DESIDOC Journal of Library & Information Technology, Vol. 43, No. 3, May 2023, pp. 169-175, DOI : 10.14429/djlit.43.3.18379

Developing "Energy Access" Ontology Using Protégé Tool

Reeta Sharma[1*] and Uma Kanjilal[2]

[1]Knowledge Resource Centre, TERI, New Delhi, India
[2]Indira Gandhi National Open University, New Delhi - 110 068, India
[*]E-mail: reetas@teri.res.in

ABSTARCT

In this internet-based world where a large amount of information exists on the World Wide Web (WWW), the information retrieval systems are mainly centred on topic-based classification. Numerous efforts and time have been consumed for relevant information extraction through major search engines if the data is not organised correctly. Ontologies proved to be an effective technique for representing and retrieving information, which is the key idea in semantic web applications. Ontologies not only help in efficient knowledge representation and information retrieval, but it also helps in mapping the hidden knowledge about a subject. This paper discusses the process and method of building ontology on the "Energy access" domain. The methodology is based on the tools used in developing the ontology. Several tools are used to create an ontology. Protégé is a most popular tool for ontology editing and for developing ontology.[1] In this paper, various aspects like super class and subclass hierarchy, creating a subclass, instances for class illustration, query retrieval process, visualisation, and graph views have been demonstrated by using protégé software.

Keywords: Knowledge representation; Information retrieval; Domain ontology; Semantic web; Energy access; Protégé

1. INTRODUCTION

The concept 'Semantic Web' was imagined and deliberated first by Tim Berners-Lee to project and develop a 'web' such that the documents are connected, and the meaning of information in web documents can also be known and established. It is defined as "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation".[1]

In contrast to the current web (information technology systems), the semantic web is used for automating, integrating, and reusing data across diverse applications. Semantic applications aim to provide data with more meaning. As a result, the Semantic web envisions expanding the existing web, making it simpler to find, share, reuse, and integrate information. By enhancing current online documents/publications with new data and metadata, it expands their functionality and makes it possible for machines to process and retrieve them autonomously.

Two complementary technologies, XML and RDF, are required to build an intelligent web, i.e., a semantic web. XML is a simple, extensible markup language that permits users to generate tags to explain web documents. RDF (Resource Description Format) turns elementary web data into structured data. RDF creates standards for XML applications to make it easier for them to interoperate and communicate with one another, increasing the integration and interoperability of data and systems. RDF and XML are used to describe the information available on the internet, i.e., metadata (data about other data). However, if we expect machines to interact with each other or share data in the real sense of the word, then semantic interoperability is essential. An accurate description is required to define various terms and their associations to achieve semantic interoperability.[2] Thus, ontology came into existence and was therefore developed in Artificial Intelligence (AI) to encourage the reuse of knowledge among complex systems by using a combination of XML and RDF technologies.

Ontology plays a vital and significant role in information access and, thus, serves as the central enabling technology in achieving the semantic web environment.

2. LITERATURE REVIEW

The term "Ontology" refers to the shared understanding of a specific field of study that can be applied as a unifying framework to enhance information sharing,

Received : 01 August 2022, Revised : 17 May 2023
Accepted : 17 May 2023, Online published : 06 July 2023

interoperability, and reuse between humans or machines.[3] Ontology takes hold of an entire set of meanings and comprises everything from controlled vocabularies, taxonomical categories, classifications of products and services, database vocabulary, and relationships between concepts.

In the last decade, the use of ontologies in information systems has become increasingly popular in various fields and was seen as a formal method for modeling the structure of a system. The skeleton of ontology consists of a hierarchy that classifies the concepts at each level and proceeds from generalised to specialised concepts. In order to represent exact information, an ontology engineer examines the entries that matter most and arranges them into concepts and relationships.[4]

Ontologies are crucial to the semantic description needed for the classification and common understanding of documents in the knowledge domain. They support the interchange of information retrieval, employ a single notion to reduce fuzzy concepts, and play a vital role in developing knowledge-based systems.[5]

One essential requirement of domain ontologies is to attain a significant coverage of the domain concepts and their relationships. These can be created by using an automatic or semi-automatic approach. Traditional text mining techniques were labor-intensive and time-consuming, and using semantics is one of the challenges in converting data to knowledge. Further, experts in Natural Language Processing (NLP) also claim that working with ontologies requires the use of a semantic approach.[5]

Techniques for automated ontology enrichment from free-text documents are provided by methodologies created in natural language processing, information extraction, information retrieval, and machine learning.[6]

The ontology techniques combine artificial intelligence and machine language to facilitate knowledge exchange. It also incorporates natural language processing and knowledge representation techniques and can be a communication channel between human beings and computers. The ontology can be further used for knowledge management and information retrieval.[7] Ontology can be built, organised, and managed manually or to some extent automatically. Depending on how much human engagement an automatism technique requires, we can distinguish between semi-automated and completely automatic ways of building, maintaining, and evolving the ontology. The majority of the techniques use semi-automatic tools designed for ontology use.[8]

3. OBJECTIVES

The objectives of the study are to:

- Examine domain knowledge in the area of 'Energy Access'
- Distinguish domain knowledge from the working knowledge
- Design and develop ontology on 'Energy Access' domain which will help TERI and other organisation to share a common understanding of the structure of information and its various concepts among people or software agents;
- Make domain assumptions clear to researcher and budding scientist and policymakers by designing and developing ontology model.

4. RESEARCH QUESTIONS

Q.1 How can an ontology of Energy Access (OEA) model be designed and developed to visualise the energy access domain?

Q.2 How does the model capture and represent 'Energy Access' domain knowledge?

5. DOMAIN ONTOLOGY ON "ENERGY ACCESS"

The "Energy access" concept does not offer itself an easy-going definition. In a report published by IEA in 2017, "Energy access is the golden thread that weaves together economic growth, human development, and environmental sustainability".[16] "Further, in a report by World Bank Group publication, the IEA defines energy access as "a household having reliable and affordable access to both clean cooking facilities and to electricity, which is enough to supply a basic bundle of energy services initially, and then an increasing level of electricity overtime to reach the regional average."

5.1 Need for Ontology in 'Energy Access' Domain

In spite of so much effort by Government and policymakers and other stakeholders, this field still needs more coverage and awareness among various stakeholders. Access to energy means many things and can be broadly classified as "Access to Energy services", "Energy supply," "Energy use," "Socioeconomic development," "Standalone energy solutions", "Illustrative technologies," etc. The "Energy access" domains can be classified into many classes and subclasses.

Poor coverage of this field and confused terminologies over the internet such as energy poverty, rural energy, sustainable energy, etc. need to be addressed with the help of universal standard and common knowledge, which is rich in semantics and represented in machine-understandable form.

Ontology in this domain will make stakeholders understand and be aware of 'energy access' and how it can be achieved by using appropriate technologies. Hence developing such kind of ontology is the need of the hour.

6. METHODOLOGY

There are number of methodologies available on the net to build the domain ontologies; however, one can modify the same according to their requirement.[8] Some of the most popular one are - Knowledge engineering, Methodology, On-To-Knowledge and TOVE etc.

However, there is no set pattern for developing ontologies, and one can build the same with viable alternatives.[8] The process of developing an ontology is largely iterative. The concepts in the ontology should represent the relationships in the determined domain of interest and be near to objects (physical or logical).

A crude initial attempt at ontology can be made, and it can then be reviewed and improved. The ontology concepts must reflect the reality of the ontology, which is a model of a genuine domain in the world. An ontology can be evaluated and debugged after it has been defined by using it in applications for addressing problems or talking with specialists on the topic. We will, therefore, almost probably need to update the original ontology. For building the OEA ontology following steps were followed:

6.1 Identify Resources and Experts

Regarding the aim of the study and finding answers to the stipulated research questions, it is required to see how researchers in databases or any other information retrieval system on the worldwide web has used these terms.

It is premised that research publications are one of the best sources to know the various aspects related to energy access field and for what purpose researchers are using to retrieve the data on the same.

6.2 Collection of Terms and Period of Study

The period considered for the study was last decade's research publication i.e. 2008-2018 initially later on literature from 2018 onwards was also explored from various printed and online resources.

6.3 Apply Knowledge Capture Techniques

Techniques like Interview, Delphi method, Brainstorming, concept map, etc. were adopted to know the various terms and research questions in Energy Access Domain.

6.4 Ontology Extraction

Ontology in EA domain was extracted from various concepts, relations, constraints, and axioms extraction.

7. ONTOLOGY CONSTRUCTION TOOLS

The purpose of utilizing a tool to create ontologies serves to ensure sustainability for the ontology life cycle and reuse. By using these tools, ontology construction can be made easier, which can be very challenging otherwise. Several online ontology-building tools support users by offering interfaces such as OntoEdit, Swoop, Onto Studio, Oiled, Sematic Works, TextEditor, Protégé, etc.[8]

Protégé is one of the most popular tools for creating and revising ontologies with a GUI (Graphical User Interface). By using taxonomies, it allows ontology developers to explain conceptual terms and how they relate to one another.[9]

7.1 Protégé and its Advantages

Protégé is an open-source ontology editor and framework developed by Standford University and free to download. It is free to download on any platform and has different plugin options. It offers a Graphical User Interface (GUI) that enables users to construct sophisticated ontologies. With the use of diagrams, Protégé's visualisation tools, including OntoViz, enable users to visualize ontologies. A large community of academic, governmental, and business users uses Protégé to create knowledge-based solutions across various industries.[10-11]

Protege allows for the creation of domain ontologies, improves meta-modeling flexibility, and allows for the customisation of data entry forms. Without being aware of or considering the output language's syntax, it typically focuses on conceptual modeling and knowledge engineering.[12]

According to a survey by Sir Jorge Cardoso of the most popular ontology editors, the market share of the protégé tool was 68.2 %, followed by Swoop, Onto Edit, TextEditor, Altova Semantic Works, Oiled, Onto Studio, etc.[13]

7.2 Protégé-OWL

OWL is a web ontology language. It is used to manage information on the web and is constructed on top of RDF (Resources Description Format). It was created and interpreted by machines. Owl and RDF are similar; however, OWL is more powerful and comprehensible than RDF. Compared to RDF, it has a larger vocabulary and a more stable syntax.[14]

A network of classes, attributes, and individuals can be used to represent the OWL ontology. Classes specify the logical features and names of the pertinent domain concepts. Properties provide associations between classes and enable assigning simple values to instances. They are also known as roles, attributes, or slots. With particular values for the properties, individuals are instances of the classes.[7]

7.3 Ontology Classes of "Energy Access"

The "Energy Access" domain is used to design a concept that describes each class's functionality. Here, the class "Energy use" is a sub-class of "Energy access," and it has three subclasses, namely (i) "households," (ii) "productive engagements," and (iii) "community facilities." The class household is again subdivided into three subclasses (i) "Access to electricity",(ii) "Access to energy for cooking solutions", and (iii) "Access to energy for space-heating solutions".[17] The classification further proceeds to a large number of classes and sub-classes. Several classes and subclasses have been created under the "Energy access" domain, but due to space scarcity, only some classes and subclasses are described elaborately in this paper.

The representation of one of the class,"Energy use," and its subclasses is depicted in Fig 1.

7.4 Class "Energy use" Representation in Portege

Proteges 5.5.0 is used to represent the Energy access domain. One can add, delete, and rename the classes by using the classes tab in Protégé, as shown in Fig 2. The orignal class hierarchy tree view by using the Ontograph plugin should look like the image shown below in Fig 2. One class, thing, is present in the empty ontology.

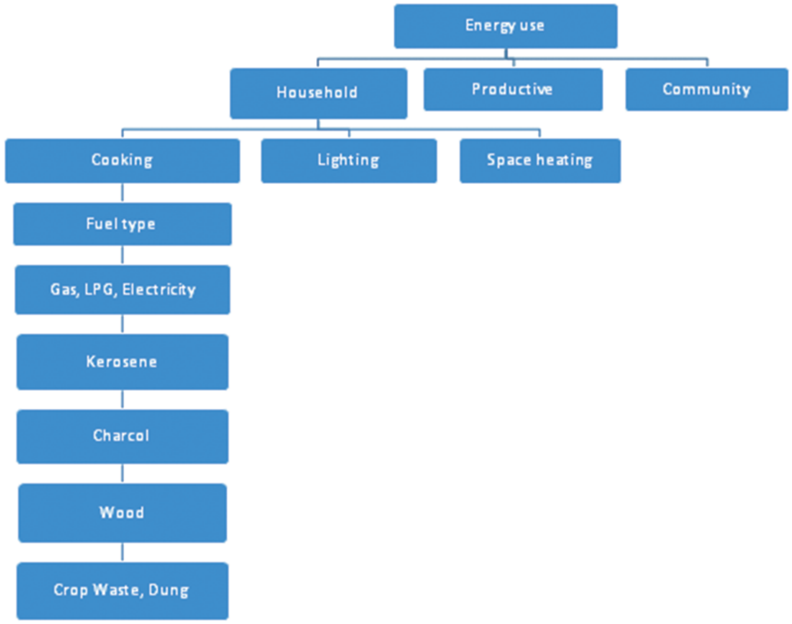

Figure 1. Representation of "Energy use" class and subclasses.

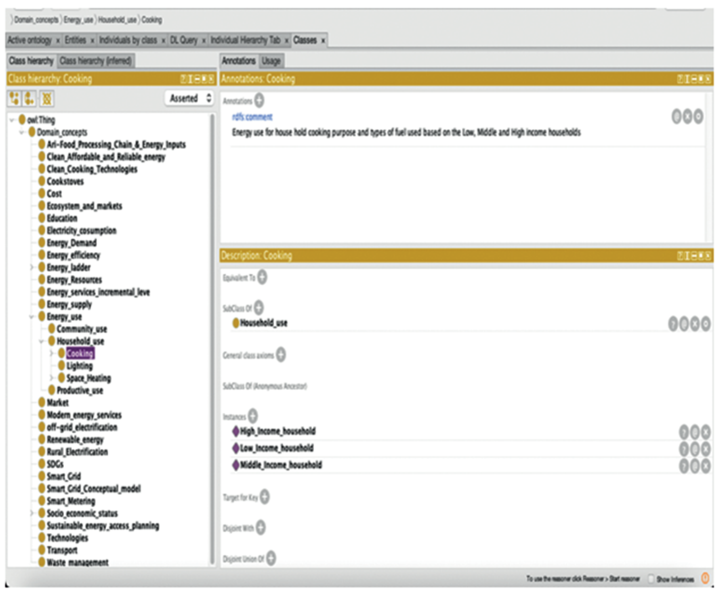

Figure 2. Representation of class hierarchy using Protégé.

The class Thing is used to refer to the collection of all individuals. Because of this, Thing is the super class of all classes.[18] The class tab is selected to add a class and subclass. This creates a new class as a subclass of the chosen class Thing.

7.5 Properties

OWL properties denote relationships between two objects. Object properties and datatype properties are two main types of properties. While data type properties define the relationship between an individual and data values, object properties are used to indicate the relationship between two individuals and link one individual to another. Annotation property is another OWL property that adds metadata to classes, individuals, and object/data type properties.[18]

The characteristics of a property are modified through the form shown in Fig 3. Adding metadata space is provided in the upper part of the protégé window, where it also displays the property's name, annotations, and so on.

This paper implements many properties. Each property has its own characteristics. Every attribute of the property is independent, and each has its features. This paper discusses some object properties; namely, has Fuel Type, Used by, is Caused By, which is shown in Fig 3.

7.6 RDF/XML Rendering

Any RDF expression's structure comprises a subject, a predicate, and an object, also known as triples. The node and arc diagram can be used to depict each triple

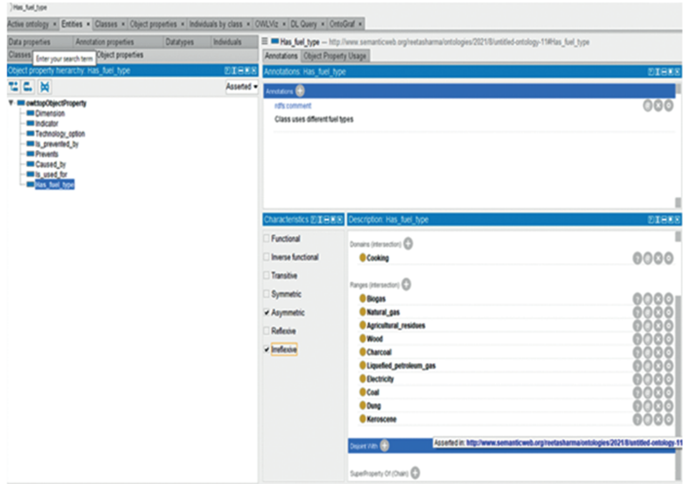

Figure 3. Object property for classes.

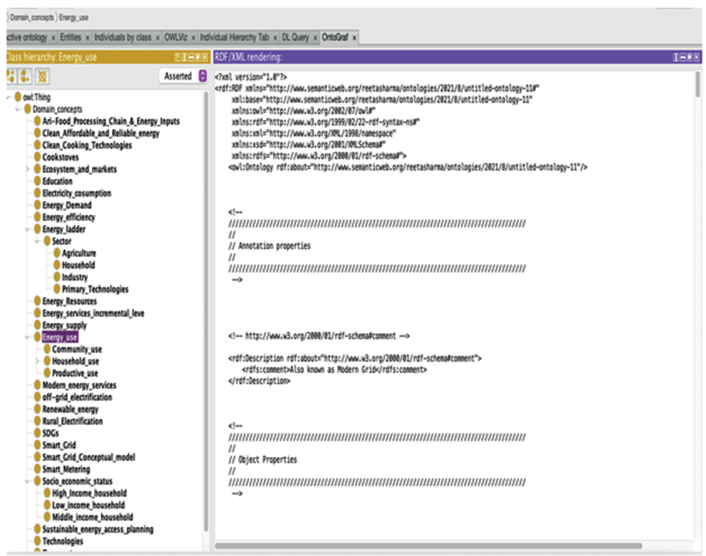

Figure 4. RDF/XML schemas.

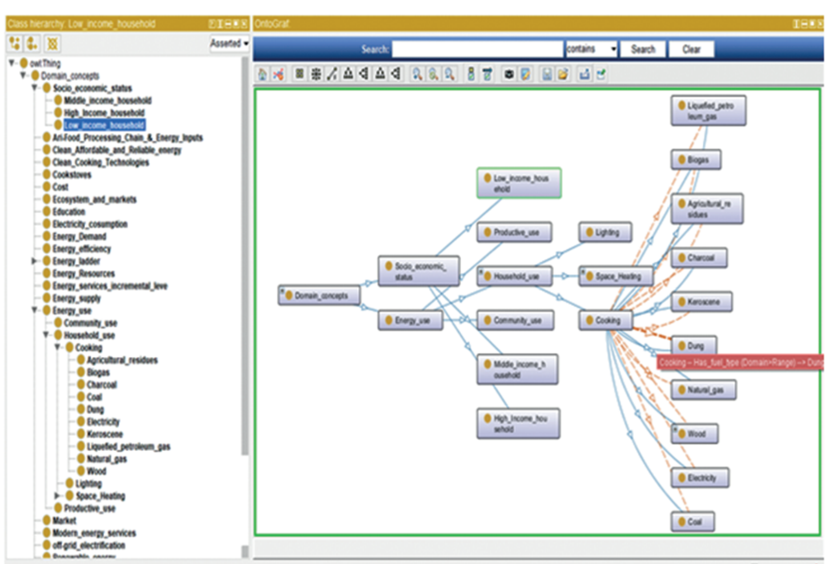

Figure 5. Ontographical view of "Energy access" ontology.

as a node-arc node link, indicating an RDF graph (A set of each triple).

The ontology used a common language XML, to express and avoid conversion between different description languages. For exchanging data, XML contains standards. RDF/XML is the most popular serialisation format for OWL ontology.[8] RDF/XML schemas are represented in Fig 4.

8. ONTO GRAPHICAL VIEW OF "ENERGY ACCESS" ONTOLOGY

Onto graphical view completely describes each class and subclasses that were created. Due to the lack of space, only a few classes and sub-classes are mentioned in the below graph, depicted below in Fig 5. This graph also depicts the relationship that exists between every class and subclasses.

9. CONCLUSIONS AND FUTURE WORK

This study aimed to extract knowledge from the "Energy Access" domain by describing a framework for ontology creation. Under this construction of the ontology framework, the researchers and budding scientists involved in the domain will achieve a mass of linguistic information and the context-based knowledge information demonstrated.

The rapidly developing field of ontology has the potential to significantly advance the management, organisation, and interpretation of information. It is essential for facilitating communications, interoperability, and content-based access and offering qualitatively new service levels on the semantic web, the next stage of web transformation. Further information professionals and subject experts are creating domain ontologies to map the scattered knowledge on a domain. The study's outcome is expected to be advantageous to researchers to gain a shared understanding of the subject. This ontology model will likely bring all relevant terminologies to one platform with more clarity in concepts and their relations. This will help researchers and budding scientists to understand this field much better.

REFERENCES

1. Berner-Lee, T; Hendler, J & Lassila, O. The semantic web. Scientific American: Feature Article. *The Semantic Web:* May 2001. https://www-sop.inria.fr/acacia/cours/essi2006/Sientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf (Accessed on 20 July 2022).
2. Giri, K. Role of ontology in semantic web. *DESIDOC Journal of Library & Information Technology*, 2011, **31**(2), 116-120. https://publications.drdo.gov.in/ojs/index.php/djlit/article/download/863/364 (Accessed on 20 July 2022).
3. Uschold, M. & Michael, G. Ontologies: Principles, methods and applications. *Knowledge Engineering Review.* 1996, **11**(2), 93136. http://www.aiai.ed.ac.uk/publications/documents/1996/96-ker-intro-ontologies.pdf (Accessed on 26 July 2022).
4. Gersberg, I.W. & Ebecken, N.F.F. A simple strategy to start domain ontology from scratch. *Int. J. Advanced Computer Sci. Applications (IJACSA),* Special Issue on Natural language processing 2014. doi: 10.14569/SpecialIssue. 2014.040107.
5. Pasapitch, C.; Nittaya, K. & Kittisak, K. On transforming the ER model to ontology using protégé owl tool. *Int. J. Computer Theory and Engineering,* 2014, **6**(6) 484-489. doi:10.7763/IJCTE.2014.V6.914.
6. Liu, K.; Hogan, W.R. & Crowley, R.S. Natural language processing methods and systems for biomedical ontology learning. *Journal of Biomedical Informatics.* 2011, **44**(1), 163-179 doi.org/10.1016/j.jbi.2010.07.006.
7. Elhefny, M.A; Elmogy, M & Elfetouh, A.A. Building OWL ontology for obesity related cancer. *9th International Conference on Computer Engineering & Systems (ICCES)*, 2014. 177-183 doi: 10.1109/ICCES.2014.7030953.
8. Vogrinčič, S. & Bosnić, Z. Ontology-based multi-label classification of economic articles.*ComSIS,* 2001, **8**(1), 111-119. doi:10.2298/CSIS100420034V.
9. Ramkumar, S. & Poorna, B. Semantic information retrieval based on domain ontology. *International Conference on Computing and Intelligence Systems,* 2015, **4**(special issue), 1387-1390. (Accessed on 30 July 2022)
10. Vindula, J. Ontology generation and visualisation with protégé. *Medium.com,* 2017. https://medium.com/@vindulajayawardana/ontology-generation-and-visualisation-with-prot%C3%A9g%C3%A9-6df0af9955e0. (Accessed on 12 July 2022).
11. Protégé. The protégé ontology editor and knowledge acquisition system. *Stanford University*, 2020. https:/protege.stanford.edu. (Accessed on 01 August 2022).
12. Kapoor, B. & Sharma, S. A comparative study ontology building tools for semantic web applications. *International Journal of Web & Semantic Technology* (IJWesT). 2010, **1**(3). doi: 10.5121/ijwest.2010.13011.
13. Cardoso, J. The semantic web vision: Where are we?. *IEEE Intelligent Systems*, 2007, **22**(5) 84-88. doi:10.1109/MIS.2007.4338499.
14. Jayanth, K.; Sujitha, R.; Priya, M.; Raju, J.; Vanmathi, C. Implementation of ontology matching using protégé. *International Journal of Computer, Applications Technology and Research*. 2013, **2**(6), 723-725. doi:10.7753/IJCATR0206.1017.
15. Mahalakshmi, S. & Saranya, G. Developing cancer ontology using Potégé-OWL. *Adalya J.,* 2020. **9**(2). https://www.adalyajournal.com/gallery/68-feb-ic142.pdf (Accessed on 21 July 2022)
16. IEA. WEO Energy Access Outlook 2017: From poverty to prosperity. *International Energy Agency,*

2017. http://www.worldenergyoutlook.org/resources/energydevelopment. (Accessed on 81 July 2022).

17. Bhatia, M. & Angelou, Ni. Beyond connections: Energy access redefined. ESMAP Technical Report; 008/15. *World Bank*, Washington, DC. © World Bank, 2015. https://openknowledge.worldbank.org/handle/10986/24368 License: CC BY 3.0 IGO. (Accessed on 24 July 2022).
18. Kalaivani, P.; Anandaraj, A. & Raja, K. An ontology construction approach for the domain of poultry science using protégé. *Int. J. Infor. Technol. Manage. Sci.,* 2011, **1**(2). https://arxiv.org/ftp/arxiv/papers/1302/1302.5417.pdf (Accessed on 01 August 2022).

CONTRIBUTORS

Ms Reeta Sharma is a noted library and knowledge management professional and expert in developing knowledge management system, digital libraries and website taxonomies. She is working as Associate Fellow with Knowledge Resource Centre at TERI and have more than 20 years of experience in knowledge management, digital libraries, academic research, project management and organising outreach and capacity development programs for various stakeholders.
Contributions to the paper: Writing and developing the paper, data interpretation and content editing.

Prof Uma Kanjilal is Pro-Vice Chancellor, Professor and Head in the Faculty of Library and Information Science in the Indira Gandhi National Open University (IGNOU) she has more than 31 years of experience in the Open and Distance Learning System. Prof. Kanjilal is one of the National Coordinators of SWAYAM (India MOOCs) and SWAYAM PRABHA (Educational DTH Channels) of the Ministry of Human Resource Development. She was a Principal Investigator (PI) of IGNOU for the National Virtual Library of India (NVLI) Project of Ministry of Culture. Her specialisation includes e-learning, multimedia courseware development, ICT applications in Libraries and Digital Libraries. She was a Fulbright Scholar in University of Illinois, Urbana Champaign in 1999-2000 where she worked on Multimedia courseware development.
Contributions to the paper: Structuring of content and review.