Multidimensional Representation of COVID-19 Data Using OLAP Information Technology

2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT) | 978-1-7281-7443-3/20/$31.00 ©2020 IEEE | DOI: 10.1109/CSIT49958.2020.9321889

Oleksii Duda
Computer Science Department
Ternopil Ivan Puluj National Technical University
Ternopil, Ukraine
oleksij.duda@gmail.com

Volodymyr Pasichnyk
Information Systems and Networks Department
Lviv Polytechnic National University
Lviv, Ukraine
vpasichnyk@gmail.com

Nataliia Kunanets
Information Systems and Networks Department
Lviv Polytechnic National University
Lviv, Ukraine
nek.lviv@gmail.com

Rzheuskyi Antonii
Information Systems and Networks Department
Lviv Polytechnic National University
Lviv, Ukraine
antonii.v.rzheuskyi@lpnu.ua

Oleksandr Matsiuk
Computer Science Department
Ternopil Ivan Puluj National Technical University
Ternopil, Ukraine
oleksandr.matsiuk@gmail.com

***Abstract*—In December 2019, an unknown disease was first reported in Wuhan (Hubei province, China), which subsequently spread around the world. COVID-19 is an infectious disease accompanied by severe acute respiratory syndrome (SARS-CoV-2) [1], which is called a "coronavirus" for the visual similarity of the pathogen to the crown [2]. On April 23 2020, the World Health Organization has identified more than 2.5 million confirmed cases of COVID-19 virus disease[3]. Considering the scope of spread and features of the pandemic caused by the COVID-19 virus, timely diagnosis of the disease of citizens is important. The first step in developing any diagnostic or treatment tools is to effectively process data collections, much of which are based on open data from national, provincial, and municipal health facilities. At the same time, the importance of IT support the processes of making operational and strategic decisions in the health sector, as a critical area of smart cities and smart regions is increasing. Therefore, developing effective and reliable means of generating and processing detailed collections of COVID-19 outbreaks in real-time is an important research area in the implementation of innovative information technology projects of the smart city and smart region class. It requires design, development and implementation of information technology and software-algorithmic tools to collect and process data globally.**

Keywords— information technology, data warehouse, data structure, analytical processing, COVID-19.

I. Analysis of the State of Research

Successful fight against the COVID-19 pandemic is the effective implementation of the two locomotives, one of which is a systematic approach to solving complex problems, and the second is modern information technology.

China first faced the COVID-19 pandemic, while making a number of system decisions based on the use of innovative information technologies [4]. Taking into account the large population and size of the country, limited time and resources, the use of effective communications, data exchange procedures and information technology were critically important.

Chinese mobile operators, together with Huawei, have built a specialized 5G network focused on supporting COVID-19 medical and treatment events and facilities, using cloud services and algorithmic software based on artificial intelligence systems [5].

In order to effectively counteract the COVID-19 pandemic, the integrated use of modern information technologies, including the Internet of Things (IoT), innovative communication networks (for example 5G), methods and tools for analytical processing of big data, artificial intelligence systems and blockchain technologies are considered in the article [6]. Kalkreuth and Kaufmann [7] submit a list of COVID-19 public online resources and open data sets on COVID-19 medical images.

Bullock and other authors [8] reviewed advanced researches on the treatment and prevention of COVID-19 using machine learning methods and artificial intelligence systems. Article [9] describes the process of collecting graphic data on X-rays images of patients diagnosed with COVID-19.

The structure of images metadata collected from open medical sources, websites, and publications is described. A group of authors [10] perform statistical analysis of metadata on clinical descriptions of pneumonia and mortality from COVID-19, and analytical overview of researches on the treatment and prevention of Covid-19 is presented in the paper [11].

An overview of the promising possibilities of using computer vision systems for effective diagnostics of COVID-19 [12] is based on the analysis of an extensive layer of publications.

The article [13] presents the results of analytical processing of geolocation data on spread of COVID-19 in real time using "Google map" [14] and "Google earth" [15].

The authors developed interactive visualization tools using Mapbox and JavaScript based on geographical reference tables [16]. As the authors note [17] large-scale collection and analytical processing of data with the proper level of privacy and confidentiality can help reduce the COVID-19 pandemic.

One of the positive examples of fighting the pandemic COVID-19 is demonstrated in Vietnam through the timely introduction of stringent measures of social distancing [18].When the neighboring countries – Taiwan and South Korea conducted massive population testing, Vietnam lacked the financial and material resources to carry out such actions. Therefore, selective testing and active prevention of the spread of infection was conducted.

The "low-cost model" currently implemented in Vietnam for containment of the COVID-19 outbreak is considered successful. The authors [19] presented a systematic review and critical evaluation of models of diagnosis and prediction of COVID-19 infection.

The use of a global model of disease transmission in metapopulation to generate predictions about the impact of travel restrictions on the national and international spread of the epidemic is described in the article [20].

Currently, 19 countries have access to data about the mobile location of citizens trying to fight the coronavirus outbreak [21], while a significant part of innovative information technologies have been successfully developed and implemented as part of projects of the smart city and smart region class.

A number of measures have been hold to protect the population and counteract COVID-19 in Ukraine. In particular, quarantine was enacted and a number of legislative acts were adopted [22], among which the Law of Ukraine "On amendments to the Law of Ukraine" On protection of the population from infectious diseases" concerning the prevention of the spread of coronavirus disease (COVID-19)" [23].

This law allows the processing of personal data without the consent of the person in order to counteract the spread of COVID-19, in the manner specified in the decision to set quarantine and on condition to use such data only for the purpose of carrying out anti-epidemic measures.

Particularly among such data: information concerning the state of health, hospitalization or isolation, surname, name, date of birth, place of residence, work (study), etc.

Within 30 days after the end of the quarantine, such data are subject to depersonalization, and if depersonalization is not possible – to destruction.

II. MULTIDIMENSIONAL REPRESENTATION OF COVID-19 DATA

One of the key components of information technology support of processes of monitoring of the situation of spread and treatment of COVID-19 is the organization of data warehouses using information technology of multidimensional data analysis – OLAP (online analytical processing), which is effectively used in the implementation of databases and data warehouses for resource networks, social-communication services in innovative information technology projects of smart cities and smart regions. The use of this technology provides an opportunity to analyze in detail the current epidemiological situation regarding COVID-19 and the treatment of infected citizens, as well as to identify trends and forecasts by comparing data belonging to different collections by origin and time range. The processes of generalization of detailed data, collected in medical institutions at different levels, that allows us to get new knowledge from the consolidated for various aspects, different types of information sets are usually used in medical OLAP systems [24].

The source datasets include information about COVID-19, health and treatment status, places of hospitalization or self-isolation, patients personal data, geotracking data, social communications, which are processed using OLAP technologies. OLAP information technology is based on a multidimensional data model that contains basic entities, in particular, the data hypercube *rel*, dimension *D*, attribute *A*, cell *X*, value *rel (D,A)* [25]. Multidimensional analysis of COVID-19 data provides research and analysis of the characteristics of municipal and open medical and sociocommunication resources, including the choice of a set of attributes used for parameterization of the information model [26-30] (Fig. 1).

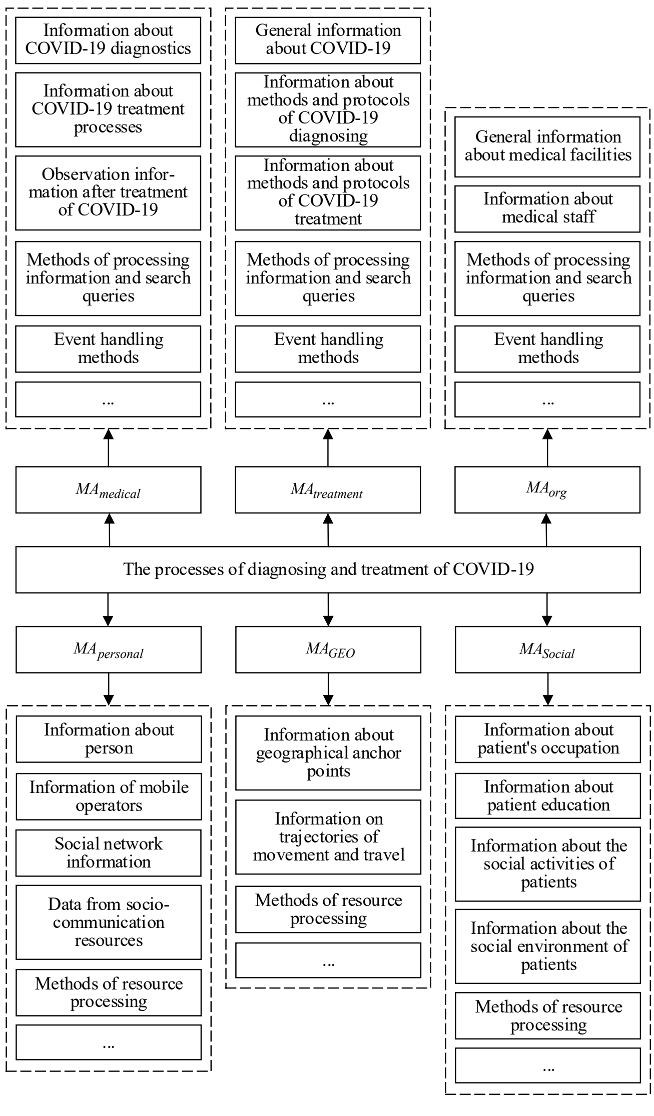

Fig. 1. Attributes of the data hypercube for information technology support of the processes of detection and treatment of COVID-19

The set of categories and attributes MA used to describe processes related to COVID-19 treatment and counteraction in conditions of smart cities and smart regions, is represented as:

$$MA = \langle MA_{medical} \cup MA_{treatment} \cup MA_{org} \cup MA_{personal} \cup MA_{GEO} \cup MA_{Social} \rangle \quad (1)$$

where $MA_{medical}$ is a category of attributes of characteristics and properties regarding general medical data about the patient's health status and treatment of COVID-19. In particular, the patient's history and appeals, information about examinations and treatment, information about results of diagnostic tests and procedures, information about medication and treatment procedures.

$MA_{treatment}$ is a category of attributes of characteristics and properties regarding methods of diagnosis and treatment of COVID-19.

MA_{org} is a category of attributes of characteristics and properties for the healthcare facility.

$MA_{personal}$ is a category of attributes of characteristics and properties in relation to personal data of a patient infected with COVID-19. According to the above Law of Ukraine [23], the depersonalization sets of information are subject to removal in 30 days after quarantine.

MA_{GEO} is a category of geolocation attributes, including data collections for geotracking monitoring.

MA_{Social} is a category of attributes regarding the patient's social communications. A significant number of attributes used to describe the processes of diagnostics and treatment of COVID-19 are synonymous and have the same meaning. These attributes can be divided into the following subcategories:

- *Metadata* is characteristics of medical metadata.
- *Docs* is characteristics of medical documentation.
- *Prov* is an information about health care providers.
- *Use* is an information on the consumption of health care.
- *Geo* is geolocation data.
- *Travel* is information about patient's travel.

The attributes of medical metadata *Meta* are shown in Table 1.

TABLE I. ATTRIBUTES OF MEDICAL METADATA

Attribute	Function
Metadata.Patient.ID	Unique patient's ID
Metadata.COVID.ID	Unique identifier of the registered case
Metadata.Offset	Number of days from the onset of symptoms and hospitalization of the patient
Metadata.Sex	Sex of the patient – Male (M), Female (F)
Metadata.Ages	Patient's age in years
Metadata.Survival	Death of the patient – Yes (Y) or no (N)
Metadata.Hospital.Id	Unique ID of the medical facility
Metadata.Critical_notes	Medical description

Medical documents of various types *Docs.Types* are components of data collections on COVID-19 treatment processes, in particular:

- Handwritten clinical records.
- Clinical records are submitted electronically.
- Emails.
- Scanned records.
- Text messages.
- Results of correspondence between medical professionals.
- Results of laboratory tests.
- X-ray images.
- Other types of medical images.
- Photos.
- Video and audio records.
- Images, captured using diagnostic and monitoring equipment.
- Patient's consent forms.

Appropriate sets of attributes and characteristics are used for each of the submitted types of medical documents. Methods of digitization and use of graphic, audio, video and medical images are provided for certain types of medical documents. Besides, data regarding health care providers (*Prov*) are components of data collections on COVID-19 treatment processes. In particular, clinics, hospitals, nursing homes and long-term care facilities, retailers and other health care providers, prevention services, financial and administrative institutions of the health care system, providers of auxiliary services and secondary care. Sets of attributes and characteristics are used for each of the submitted types of health care providers. The components of data collections about patients infected with COVID-19 are geographical encryption attributes in Table 2.

TABLE II. ATTRIBUTES OF GEOGRAPHICAL DATA ENCRYPTION

Attribute	Function
Geo.*PointID*	Unique identifier of the geographical anchor point
Geo.Latitude	Latitude
Geo.Longitude	Longitude
Geo.Resolution	Spatial representativeness of "latitude" and "longitude". A "dot" indicates that a specific location is represented by the given coordinates.
Geo.Traectory	Travel trajectory. Can be used to search for unidentified social contacts of a patient infected with COVID-19
Geo.Country	Country
Geo.Regio	Elements of the administrative-territorial hierarchy. Region/District/Territorial community.
Geo.City	Settlement
Geo.Location	Address, location

When identifying social contacts during the incubation period of COVID-19, the patient's travel history is important. Components of data collections regarding travels *Travel* :

- *Travel.ID* is an unique travel identifier.
- *Travel.History.Dates* is recorded start dates, key moments, and end dates of the patient's travel.
- *Travel.History.Locations* is a description of a recently recorded travel history.
- *Travel.Additional_information* is any additional information that may be useful regarding travel or patient when establishing social contacts during the COVID-19 incubation period. Each of the attributes for a separate category

$$Medical = \langle Medical_{Meta}, Medical_{Docs}, Medical_{Prov}, Medical_{Use}, Medical_{Geo}, Medical_{Travel} \rangle \quad (2)$$

provides a description of the relevant characteristics of the processes taking place in medical facilities, and together – fully characterize the processes of providing services for the diagnosis and treatment of COVID-19. In the process of practical implementation of the designed information technology platform, the contexts of constructing measurements of data hypercubes in accordance with the stages of data processing will be generated. When forming the database structure for information technology support of the processes of diagnostics and treatment of COVID-19, the speed of the course of processes of data growth should be taken into account. The basic tuple of tables of a database prototype consists of more than twenty entities which contain more than 55 characteristics. In addition, a tuple that includes more than seven entities containing more than twenty characteristics is automatically generated for each patient added to the system. Entities descriptions and characteristics are generated separately for storing the results of analytical processing of data accumulated in the system.

CONCLUSIONS

Within further research, the above attributes will be used to build a multidimensional information model of the provision and consumption of medical services and the implementation of measures for the treatment of COVID-19. When building an information model, data hypercubes will be analyzed, along with data sampling for specific values of a fixed set of dimensions. Formed as prototypes of data warehouse, based on constructed hypercubes, which will be used for monitoring, analysis of medical services provision and forecasting in the context of overcoming the global civilizational challenge that is, the COVID-19 pandemic, is the result of research. These data warehouses will be used to develop prototypes of information systems and software-algorithmic complexes integrated into the projects of the smart city and smart region class.

REFERENCES

[1] Paules Catharine I., Hilary D. Marston, and Anthony S. Fauci, "Coronavirus infections—more than just the common cold," in Jama, vol. 323.8, 2020, pp. 707-708.

[2] Chen, Yu, Qianyun Liu, and Deyin Guo. "Emerging coronaviruses: genome structure, replication, and pathogenesis," in Journal of medical virolog, vol. 92.4, 2020, pp. 418-423.

[3] WHO. Coronavirus disease (COVID-19) outbreak situation, https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[4] Make time: how China used 5G to combat coronavirus. https://www.epravda.com.ua/publications/2020/04/13/659305/.

[5] HUAWEI CLOUD: Fighting COVID-19 with Technology. https://www.huawei.com/en/press-events/news/2020/4/fighting-covid-19-with-technology.

[6] Ting, Daniel Shu Wei, Lawrence Carin, Victor Dzau, and Tien Y. Wong, "Digital technology and COVID-19," in Nature Medicine, vol. 26, 2020, pp. 459–461.

[7] Kalkreuth Roman, and Paul Kaufmann. "COVID-19: A Survey on Public Medical Imaging Data Resources," https://arxiv.org/abs/2004.04569

[8] Bullock Joseph, Alexandra Luccioni, Katherine Hoffmann Pham, Cynthia Sin Nga Lam, Miguel Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," https://arxiv.org/abs/2003.11336

[9] Cohen Joseph Paul, Paul Morrison, and Lan Dao. "COVID-19 image data collection," https://arxiv.org/abs/2003.11597.

[10] Shangxia Jiang, Yueming Wu, Tianzheng Lou, Junlong Xu, Yu Zhang, Hu Chen, Hewei Xu, "A Meta-analysis of Clinical Characteristics and Mortality COVID-19 Pneumonia," https://www.researchsquare.com/article/rs-18723/v2.

[11] Mohamad Chahrour, Sahar Assi, Michael Bejjani, Ali A. Nasrallah, Hamza Salhab, Mohamad Fares, Hussein H. Khachfe, "A bibliometric analysis of Covid-19 research activity: A call for increased output," https://www.cureus.com/articles/29507-a-bibliometric-analysis-of-covid-19-research-activity-a-call-for-increased-output.

[12] Anwaar Ulhaq, Asim Khan, Douglas Gomes, Manoranjan Paul "Computer Vision for COVID-19 Control: A Survey," https://arxiv.org/abs/2004.09420.

[13] Xu, Bo, et al. "Epidemiological data from the COVID-19 outbreak, real-time case information," in Scientific data, vol. 7, 2020, pp. 1–6.

[14] Google maps. https://www.google.co.uk/maps/.

[15] Google Earth. http://www.google.co.uk/intl/uk_uk/earth.

[16] HealtMap | COVID-19. https://www.healthmap.org/ncov2019/

[17] Ienca Marcello, Effy Vayena, "On the responsible use of digital data to tackle the COVID-19 pandemic," in Nature Medicine, vol. 26, 2020, pp. 463–464.

[18] The Secret to Vietnam's COVID-19 Response Success, https://thediplomat.com/2020/04/the-secret-to-vietnams-covid-19-response-success/

[19] Wynants Laure, et al. "Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal, https://www.bmj.com/content/369/bmj.m1328

[20] Chinazzi Matteo, et al., "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," in Science, vol.368, iss. 6489, 2020, pp. 395-400.

[21] 19 countries track mobile location to fight COVID-19, https://www.gpsworld.com/19-countries-track-mobile-locations-to-fight-covid-19/.

[22] Legislation concerning COVID-19, https://zakon.rada.gov.ua/rada/main/g23.

[23] LAW OF UKRAINE. On Amendments to the Law of Ukraine "On Protection of the Population from Infectious Diseases" on Prevention of the Spread of Coronavirus Disease (COVID-19), https://zakon.rada.gov.ua/laws/show/555-20.

[24] Gupta Yogesh Kumar, "Aspect of Big Data in Medical Imaging to Extract the Hidden Information Using HIPI in HDFS Environment." Advancement of Machine Intelligence in Interactive Medical Image Analysis. Springer, Singapore, 2020, pp. 19–40.

[25] Carlo Combi, Barbara Oliboni, Giuseppe Pozzi, Alberto Sabaini, Esteban Zimányi, "Enabling instant-and interval-based semantics in multidimensional data models: the T+ MultiDim Model," in Information Sciences, vol. 518, 2020, pp. 413–435.

[26] A. Rzheuskyi, N. Veretennikova, N. Kunanets, V. Kut, "The information support of virtual research teams by means of cloud managers," in International Journal of Intelligent Systems and Applications, vol. 10(2), 2018, pp. 37–46.

[27] J. Stolarek, P. Lipiński, "Improving watermark resistance against removal attacks using orthogonal wavelet adaptation," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012, 7147 LNCS, pp. 588-599.

[28] V. Tomashevskyi, A. Yatsyshyn, V. Pasichnyk, N. Kunanets, A. Rzheuskyi, "Data warhouses of hybrid type: features of construction," in Advances in Intelligent Systems and Computing II (AISC), vol. 938, 2019, pp. 325–334.

[29] A. Rzheuskyi, N. Kunanets, V. Kut, "Methodology of research the library information services: The case of USA university libraries" in Advances in Intelligent Systems and Computing, vol. 689, 2017, pp. 450–460.

[30] N. Boyko, O. Basystiuk, N. Shakhovska, "Performance evaluation and comparison of software for face recognition, based on dlib and opencv library", in 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), 2018, pp. 478-482.