



Covid19-IBO: A Covid-19 Impact on Indian Banking Ontology Along with an Efficient Schema Matching Approach

Archana Patel¹ · Narayan C. Debnath¹ · Ambrish Kumar Mishra² · Sarika Jain³

Received: 19 March 2021 / Accepted: 26 August 2021 / Published online: 15 October 2021
© Ohmsha, Ltd. and Springer Japan KK, part of Springer Nature 2021

Abstract

The exponential spread of Covid-19 is not only a serious concern for public health but has also severely affected the global economy. India is not an exception. The banking sector must plan innovatively in a wide range of scenarios focusing upon Covid-19 specific requirements. It becomes essential to examine the impact of Covid-19 on the performance of the Indian banking sector and take focused initiatives at both the tactical and the strategic levels. This paper offers the Covid-19 Impact on Banking Ontology (Covid19-IBO) that provides semantic information about the impact of Covid-19 on the banking sector of India. The developed ontology has been verified and validated and has been made available on the Linked Open Data cloud. It can be utilized to annotate the related data to provide meaningful insights. The Covid-19 ontologies already available have some overlapping information that causes redundancy. Unified integration of these ontologies is required to operate upon them unambiguously. It becomes reasonable to develop a matching approach to link all these ontologies semantically. We, therefore, also provide a schema matching approach with reasonable results to map the Covid-19 ontologies.

Keywords Big data · Ontology · Schema matching · Indian banking · Covid-19 · Evaluation

Abbreviations

GDP	Gross domestic product
RBI	Reserve bank of India
AI	Artificial intelligence
I/O	Input/output
IT	Information technology
OBO	Open biological and biomedical ontologies
W3C	World wide web consortium

✉ Ambrish Kumar Mishra
mailto:mailofambrish@gmail.com

Extended author information available on the last page of the article

OWL	Web ontology language
ATM	Automated teller machine
GoI	Government of India
ETB	Existing to bank
NTB	New to bank
PSL	Priority sector lending
NPA	Non-performing asset
FOAF	Friend of a friend ontology
SQL	Structured query language
DB	Data base
SPARQL	Protocol and RDF query language
IRI	Internationalized resource identifier
OWL2DL	Web ontology language2 description logic
OWL2RL	Web ontology language2 rule language
OWL2EL	Web ontology language2 existential quantification
	OWL2QL: Web ontology language2 query language
T-Box	Terminological-box

Introduction

The outbreak of infectious diseases shapes our society and create socio-economic problems in a way that no other phenomena does. Some most notable epidemic outbreaks in the human history include the HIV/AIDS, the 1918 Spanish Flu, SARS (Severe Acute Respiratory Syndrome), Zika, and Ebola. These pandemics affected a limited geographical area and with limited losses. Covid-19 has spread exponentially as a full-blown pandemic more than any other pandemic in history and has economic impacts both globally and for India. Many experts have even called this global economy disruptor a Black Swan event, an event that is unpredictable and has serious and extreme consequences in all aspects socially, economically, as well as politically. Restrictions like Nationwide lockdowns, shutting down public places and transport, social distancing, and work from home are all changing the behavior and life style of people. This in turn disrupting the physical operations and supply chain, posing big challenges to the financial institutions. Moreover, the stress built-up among non-banking financial companies (NBFCs) have limited their capacity to lend, further adding fuel to the fire.

The Covid-19 pandemic has hit the world at such a time when India was facing a negative growth in GDP and was trying to get on track by recovering at a slow rate. Due to this pandemic, the recovery process is severely impacted. A recent report published by the Reserve Bank of India indicates that in the last two quarters, 19 Indian sectors are being adversely impacted by this pandemic resulting in the stress of debt having a value of 1.552×10^7 million [1]. The micro, small, and medium enterprises (MSME) and retail loans are predicted to be at the most risk. Table 1 shows the adverse impact on % share in banking sector debt and debt in million, respectively.

Table 1 Impact assessment in percentage and corresponding value of Debt in Rs. [1]

Impact on banking sector	% Share in banking sector debt	Debt in million
Impacted by Covid-19 but not stressed pre-Covid-19	29	1.552×10^7
Impacted by Covid-19 but stressed pre-Covid-19	42	2.22×10^7
Not impacted by Covid-19	26	1.354×10^7
Not impacted by Covid-19 but stressed pre-Covid-19	3	1.51×10^6

Banks are considered as the heart of the financial system as they are the intermediaries between the borrowers and the depositors. To keep the financial system up and running, the stability of banks is of utmost importance. The banking sector is the one that is predicted to recover the last from this pandemic. It has become challenging for traditional banks to recalibrate and manage changes in liquidity. Multiple other challenges have come to the forefront including digital banking, credit risk, crisis management, and cyber threat. In addition to other challenges, the major technological challenge is the inability to access systems and data because of various constraints. Protecting and driving the existing businesses requires multi-dimensional initiatives in the post-Covid-19 scenario. The financial sector must plan innovatively in a wide range of scenarios focusing upon Covid-19 specific requirements. It becomes essential to examine the impact of Covid-19 on the performance of the Indian banking sector to take precautionary measures in the future. Focused initiatives are required at both the tactical (to address immediate concerns) and the strategic levels (to recalibrate business models) for the economic rejuvenation in the long run. The fast and effective immunization of the India's large population can be a key to avoid repeated disruptions.

The Reserve Bank of India (RBI) has already introduced various measures to bridge over the catastrophic situation that Covid-19 has led us to, like giving relief to the lending institutions. In addition to what RBI is doing, new technological advances like artificial intelligence (AI) and semantic computing can go a long way in solving many such problems. In response to the Covid-19 pandemic, banks need to make its operations 100 percent digital as early as possible. They are required to adjust the way they operate according to the changing demands of the situation. By making judicious use of technologies, the financial institutions can realize transformative cost reduction across business lines. The existing artificial intelligence technologies should be leveraged to their full potential to address the aforementioned challenges. To the best of our knowledge, one side of the coin, i.e., the statistical approaches is doing well while a little focus is being laid down on the benefits that could be reaped from the other side of the coin, i.e., the symbolic approaches to artificial intelligence. AI can support all the aspects of banking but the problem of data silos remains a challenge. What banks call “Single View of Customer” or 360° view of the customer means a unified space with all the customer's information regarding the relation of this customer with this bank including his product holdings, any complaints and responses, and any past interactions through any channel. This will help the banks better sell to their customers. Semantic computing by virtue

of semantic intelligence technologies (SITs) has the required potential to integrate the existing data silos and can prove as powerful lever for improving efficiency. The choice is ours; either continue to be terrified by the impacts that Covid-19 has brought in, or cash the same as an opportunity to leverage technology to combat the situation. Especially for financial institutions like the banking sector, artificial intelligence could offer better data integration through semantic knowledge representation approaches.

Many reports containing plenty of data stating the effects of this pandemic in the same context are available in the public domain. The investigation of the impact of Covid-19 from a large amount of distributed and unstructured data is very vital so that we can use those results to prevent the downfall of the economy and in turn minimize the pandemic effect. Various websites provide this Covid-19 data: <https://www.indiabudget.gov.in/economicsurvey/>, <https://dea.gov.in/>, <https://finmin.nic.in/>, <https://www.who.int/emergencies/diseases/>, and <https://www.mygov.in/covid-19/>. But these websites offer a static representation of Covid data which is largely unstructured (text, audio, video, image, newspaper, blogs, etc.) creating a major problem for the users to analyze, query, and visualize the data. The data integration task gets highly simplified by the incorporation of knowledge organization systems (taxonomy, vocabulary, ontology) as background knowledge. Storing the knowledge using the semantic data models enhances the inference power also. Various attempts have already been made to store the Covid-19 information even semantically in a machine-understandable manner using symbolic AI approaches, resulting in many different ontologies about this domain. But none of these ontologies contain information about the performance of the Indian banking sector during Covid-19.

This paper offers the Covid-19 Impact on Banking Ontology (Covid19-IBO) that provides semantic information about the impact of Covid-19 on the banking sector of India. Covid19-IBO can be utilized to annotate the related data to provide meaningful insights. The Covid-19 ontologies already available have some overlapping information that causes redundancy. Unified integration of these ontologies is required to operate upon them unambiguously. It becomes reasonable to develop a matching approach to link all these ontologies semantically. We, therefore, provide a schema matching approach to map the Covid-19 ontologies. The major contributions of the paper are listed below:

- Development of Covid19-IBO, an ontology highlighting the impact of Covid19 on the Indian banking sector
- Evaluation of the Covid19-IBO ontology using quantitative and qualitative approaches to determine its efficacy
- Proposing a schema matching approach for Covid-19 ontologies and discussing its performance

This study is again essential because this study will be used as a torchbearer if any of the pandemic like Covid-19 outbreaks in the future. The rest of the paper is presented in five sections. Section 2 focuses on the background information of the proposed work and lists down some previous contributions in this direction. Section 3 shows the development of the Covid19-IBO. Section 4 presents the evaluation

of Covid19-IBO. Section 5 poses a schema matching approach for Covid-19 ontologies and discusses the results and the last section concludes the paper.

Background Information

The Covid-19 pandemic has adversely impacted the Indian economy. To control the flow of the virus, the Government of India has announced a nationwide lockdown and various policies to help the people. All this enormous data about the pandemic is available on the internet to analyze the impact of the Covid-19 outbreak, the available data are mainly in text format, audios, videos, images, newspapers, blogs, etc. Moreover, these available data are highly unstructured and distributed in their own data silos, which creates a major problem for the users to analyze, query, and visualize the data. Dev and Sengupta [2] have examined the condition of India's economy before the Covid-19 pandemic as well as all the policies that have been made so far. The authors Rakshit and Basistha [3] have published an article about the economy of India by considering the Covid-19 outbreak as a man-made disaster that means human tragedy. They have discussed three vital research questions: (a) the effect of Covid-19 outbreak on the economy of India as well as detailed analysis of those sectors that are badly impacted by Covid-19 (b) the effect of Covid-19 outbreak on the relationship between China and India (c) the performance of the Indian health system during this pandemic. Kanitkar [4] has claimed that the loss of India's economy during the Covid-19 pandemic is about 10–30% of its GDP. This loss has been calculated via a linear I/O model. The author has also explained the emission of CO₂ from the power sector and the supply–demand of electricity. Demirguc-Kunt et al. [5] have examined the performance of the banking sector during the Covid-19 outbreak as well as consider the bank stock prices all over the world. They have used the global database to determine the role of the financial policy on the performance of bank stocks. The World Health Organization (WHO) provides a multilingual Covid-19 database that updates regularly and comprises all the information about Covid-19 [6, 7]. Kousha and Thelwall [8] have offered access to the scholarly databases to the people so that they can identify and discover the vital studies from the Covid-19 related publications such as Facebook, news, tweets, databases, and many other places.

Ontology as a Semantic Model: To respond effectively to emergencies such as public health, we need to share information across various disciplines and IT systems [9]. The digital technologies for analyzing the data will be possible to work effectively, if the underlying data model is itself robust. This is the place where ontologies offer excellent services and overcome the problem of interoperability. Ontology encodes knowledge in a machine-understandable manner in three ways: (a) listing the concept of a domain; (b) explaining the meaning (semantic) of these concepts by attaching appropriate properties and relations; and (c) facilitating annotating the data with a consistent, powerful, and reliable vocabulary. Let us take a simple concept namely *Allahabad Branch bank* and build its knowledge. First, we need to collect more concepts about it that can be collected by the following statement:

- 1. Allahabad Branch bank is in Faizabad district
- 2. Faizabad district is in Uttar Pradesh (UP)
- 3. Faizabad district is same as Ayodhya district
- 4. Ram’s Place is in the Ayodhya district
- 5. Ayodhya district is near the Saryu River

These facts are represented by Fig. 1 (knowledge graph depicts all these statements). This graph of domain knowledge represents the statements in the form of $A \rightarrow B$ where A and B are the concepts and the arrow \rightarrow explains the relationship between A and B. For example, the statement *Allahabad Branch bank* $\xrightarrow{\text{is_in}}$ *Faizabaddistrict* explains that is_in relationship exists between Allahabad branch bank and Faizabad district. A knowledge graph is a very powerful way of representing the concepts of a domain and simulates human knowledge. But there are certain problems like the heterogeneous sources of knowledge, inconsistencies in domain concepts across datasets, and data integration. All these problems can be efficiently handled if the concepts in knowledge graph adhere to some ontology. By mapping the knowledge graph with standard vocabularies (ontologies), the machine can do reasoning as human beings do. The ontologies formally represent and precisely define concepts and their relationships. They are put on the Linked Open Data Cloud for interoperability and reusability. They improve the semantic content of the data and link datasets at a schema level.

Figure 1 depicts the use of vocabularies to structure data in a knowledge graph and to add meaning to it. All the concepts of the knowledge graph are mapped to some class in the ontology and are called an instance of that class. For example, Allahabad Branch bank is an instance of a class Bank which is defined in Covid19-IBO ontology. Defining a class allows us to add more meaning to the concepts by assigning relations that connect a class to another class. Ontology

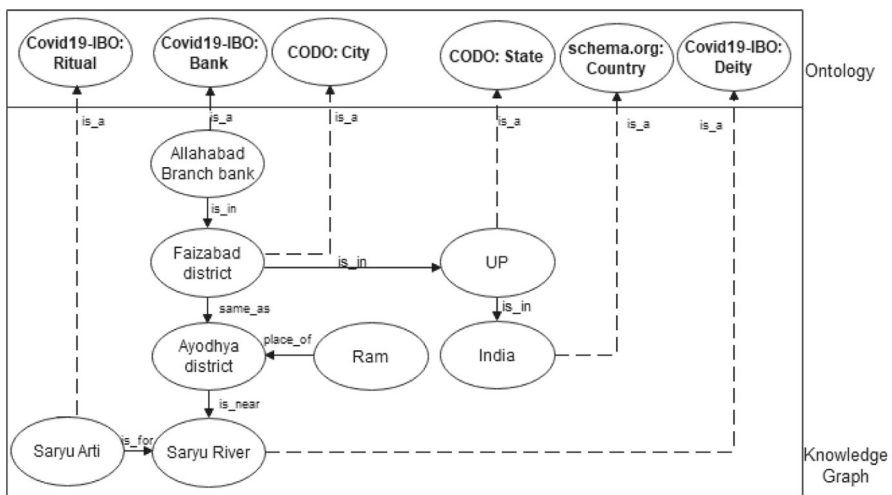


Fig. 1 Mapping of knowledge graph with an ontology

also allows attaching semantic with the relations. For example, we can make a relation as transitive, reflexive, complementary, disjoint, and many more. All the relationship that imposed on the classes must be automatically imposed on the instances of that classes. These relations provide additional semantic to the ontology by which machines deduced more knowledge similar to human beings.

The latest standard for representing ontology is OWL which is defined by W3C. OWL is designed to encode rich knowledge about the entities and their relationships in a machine-understandable manner. OWL is based on the Description Logic (DL) which is a decidable subset (fragment) of First-Order Logic (FOL) and it has a model-theoretic semantics. Along with the databases, various ontologies also have been developed to extract hidden and semantic information. Many bioinformaticians and biologists have argued that if ontologies are designed in a coordinated manner, then they will become an effective tool for data sharing [9]. The Open Biomedical Ontologies (OBO) Foundry is the most successful approach for the development of the coordinated ontology. OBO Foundry has the following ontology design principles:

1. Ontologies should have a common space for identifiers
2. Ontologies' syntax should be well-specified
3. Ontologies should be reuse and available publicly
4. Ontologies should be developed in a collaborative and modular manner
5. Ontologies should use unique relations between their entities and should have a specified scope
6. Ontologies should have a common top-level architecture

The OBO foundry principles offer a 'hub and spokes model' for the ontology development and its extension (new versions), where a 'hub or core' ontology provides a basis construct or entities for the extension ontologies.

Existing Covid-19 Ontologies: There are a total of six ontologies available about Covid-19 information. These all ontologies are listed below:

1. **Infectious Disease Ontology (called IDO):** IDO is an interoperable ontology that contains the domain information about an infectious disease where entities are related to the clinical and biomedical aspects of the disease [10]. IDO follows the ontology design principle of OBO Foundry and provides a 'hub and spokes model' for the development of the other ontologies. The backbone of the IDO ontology is entity 'disease' which is extracted from the OGMS ontology [9]. IDO has well-tested and offers a reliable starting point for the other ontologies. Therefore, various ontologies reuse the entities of IDO as needed. The extension ontologies of IDO concerning Covid-19 data are listed below and all these extension ontologies follow the principle of OBO foundry. The latest version of IDO is released in August 2020
2. **Virus Infectious Disease Ontology (VIDO):** VIDO extends the IDO ontology to add the virus-specific entities and provides concrete information about the domain of virus disease. Particularly, VIDO inherits the entities from the IDO by

adding the term 'virus' to create a subclass and the logical and textual information about the classes are adjust accordingly. VIDO also uses the entities from the OBO Foundry ontologies like VIDO imports SARS-CoV-2 proteins term from the Protein Ontology. The latest version of VIDO is released in August 2020

3. **Coronavirus Infectious Disease Ontology (CIDO):** IDO offers a good foundation for the development of pandemic-related ontologies such as Covid-19. CIDO extends the IDO to develop coronavirus-specific ontology which makes it a more generalized ontology as compared to VIDO. CIDO familiarizes eight entities that basically belong to the coronavirus domain and it also imports entities from the other ontologies similar to the VIDO. The latest version of CIDO is released in May 2021
4. **Covid-19 Infectious Disease Ontology (IDO-Covid-19):** IDO-Covid-19 is the most specific version of CIDO which covers the information related to Covid-19 and its cause SARS-CoV-2. IDO-Covid-19 follows the design principle of OBO Foundry and extends the CIDO in the same manner as CIDO extends VIDO and VIDO extends IDO. IDO-Covid-19 also extracts the terms from other ontologies like the term SARS-CoV-2 is imported from the NCBITaxon [10]. The latest version of VIDO is released in August 2020

The IDO ontology provides a repository of the entities that are used in infectious disease research and offers a useful foundation for the other ontologies. The relationships among these ontologies (IDO, VIDO, CIDO, and IDO-Covid-19) are depicted in Fig. 2. Line (—) shows the entities of the corresponding ontology; for example, IDO has four entities: infection, infectious disorder, infectious disease, and infectious disease course. All four ontologies (IDO, VIDO, CIDO, and IDO-Covid-19) are connected by an arrow (⇒), where the lower ontology extends and imports the needed entities (not all entities and axioms) from the other ontologies higher up in the hierarchy. By default, the arrow (→) shows is_a relationship unless no relationship is stated.

5. Dutta and DeBellis [11] have published the ontology as a data model namely **COViD-19 Ontology for the case and patient information (called CODO)** on the web as a knowledge graph that provides the information about the Covid-19 pandemic. The primary focus of the CODO ontology is to describe the Covid-19

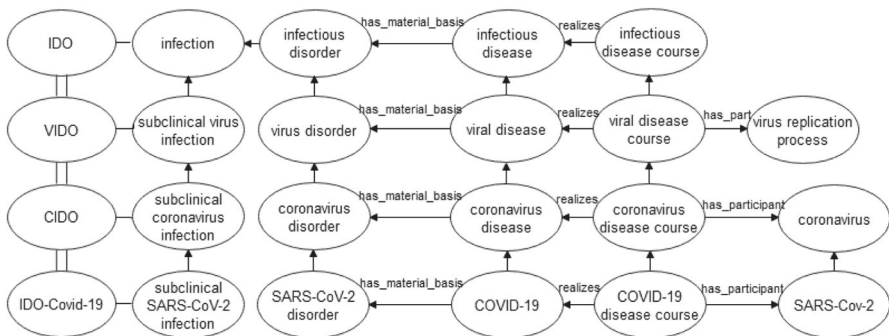


Fig. 2 Relationships among IDO, VIDO, CIDO, and IDO-Covid-19

cases and Covid-19 patient data. The format of CODO ontology is based on the OWL ontology language

CODO is designed by YAMO methodology which contains ten steps [11] namely (a) Domain Specification: defines the domain or scope of the ontology, (b) Domain Footprint: defines intention or aim of the ontology, (c) Knowledge Acquisition: collection of information from different sources, (d) Knowledge Formulation: analyze the collected knowledge to find out concepts or elements according to the scope of the ontology, (e) Knowledge Production: finds the relationships between the extracted concepts or entities, (f) Term Standardization: standardizing the term by selecting a more appropriate term from other corresponding available terms, (g) Knowledge Ordering: organize the term within the array, (h) Knowledge Modeling: model the various facets like, classes, their relationship, etc., (i) Knowledge Formalization formalize the term using formal logic languages, (j) Evaluation: measure the quality and standard of the ontology to know that does ontology fulfill the scope for which it is built. CODO ontology is developed using protégé tools and supports FAIR principles, where F = Findable, A = Accessible, I = Interoperable, R = Reusable.

6. **COVID-19 surveillance Ontology (COVID19):** This is an application ontology for Covid-19 and can be used in various domains. The goal of this ontology is to offer Covid-19 cases and their related respiratory information by accessing data from different medical records systems. However, this ontology is designed as a taxonomy having 32 classes only [12]. The ten foundational concepts of COVID19 ontology are: Definite COVID-19, COVID-19 confirmed by a lab test, SARS-CoV-2 detected, Probable Covid-19, Clinical codes, Possible COVID-19, Suspected COVID-19, Under investigation, Exposure, COVID-19 excluded. The COVID19 ontology is developed by the protégé tool whereas its format is based on OWL language.

The methodology that was used for the development of COVID19 ontology has three steps: (a) Ontology layer: defines relevant concepts according to the scope of the ontology, (b) Coding layer: all defined concepts are encoded in a hierarchical manner, (c) Data extract validation: systematically validate the ontology to check that ontology is well meeting the requirements of the domain [13]. The most popular domain-specific use cases of this ontology are public health, clinical research and informatics, and virology.

These available ontologies offer semantic information about the Covid-19. None of these ontologies contain information about the performance of the Indian banking sector during Covid-19 whereas the banking sector of any country plays a vital role in the economy of that country. Therefore, investigation of the impact of Covid-19 from a large amount of distributed and unstructured data is very vital so that we can use those results to prevent the downfall of the economy and in turn minimize the pandemic effect. This study is again essential because this study will be used as a torchbearer in the future if any of the pandemic impacts like Covid-19. Hence, our proposed Covid19-IBO ontology is unique in its scope. We have imported some concepts about the Covid-19 disease from the CODO and IDO-Covid-19 ontologies as these ontologies have enough information about the Covid-19 data.

Development of Covid19-IBO Ontology

Ontology is a knowledge representation model that encodes knowledge in the form of classes, relationships, properties, instances, and axioms [14]. Ontology has a reasoning capacity that infers hidden knowledge as a human does. Therefore, ontology is being utilized in various applications to structure real-life data [15]. The proposed Covid19 Impact on Banking Ontology (Covid19-IBO) contains a structured knowledge about the Indian banking sector. Using this information, machines become more capable of inferring much practical and relevant knowledge of the past, present, and future situation, which will help people by strengthening their decision-making process strategically and tremendously. As of now, various methodologies for the development of ontology are available [16]. The four most famous ontology development methodologies namely TOVE, Enterprise Model Approach, METHONTOLOGY, and KBSI IDEF5 are listed below:

1. **TOVE (Toronto Virtual Enterprise):** This methodology is proposed based on experience acquired from the development of the TOVE project [17]. It consists of six steps namely (a) Motivation Scenarios: frame set of problems that motivate us to develop the ontology, (b) Informal Competency Questions: according to the motivation scenarios, frame the competency questions that must be answered by ontology, (c) Terminology Specification: entities or concepts like, classes, properties, relations, and instances must be specified, (d) Formal Competency Questions: formalize the requirement of an ontology on the basis of defined terminology, (e) Axiom Specification: axioms must be express the competency questions, (f) Completeness Theorems: defines the completeness of the ontology by verifying the competency questions
2. **Enterprise model approach:** This methodology is proposed based on experience acquired from the development of enterprise ontology [18]. It consists of four steps namely (a) Identify Purpose: define the level of formality that ontology should be described, (b) Identify Scope: define the list of specifications that contains complete information which should be cover by ontology. This can be done by framing competency questions, motivations scenarios, and brainstorming, (c) Formalization: encode all the information like concepts, properties, relations, and instances, (d) Formal Evaluation: after the development of the ontology, need to evaluate it to know the quality of the ontology. This can be achieved by checking the ontology against the scope and purpose
3. **Methontology:** This methodology consists of seven steps namely (a) Specification: identify the purpose of the ontology including scope, motivation scenarios, competency questions, intended users, formality level, and so on. The output of this step is a specification document (b) Knowledge Acquisition: this step occurs parallel with step1. The expert of the domain and knowledge sources are used to analyze the text, (c) Conceptualisation: all the concepts like classes, properties should be identified, (d) Integration: for reusability purpose, definitions from other available ontologies should be incorporated, (e) Implementation: ontology

- should be formally encoded in any language, (f) Evaluation: check incompleteness, inconsistency, and redundancies as per guidelines of [19], (g) Documentation: make a document that describes all the activities of the developed ontology
4. **KBSI IDEF5:** This methodology has a general procedure that helps in the creation, modification, and maintenance of the ontologies. It has five steps namely (a) Organising and Scoping: identify scope, purpose, objective, requirement, and context of the ontologies, (b) Data Collection: data are collected by different sources using KA techniques like an expert interview, (c) Data Analysis: all the extracted or collected concepts are analyzed to know the entities according to the scope of the ontology, (d) Initial Ontology Development: preliminary ontology is developed based on the identified concepts and their relationship, (e) Ontology Refinement and Validation: the concepts of the preliminary ontology is iteratively refined and validate to fulfill them in the scope of the ontology.

Figure 3 summarizes the steps of all four ontology development methodologies described above. TOVE and Enterprise Model Approach are stage-based models whereas METHONTOLOGY and KBSI IDEF5 are evolving prototype models. Both models have strong and weak points. A stage-based model is good if the purpose and requirements for building the ontology are clear at the outset whereas evolving prototype model is applicable when the purpose for building ontology is not clear. Ontology development is a creative process and no two ontologies developed by a different person would be the same [20]. It is quite clear that developing ontologies is a matter of craft skill rather than an understood engineering process. Hence, there is no single correct ontology design methodology that can be applied to all the domain ontologies. [16].

After studying above mentioned methodologies, we select five phases for the development of Covid19-IBO ontology. These phases are Scope Determination, Extraction of the Concept, Organization of the Concept, Encoding, and Evaluation.

(i) *Scope Determination:* In scope determination, we fix the scope and boundary of the Covid19-IBO ontology. For this purpose, we draw some motivation scenarios that provide a better understanding of the scope and objectives of the Covid19-IBO and helps to build the correct ontology according to the user’s need. Some motivation scenarios in the form of problem statements are listed below:

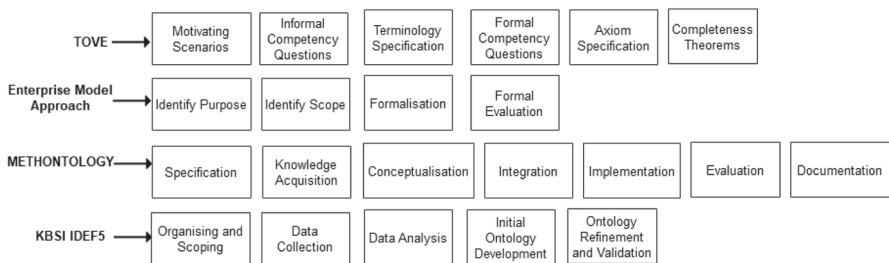


Fig. 3 The most famous ontology development methodologies and their steps

Scenario 1. What precautionary measures may be taken to prevent loss to the banking sector by the Covid-19 pandemic?

To prevent the loss, there should be a review of the banking portfolio, i.e., a new Assets Quality Review (AQR) should be created which may contain the best creative way under which all banks evergreen loans. The central bank must be able to detect the probable defaults in the early phase; this will provide strength to the former to prevent loss and its impact the least on the banking sector. By estimating the probable loss, the central bank may set a toolkit to focus on remedial procedures also. The RBI must make bank auditors responsible and impose some penalties on them if they are unable to perform their task (whether evergreening is observed). Recapitalization exercise should be made mandatory immediately after AQR. Debt recovery tribunals minimize the evergreening of lending.

Scenario 2. Which type of challenging conditions may arise for the banking industry if they are seen playing out in upcoming years? How will they impact the current business environment? What should be our business plan for eventualities like this?

In the current challenging environment high credit risk, loan repayment, liquidity infusion, digital optimization, contactless customer service and time instability due to mutated Covid-19 forms are the major problems. To overcome these problems, the banking sector must set their priority to operate with minimum contact by enabling their delivery channels to deliver their products online, i.e., they must put their efforts towards digitization and modernization. Banks must update their policies regarding loan forbearance and focus on a tailored approach to get the repayment of loans disbursed. To minimize the loan defaults, banks may review the creditability of industries as well as individuals, also, banks must opt for forecasting models to early assessment of credit risk arouse over a period. Flexibility in the work model is the need of the hour in which technical up-gradation, AI (Artificial Intelligence), Public Cloud, etc. play a vital role that will make a digital base for future banking needs by cost optimization.

Scenario 3. In the current challenging environment:

- (a) What are the capabilities in which we should invest?
- (b) Before investing in capabilities at this challenging time, the banking sector must review and manage the portfolio of current risks by differentiating it in geographical, sectorial, and segmental domains with the help of technology. The traditional products of the banking sector must be customized, as per Covid-19 particular needs of customers. Apart from it, some innovative products like ecosystem financing must be deployed. This sector must be ready for potential mergers and acquisitions to support quick capability building.

What digital capabilities do we need to build to support social distancing needs and enable innovative models?

The banking sector must be friendly with contactless banking options viz advanced ATMs (ATMs enabled with Audiovisual response), doorstep banking and robust digital channels, etc. For the collections, enhance efficiency through analytics-led models as well as ecosystem partnerships. To support social distancing, banks must work to provide strength to IT infrastructure using cloud-

based latest technologies to provide quick support and service, it also enables remote operations of banks as well as innovative latest need-based product offerings

Based on these motivation scenarios, we make a list of 27 competency questions (CQs) that explain the requirements of the ontology. These questions also help in the evaluation of ontology. Some selected competency questions are mentioned below:

- CQ1. Which bank and sector suffered most from this pandemic?
- CQ2. How many bank employees have lost their job during a pandemic?
- CQ3. How many new accounts, loans, credit cards are being issued from starting of the pandemic?
- CQ4. What are the different types of banks operating in India?
- CQ5. How to increase the profit of the Indian Banking sector?
- CQ5. How to increase the profit of the Indian Banking sector?
- CQ5. How to increase the profit of the Indian Banking sector?
- CQ7. What are the lending schemes started by the Indian government during the Covid-19 pandemic?
- CQ8. What type of assistance provided by the Indian government to the farmers during the Covid-19 pandemic?
- CQ9. What are the different types of accounts operated by Indian banks during the Covid-19 pandemic?
- CQ10. The name of the Covid-19 vaccine approved by the Indian government for the bank employees?
- CQ11. What is the different type of NPA in the banking sector?

The formulated ontology should be capable to provide the answer to these questions by encoding the appropriate information in the form of entities or concepts (classes and properties: data properties, object properties, annotation properties). The detailed description is mentioned in Sect. 4 (Covid19-IBO evaluation).

(ii)*Concept Extraction*: In this phase, we extract the concepts or entities for the Covid19-IBO ontology. It offers a lightweight scope of representations that reflects various important and required concepts about the impact of Covid-19 on the Indian Banking sector and relations among the concepts. For the development of the Covid19-IBO ontology, we extract the entities from different sources (as mentioned below) according to the need and scope of the domain.

- (a) Research articles related to Indian economy and Covid-19 pandemic, available on conferences, journals, and book chapters that are listed in the reference list
- (b) Existing ontology repositories/portals like Bio portal, EMBL-EBI, Agro portal
- (c) Articles on websites (Like Wikipedia, blogs, different sites, and so on)

- (d) Covid-19 and Indian economy related Databases (<https://covid19.who.int/>, <https://www.mygov.in/covid-19>, <https://dea.gov.in/>, <https://www.rbi.org.in/>, <https://pib.gov.in/AllRelease.aspx?MenuId=3>, etc.).
- (e) Covid-19 ontologies (CODO, CIDO, VIDO, IDO, COVID-19, and, etc.)
- (f) Conducting interviews with an expert person like a doctor, Auditors
- (g) Various Covid-19 Reports, GoI press releases about Covid-19 impact on Indian economy, monthly economic report of India

These sources produce a list of potentially relevant entities (like classes, properties, and instances). All the extracted entities from the sources are stored in the excel sheet (available on GitHub: <https://bit.ly/3x6oECS>) and appropriate entities are selected after examining all their possible synonyms. Some of the most significant entities (relevant to Covid-19 and the Indian banking sector) are mentioned below:

Entities: Covid-19

Disease, Viral disease, Infectious disease, Covid-19, Covid care centre, Covid-19 Diagnosis, Covid Hospital, Doctors, Nurses, Covid-19 symptoms, Covid-19 post disease, Black fungus, White fungus, Patient Details, No of Patient, No of Patient dead, No of Patient recover, Covid vaccine, Covishield, AstraZeneca, Sputnik.

Entities: Indian Banking

Account, Bank, Risk Loss Factors, Statistics, Loan, Saving, Deposit, Current, Commercial Bank, Banking Relationship, Cooperative Bank, Regional Rural Bank, Credit, NPA, Cluster, Zone, ETB, NTB, Doubtful, Standard, SubStandard, ExchangeRate, Market, Operational, Political, Market, Lending, AtmNirbhor, PSL.

(iii) *Concept Organization*: This phase aims to organize the extracted concepts, properties, and instances in a hierarchal manner (parent–child relationship). We follow three steps (as mentioned below) for the organization of the extracted concepts:

- (a) *Analysis*: The extracted concepts or entities from the different sources are analyzed based on the characteristic of each entity and then breaking them into their element-level entities. All the entities are grouped according to their similarity. For example: analyzing the entities CommercialBank (operating at large-scale with profit as the base), CooperativeBank (operating at small-scale with service as the base), and RegionalRuralBanks (operating at regional level by providing basic banking and financial services) according to their characteristic reveal that all these entities have a common point and can be made as subclasses of the class Bank (a financial institution)
- (b) *Knowledge Synthesis*: All the elemental entities are organized hierarchically by defining the relationship between the extracted entities. The below-mentioned example shows the subclass relationship (the indentation depicts the subclass relationship that generates hierarchy) between the classes. We also specify the axioms that imposed restrictions on the definition and interpretation of the entities. Axioms are defined according to the competency questions and motivation scenarios

Lending

- AtmNirbhor
- Corporate_Lending
- Govt_Schemes
- PSL
- Retail

(c) **Reuse and Standardization:** As various ontologies are available for the Covid-19 disease so it is reasonable to use the entities of these ontologies when applicable. For the Covid19-IBO ontology, we have imported the concept from Schema.org (use for modeling common concepts like location), OBO (use for modeling clinical findings and symptoms), and FOAF (use for modeling person class and related properties) vocabularies. We have also reused some concepts from CODO and IDO-Covid-19 ontologies which are depicted in Fig. 4. The reused entities of CODO and IDO-Covid-19 ontologies are highlighted by green and blue color, respectively, in Fig. 4. These entities are encoded in the Covid19-IBO using IRI which is internationalized resource identifier. Reusing the concept via IRI help to avoid multiple interpretations of entities in the ontology

(iv) *Encoding of Ontology:* In this phase, we encode the ontology either using tools (offline and online tools) or crafting it by hand. Encoding of the ontology involves structuring and modeling the domain knowledge produced in the preceding

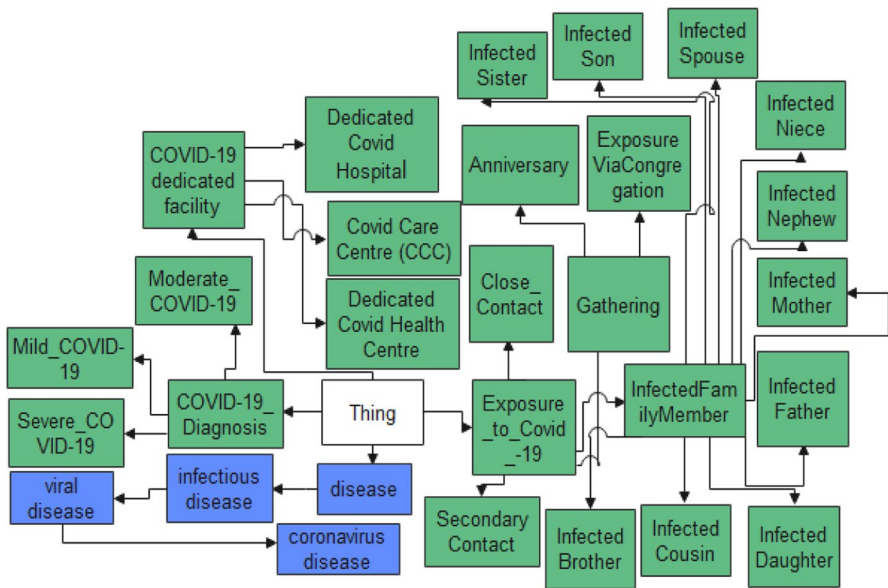


Fig. 4 Reused entities in Covid19-IBO from CODO and IDO-Covid-19 ontologies

phases. The Covid19-IBO ontology is encoded using web ontology language-Description Logic (OWL-DL) in protégé 5.5.0 ontology editor that is developed at Stanford University [21]. Protégé is open source [22] and has a very interactive user interface. Users can encode the ontology in the protégé tool without knowing the ontology languages. Protégé tool allows exporting the ontology in data formats like RDF/XML, N-Triples, OWL/XML, OBO, JSON-LD, Turtle, Manchester OWL, and LaTeX. These data formats are compatible with each other. Protégé tool supports the pellet reasoner that we use to check the consistency of the developed ontology.

The development of an ontology is not the same as defining the classes and properties in object-oriented programming. The ontology designer takes design decisions on the basis of the structural properties of a class whereas, in object-oriented programming, these decisions are made on the basis of operational properties of a class. Consequently, in ontology and object-oriented programming, the class structure and relations among the classes are different in a similar domain.

Defined Classes: The classes of the ontology define the concept in a domain, like: a class of Bank_Account represents all the type of bank account that a customer can hold. A class can have one or more subclasses that represent specific concepts in a domain as compared to the superclass (which is the most generalized class), like the class Bank_Account can have four subclasses namely Current, Deposit, Loan, and Saving. There are three possible approaches to build the ontology hierarchy: (a) Top-down approach: to define most general classes then divide them into more specific classes, (b) Bottom-up approach: to define most specific classes then grouping them into more general classes, (c) Combination approach: it is a combination of both top-down and bottom-up approaches. First, define more silent classes then make them general and specific classes as per requirement. We follow the

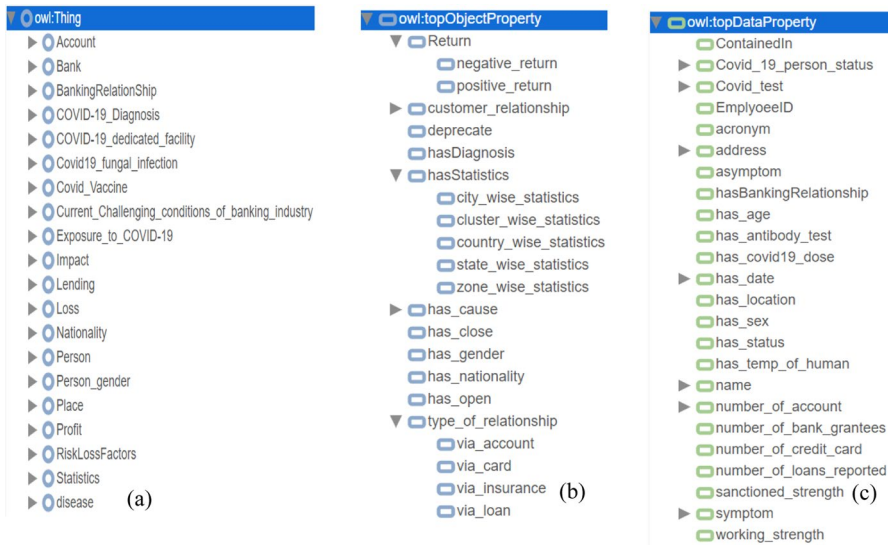


Fig. 5 Covid19-IBO ontology: (a) Classes (b) Object properties (c) Data properties

combination approach to develop the Covid19-IBO ontology. In the class hierarchy, Classes can have restrictions on themselves like Equivalent to, SubClass of, Member, Target for key, Disjoint with, Disjoint Union of, etc. Figure 5 (a) depicts the class hierarchy of Covid19-IBO ontology.

Defined Properties: We must define the internal structure of the classes because the classes alone do not offer adequate information to answer the competency questions defined in phase 1. We already defined the classes from the collected entities which are mentioned in phase 2. Most of the remaining entities become properties of these classes. There are two types of properties namely object properties (relate pairs of individuals) and data properties (relate individuals with literal data like string, datetimes, etc.). All the properties must have domain and range that imposed restrictions on the individuals as well classes. Figure 5 (b), and (c) shows the classification of the object, and data properties, respectively. All object and data properties have domain and range for example, Axis Bank (Domain) has_open (Object Property) Saving Account (Range); PNB Bank (Domain) has_status (Data Property) NPA (Range).

The design goal of Covid19-IBO ontology is to keep it small, easy to grasp, and reuse the existing concepts of available ontologies as well as expressive enough to cover the domain. The Covid19-IBO ontology consists of 160 classes and 77 properties (34 object properties and 43 data properties). The latest version of Covid19-IBO ontology is Covid19-IBO1.5 which is publicly available on the Bio portal, link: <https://biportal.bioontology.org/ontologies/COVID19-IBO>.

v. *Ontology Evaluation:* We evaluate the developed ontology to know its quality, completeness, and correctness according to the scope. The detailed evaluation of Covid19-IB is presented in the next section of this paper.

Covid19-IBO Evaluation

There are various ontology evaluation approaches [23–25] available in the literature. These approaches are divided into four categories as mentioned below:

- **Technology-based approach:** It investigates the structural characteristics (syntax, formal semantic, and consistency) of an ontology. Therefore, it focuses on the correctness and usability of the ontology rather than the quality of the content and applicability of the ontology
- **Quality-based approach:** It measures the quality of the ontology by detecting formal and semantic redundancies, inconsistencies, and missing definitions of the stored entities. The quality measurement of the ontology depends on the set of predefined metrics and sometimes these metric's value needs human efforts
- **Data-driven approach:** It focuses on the usability of the ontology. This approach compares the ontology with the related data sources to check that ontology has enough entities or not. This approach aims to examine the coverage of ontology in a particular domain. This approach does not check the clarity and correctness of the ontology

- **Application-based approach:** This approach aims to evaluate the developed ontology using it in a specific domain. The result completely depends on the domain where the ontology is used, therefore, this approach is not considered for general evaluation

However, there are various ways to evaluate the ontology but there is no ideal solution found yet as each approach has its limitations. In this paper, we follow a combined approach that given by authors Mehla and Jain [26] to evaluate the ontology. They have categorized the ontology evaluation approaches into two groups namely Verification (that means Building an ontology correctly) and Validation (that means building the correct ontology).

Verification of Covid19-IBO ontology: For verification, we use Quantitative and Qualitative approaches [27]. The quantitative approach measures the quantity of the attributes in the ontology whereas the qualitative approach checks the quality of the ontology. We use metric-based and criteria-based approaches to evaluate the ontology quantitatively and qualitatively, respectively. The metric-based approach calculates the value of attributes that show the richness of the ontology for example, the value of depth and breadth reflects the information about the total number of levels and the maximum number of nodes at the level in the ontology, respectively. But this approach did not detect anomalies inside the ontology whereas the criteria-based approach checks the Consistency, Conciseness, Completeness, Correctness, and Clarity of the ontology by detecting anomalies inside the ontology.

(a) **Qualitative evaluation:** We evaluate the ontology qualitatively using an Ontology Pitfalls Scanner (OOPs) tool [28] which is a criteria-based approach. OOPs shows the pitfalls or error under the three categories namely minor (not very serious pitfall), important (not critical pitfall but it is important to correct it), and critical (need to remove this pitfall) pitfalls. Currently, OOPs contains 41 pitfalls (starting from P01 to P41) along with their descriptions. These pitfalls state different perspectives of an ontology, therefore, these pitfalls are classified into two groups [29]:

- **Classification by dimension:** Pitfalls are classified according to the dimension as defined by authors Gangemi et al. [30]. These dimensions are *Structural Dimension* (focus on the ontology syntax and semantics. It includes five aspects namely modeling decision, real-world modeling, No inference, Wrong inference, and ontology language), *Functional Dimension* (focus on the use of an ontology i.e., ontology conceptualization. It includes two aspects namely requirement completeness and application context), *Usability-Profiling Dimension* (focus on the communication context of an ontology. It includes two aspects namely Ontology understanding and Ontology Clarity).
- **Classification by Evaluation Criteria:** OOPs have three criteria to check the quality of the ontology. These criteria are *Consistency* (to check that ontology is developed as per specification and enables reasoner to infer knowledge), *Completeness* (to assure all the required knowledge is available inside the ontology), *Conciseness* (to ensure that no irrelevant information encoded inside the ontology that increases the size of ontology unnecessarily).

Table 2 Obtained pitfalls in ontologies

Ontologies	Pitfalls													
	Important								Minor					
	P10	P11	P24	P25	P30	P34	P38	P41	P04	P07	P08	P13	P20	P22
IDO	×	24	×	×	1	×	1	1	1	×	×	7	4	1
VIDO	×	24	×	×	1	×	1	1	1	×	×	7	4	1
CIDO	1	300	225	×	6993	×	1	1	1	×	7833	349	×	1
IDO-Covid-19	×	24	×	×	1	×	1	1	1	×	×	7	4	1
CODO	1	58	4	4	2	7	1	×	1	×	58	38	3	1
CoVID19	1	×	×	×	×	×	1	1	4	×	×	×	×	×
Covid19-IBO	1	1	×	×	2	×	×	1	6	1	20	31	×	1

Table 2 shows the pitfalls that are available in the existing ontologies namely IDO, VIDO, CIDO, IDO-Covid-19, CODO, and COVID19 along with Covid19-IBO ontology. The number (like 1, 3, 4,...) in Table 2 shows the total number of cases corresponding to the mentioned pitfall for example, IDO has 24 cases of pitfall P11. We have listed only those pitfalls in Table 2 for which atleast one case was discovered for atleast one of these ontologies. The sign × reflects that the mentioned pitfall was not reported against that ontology. In all these ontologies, there is no critical pitfall; only important and minor pitfall are presented.

These pitfalls hamper the quality (Usability-Profiling, Structural, Functional as per dimension aspect; Completeness and Consistency as per criteria aspect) of the ontology. The obtained pitfalls of the ontologies are P04 (shows that unconnected elements inside ontology are presented), P07 (shows that different concepts are merged in the same class), P08 (shows that annotation properties are missing), P10 (shows that disjoint axioms are missing), P11 (shows that domain and range of data and object properties are not defined), P13 (shows that inverse relationship is not defined), P20 (shows that annotation properties are misused), P22 (shows that ontology has different naming conventions), P24 (shows that element of ontology has recursive definitions), P25 (shows that relationship defined as inverse to itself), P30 (shows that equivalence class is not defined explicitly), P34 (shows untyped class), P38 (shows that no owl:Ontology tag is defined), and P41 (shows that no license is declared) [29]. The obtained pitfalls according to the groups (Dimension and Criteria) are mentioned below:

Dimension	Criteria
P07, P20, P22, P38, P41, P08, P38, P41 → Usability-Profiling	P04, P10, P11, P13 → Completeness
P11, P13, P24, P25, P30, P34 → Structural	P07, P24 → Consistency
P04, P10 → Functional Dimensional	

Table 3 Metrics value of the available ontologies with respect to Covid19-IBO

Metrics	Ontologies						
	IDO	VIDO	CIDO	IDO-Covid-19	CODO	COVID19	Covid19-IBO
Axioms	4103	4647	104,336	5018	2047	165	714
Logical axioms count	833	956	17,367	1032	917	32	352
Class count	362	429	7564	486	91	33	160
Object property count	43	43	409	43	73	0	34
Data property count	0	0	18	0	50	0	43
Attribute Richness	0.0	0.0	0.056	0.0	0.549	0.0	0.294
Inheritance Richness	1.237	1.258	1.547	1.242	1.010	0.969	0.98
Relation Richness	0.384	0.360	0.873	0.340	0.471	0.0	0.196
Absolute cardinality							
Root	1	1	1	1	1	1	1
Leaf	249	295	403	306	67	28	111
Sibling	362	429	1657	486	91	33	160
Depth:							
Absolute	2551	3100	1953	3588	327	93	545
Average	7.046	7.226	3.052	7.382	3.593	2.818	3.40
Maximal	10	11	38	13	6	3	7
Breadth:							
Absolute	362	429	7564	486	91	33	160
Average	3.175	3.177	3.67	2.685	3.64	5.5	3.244
Maximal	17	18	41	18	13	11	19
Total number of paths	362	429	7564	486	91	33	159

We have removed all the minor and important pitfalls from the Covid19-IBO ontology by enhancing it; however, other existing ontologies still have these pitfalls.

Quantitative Evaluation: We evaluate the ontology quantitatively using the OntoMetric tool [31], which is a metric-based approach. It divides the features into five metrics namely Schema, Instance, Base, Graphs, and Individual axioms. Table 3 indicates the important magnitude of some metrics of ontologies namely IDO, VIDO, CIDO, IDO-Covid-19, CODO, and COVID19 along with Covid19-IBO ontology. These magnitudes are calculated with the help of the OntoMetric tool that shows the richness of the ontology. The metrics description is mentioned below:

- The metrics namely Axioms (they state what is true in a domain), Logical axioms count (they affect the logical meaning of the axioms), Class count (shows a total number of classes in ontology), object property count (shows a total number of object properties that links individuals to individuals), and data property count (shows a total number of data properties that link individuals

to data values) are the *base metrics* that express the quantity of the ontology elements.

- The metrics namely attribute richness (shows the attributes defined for each class), inheritance richness (shows the distribution of information across different levels), relation richness (shows the diversity of types of relations) are the most significant *schema metrics* that reflect the design of the ontology.
- The remaining metrics related to the cardinality (shows graph-related specific elements), depth (shows cardinality of existing paths), and breadth (shows cardinality of levels) of the ontology. The absolute root cardinality (shows a number of the root node), absolute leaf cardinality (shows a number of the leaf nodes), absolute sibling cardinality (shows a number of sibling nodes), absolute depth (shows cardinality of each path), average depth (shows vertical modeling of hierarchy), maximal depth (shows the height of the graph), absolute breadth (shows cardinality of each level), average breadth (shows horizontal modeling of the hierarchy), maximal breadth (shows the breadth of the graph), and the total number of the path are the graph metrics which calculates the structure of the ontology.

Table 3 depicts that Covid19-IBO ontology has a good level of information. The available ontologies namely IDO, VIDO, CIDO, IDO-Covid-19, CODO, and COVID19 cannot be use in our proposed domain despite having enough information (or in some cases, these ontologies have more metric values as compared to Covid19-IBO). These ontologies contain the information related to Covid-19 whereas Covid19-IBO ontology describes the complete information about the impact of Covid-19 on the performance of the Indian banking sector as well as it also contains some information about the Covid-19 disease as needed.

Validation of Covid19-IBO Ontology

As we described earlier, validation means building the correct ontology. To validate the Covid19-IBO ontology, we use competency questions, available validators (W3C RDF validator, OWL validator, RDF Triple-Checker, and Vapour), and subjective testing.

(a) Competency Questions

The set of competency questions describe the requirement of the ontology as per the specified domain as well as check the completeness of the developed ontology. A set of 27 competency questions (some of them are mentioned in Sect. 3) that cover the domain of ‘Impact of Covid19 on Indian Banking sector’ are provided in natural language (English) and are translated into SPARQL queries. The SPARQL query engine is almost similar to OWL just like SQL is to relational DB. The underlying structure of OWL is graph-based rather than tables, therefore, SPARQL constructs graph patterns to infer the knowledge from the ontology. SPARQL integrates the data from multiple data sources, therefore, the SPARQL query starts with a list of namespaces found in IRI. Each query is run on the Covid19-IBO ontology with

Table 4 SPARQL queries and their results

Query Related to New Concepts												
1. Which type of challenging conditions may arise for the banking industry if they are seen playing out in upcoming years?												
prefix onto: <http://www.semanticweb.org/archana/ontologies/2021/5/untitled-ontology-6#> SELECT ?CD WHERE {?CD rdfs:subClassOf onto:Current_Challenging_conditions_of_banking_industry .}												
<table border="1"> <thead> <tr> <th colspan="2" style="text-align: center;">CD</th> </tr> </thead> <tbody> <tr><td>Liquidity_infusion</td></tr> <tr><td>Contactless_customer_service</td></tr> <tr><td>High_credit_risk</td></tr> <tr><td>Loan_repayment</td></tr> <tr><td>Digital_optimization</td></tr> <tr><td>time_instability</td></tr> </tbody> </table>		CD		Liquidity_infusion	Contactless_customer_service	High_credit_risk	Loan_repayment	Digital_optimization	time_instability			
CD												
Liquidity_infusion												
Contactless_customer_service												
High_credit_risk												
Loan_repayment												
Digital_optimization												
time_instability												
2. What precautionary measures may be taken to prevent loss to the banking sector by the Covid-19 pandemic?												
prefix onto: <http://www.semanticweb.org/archana/ontologies/2021/5/untitled-ontology-6#> SELECT distinct ?p ?o WHERE { {onto: New_Assets_Quality_Review ?p ?o . ?p rdf:type owl:AnnotationProperty.} union {onto: Detect_probable_defaults_in_early_phase ?p ?o . ?p rdf:type owl:AnnotationProperty.} union {onto: Bank_Audit ?p ?o . ?p rdf:type owl:AnnotationProperty.} }												
<p>To prevent the loss, there should be a review of the banking portfolio i.e., a new Assets Quality Review (AQR) should be created which may contain the best creative way under which all banks evergreen loans. Recapitalization exercise should be made mandatory immediately after AQR. The central bank must be able to detect the probable defaults in the early phase; this will provide strength to the former to prevent loss and its impact the least on the banking sector. By estimating the probable loss, central bank may set a toolkit to focus on remedial procedures also. The RBI must make bank auditors responsible and impose some penalties on them if they are unable to perform the tasks (whether evergreening is observed).</p>												
Query Related to Imported Concepts from CODO, Schema.org, and FOAF ontologies												
3. Suspected causes of catching COVID-19.												
prefix arch: <http://www.isibang.ac.in/ns/codo#> SELECT ?SC WHERE {?SC rdfs:subClassOf arch:Exposure_to_COVID-19 .}												
<table border="1"> <thead> <tr> <th colspan="2" style="text-align: center;">SC</th> </tr> </thead> <tbody> <tr><td>InfectedViaJob</td></tr> <tr><td>Gathering</td></tr> <tr><td>SecondaryContact</td></tr> <tr><td>ContactWithInfectedBankingStaff</td></tr> <tr><td>Close_contact</td></tr> <tr><td>InfectedNeighbor</td></tr> <tr><td>InfectedFamilyMember</td></tr> <tr><td>InfectedDomesticHelp</td></tr> <tr><td>InfectedCo-Customer</td></tr> </tbody> </table>		SC		InfectedViaJob	Gathering	SecondaryContact	ContactWithInfectedBankingStaff	Close_contact	InfectedNeighbor	InfectedFamilyMember	InfectedDomesticHelp	InfectedCo-Customer
SC												
InfectedViaJob												
Gathering												
SecondaryContact												
ContactWithInfectedBankingStaff												
Close_contact												
InfectedNeighbor												
InfectedFamilyMember												
InfectedDomesticHelp												
InfectedCo-Customer												
4. How do the customer's name and their address stored in the Indian banking system?												
prefix foaf: <http://xmlns.com/foaf/0.1/> prefix schema: <https://schema.org/> SELECT ?na ?add WHERE { { ?m rdfs:subPropertyOf foaf:name } union { ?s rdfs:subPropertyOf schema:address } }												
<table border="1"> <thead> <tr> <th>na</th> <th>add</th> </tr> </thead> <tbody> <tr><td>last_name</td><td>postalCode</td></tr> <tr><td>middle_name</td><td>streetAddress</td></tr> <tr><td>first_name</td><td>addressLocality</td></tr> </tbody> </table>		na	add	last_name	postalCode	middle_name	streetAddress	first_name	addressLocality			
na	add											
last_name	postalCode											
middle_name	streetAddress											
first_name	addressLocality											

the help of the protégé tool to test that all requirements have been fulfilled or not. All the missing entities are added into ontology as per failed queries that did not get the correct answers from the ontology. Thus, these questions help us to know whether the ontology is complete in itself or not. As mentioned in the literature, the motivation scenarios are one of the most powerful keys to know the boundary of the developed ontology. We have set three motivation scenarios as mentioned in the previous section. These motivation scenarios are also encoded inside the ontology (queries 1 and 2 of Table 4 are related to the motivation scenarios). The below-mentioned Table 4 shows few SPARQL queries and their results that are inferred from the Covid19-IBO ontology. Query 3 shows that the concepts are imported from the CODO ontology whereas query 4 explains that the concepts from schema.org and foaf are successfully encoded. We use prefix onto: for the Covid19-IBO ontology; prefix arch: for the CODO ontology; prefix schema: for schema.org; prefix foaf: for the FOAF ontology. We use the SPARQL query module of the protégé 5.5.0 tool to run the SPARQL query and Pellet reasoner to check the consistency of the developed Covid19-IBO ontology. All the SPARQL queries are successfully executed over Covid19-IBO ontology that proves that ontology is consistent.

(b) Validators

We validate the Covid19-IBO ontology with the existing validators namely W3C RDF validator [32], OWL validator [33], RDF Triple-Checker [34], and Vapour [35]. The W3C RDF validator and OWL validator ensure that ontology is syntactically correct concerning RDF and OWL syntax, respectively, whereas RDF Triple-Checker and Vapour are used to check the dereferenceability of namespaces in the document. The process to retrieve a representation of a resource identified by a URI is called dereferencing that URI [36]. The process of dereferencing a URI is vital on the web.

- W3C RDF validator is developed at HP-Labs in Bristol by Jeremy Carroll. It is the most popular and widely used tool for the validation of RDF documents. It validates the RDF document by tracking the RDF issues and shows a warning message when an error occurred. W3C RDF validator shows the number of tuples (subject–object–predicate) that are encoded in the ontology as well as its graphical representation. This validator aims to ensure that the document is syntactically valid.

Validating Covid19-IBO: The W3C RDF validator extracts all the 714 tuples of Covid19-IBO ontology successfully. No warning message was generated during validation of the Covid19-IBO ontology that proves that ontology is well built and syntactically correct. The information about the extracted tuples via RDF validator is available on the link: <https://bit.ly/3vMhzGC>.

- OWL validator is developed by the University of Manchester. OWL validator validates the ontologies against OWL2, OWL2DL, OWL2RL, OWL2EL, and OWL2QL profiles. This validator aims to ensure that all the concepts and properties are specified as per the W3C standard. OWL validator shows an error message with the detailed report when an input ontology does not support the selected profile.

RDF Triple-Checker

URI/URL:

Output: HTML JSON

Loaded 714 triples

Namespaces

Namespace	Looks Legit?
http://www.w3.org/1999/02/22-rdf-syntax-ns#	Matched common namespace. Prefix rdf
http://www.w3.org/2000/01/rdf-schema#	Matched common namespace. Prefix rdfs
http://www.w3.org/2002/07/owl#	Matched common namespace. Prefix owl

Terms

Count	Type	Namespace	Term	Looks Legit?
238	property	http://www.w3.org/1999/02/22-rdf-syntax-ns#	type	OK - Looks good.
42	property	http://www.w3.org/2000/01/rdf-schema#	subPropertyOf	OK - Looks good.
76	property	http://www.w3.org/2000/01/rdf-schema#	domain	OK - Looks good.
76	property	http://www.w3.org/2000/01/rdf-schema#	range	OK - Looks good.
158	property	http://www.w3.org/2000/01/rdf-schema#	subClassOf	OK - Looks good.
4	property	http://www.w3.org/2000/01/rdf-schema#	comment	OK - Looks good.
1	class	http://www.w3.org/2002/07/owl#	Ontology	OK - Looks good.
1	class	http://www.w3.org/2002/07/owl#	AnnotationProperty	OK - Looks good.
34	class	http://www.w3.org/2002/07/owl#	ObjectProperty	OK - Looks good.
43	class	http://www.w3.org/2002/07/owl#	DatatypeProperty	OK - Looks good.
159	class	http://www.w3.org/2002/07/owl#	Class	OK - Looks good.

Fig. 6 Result of Covid19-IBO ontology over RDF Triple-Checker Validator

Validating Covid19-IBO: The Covid19-IBO ontology has OWL2, OWL2DL, and OWL2RL profiles and these profiles are successfully validated via OWL validator.

- RDF Triple-Checker is used to check the dereferenceability of namespaces associated with classes and properties in a document. This tool also helps to find the errors (like typos) in RDF data. The RDF Triple-Checker aims to examine that the URI of the classes and properties are defined in the namespace and the namespace corresponds with a well-known prefix.

Validating Covid19-IBO: The Covid19-IBO ontology is loaded on RDF Triple-Checker via URI. Figure 6 shows that all the 714 triples are extracted successfully and all the classes and properties have a well-defined prefix.

- Vapour, a linked data validator is also used to check the dereferenceability of a specified URI. It uses the content negotiation and determines that the specified URI is either an information resource or a non-information resource. It has a spe-

cial feature that allows checking the dereferenceability of the terms and namespaces.

Validating Covid19-IBO: The Covid19-IBO ontology has successfully passed all the test cases of the Vapour validator tool.

(c) Subjective Testing for Covid19-IBO

To test the developed ontology, we have adopted subjective testing suggested by industry experts. Several important parameters have been considered for the rating of this ontology by the users on the scale of 1 to 10, where 1 stands for “not at all satisfying these parameters” and 10 for “best possible satisfaction”. The parameters incorporated are Adoption of properties as compared to existing ontologies, User friendly (Ease of adoption by a user), Future use (How ontology can be used in the future, e.g., during the vaccination process), Relevance of current time (Referring to the current situation of Covid-19 pandemic), Benefits of ontology, Impact on economy and Impact on society (How this ontology can be used by students, research scholars, professionals, Industry personnel, etc.).

Thirty-nine users participated in the subjective testing by filling the google form. Among the 39 users, 18 are female and 21 are male. The average age of the participants is 28.05 whereas the minimum and maximum age of participants is 21 and 52, respectively. The participated users belong to different occupations like faculty, engineers, students, and scholars, etc. Users have also provided additional comments on the developed ontology like very useful, incorporate all the facts of the Indian economy, ontology is very much required to analyze the impact of Covid-19 on the Indian banking sector and so on that shows accuracy and importance of Covid19-IBO ontology. The user’s results are shown in Fig. 7 based on the different parameters suggested by experts.

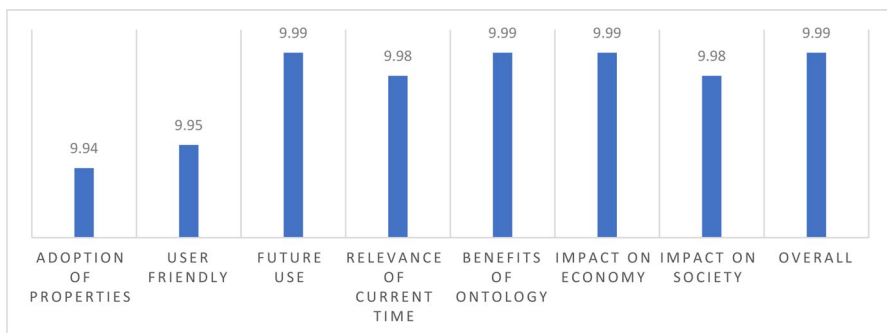


Fig. 7 Ranking of Covid19-IBO

A Schema Matching Approach for Covid-19 Ontologies (SMA-Covid19)

Various ontology matching systems are available to match the entities of the ontologies. These matching systems detect the relationship (R) that can be equivalence, less general, more general, and disjoint. The relationship between the entities is represented in the form of tuple $\langle e1, e2, R \rangle$ that is called correspondence. The performance of these matching systems is checked based on the various performance parameters namely precision, recall, F-measure, G-mean, and run time. Precision explains the correctness of the system using the formula (total correct obtained correspondences divided by total obtained correspondences), recall measures the completeness of the system using the formula (total correct obtained correspondences divided by reference alignment), F-measure explains the accuracy of the system by calculating the harmonic mean of precision and recall, G-mean is the geometric mean of precision and recall [37]. The running time is the execution time that system takes to match the ontologies. The Covid19-IBO ontology semantically offers complete and precise information about the impact of Covid-19 on the Indian banking sector [38]. Now, it is very important to map the existing Covid-19 related ontologies (CODO, VIDO, IDO-Covid-19, IDO, and CIDO) with Covid19-IBO to know the overlapping information as Covid19-IBO also contains some information about Covid-19 disease. To map the Covid-19 ontologies, we propose a schema matching approach to find out the equivalence relationship between the entities of T-Box (contains all the classes of the ontology).

The proposed approach undergoes three phases, namely preprocessing phase, actual matching phase, and evaluation phase. In the preprocessing phase, we select source ontology (O_S) and target ontology (O_T) from the repository of input ontologies. After selecting the source and target ontologies, fetch all the labels

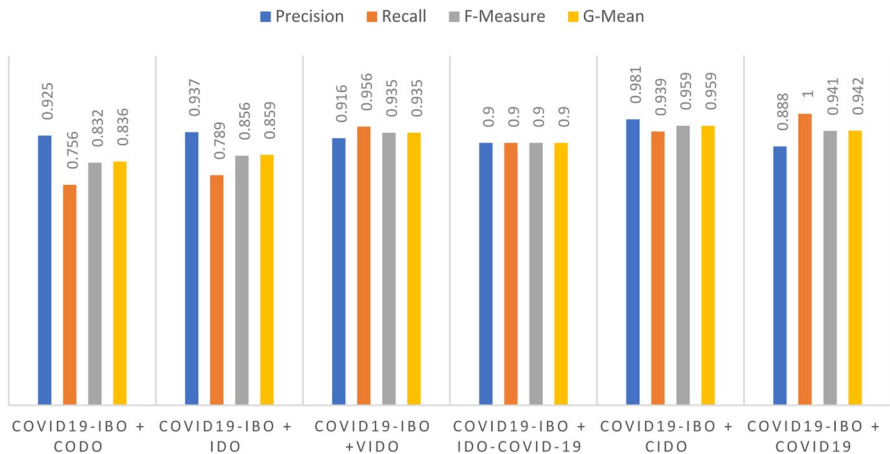


Fig. 8 Performance of SMA-Covid19 algorithm over Covid-19 ontologies

with the help of the Id of the node in both ontologies. The matching phase detects 1:1 mapping between the concepts of the selected source and target ontologies using two matchers namely Levenshtein and synonym. Levenshtein matcher calculates the similarity between the strings (labels) and synonym matcher fetches all the synonym of the concept from the synonym.com dictionary. Synonym matcher provides the background knowledge for the Covid-19 ontologies. All the synonyms of the disease domain are locally stored (.txt) in the system using APIs. So, at the run time, we directly fetch synonym from the file instead of synonym.com that provides freedom from necessary internet connectivity and enables us to match the ontologies in offline mode. The final similarity value between the concepts is assigned by taking the average value of both matchers. The third phase evaluates the performance of the schema level matching (T-Box) by calculating the value of the performance parameters (precision, recall, F-measure, G-mean, and run time).

Result and Discussion

To test our approach, we have selected different Covid-19 ontologies namely CODO, IDO, VIDO, IDO-Covid-19, CIDO, and run the SMA-Covid19 algorithm on these ontologies with Covid19-IBO as the source ontology. Figure 8 shows the matching results over these ontologies (Covid19-IBO&CODO, Covid19-IBO&IDO, Covid19-IBO&VIDO, Covid19-IBO&IDO-COVID19, Covid19-IBO&CIDO, and Covid19-IBO&COVID19) in term of precision, recall F-measure, and G-mean.

- *Covid19-IBO & CODO*: The reference alignment contains 82 correspondences. The SMA-Covid19 algorithm detects 67 correspondences in which 62 correspondences are correct. The obtained precision, recall, F-measure, and G-mean are 0.925, 0.756, 0.832, and 0.836, respectively, in the dataset of Covid19-IBO and CODO. The running time of the SMA-Covid19 algorithm is 0.094 s.
- *Covid19-IBO & IDO*: The reference alignment contains 19 correspondences. The SMA-Covid19 algorithm detects 16 correspondences in which 15 correspondences are correct. The obtained precision, recall, F-measure, and G-mean are 0.937, 0.789, 0.856, and 0.859, respectively, in the dataset of Covid19-IBO and IDO. The running time of the SMA-Covid19 algorithm is 0.62 s.
- *Covid19-IBO & VIDO*: The reference alignment contains 23 correspondences. The SMA-Covid19 algorithm detects 24 correspondences in which 22 correspondences are correct. The obtained precision, recall, F-measure, and G-mean are 0.916, 0.956, 0.935, and 0.935, respectively, in the dataset of Covid19-IBO and VIDO. The running time of the SMA-Covid19 algorithm is 0.71 s.
- *Covid19-IBO & IDO-Covid-19*: The reference alignment contains 30 correspondences. The SMA-Covid19 algorithm detects 30 correspondences in which 27 correspondences are correct. The obtained precision, recall, F-measure, and G-mean are 0.9, 0.9, 0.9, and 0.9, respectively, in the dataset of Covid19-IBO and IDO-COVID19. The running time of the SMA-Covid19 algorithm is 0.83 s.

- *Covid19-IBO & CIDO*: The reference alignment contains 116 correspondences. The SMA-Covid19 algorithm detects 111 correspondences in which 109 correspondences are correct. The obtained precision, recall, F-measure, and G-mean are 0.981, 0.939, 0.959, and 0.959, respectively, in the dataset of Covid19-IBO and IDO-Covid-19. The running time of the SMA-Covid19 algorithm is 185.3 s.
- *Covid19-IBO & Covid-19*: The reference alignment contains eight correspondences. The SMA-Covid19 algorithm detects nine correspondences in which eight correspondences are correct. The obtained precision, recall, F-measure, and G-mean are 0.888, 1, 0.941, and 0.942, respectively, in the dataset of Covid19-IBO and COVID19. The running time of the SMA-Covid19 algorithm is 0.25 s.

The obtained result shows that the proposed approach has reasonable results over the Covid-19 ontologies. Now with the help of the SMA-Covid19 algorithm, we can detect the overlapping information among the available Covid-19 ontologies in the absence of reference alignment (reference alignment given by the expert of the domain).

Conclusion and Future Work

The positive economic growth of any country reflects the financial soundness and increased purchasing power of that country. To know the impact of Covid-19 on the Indian banking sector, we have presented Covid19-IBO ontology by analyzing existing data that is available in different formats. The Covid19-IBO ontology reused already existing concepts of the CODO, IDO-Covid-19 FOAF, and schema.org. The developed Covid19-IBO ontology has been evaluated by the OntoMetric tool, OOPs pitfall scanner, competency questions, validators, and subjective testing. The evaluation's results show that ontology is well built as per the specified scope. Our ontology indicates that the banking sector must focus on digitization, risk management, and collection. The exact impact of the moratorium period and challenges in operations, implementing innovative banking models, etc. depends on the pandemic time that is not yet determined due to mutation of the Covid-19 virus. This sector must be focusing on the evaluation of rapidly altering situations and emerging situations based on various lead indicators and the latest research. We have also provided a schema matching approach (SMA-Covid19) that maps the proposed Covid19-IBO ontology with other accessible and available Covid-19 ontologies. Covid19-IBO ontology provides strength to the other Covid-19 ontologies by incorporating information about the impact of the pandemic in the Indian banking sector and it provides significant advantages compared to other Covid-19 ontologies. The current version of Covid-IBO ontology is available on the bio-portal for public use. The future work of this paper is to focus on the development of the widget that offers semantic services like visualization, annotation, mapping, etc.

References

- Hindustan Times (2020) URL: <https://www.hindustantimes.com/india-news/70-of-banking-sector-debt-affected-by-covid-19-s-impact/story-MAYiYZWz5NE6Pijm7XQNSJ.html>
- Dev, S. M., Sengupta, R.: Covid-19: Impact on the Indian economy, Indira Gandhi Institute of Development Research, Mumbai (2020)
- Rakshit, B., Basistha, D.: Can India stay immune enough to combat COVID-19 pandemic? An economic query. *J. Public Aff.* **20**(4), e2157 (2020)
- Kanitkar, T.: The COVID-19 lockdown in India: Impacts on the economy and the power sector. *Global Transitions* **2**, 150–156 (2020)
- Demirguc-Kunt, A., Pedraza, C.: Ruiz-Ortega, Banking sector performance during the covid-19 crisis (2020)
- WHO Coronavirus disease (COVID-19) situation reports, (2020) URL: <https://datasetsearch.research.google.com/search?q=query=covid19&docid=g3oiDuHtkLygNkWHAAAAA%3D%3D>
- Eular Covid-19 database, 2020. URL: https://www.eular.org/eular_covid19_database.cfm
- Kousha, K., Thelwall, M.: COVID-19 publications: Database coverage, citations, readers, tweets, news, Facebook walls Reddit posts. *Quantitative Science Studies* (2020). https://doi.org/10.1162/qss_a_00066
- Babcock, S., Cowell, L.G., Beverley, J., Smith, B.: The infectious disease ontology in the age of COVID-19. *J Biomed Semantics* (2020). <https://doi.org/10.1186/s13326-021-00245-1>
- BioPortal, 2020. URL: <https://bioportal.bioontology.org/ontologies/IDO>
- B. Dutta, M. DeBellis, CODO: An Ontology for Collection and Analysis of Covid-19 Data, (2020)
- Bio portal. URL: <https://bioportal.bioontology.org/ontologies/COVID19/?p=summary>
- de Lusignan, S., Liyanage, H., McGagh, D., Jani, B.D., Bauwens, J., Byford, R., Hobbs, F.R.: COVID-19 Surveillance in a Primary Care Sentinel Network: In-Pandemic Development of an Application Ontology. *JMIR Public Health Surveill.* **6**(4), e21434 (2020)
- Web Ontology Language, URL: https://en.wikipedia.org/wiki/Web_Ontology_Language
- Jain, S., Patel, A.: Smart ontology-based event identification. In 2019 IEEE 13th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoc), pp. 135–142. IEEE. (2019)
- Jones, D., Bench-Capon, T., Visser, P.: Methodologies for ontology development. (1998)
- Gruninger, M., Fox, M. S.: The design and evaluation of ontologies for enterprise engineering. In Workshop on Implemented Ontologies, European Conference on Artificial Intelligence (ECAI) (1994)
- Uschold, M., King, M.: Towards a methodology for building ontologies. *Artificial Intelligence Applications Institute, University of Edinburgh, Edinburgh* pp. 19–1. (1995)
- Gómez-Pérez, A.: Towards a framework to verify knowledge sharing technology. *Expert Syst. Appl.* **11**(4), 519–529 (1996)
- Noy, N. F., McGuinness, D. L.: *Ontology development 101: A guide to creating your first ontology* (2001)
- Protégé (2000). The Protege Project. <http://protege.stanford.edu>
- Musen, M.A.: The protégé project: a look back and a look forward. *AI matters* **1**(4), 4–12 (2015)
- Brank, J., Grobelnik, M., Mladeni, D.: A survey of ontology evaluation techniques (2005)
- Yu, J., Thom, J.A. and Tam, A.: November. Ontology evaluation using Wikipedia categories for browsing. In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (pp. 223–232). ACM (2007)
- Tartir, S., Arpinar, I.B., Sheth, A.P.: Ontological evaluation and validation. In *Theory and applications of ontology: Computer applications*, pp. 115–130. Springer, Netherlands (2010)
- Mehla, S., Jain, S.: Development and evaluation of knowledge treasure for emergency situation awareness. *Int. J. Comput. Appl.* **43**, 1–11 (2019)
- Jain, S.: Meyer V Evaluation and refinement of emergency situation ontology. *Int. J. Inform. Educ. Technol.* **8**(10), 713–719 (2018)
- OOPS!, URL: <http://oops.linkeddata.es/>
- Patel, A., Jain, S., Debnath, N.C., Lama, V.: InBiodiv-O: an ontology for Indian biodiversity knowledge management. (2021). [arXiv:2108.09372](https://arxiv.org/abs/2108.09372)
- Gangemi, A., Catenacci, C., Ciaranita, M., Lehmann, J.: *Modelling ontology evaluation and validation*, pp. 140–154. Springer, Berlin (2006)
- OntoMetrics, URL: <https://ontometrics.informatik.uni-rostock.de/ontologymetrics/>

32. RDF Validator, URL: <https://www.w3.org/RDF/Validator/documentation>
33. OWL2 Validator, URL: <http://mowl-power.cs.man.ac.uk:8080/validator/>
34. RDF Triple-Checker, URL: <http://graphite.ecs.soton.ac.uk/checker/>
35. Vapour, URL: <http://linkeddata.uriburner.com:8000/vapour>
36. Dereferencing HTTP URIs, URL: <https://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14>
37. Espíndola, R.P., Ebecken, N.F.: On extending f-measure and g-mean metrics to multi-class problems. *WIT Transactions on Information and Communication Technologies* **35**, 10 (2005)
38. Mishra, A. K., Patel, A., Jain, S.: Impact of Covid-19 outbreak on performance of Indian Banking Sector. *CEUR Workshop Proceedings*. **2786**, 368–375. International Semantic Intelligence Conference (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Archana Patel¹ · Narayan C. Debnath¹ · Ambrish Kumar Mishra² · Sarika Jain³

Archana Patel
archanamca92@gmail.com

Narayan C. Debnath
narayan.debnath@eiu.edu.vn

Sarika Jain
jasarika@nitkkr.ac.in

¹ Department of Software Engineering, School of Computing and Information Technology, Eastern International University, Binh Duong New City, Vietnam

² Office of DGA (Energy) Comptroller and Auditor General (CAG), Delhi, India

³ Department of Computer Applications, National Institute of Technology, Kurukshetra, Haryana, India