JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Review of Epidemic Analysis Ontologies

[1]Asmae EL KASSIRI, [2]Fatima-Zahra BELOUADHA
[1]Professor, [2]Professor
[1,2]Computer Sciences Department,
[1,2]Mohammadia School of Engineers, Mohammed V University in Rabat, Morocco

Abstract: On the one hand, Machine Learning has proven its efficiency to analyze and predict epidemics evolution if the adequate data is used. On the other hand, Ontologies are a strong way to unify the understanding of information among people or software agents. In this context, we study the use of Ontologies to guide the data scientists and the data engineers to adopt the more meaningful data for the Machine Learning models aiming analyzing and predicting the epidemics evolution in general by taking Covid-19 as an example.

Index Terms **- Covid-19, Ontology, Analysis, Machine Learning, Epidemic**

I. INTRODUCTION

Throughout history, humans were forced to combat many epidemics using all possible resources and techniques. Naturally, during this last epidemic caused by the Covid-19 virus, the scientific community was systematically oriented to use the Information and Communication Technologies (ICT) to affront this epidemiology.

Being one of the most successful techniques for prediction, Machine Learning (ML) has been widely used in the healthcare domain for different objectives, mainly to analyze disease spread and to enhance the automatic diagnostic and prognostic. In the context of COVID-19, many studies have used the ML for many objectives. Some were focused on predicting the outbreak, others on identifying positive patients using radiological images or laboratory test results, while some others aimed at predicting disease progression, outcomes (recovery and mortality), length of stay, and the number of days spent in the intensive care unit (ICU) [1].

At the beginning of an epidemic, it is very interesting to be able to quickly evaluate and predict the situation's evolution in the objective to react with efficient and fast decisions. The Covid-19 experience has made humanity awake to be more creative and awesome by exploiting every existent tool and technology to save humanity. At the crisis, there were no medications or vaccinations. The unique available solution was to impose many precautions like confinement or congestion control or isolation, but it was necessary to evaluate those decisions efficiency to limit the virus propagation. In this context, ML tools have been widely exploited specially to predict the outbreak evolution and the efficiency of those precautions.

From this experience, we have learned that we need a preliminary preparation for possible similar situations in the future. In this context, we think that we need a common presentation of the more influencing coefficients on the epidemic propagation to facilitate their understanding and their reuse in the future. Thus, we have opted for an ontology.

Ontology is a way to represent a common understanding of a domain [2]. The objective is to unify the understanding of information among people or software agents [3]. Many ontologies were proposed in the Covid-19 context to collect data and reuse it to allow discovering new knowledge [4].

Therefore, the result of this work is a review of the most recent and important Ontologies for epidemics Analysis in the objective to give to the data-scientists and data-engineers a unified and semantic model to semantically present the needed data to analyze the more influencing factors on an epidemic propagation.

The rest of this article is constituted from four other sections. The first one is dedicated to presenting the materials and methods used to accomplish this review; the second one to represent and discuss our results; and finally, we conclude and present some limitations.

II. RELATED WORKS

Liu et al. have addressed a review of ontologies designed for crisis management [5]. By analyzing the synthesis of their reviewed ontologies (Fig. 1), we can deduce that they were not interested by verifying if statistical data, which can be used for crisis analysis, were covered.

Ontology Name	Use Case	Role of the Ontology
MOAC	crowdsourcing linked open data for humanitarian information management (Ortmann et al, 2011)	It provides a shared vocabulary for incident reporting as linked open data.
NNEW weather ontology	4-D Wx Data Cube – semantic search	It enables the 4-D Wx Data Cub to "translate" between terms to return all semantically similar data and discover resources without exact keyword match via ontology alignments.
FOAF	MOAC	The FOAF vocabulary (foaf:agent) is used in the MOAC ontology to define the class "ShelterAgency" (a humanitarian partner involved in shelter related humanitarian response).
SIADEX	planning system (de la Asunción et al., 2005)	It serves as a knowledge base for providing knowledge required by the planning process.
ISyCri	reasoning about the treatment processes (Benaben et al., 2008)	Can be injected into Protégé for reasoning. Through using Semantic Web Rule Language (SWRL), emergent process elements (e.g. potential actors, preventive services, curative services) are deducted from the reasoning procedure.

Fig. 1. Use Cases of the Ontologies for Crisis Management and the Roles of the Ontologies [5]

Yousefianzadeh and Taheri have addressed a review of the proposed ontologies by the National Center for Biomedical Ontology [6]. Actually, aware of the ontologies power in the battle against COVID-19 epidemic, the National Center for Biomedical Ontology has proposed an open registration site for biomedical ontologies entitled BioPortal [3] [4] to publish the proposed ontologies. Thus, Yousefianzadeh and Taheri have studied and compared the five proposed ontologies at the time of their article publication. Actually, at the time of writing this article, the number of proposed ontologies stands at 16. They simply cited the objective of each ontology and its potential uses.

In fact, by investigating works related to ontologies designed to gather and represent data for crisis analysis by Machine Learning (ML) techniques, we have discovered that no work was proposed in literature for this objective. Thus, we have decided to conduct this investigation in the objective of verifying if any existent ontology allows guiding ML analysis for epidemics analysis and especially for COVID-19 analysis. As a result, we have decided to review the more recent proposed ontologies in order to verify their ability to give a common way for data presentation that can be used to analyze the epidemic situation and make adequate ML analysis in order to predict the epidemic evolution.

III. Materials and Methods

To elaborate the epidemic ontologies review, we have adopted a method based on the SLR (Systematic Literature Reviews) approach [10] and inspired from [11] and [12]:

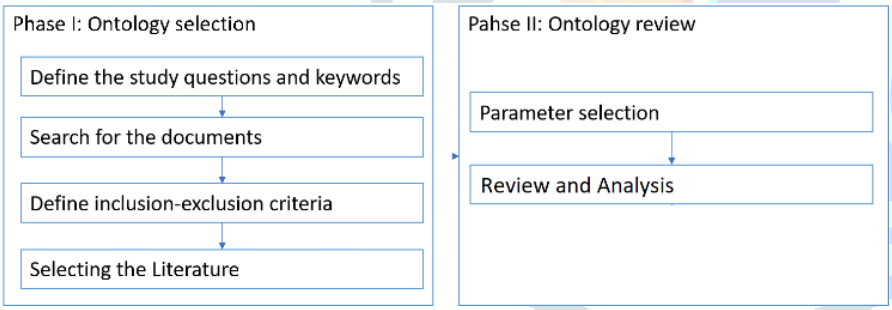

Fig. 2. The review method

The aim of the first phase is to select the more relevant ontologies of our review. Thus, we have:

2.1 Defined the study questions and keywords:

Table 1 The Defined questions and keywords

Questions	Keywords
Q1. What are the ontology-based models for epidemics? Q2. Are epidemics analysis represented using ontologies? Q3. Are needed data for epidemics analysis represented using ontologies? Q3. What quantitative measures are presented by ontologies?	K1. Epidemic K2. Pandemic K3. COVID-19 K4. Ontology

2.2 Searched the documents:

We have chosen Google Scholar to search for the defined keywords using the string search "Epidemic Ontology" OR "COVID-19 Ontology" OR "Pandemic Ontology".

As a result, we have retrieved 164 documents. After removing not relevant and duplicated documents, the titles were reduced to 19.

2.3 Defined inclusion-exclusion criteria:

Table 2 The criteria of inclusion-exclusion documents

Criteria categories	Inclusion	Exclusion
Publication status	Published in journals and conference	Unpublished
Availability of Description	Available explicit description	Not available
Relevance	Answers the research questions	Doesn't answer the questions
Language	Written in English Language	Other languages
Function	Domain and Application Ontology	Global Ontologies
Perspective	Engineering point of view	Philosophical and anthropological point of view

2.4 Select the Literature:

After applying the above criteria, the works were reduced to 3:

Table 3 The Selected Works

Ontology	Title	Author	Year of Publication	Objective
CIDO [13]	CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis	He, Y.; Yu, H.; Ong, E.; Wang, Y.; Liu, Y.; Huffman, A.; Huang, H. H.; Beverley, J.; Hur, J.; Yang, X., Chen, L.; Omenn, G. S.; Athey, B.; Smith, B.	2020	To create a knowledge base for standard and logical representation of heterogeneous coronavirus knowledge
CODO [14]	CODO: An Ontology for Collection and Analysis of Covid-19 Data	Dutta, Biswanath; DeBellis, Michael	2020	To provide a standards-based open-source model that facilitates the integration of data from heterogeneous data sources
CEMO [9]	Covid19 Epidemiology Ontology	Rosinach, Núria Queralt; Schofield, Paul; Hoehndorf, Robert; Weiland, Claus; Schultes, Erik; Bernabé, César Henrique; Roos, Marco	2020	To make epidemiological quantitative data for monitoring the COVID-19 outbreak machine-readable and interoperable to facilitate its exchange, integration and analysis, to eventually support evidence-based rapid response

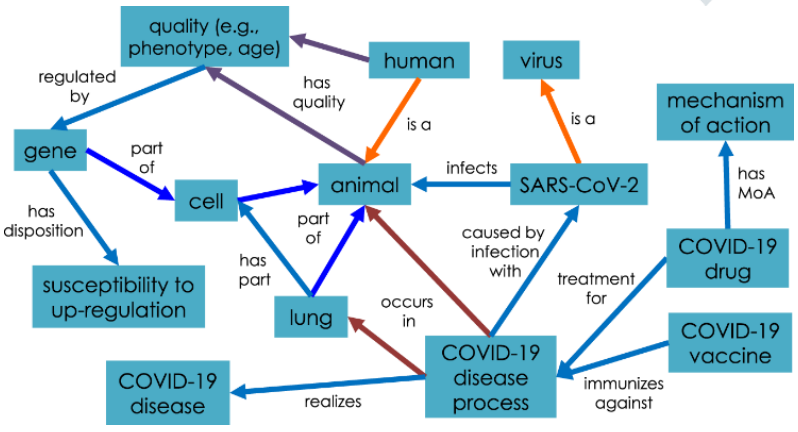

Fig. 3. A short model of the CIDO ontology [13]

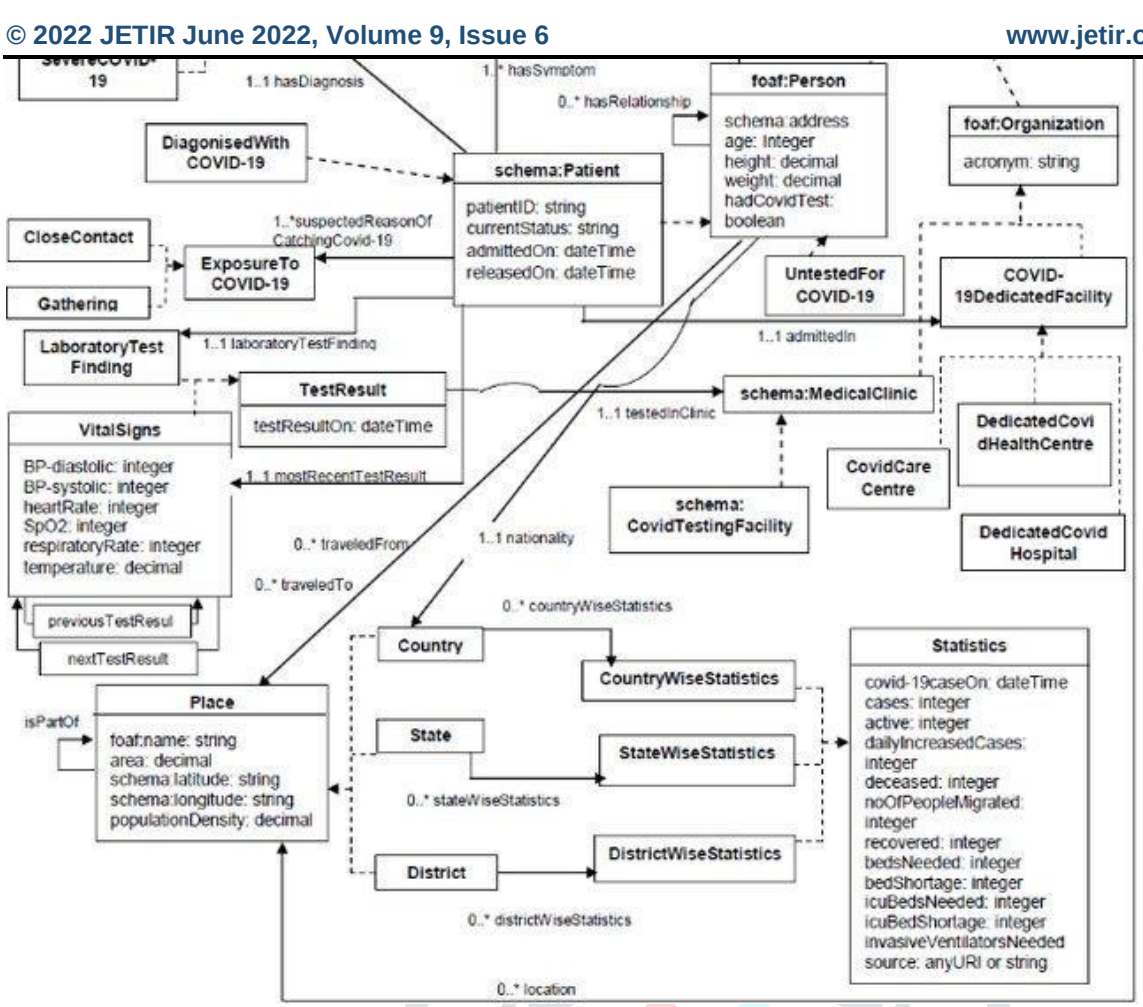

Fig. 4. An Overview of the CODO Ontology [14]

Fig. 5. A short model of the CEMO ontology [9]

IV. RESULTS AND DISCUSSION

Results

This section aims to evaluate the selected ontologies and their ability to respond to the defined objectives.

3.1 Parameter selection

Our objective is to evaluate the selected ontologies ability is to check the existent ontologies ability to evaluate the pandemic situations with concrete parameters to ensure accurate predictions. Thus, we will check if the selected ontologies are able to:

- C1. Represent the whole cycle of a pandemic (beginning of a pandemic, lockdown, end of lockdown, end of a pandemic, etc.);
- C2. Represent the different categories of population (infected, recovered, death, vaccinated, etc.);
- C3. Represent the measured statistics of the situation (mean, average, etc.);

3.2 Review and Analysis

The below table is summarizing the review of the three selected ontologies:

Table 4 A Review of the selected ontologies

	C1	C2	C3
CIDO [13]	COVID-19 disease and COVID-19 disease process to present the process evolution of COVID-19	--	--
CODO [14]	--	The people categories are partially presented as properties in the statistics entity	The Statistics entity and all its inherited entities to represent the statistical measures, and they are related to a Country, a State or a District
CEMO [9]	The role can be an evaluator role (stato:evaluant role), realized in (ro:realized-in) a process (cemo:process) having a type (stato:essay) and an output (ro:has-output)	The people entity (cemo:entity), has a type (ido.organisation population"), a quality (cemo:quality, iao:has quality) and a role (cemo:role, iao:has-role). The role denotes (iao:denotes) an identifier (cemo:identifier, appolo_sv:identifier); The quality has the type (ido:quality) and allows concretizing an observation (ro:concretizes);	The measures (iao:measurement type) are expressed in a unit (cemo:unit, ro:has-unit, iao:measurement label), with an error (cemo:error, cio:has-confidence level) and a value (stato:has value)

Discussion

By analyzing the selected ontologies during phase II of our review, we can note that:

- The CIDO ontology is just presenting the evolution of the virus but not the evolution of the pandemic situation by its different steps and associated statistics;
- The CODO ontology is able to present correctly the statistics associated to epidemics but not their cycles;
- The CEMO ontology is able to present the whole cycle of a pandemic, the different categories of population, and the measured statistics of the situation.

Thus, we can conclude that the CEMO ontology [9] is the closest to our objective aiming to select epidemiological quantitative data for monitoring the outbreaks. It allows presenting epidemics context and their associated necessary data that can be easily exploited in ML analysis.

V. CONCLUSION AND RECOMMENDATIONS

The principal objective of this study was to evaluate the existing ontologies provided in the context of COVID-19 and epidemics ontologies to identify the most adaptive to ML analysis. We have found that CEMO ontology [9] is the most suited to this objective.

Whatever, it will be useful to extend the CEMO ontology in order to become able to represent all the characteristics associated to each period of the epidemic cycle like the weather conditions (heat, rains, etc.), the economic conditions (unemployment, price fluctuations, etc.), the vaccination usage and evaluations, etc.

VI. ACKNOWLEDGMENT

This paper was written within the scope of a COVID-19 project supported by the supervisory ministry MENFPESRS and the CNRST of Morocco to predict the epidemics spread.

REFERENCES

[1] H. B. Syeda, M. Syed, K. W. Sexton, S. Syed, S. Begum, F. Syed, F. Prior and F. Y. Jr, "Role of Machine Learning Techniques to Tackle the COVID-19 Crisis: Systematic Review," JMIR medical informatics , vol. 9(1), p. e23811, 11 Jan 2021.

[2] D. Marcela, S. J. M. Cavero and E. M. Martínez, "THE ROAD TOWARD ONTOLOGIES," in Integrated Series in Information Systems book series, vol. 14, Springer US, 2007, pp. 3-20.

[3] N. F. Noy and D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology," Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report, vol. 32, 2001.

[4] A. Taheri and O. Yousefianzadeh, "COVID-19 Ontologies and their Applications in Medical Sciences: Reviewing BioPortal," Applied Health Information Technology, vol. 1, pp. 30-35, 2020.

[5] S. Liu, C. Brewster and D. Shaw, "Ontologies for Crisis Management: A Review of State of the Art in Ontology Design and Usability," Proceedings of the Information Systems for Crisis Response and Management conference (ISCRAM 2013), p. p.34, 12-15 May 2013.

[6] O. Yousefianzadeh and A. Taheri, "COVID-19 Ontologies and their Application in Medical Sciences: Reviewing BioPortal," Applied Health Information Technology, vol. 1, no. 1, pp. 30-35, 2020.

[7] D. Pati and L. N. Lorusso, "How to Write a Systematic Review of the Literature," The Center for Health Design HERD, vol. 11, no. 1, pp. 15-30, 2017.

[8] L. Alejandra, G. Camacho and S. NiceAlves-Souza, "Social network data to alleviate cold-start in recommender system: A systematic review," Information Processing & Management, vol. 54, no. 4, pp. 529-544, 2018.

[9] U. Varadarajan and B. Dutta, "Models for Narrative Information: A Study," arXiv, no. arXiv:2110.02084, p. N. 2110.02084, 2021.

[10] [Y. He, H. Yu, E. Ong, Y. Wang, Y. Liu, A. Huffman, H. H. Huang, J. Beverley, J. Hur, X. C. L. Yang, G. S. Omenn, B. Athey and B. Smith, "CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis," Scientific data, vol. 7, no. 1, p. Num 181, 2020.

[11] B. Dutta and M. DeBellis, "CODO: An Ontology for Collection and Analysis of Covid-19 Data," in Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, 2020.

[12] N. Q. Rosinach, P. Schofield, R. Hoehndorf, C. Weiland, E. Schultes, C. H. Bernabé and M. Roos, "Covid19 Epidemiology Ontology," BioHackrXiv, vol. Virtual Conference 2020, 11 August 2020.