al., 2017). NoSQL databases also can be successfully used to processing geospatial data in ITS systems (Detti et.

al., 2017). The column-oriented databases (CODB) constitute one of the NoSQL database categories (Idreos et. al., 2012). An important feature of the column-oriented model is the potentially high efficiency of operation on data aggregated in columns (Sun et. al., 2016). This is a prerequisite for its use in order to analyze the measurement data of the electric vehicle characteristics, in which the data record operation is usually performed once, however, the read operation can be repeated many times. The objective of the article is to study the impact of using the column-oriented data structures on the efficiency of the analysis of the electric vehicle's selected operational characteristics obtained via telemetry.

2. Structure of the electric vehicle's measurement characteristics

The electric vehicle's measurement characteristics can be divided into three categories. The first of them includes parameters typical for most land vehicles: instantaneous measurements of speed, acceleration and coordinates in the GPS system, and the covered distance. The second category includes characteristics specific for an electric vehicle: instantaneous measurements of voltage and current in the vehicle supply system, information on the battery status from the BMS (*Battery Management System*) system, etc. The third category can refer to characteristics not directly related to the vehicle traction system, such as temperature measurements of individual elements, travel time, etc.

In the telemetry system described in (Tomczuk et. al., 2017), which was used for further studies, the measurements of the current vehicle characteristics are made with a constant frequency f (in the considered example of the value $f = 10 \text{ Hz}$). For each measurement, its number i , voltage U_i and current I_i in the vehicle power supply, as well as instantaneous speed v_i , instantaneous acceleration in relation to the direction of movement A_i , GPS coordinates $\{Lon_i, Lat_i\}$, distance S_i covered from the moment of recording, temperature of the engine Ts_i , battery Tb_i and converter Tp_i are recorded. Selected measurement features with sample values are presented in Table 1.

Fig. 1 presents a graphic visualization of the selected measurement characteristics of the electric vehicle.

Table 1. Measurement data structure with sample measurement values

Measurement number i	Battery voltage U_i , [V]	Battery current I_i , [A]	Car speed v_i [km/h]	Car acceleration, A_i [m/s^2]	Distance S_i , [m]	Temperature Tb_i [$^{\circ}C$]	GPS Coordinates	
							Lon_i	Lat_i
19290	41.84	66.33	18.425	0.944	606	24.02	21.0057933	52.221908
19300	44.85	29.67	22.371	0.797	611	24.03	21.0057937	52.221908
19310	41.79	65.74	17.359	-1.229	616	23.93	21.0057942	52.221908
19320	42.75	52.36	31.322	3.385	624	24.11	21.0057950	52.221908
19330	43.08	52.94	28.473	-0.766	633	24.08	21.0057990	52.221908
...

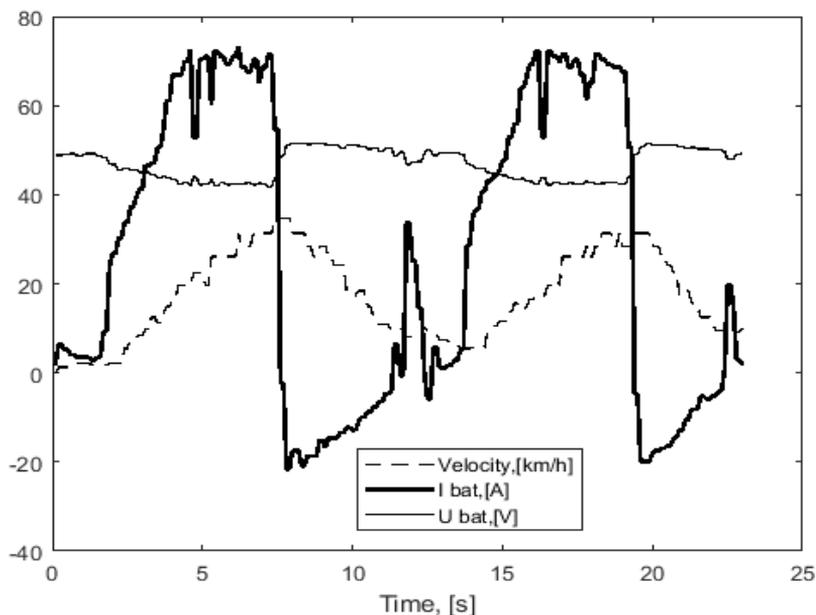


Fig. 1. Sample visualization of the vehicle's measurement characteristics

In order to record the vehicle measurement characteristics in the database, it is important to define a logical entity-relationship data model (ER model), and then, to implement it in the form of a physical data model according to the ultimately selected database. The basis for the logical data model of the computer telemetry system of electric vehicles is comprised of three basic entities: *Vehicle* – representing individual units of electric vehicles, *Trip* – measurements made while covering the designated

route by the selected vehicle and *Measurement* – single measurement characteristics (Fig. 2).

The physical implementation of the data model in the form of the ultimately selected database structures require, among others, the determination of the data types of individual attributes of each entity. Table 2 shows the data types of the Measurement entity and their size in bytes in the implementation of the Microsoft SQL Server database that was used in the studies.

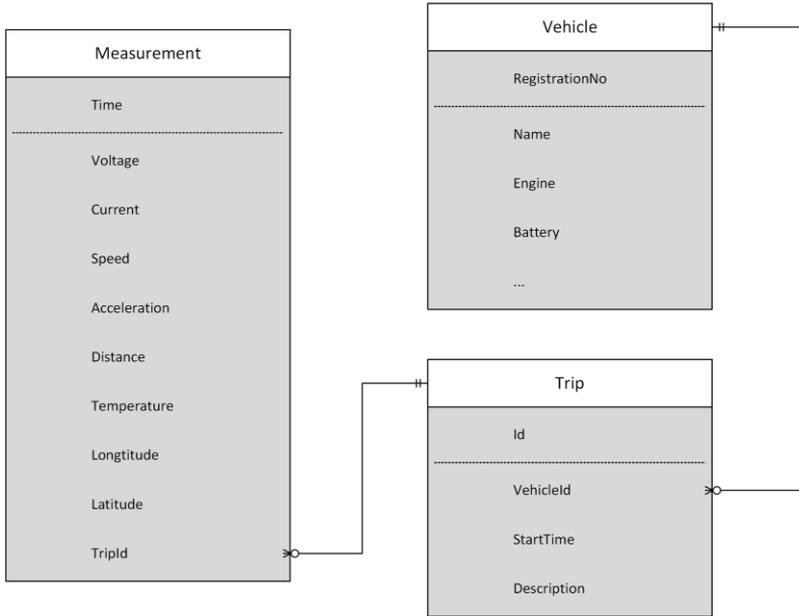


Fig. 2. ER data model of the vehicle’s measurement characteristics

Table 2. Data types of individual measurement fields and their size

Field	Data type	Size in bytes
Measurement number	Integer	4
Time	Datetime	8
Voltage	Decimal	5
Current	Decimal	5
Speed	Decimal	9
Acceleration	Decimal	9
Distance	Integer	4
Temperature	Decimal	5
GPS Coordinate	Decimal	9

3. Functions used for data analysis

The analysis of the electric vehicle energy consumption while covering the route requires calculations on the unit measurements collected in the database. The following functions are important in terms of the analyzed energy consumption problem:

- a) calculation of the minimum $P_{min}(a, b)$ or maximum $P_{max}(a, b)$ value of the selected measurement characteristics $P \in \{U, I, v, A, T\}$ (voltage, current, speed, acceleration, temperature, etc.) within a specified range $(a, b), a < b$

- b) calculation of the average expected value of the selected characteristics in the range:

$$P_{avg}(a, b) = \frac{\sum_{i=a}^b P_i}{b - a}$$

- c) calculation of the estimated electricity consumption in the range

$$W(a, b) = \sum_{i=a}^b \frac{U_i \times I_i}{f}$$

- d) calculation of the route S on the basis of the speed and acceleration measurements

$$S(a, b) = \sum_{i=a}^b \left(\frac{v_i}{f} + \frac{A_i}{2f^2} \right)$$

The measurements of the operational characteristics of a single-person electric vehicle obtained during a test drive on the open road constituted a set of source data. Microsoft SQL Server 2014 Developer Edition version was used as a database, which as a relational database, at the same time, offers the support for the column-oriented data structures (Larson et. al., 2012). As server was used computer with the Intel

Core i5-4300 processor, 8 GB internal memory and the SSD hard disk with a capacity of 500 GB.

The idea of the experiment was to run analytical packages of functions on identical measurement data sets stored in the form of a relational structure and a column-oriented database. The following data structures were analyzed: *Default* – default disordered tabular database structures, *RS (RowStore)* – indexed relational database data structure, *CS (ColumnStore)* – column-oriented data structure. The package contained $n = 100$ commands that call each of the above-presented functions on randomly generated arguments. At the launch of the package, the execution time on each of the tested data structures and processor loading were recorded. The experiment with the launch of each package was repeated $k = 30$ times in order to obtain statistically representative results. The database server before the start of each measurement was restarted. In order to minimise the impact of optimisation algorithms built in the database management system on the experiment result, the cache memory and mechanisms for collecting the statistics of the execution of queries were deactivated.

4. Experiment results

At the first stage of the studies, the time of executing the function $V_{min}(a, b)$ returning the minimum value of the vehicle speed in the determined time interval was calculated. Fig. 3 presents a graph of relationships of the time of executing the package of functions with a size of $n = 100$ for individual data structures. As it can be observed, in case of searching for the minimum (or maximum) value of a single measurement, the use of a regular index *RS* on the column *V* allows to shorten the operation execution time to ~33% of the reference time (structure *Default* was adopted as a reference point), and the use of the column-oriented form of the *CS* data representation is less effective in this case and it shortens the operation time to ~54%. Assuming that the search time of a given measurement interval (a, b) is similar for all compared structures, the greater efficiency of the structure application *RS* results from sorting the values V_i according to the index, which in the search allows to significantly reduce the number of the analyzed data records.

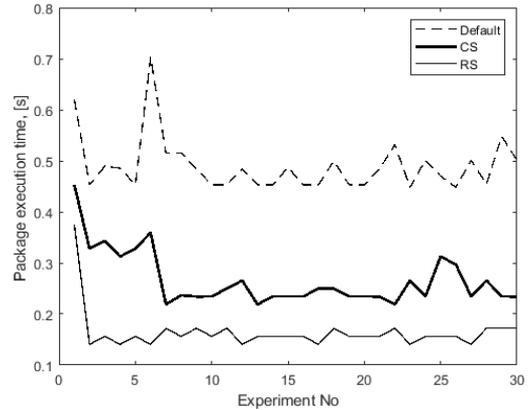


Fig. 3. Function $V_{min}(a, b)$ package execution time

However, in case of aggregate functions (sum, average, etc.), which operate on data sets, the effectiveness of using the regular index *RS* aims to zero. Fig. 4 presents the measurement results of the time of executing the package of functions $V_{avg}(a, b)$, which calculates the vehicle speed in the time interval determined by arguments a and b . As it can be noticed, the time of calculating the function value for the default data structure *Default* was slightly extended by adding the time of adding up the values. The introduction of the regular index *RS* in this case resulted in a symbolic increase in efficiency ~2%. The column-oriented index *CS*, in turn, proved to be a much more effective solution showing almost double reduction of the calculation time.

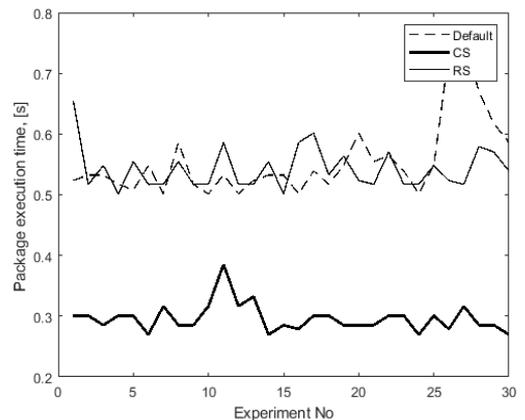


Fig. 4. Function $V_{avg}(a, b)$ package execution time

Subsequently, the experiment with the measurement of the time of executing the package of functions $W(a, b)$ of estimating the electricity consumption for the time interval (a, b) was carried out. As before, the time of performing the operation on the default data structure in the base was chosen as a reference point. According to the analysis of the experiment results (Fig. 5), it can be concluded that with an increase in the number of columns occurring in the query, there is the time extension of the calculation execution for each of the tested structures. The application of indexes in the structure *RS* is not recommended in this case because it increases the total size of data files and it can contribute to the calculation time increase even in comparison with the structure *Default*. The use of the column-oriented data structure *CS* allowed to reduce the execution time of the queries by an average of ~30%. Therefore, it is possible to conclude that the effectiveness of using the column-oriented data structures depends on the number of columns in the query.

The next analysis concerned the time of executing the function of the $S(a, b)$ vehicle route calculation on a randomly selected section. The experiment results are shown in Fig. 6. The column-oriented data structure *CS* in this case shows ~12% reduction of the calculation time compared to *Default*.

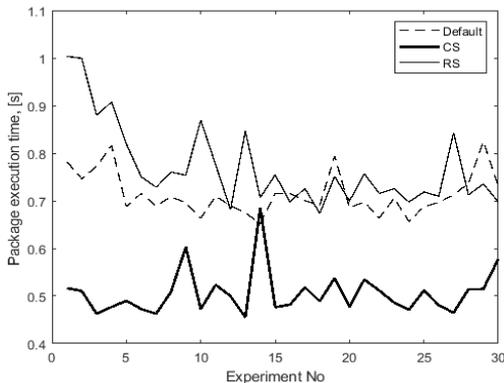


Fig. 5. Function $W(a, b)$ package execution time

By analyzing the causes of the efficiency reduction in comparison with the previous experiments, it is important to pay attention to the size of data structures representing individual attributes of the *Measurement* class. In order to store attributes *Velocity* and *Acceleration*, a 9-byte data type *Decimal* was applied, and in case of attributes with less precision

Current and *Voltage*, the *Decimal* structure size was 6 bytes. It allows to conclude that the efficiency of data processing in the column-oriented structure decreases with an increase in the size of the analyzed fields.

During the experiments, the system load statistics were also collected. Fig. 7 presents a graph of the load of a central processing unit (CPU) for individual data structures during the launch of the package of functions $W(a, b)$. As it can be noticed, the use of the column-oriented data structures allows to reduce the processor unit load, at the same time, reducing the mean square deviation from this value. From a practical point of view, it means the reduced, more even and predictable load of the processor in the process of data processing on the column-oriented structures.

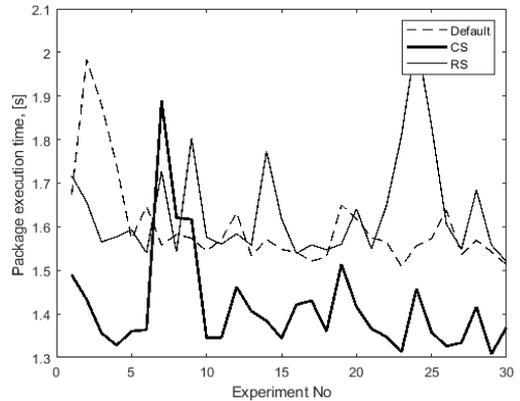


Fig. 6. Function $S(a, b)$ package execution time

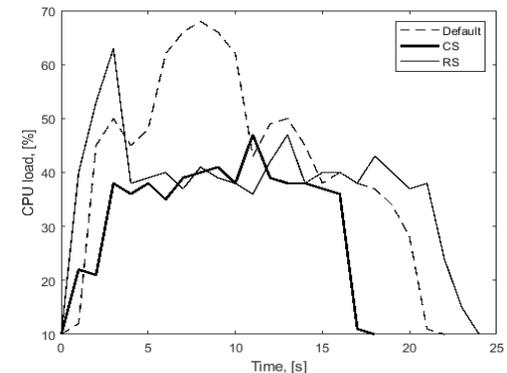


Fig. 7. Comparison of the load of the CPU calculation unit during the measurement data processing for various database structures

A summary of the time of executing the packages of the analyzed functions on individual data structures in relation to the default structure was presented in Table 3.

Table 3. Relative execution time summary of the packages of functions

Function	Data structure		
	Default	Column store (CS)	Row store (RS)
$V_{min}(a, b)$	100%	54%	33%
$V_{avg}(a, b)$	100%	53%	98%
$W(a, b)$	100%	71%	108%
$S(a, b)$	100%	88%	102%

5. Conclusions

The modern IT systems in transport operate on large data sets. The intensive implementation of electric vehicles requires the analysis of these operational characteristics in motion. The measurement data obtained during control passages constitute a source of data for such an analysis. In view of the large volume of collected characteristics, which in specific cases, may include many vehicles, e.g. in public transport, an important aspect is the efficiency of data processing in the system.

The article proposes an approach for increasing the efficiency of the analysis of operational characteristics of the electric vehicle by using the modern column-oriented data model. The logical model of the measurement system was characterized, the most important entities and their attributes, as well as their representation in the physical structure of the database were determined. The functions calculating the average value of the tested characteristic, which are important in the perspective of the data analysis, as well as the estimated energy consumption and vehicle distance were formulated. The measurements of the time of executing the packages of individual functions on the relational and column-oriented database structures were carried out.

By analyzing the experiment results, it can be concluded that the use of the column-oriented data structures in most cases allowed to shorten the time of executing the queries in comparison with the relational structure. There was also a relationship between the number of columns applied in the query and their total size in bytes and the execution time of the query: the increase of both mentioned features

reduces the efficiency of using the column-oriented structures. There was also a decrease in the load on the computer system during data processing on the column-oriented structures.

In conclusion, the use of the column-oriented databases in the measurement data processing of operational characteristics of electric vehicles is justified and it can bring measurable effects in the form of shortening the expected time of executing the queries as well as contribute to reduce the system load, taking into the above-presented limitations.

References

- [1] ABRAMOVA, V., BERNARDINO, J. & FURTADO, P., 2014. Which NoSQL Database? A Performance Overview. *Open Journal of Databases*, 1(2), 17-24.
- [2] CHOROMAŃSKI, W., DASZCZUK, W., DYDUCH, J. et al., 2014. PRT (Personal Rapid Transit) Network Simulation. In: *Proceedings of the 13th World Conference on Transportation Research*, Rio de Janeiro, July 2013, arXiv:1711.07325.
- [3] CZEREPICKI, A., 2016. Application of graph databases for transport purposes. *Bulletin of the Polish Academy of Sciences. Technical Sciences*, 64 (3), 457-466, DOI: 10.1515/bpasts-2016-0051.
- [4] DETTI, A., ORRU, M., PAOLILLO, R. et al., 2017. Application of Information Centric Networking to NoSQL databases: The spatio-temporal use case. In: *2017 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*, Osaka, 2017, 1-6, DOI: 10.1109/LANMAN.2017.7972131.
- [5] IDREOS, S., GROFFEN, F., NES, N., MANEGOLD, S., MULLENDER, S. & KERSTEN, M., 2012. MonetDB: Two Decades of Research in Column-oriented Database Architectures. *IEEE Data Engineering Bulletin*, 35 (1), 40-45.
- [6] JACHIMOWSKI, R., GOŁĘBIOWSKI, P., IZDEBSKI, M., PYZA, D. & SZCZEPAŃSKI, E., 2017. Designing and efficiency of database for simulation of processes in systems. Case study for the simulation of warehouse processes. *Archives of Transport*, 41(1), DOI: 10.5604/01.3001.0009.7380.
- [7] JING, H., 2011. Survey on NOSQL Databases. In: *Proceedings 6th International Conference*

- on *Pervasive Computing and Applications*, Port Elizabeth, 363-366, DOI:10.1109/ICPCA.2011.6106531.
- [8] KEMP, G., VARGAS-SOLAR, G., FERREIRA DA SILVA, G., GHODOUS, P., COLLET, C. & AMALYA, P., 2016. Cloud big data application for transport. *International Journal of Agile Systems and Management*, 9, 232-250, DOI: 10.1504/IJASM.2016.079940.
- [9] LARSON, P., HANSON, E. & PRICE, S., 2012. Columnar Storage in SQL Server 2012. *IEEE Data Engineering Bulletin*, 35(1), 15-20.
- [10] SUN, S. & LIU, C., 2016. Application of improved storage technology in Intelligent Transportation System. *Advances in Transportation Studies*, 3, 51-60, DOI: 10.4399/978885489937705.
- [11] TOMCZUK, P., CZEREPICKI, A., KONIAK, M. & JASKOWSKI, P., 2017. The Design of the Telemetric System for Recording Operating Parameters of Electric Vehicles. *Logistics and Transport*, 1, 47-54.
- [12] THOMAS, G., ALEXANDER, G., SASI, P.M., 2017. Design of high performance cluster based map for vehicle tracking of public transport vehicles in smart city. In: *2017 IEEE Region 10 Symposium (TENSYP)*, Cochin, 1-5, DOI: 10.1109/TENCONSpring.2017.8070027.
- [13] VELA, B., CAVERO, J. M., CÁCERES, P., SIERRA, A. & CUESTA, C. E., 2017. Defining a NoSQL Document Database of Accessible Transport Routes. In: *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, Exeter, 1125-1129, DOI: 10.1109/iThings-GreenCom-CPSCom-SmartData.2017.172.
- [14] ŻOCHOWSKA, R. & SOCZÓWKA, P., 2018. Analysis of selected transportation network structures based on graph measures. *Scientific Journal of Silesian University of Technology. Series Transport*, 98, 223-233. ISSN: 0209-3324. DOI: <https://doi.org/10.20858/sjsutst.2018.98.21>.