

UDC 621.391

INDUSTRY 4.0 AND MODERN APPROACHES FOR BIG DATA ANALYTICS

Andriy O. Luntovskyy

BA Dresden University of Coop. Education Saxon Study Academy Dresden, Germany

Larysa S. Globa

National Technical University of Ukraine

Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine

Background. With the acceleration of industrial development, i.e. with the new “Industry 4.0”, structuring and processing of acquired voluminous and heterogeneous data become considerably more complicated and represents an important scientific-practical problem. Cyber-PHY, IoT, sensor networks, robotics, multiple real-time applications can generate large arrays of unmanaged, weakly structured and non-configured data of various types, known as “Big Data”. However, the problem of “Big Data” is very hard to solve nowadays.

Objective. The purpose of the presented in this paper research is to analyze the sources of Big data, determine their main characteristics and suggest the ways for overcoming the growing dimension of Big data.

Methods. In contradistinction to traditional ways of the Big data problem solving, when only dealing with certain empirical approaches and models, the paper proposes to introduce ontologies for describing data groups, use compression of data volumes into knowledge that significantly reduces their volumes and improves understanding of their sense.

Results. The effectiveness of given solutions is confirmed by the best known practices and our own case studies aimed at overcoming this well-known complex problem.

Conclusions. To overcome the problem of “Big Data” there is no single universal solution. The analysis shows that the solution can be found by introducing ontologies, determining the mutual influences and correlations between the data, thus gaining knowledge based on a huge amount of data.

Keywords: Big Data; Industry 4.0; 5G; IoT and Robotics; Blockchain; Analytics and Data Mining; Ontology; Cloud and Fog; Veracity.

I. MOTIVATION

Nowadays, “Big Data” accumulation is typical for trading and marketing, electronic payments, production process, for the traffic from mobile providers, international justice and forensics, for public fiscal authorities, pharmaceutical and advertising industry. A large number of research institutes, organizations and universities accumulate, store and process large amounts of technical and scientific information. The sources can be primarily classified as follows [1-4]:

- Data from sensors and production capacities (Smart Manufacturing)
- Navigation data from GPS and other navigation systems, e.g. Galileo
- Data from GIS (Geographic Information Systems)
- Traffic and telematics data (IoT)
- Robotics data (navigation, sensing, measurements)
- QoS management data and traffic data from mobile providers
- Power and renewable energy stations (data for smart metering)
- Data from scientific simulations and parallel computing (clusters, grids)

- Imaging data of patients from CT and MRT (Computer and Magnetic Resonance Tomography) in healthcare
- Marketing information from cookies, social networks and Web2.0 semantic networks
- Collected personal data of citizens and much more.

In the conditions of modern industrial development of “Industry 4.0” there are even more “Big Data” sources: home automation, patient health data, M2M and robotic data, business intelligence, pharmacological research, networking and experimental data.

The mobile networks and aps for the 5G will definitely take an active part in the process of receiving and processing of large data amounts [2,5-7] too. By the year 2020, the new 5G networks will use more than 50-100 billion sensors to download comprehensive information about how we interact with things that surround us or even are inside of us?!

As one of the most interesting further topics the Blockchain technology occurs. Nowadays, it is exponentially increasing and enables modern cryptocurrencies (Bitcoin, Monero, Ethereum, to name a few). The Blockchain technology and its associated applications like cryptocurrencies are so-called “resource eaters” due to their enormous energy and memory

consumption. Large amounts of chained crypto blocks are surely causing “Big Data problematics”.

Long-term mass data retrieving and processing requires large amounts of memory: decentralized, i.e. on private hosts and even on mobile devices, as well as centralized, in data centres and clouds. This means Big Data shortcoming too.

Another general problem is the economic and technological recognition of crypto-technologies, processing of mass tests for the validity of blocks, the prevention of criminal activities (extortion, fraud) based on Blockchain [2,18,26].

The further paper content was built as follows:

1. State-of-the-art approaches are analysed (Section II).
2. Big data shortcomings for important areas (best practices/ case studies 1-5) are discussed like WSN, health care, IoT and robotics (Sections III-V).
3. Analytics placement options to overcome Big Data are offered (cloud and fog based).
4. Regular paradigms for Big Data are overviewed (Section V) as well as empirical Data Analytics for Big Data are discussed (Section VI).
5. Finalising, conclusions and outlook have been done.

II. STATE-OF-THE-ART

Huge and heterogeneous data volumes (approx. 100PB to 100EByte) are currently so immobile that their access and management become very efficient lately without the use of special techniques [1-3]. Nowadays, Big Data are too complex for their efficient processing with classical manual methods for data structuring. Herewith recognition of the problem is given by one of the world's leading research and consulting company as very serious: the company Gartner Inc. notes “the problem of Big Data as one of the most important trends of IT-infrastructure development along with virtualization and energy efficiency of IT”. Under use of convenient retrieving, archiving and analytics methods for Big Data from [1-3] like databases and data warehouses, electronic sheets, formatted texts, CSV-data, Web- and XML-documents, graphical, audio and video documents data become faster and not suitable for anything and inadequate. To solve this problem the computational techniques of so called “Data Mining” inter alia with ontologies, data pattern recognition, fuzzy logic etc. are used. These techniques can correspondingly be based on cloud analytics and disassemble the complexity and heterogeneity [1-13]:

1. Taxonomies and predictive classification.
2. Preliminary clustering and compression.
3. Predictive deviation detection (based on regular statistics).
4. Descriptive association rule discovery (fuzzy based).
5. Ontologies and knowledge data bases (KDB).
6. Regression Models.
7. Sequential Pattern Discovery (Machine Learning).

Otherwise, such a large amount of data remains unstructured and will be in future characterized by the over-proportional management complexity. Additionally, significant increase of network traffic can be forecasted during the data processing under the use of typical modern solutions like Cloud

Computing with heterogeneity of networks, replication across multiple geographically-distributed computing nodes [2, 6, 7, 10].

III. The 6V PREVENTION FACTORS

However, there are six factors, which affect the complication or prevention of the use of the regular and/or (cloud-based) analytical methods (Fig. 1). They build so-called 6V paradigm which distinguishes from known 5V or even 10V-paradigms [1,3]:

1. Volume: the amount of generated and stored data is moving from 100PB up to 100EByte nowadays.
2. Velocity: it means the frequency and speed of data acquisition and processing (from slow batch techniques to real time with strong limitation on reaction time or latencies).
3. Variety: multiple codec types of multimedia data. This helps people who analyse it to effectively use the resulting insight.
4. Veracity: The quality of captured data can vary greatly, affecting accurate analysis.
5. Violation: The discussed data are often scattered, without any clustering or any structuring.
6. Value: Inconsistency of the data sets can negatively affect the processes of data handling and management.

Unfortunately the depicted V-factors for Big Data are growing faster than the performance of their analysis (in Mbyte/s or in GFLOPS) via classical computational techniques. Therefore, for Industry 4.0, IoT systems and M2M not only improvements in networking but also in analytic blocks became very significant [1-7,27,29], which aimed at decreasing of data streams and their algorithmic complexity.

Terabyte	Real time	Tables, sheets	Text, Social	Structured	RDB
Petabyte	Quasi real time	Photos	Videos	Clustered	Warehouses
Exabyte	Periodic	Graphics	Web, XML	Poly-structured	Ontologies
Zettabyte	Batch	Audios	Mobile Data, Sensors	Scattered	KB, FKB
Volume	Velocity	Variety	Veracity	Violation	Value

Fig. 1. The own 6V paradigm

IV. UNMANAGED AND UNSTRUCTURED DATA IN INDUSTRY 4.0

The goal of clustering and further processing of unmanaged and unstructured data from Industry 4.0 is the reducing of costs for their storage, better traceability, forecasting in process trends, workflow and production process optimization, data security and human private sphere protection. The paper discusses the Best Practises and Case Studies aimed at overcoming of the Big Data problematics if possible under the use of “data compression” via their transformations (Fig. 2).

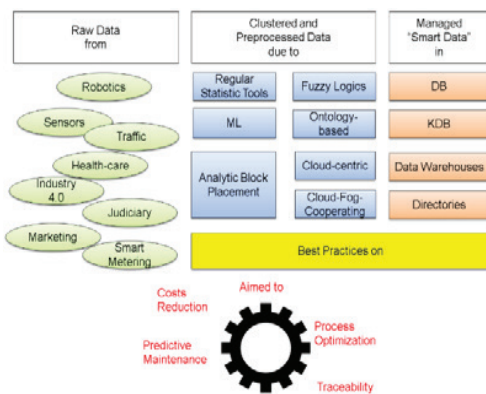


Fig. 2. Method classification: what is the aim of the clustering and further processing of unmanaged and unstructured data?

The “heavy-weighting approaches” involving precise analytical techniques and expensive specialized software are used for this aim. On the other hand, there is the opportunity to solve the Big Data problem under use of some “light-weighting approaches” based on agility: freeware, multipurpose techniques, minimal challenges on the personnel training and competencies. The paper examines the techniques and case studies of both topics. The “heavy-weighting approaches” (ontologies, knowledge bases, fuzzy logic and fuzzy knowledge bases) are compared to “light-weighting” ones. The existing reference solutions are discussed below.

So-called “Data Scientists” need to answer the following questions [8-16]:

1. How is the acquired data available?
2. When and with which periodicity the data survey acquisition and accumulation must be done?
3. In which way must data sampling be done?
4. How can the data be accessed?
5. Which formats and codecs are required to apply within data evaluation system or for processing in cloud analytics blocks?

The typical data acquisition and evaluation models include sampling, accumulation and survey processes and are characterised by the frequency of sampling: continuous, periodically with the isochronous sample and event-driven as well as via the communication type between the sensors and actuators: push model vs. pull model [8-16]. Mostly, the collected data is large unstructured and unmanaged. Just let us give an example.

Example 1: Data Analytics for „Smart Data”

Depending on the event frequency, event driven periodic, permanent, and the behaviour of the sensors (push, pull), a sensor survey and data accumulation are performed. Let WSN consist of 15000 sensors. Each sensor can transfer a short telegram up to 100 Bits. Thereby:

- The survey for each sensor is conducted 20 times per hour: 2000 Bits/h
- $x 24h = 48000\text{Bits}/\text{daily} = 6000 \text{ Bytes}/\text{daily}$ for each sensor
- In general, an average sensor wireless network accumulates experimental data for 3 years $x 365$ days:

- $x 365 x 3 = 6.57 \text{ MB}$ for each of the sensor
- The overall-data for the mentioned network:
 $6.57\text{MBytes} x 15000 \text{ sensors} = 100 \text{ GB}$ of raw data!

Further research of the data accumulated in this way can be carried out by manual and automated methods both (Fig. 3).

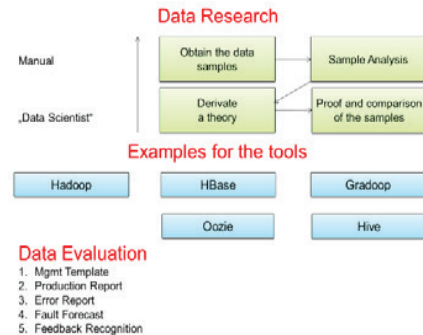


Fig. 3. Data Analytics: Manual and Automated

As a result of such data evaluation research the following common documents can be issued [15,16]:

1. Management Template.
2. Production Report.
3. Error Report.
4. Fault Forecast.
5. Feedback Recognition.

Often the data research cannot be completely studied or their trends recognized due to the complexity and heterogeneity. The future “Big Data – Smart Data” transition is given in Fig. 4.

The overcoming of myths and misconceptions is possible due to use of Machine Learning (ML). The (cloud-based) ML system enables artificial creation of knowledge from the obtained voluminous experimental data in background mode. An artificial system is “learned” from samples and examples and can summarize them after the completion of the study and evaluation (training) phase. The ML system recognizes templates and trends in research data. Thus, the ML (cloud) system can also evaluate data on representativeness and compliance.

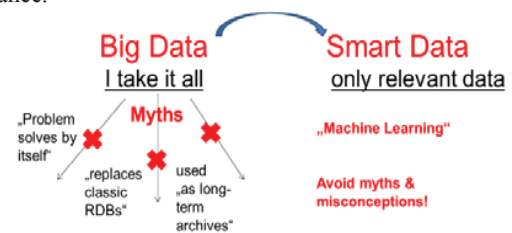


Fig. 4. “Big Data – Smart Data” transition

V. BIG DATA PROBLEMATIC FOR IOT AND ROBOTICS

The number of robots that co-operate with a person, with machines and with each other in various fields of industry and economy, everyday life and the entertainment industry, social pedagogy is growing rapidly. Let us give another example.

Example 2: Analytics Placement Options to overcome Big Data

The wide deployment of robots and robotics data create the right conditions for Big Data too. Up-to-date robotic platforms (Fig.5) are able to use virtualization technologies and virtual robots over the network [2,6,17-22].The up-date IoT and robotic applications can be classified into three groups. They solve the Big Data shortcomings in better way due to intelligent distribution of the Data Analytics [2, 6, 18, 22]:

- (1) Conventional Robots.
- (2) Cloud-Centric Robots.
- (3) Distributed (Fog-Cloud-cooperating) Robots.

Only the cloud-centric solution (2) and further distributed, fog-cloud-cooperating solution (3) both are able to overcome the discussed problem nowadays and in the future.

The analytic blocks, migration agents as well as further adaptive interfaces are delegated to the clouds and, possibly, after pre-processing and clustering backwards to the so-called “fog” under use of the mentioned solutions and protocols. The virtual analytical components (middleware, web services, and mobile software agents) are placed in the cloud and fog environment. Virtual cloud and fog solutions contain software components for the robots that implement rebootable (virtual) business processes [18-22]. Thus, we are talking about replaceable and customizable robotic algorithms (Fig. 5) in various fields of application (industry, medicine, communication and telecommunication, entertainment) which can collect, process and retrieve voluminous heterogenic Big Data in the given area.

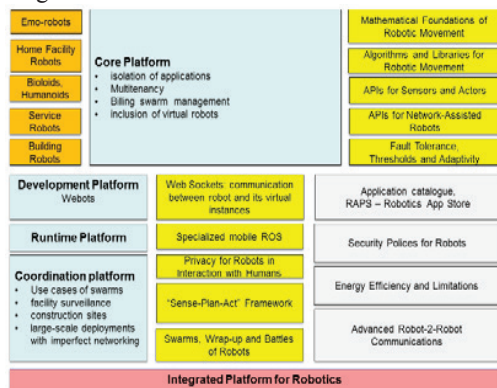


Fig. 5. The properties of the future robotics platforms

VI. REGULAR PARADIGMS FOR BIG DATA

For the processing of Big Data the usual statistical paradigms can be deployed like S, R, SPSS, Oracle R, SAP Hana, IBM SPSS, Netezza, and Grafana [23-25].

R Language (1992) as well as R Project environment are aimed to the processing of significant statistic data (up to 10^6 samples), calculation of the statistic moment of the probability distribution as well as their visualisation under GNU license. In the integrated R environment there are standard CLI and GUI. For data exports, graphic visualisation and report output all popular formats can be used: HTML, XML, LaTeX, PDF, MS Office. There are a lot of further R-editors and R-plugins, e.g. for Eclipse, Python, Gretl for all significant OS.

SPSS (IBM 1968) is a trademark of IBM, which develops and distributes software for statistics and data analysis (IBM

SPSS Statistics v25, PASW – Predictive Analysis SoftWare for MS Win, GNU/Linux, Mac OS X).

R Language is interoperable to a lot of modern packages for Big Data processing like SPSS, SAS, Platform Symphony, Tableau, SAP Hana, IBM Netezza.

Recently, the open platform Grafana can be used for real-time visualizations of fast processes of inter alia network traffic data and real-time sensor data. The Grafana system is designed to construct dashboards and graphs and is a complex multi-level open source web application based on the Apache License. The Grafana web application supports, in particular, further server components such as Graphite, InfluxDB, Opentsdb (as the backend). Grafana is also suitable as an open platform for efficient analytics and monitoring of Big Data.

And, last but not least, the Big Data Appliance is NoSQL-Cluster from application servers for massive-parallel analysis based on the integrated R tools and Apache Hadoop.

VII. EMPIRICAL DATA ANALYTICS: CASE STUDIES / BEST PRACTICES

Example 3: Ontology-Based Medical Imaging and Data Analytics

Medical formats for CT and MRT applications are based on the open standard DICOM (Digital Imaging and Communications in Medicine) founded by Siemens and Philips. DICOM stores, manages or transmits the compressed patient Big Data. DICOM standardizes both the data storage format and communication protocol, which is represented by digital medical images, data for their segmentation, surface definition or registration. The DICOM format is also the basis for archiving (Picture Archiving and Communication System, PACS) in the clinics and hospitals.

The ontology-based framework for data analysis [3,6] process enormous mass of patient data, which is stored in a variety of formats: handwritten texts, body imaging (CT, MRT), tests, forms with laboratory results, genomic data. Often such data should be processed in a few steps. Unstructured and unmanaged data must be converted to structured formats through their classification (taxonomies) and ontology creation. The given method is a so-called “heavy-weighting”[3,4,8].

Example 4: A Light-Weighting Approach

Let us examine a system example for overcoming Big Data complexity which was developed via TIQ Solutions in Leipzig [3,15,16]. The mentioned reference solution of TIQ Leipzig is represented below. The system is based on “a light-weighting approach” and is oriented for processing of Big Data in the domain of Business Analytics. The system architecture is depicted in Fig. 6. The most important quality criteria of the referred solution are wide scalability and suitability for the Linux-clusters. The system allows you to connect to the conventional DB. The complex poly-structured and redundant retrieved data can be processed with higher performance within an enterprise or institution data centre or a cluster. Some modules implemented in Java allow even real-time control. The expenditures in the form of investments CAPEX (Capital Expenditures for hardware, cable infrastructure, premises) and OPEX (Operational Expenditures for licenses, personnel,

electricity, ongoing maintenance) are significantly reduced on the basis of proven freeware components. Therefore, the solution is relevant for so-called middle-range enterprises.

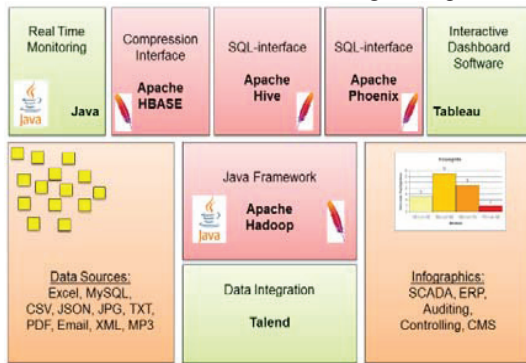


Fig. 6. Architecture Example for Big Data processing

The discussed architecture is characterized by agility, possesses no rigidity and offers only small license costs by the use of proven freeware. The components of the architecture [3,15,16] are as follows:

1. Apache Hadoop (2008) was implemented in Java and represents a freeware framework for scalable distributed applications under the use of Google-similar algorithms and GFS (Google File System), which supports computing with significant data (Petabytes-area) on the computer clusters.
2. Apache HBASE is a scalable DB for management of Big Data within the Hadoop clusters.
3. Apache Phoenix is a massively paralleled RDB-Engine with OLTP concept for Hadoop with Apache HBase as back-store. SQL queries are processed, then compiled into the series of HBase scans and are orchestrated to produce the regular JDBC results.
4. Apache Hive: a DBMS for Hadoop, aimed at data aggregation, queries and analysis with a SQL-similar interface. The data can be retrieved in diverse DB and file systems.
5. Tableau (2003): software for data visualization and reporting.
6. Talend: cloud-based software for Big Data integration.
7. SCADA (Supervisory Control and Data Acquisition): the computer-assisted monitoring and control of technical processes.

Example 5: An Approach of Big Data complexity overcoming for Telecom Operator

The telecom companies process huge multimedia data amounts. Unfortunately, such highly dynamic sphere is characterized by an essential lag in the discussed problematics. Telecom data require involving of special approaches, methods, algorithms and tools for Big Data analytics and processing:

- call options (different parameters of customer calls)
- customer profiles (personal customer data, duration of customer calls and other used services, billing records)
- network data (different parameters, operations, infrastructure parameters and technical data)

- data from financial reports, questionnaires, advertising, company plans, users' applications etc.

Therefore, the following data taxonomies and data scales should be considered:

- amount of records
- amount of calling (and called) customers
- diversity in calling behaviour
- amount of used services and services delivery options (characteristics)
- transmitted video-volumes.

But every analytic request uses just a part of all these data. In this regard, there is a need to systematize and classify data from various sources. Ontological models are used as a problem solution [27-29]. This is the first step for Big Data overcoming.

The second step is loading the data to execute the query. This requires a dynamic workflow generation mechanism based on the ontology of data, knowledge and computational components. One of the approaches to designing such kind of mechanism is discussed in [9].

The third important step is fuzzy knowledge base (FKB) design for possibility to use not very hard computational process during the telecoms operation. The huge amount of statistics can be classified and clustered using the algorithms considered in [27]. But every cluster is only one fuzzy rule. The rule set is a FKB, but to check its correctness is possible using so-called meta-graphs. FKB correctness checking is discussed in [10]. The obtained FKB can be used as the basis for a decision block [29] for IoT or other telecoms platforms (Fig.7).

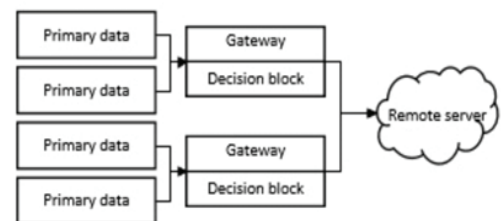


Fig. 7. Cloud architecture with a decision block [29]

VIII. CONCLUSIONS AND OUTLOOK

The given work is dedicated to Big Data sources in such modern issues like Industry 4.0, Robotics, IoT, 5G, Blockchain and best practices for analytics. The paper discusses the Best Practises and Case Studies aimed at overcoming the Big Data problematics if possible under use of "data compression" via their transformations:

1. Own 6V-paradigm on Big Data distinguishing is formulated.
2. Primarily the state-of-the-art approaches are analysed. The most common techniques classification and overview is given. So-called "heavy-weighting" (ontologies, FKB, meta-graphs) and "light-weighting" approaches are examined.

3. The Big data shortcomings for important areas (examples 1-5) were discussed like WSN, healthcare, telecom companies and robotics.
4. Regular solution paradigms for Big Data are overviewed (e.g. R Tools, SPSS, Grafana).
5. The (cloud and fog based) analytics blocks placement options to overcome Big Data are offered.
6. Best Practices for Empirical Data Analytics for Big Data are discussed.

This work can be qualified as a “work-in-progress”. The authors are looking for new efficient method combinations to overcome the above mentioned 6V-paradigm.

ACKNOWLEDGEMENT

Authors' acknowledgements to BA Dresden, the PhD students and to the colleagues J.Spillner with ZHAW Zurich, J.Pejas, A.Cariov, W.Rogoza, A.Konys with ZUT Univ. of Technology Szczecin, W.Dargie with Dresden Univ. of Technology, D.Blankenbergh with TIQ Solutions Leipzig for technical support, inspiration and challenges by fulfilling of this work.

REFERENCES

- [1] M.Ulema. Big Data and Telecommunications Telecom Analytics, Tutorial at Int. IEEE Conf. BlackSeaCom-2016, Varna, Bulgaria.
- [2] A.Luntovskyy, J.Spillner. Architectural Transformations in Network Services and Distributed Systems: Service Vision. Case Studies, Springer Nature, 2017, 344p. (ISBN: 9-783-6581-484-09).
- [3] L.Globa, I.Svetsynska, A.Luntovskyy. Case Studies on Big Data, Journal of Theoretical and Applied Computer Science, JTACS, Polish Academy of Science, Gdansk, No. 2, 2016, ISSN 2299-2634, 10p.
- [4] A.Konys, W.Rogoza. Big Data and Ontologies. Talk at ACS Int. Conf. 2016 in Międzyzdroje, Oct. 2016, 3p.
- [5] A.Luntovskyy. Advanced Software-Technological Approaches for Mobile Apps Development, 14th Int. IEEE TCSET-2018 Conf., Lviv-Slavske, 2018, 6 p. (IEEE Xplore: <https://ieeexplore.ieee.org/document/8336168/>), DOI: 10.1109/TCSET.2018.8336168.
- [6] A.Luntovskyy. SLMA and Novel Software Technologies for Industry 4.0, 21-st Int. Conf. ACS-2018, Szczecin-Międzyzdroje, 2018, in: J.Pejaś, I.El Fray, T.Hyla, J.Kacprzyk (eds.). Advances in Soft and Hard Computing, Springer Int., 12p. (Part of the AISC book series, vol. 889, DOI: <https://doi.org/10.1007/978-3-030-03314-9-16>, ISBN: 978-3-030-03313-2).
- [7] A.Luntovskyy, D.Guetter, M.Klymash. Up-to-date Paradigms for Distributed Computing, Int. IEEE Conf. AICT-2017, Lviv, pp. 113-119 (IEEE Xplore), ISBN: 978-1-5386-0638-4, DOI: 10.1109/AIACT.2017.8020078.
- [8] E.Kuiler. From Big Data to Knowledge: An Ontological Approach to Big Data Analytics, Review of Policy Research, Vol.31, No.4 (2014).
- [9] L.Globa, R.Novogrudska, A.Schill. The approach to engineering tasks composition on knowledge portals, Proc. of SPIE, Vol. 10445, 2017: <https://www.scopus.com/record/display.uri?eid=2-s2.0-85058990971&origin=resultslist&sort=plf-f&src=s&sid=15ac89516ace1e66c12de1910>.
- [10] L.Globa, M.Ternovoy, O.Shtogrina, O.Kryvenko. Based on force-directed algorithms method for metagraph visualization, Advances in Intelligent Systems and Computing, vol.342, 2015, pp. 359-369.
- [11] S.Russell, P.Norvig. Artificial Intelligence: A modern approach, New Jersey, Upper Saddle River, 2010.
- [12] M.Jones. Artificial Intelligence: A Systems Approach, Hingham, Massachusetts, New Delhi, Infinity Sci. Press LLC, 2008.
- [13] B.Marr, J.Wiley. Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance, Sons Ltd, 2015.
- [14] IBH Reports, IBH Dresden Workshop on 23.3.2017(in German).
- [15] D.Blankenbergh. Big Data in der Industrie 4.0, TIQ Solutions Leipzig, Workshop an der IBH Dresden, 13.3.2018(in German).
- [16] TIQ Solutions Leipzig: <https://www.tiq-solutions.de/>.
- [17] A.Luntovskyy, D.Guetter. Allgegenwaertige Vernetzung: Industrie 4.0, Internet der Dinge, Fog Computing u.v.m ! in BA SaxonyMagazine“Wissen im Markt”, Vol. 2/2018, Glauchau, 8p. (in German).
- [18] A.Luntovskyy, M.Klymash. Software Technologies for Mobile Apps, Apps for Fog Computing, Robotics and Cryptoapps, Lviv, 2019, 247 p. (Monograph, in Ukrainian).
- [19] A.Luntovskyy. Modern Apps and Platforms for the Robots, Int. Conf. on Theoretical and Applied Aspects of Modern Infocommunications, Vyshkiv-Carpathians-2019, Jan. 2019, 4p.
- [20] J.Spillner. Serverless Cyber-Physical Applications, Keynote Talk at Workshop "Science meets Industry", 2018, Fraunhofer-Institute Centre Dresden, 20 p.
- [21] A.Luntovskyy et al. Intelligent Networking and Bio-Inspired Engineering, Int. IEEE Conf. UkrMiCo-2017, Odessa, Ukraine, 4p.(IEEE Xplore: <http://ieeexplore.ieee.org/document/8095421/>).
- [22] Int. Federation of Robotics:<http://www.ifr.org/>.
- [23] R-Project: <http://www.R-project.org/>.
- [24] SPSS:<http://www.ibm.com/analytics/de/de/technology/spss/>.
- [25] Grafana: <http://docs.grafana.org/>.
- [26] A.Luntovskyy, D.Guetter. Cryptographic Technology Blockchain and its Applications, Proc. of UkrMico-2018, Odessa, Ukraine, Springer Int., 2019, 15 p.
- [27] L.Globa, R.Novogrudska, A.Koval. Ontology Model of Telecom Operator Big Data, Int. IEEEConf. BlackSeaCom-2018: <https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=8419227>.
- [28] L.Globa, V.Kurdecha, I.Ishchenko, A.Zakharchuk. Fuzzy Logic Usage for the Data Processing in the IoTNetworks, Proc. of OSTIS-2018, Minsk, Belarus, p. 287-292.
- [29] L.Globa, V.Kurdecha, I.Ishchenko, A.Zakharchuk, N.Kunieva. The Intellectual IoT-System for Monitoring the Base Station OoS, Int. IEEE Conf. BlackSeaCom-2018:<https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=8419227>.

Лунтовський А.О., Глоба Л.С.

Індустрія 4.0 та сучасні підходи до аналітики «Великих даних»

Проблематика. Враховуючи швидкий промисловий розвиток, появу нової «Індустрії 4.0», структурування і обробка отримуваних об'ємних і різномірних даних значно ускладнюються і являють собою важливу науково-практичну проблему. Cyber-РНУ, IoT, сенсорні мережі, робототехніка, безліч додатків реального часу можуть генерувати великі масиви некерованих, слабо структурованих і неконфігурованих даних різних типів, відомих як «Великі дані». Однак проблему «Великих даних» в даний час дуже складно вирішити, тому що поки немає прямих способів або готових рішень для її подолання.

Мета досліджень. Метою досліджень, представлених в роботі, є аналіз джерел великих даних, визначення основних їх характеристик, а також шляхів подолання наростаючої розмірності великих даних.

Методика реалізації. На відміну від традиційних шляхів вирішення проблеми великих даних, коли розглядають тільки певні емпіричні підходи та моделі, в статті пропонується ввести онтології для опису груп даних, використовувати стиснення обсягів даних до рівня знань, що істотно зменшує їх обсяги і покращує розуміння їхнього змісту.

Результати досліджень. Ефективність розглядуваних рішень підтверджується кращими відомими практиками і власними дослідженнями, спрямованими на подолання цієї складної проблеми.

Висновки. Для подолання проблеми «Великих даних» немає єдиного універсального рішення, проведений аналіз показує, що рішення можна знайти за рахунок введення онтологій, визначення взаємних впливів і зв'язків між даними, таким чином, отримуючи знання на основі великого обсягу даних.

Ключові слова: Великі дані, індустрія 4.0; 5G; IoT та робототехніка; Blockchain; аналітика і інтелектуальний аналіз даних; онтологія; хмари і туман; достовірність.

Лунтовський А.О., Глоба Л.С.

Индустрия 4.0 и современные подходы к аналитике «Больших данных»

Проблематика. С ускорением промышленного развития, новой «Индустрией 4.0», структурирование и обработка полученных объемных и разнородных данных значительно усложняются и представляют собой важную научно-практическую проблему. Cyber-РНУ, IoT, сенсорные сети, робототехника, множество приложений реального времени могут генерировать большие массивы неуправляемых, слабо структурированных и неконфигурированных данных различных типов, известных как «Большие данные». Однако проблему «Больших данных» в настоящее время очень сложно решить, так как пока что нет прямых способов или готовых решений для ее преодоления.

Цель исследований. Целью исследований, представленных в работе, является анализ источников больших данных, определение основных их характеристики, а также предложение путей преодоления нарастающей размерности больших данных.

Методика реализации. В отличие от традиционных путей решения проблемы больших данных, когда имеют дело только с определенными эмпирическими подходами и моделями, в статье предлагается ввести онтологии для описания групп данных, использовать сжатие объемов данных в знания, что существенно уменьшает их объемы и улучшает понимание их смысла.

Результаты исследований. Эффективность таких решений подтверждается лучшими известными практиками и собственными примерами, направленными на преодоление этой сложной проблемы.

Выводы. Для преодоления проблемы «Больших данных» нет единственного универсального решения, проведенный анализ показывает, что решение можно найти за счет введения онтологий, определения взаимных влияний и связей между данными, таким образом, получая знания на основе большого объема данных.

Ключевые слова: Большие данные; индустрия 4.0; 5G; IoT и робототехника; Blockchain; аналитика и интеллектуальный анализ данных; облака и туман; достоверность.