HBVO: Human Biological Viruses Ontology

Sheikh Kashif Raffat[1]*, Mohd. Shahab Siddiqui[2], Mohd. Siddiq[1], Zubair A. Shaikh[3] and Abdul Rahman Memon[4]
[1]Department of CS and IT, Federal Urdu University of Arts, Sciences and Technology, Karachi, PAKISTAN
[2]HIIT, FEST, Hamdard University, Karachi, PAKISTAN
[3]National University of Computer and Emerging Sciences, Karachi, PAKISTAN
[4]FEST, Hamdard University, Karachi, PAKISTAN

Abstract

Biological viruses have recently received a lot of attention especially in sub continent due to some spectacular effects of infections like bird flu, dengue and swine flu. This problem generated a requirement to classify these viruses in some formal form. Therefore, we are proposing an ontology for Human Biological Viruses Ontology (HBVO) that covers all the viruses that belongs to human. The proposed ontology is developed by using the principles of Open Biological Ontologies and will be available in the format of OBO. It can be viewed by using OBO-Edit. To develop HBVO we used the taxonomy developed by the International Committee on Taxonomy of Viruses.

Introduction

Ontologies got more attention and popularity among bioinformatics researchers for representing relationships among biological concepts. For the development of ontology, the Open Biomedical Ontologies (OBO)[1-2] (http://www.obofoundry. org/) library is provides an umbrella to share different biological and medical knowledge, on one common platform and format.

Biological viruses are classified in terms of orders, families, sub-families, genus and species. This classification was given by the International Committee on Taxonomy of Viruses (ICTV) (http://www.ictvonline.org/). Also the Baltimore Classification (http://www.virology. net/Big_Virology/ BV Family Group.html) classifies viruses into families depending on their type of genome (nucleic acid). Virus host belongs to different communities like vertebrates, invertebrates, plants, bacteria, algae, fungi and protozoa. This generates a requirement to define a comprehensive ontology for biological viruses. Therefore, Biological Viruses Ontology (BVO) is initiated in this research work.

As a first component, Human Biological Viruses Ontology (HBVO) will classify viruses belongs to human community[3]. Human can be the host of DNA/RNA vertebrate viruses. Some viruses like dengue in Europe and Australia still do not have any host to it. The impact of biological viruses on any (virus host) community[4], mainly human communities are devastated.

The aim of HBVO is to support an integrated conceptual framework of BVO with a structured and controlled vocabulary to describe and categorized the biological viruses. To relate the nucleic acid with BVO we used Gene Ontology (GO), Sequence Ontology (SO), Chemical Entities of Biological Interest (CHEBI), Disease Ontology (DOID) and RNA Ontology (RNAO), they have codes like GO:0019021 (DNA viral genome), SO:0000352 (DNA), CHEBI:16991 (DNA) etc. Previous ontologies like DOID (human disease) and IDO (infectious diseases) cover viruses of their own areas. But in BVO, we shall cover 4379 viruses in different species[5] classified by ICTV. This ontology will allow better querying and handling of viruses in future.

The *International Committee on Taxonomy of Viruses* (ICTV) is a committee of the virology division of the International Union of Microbiological Societies and governed by the statutes agreed with the virology division. It begins to devise and implement rules for the naming and classification of viruses early in the 1990s. The viral classification starts at the level of order and moves towards species with the suffixes as: [Order: -virales, Family: -viridae, Sub-family: -virinae, Genes: -virus]

Table-1
Evolution of taxonomy of viruses by ICTV

Year	Order	Family	Subfamily	Genera	Species
2009	6	87	19	348	2285
2008	5	82	11	307	2078
8th Report	3	73	11	289	1898

Taxonomies till 8th report issued with 12 digit decimal codes [2 digits for order, 3 digits for family, 1 digit for subfamily, 2 digits for genus, 1 digit for species, 3 digits for type species]. (http://www. ictvdb.org/Ictv/index.htm). Later on they updated the taxonomy at (http://www.ictvonline.org/virus Taxonomy.

asp? version = 2009) and abolished the assigned codes. Every year taxonomic proposals reviewed by ICTV[6-7] and approved by Executive Committee (EC) to introduce new species, genera, sub-families, families and orders, number of recognized taxonomy issued by ICTV in different years (http://www.ictvonline.org/virusTaxInfo.asp) as shown in table-1. Ontology will help to handle this situation more easily in future also.

Material and Methods

Virus: A virus (biological or non-biological) can be defined as a self reproducing entity, which uses a mechanism for reproducing itself. It needs a host to exist and can not exist by itself. The word is from Latin 'virus' referring to 'poison', in 1392 used first time in England (http://www.oed.com). Its origin is from Latin virulentus (poisonous), which means "agent that causes infection disease" (http://www.etymonline. com/). Viruses, viroid and prions are smallest infections biological entities composed of protein coat and a nucleic acid core, for transcription and replication depends on their host (vertebrate, invertebrate, plants etc). The ICTV currently has classified 4379 viruses from different species[5].

Altan-Bonnet define virus as hijacker[8], which hijacks a key enzyme from its host cell and create an ideal lipid environment to replicate. If we classify viruses according to nucleic acid then it has two major categories of DNA and RNA, both have same method for replication in any host.

For transcription and replication of the viruses, it will enter in the host cell (animal, plant, etc), hijack key enzyme, transcript and replicate, translate and produce viruses.

Ontology: An ontology[9] is a formal representation of knowledge as a set of concepts within a domain, and the relationship between those concepts. It is used to reason about the entities within that domain, and may be used to describe the domain. It is a classification technique[10] to classifying the scientific data in a controlled manner. Basically ontology is a controlled vocabulary of well-defined terms with specified relationships between them, capable of interpretation by both human and computers.

Ontologies are developed in many tools like OBO-Edit (mainly used for Biological Ontologies), Protégé (developed by the Stanford University, USA) and TODE[11] (developed by the National University of Computer and Emerging Sciences, Karachi, Pakistan). There exists some other ontologies of related areas that we have studied like Community Ontology[4] and Human Community Ontology[3].

Every thing in the world needs to be classified according to their properties. Viruses are also classified into different systems according to properties such as size, types of nucleic acid (http://www.virology.net/Big_Virology/BVFamily Group.html),

structure, host and immunological characteristics. Any newly discovered virus must belong to any of group if classification scheme exist, else it may duplicate and increase redundancy. The existing classification systems of viruses are Hierarchical Virus Classification System (http://www.nlv.ch/ Virologytutorials/ Classification.htm), Baltimore Classification System and ICTV Classification.

The Open Biomedical Ontology (OBO) library contains a wide range of ontologies of biological domain. OBO aims to unite these bio-ontologies[12] under one umbrella[1-2]. The ontologies in the OBO library are defined in OBO format (http://ontologenesis.knowledgeblog.org/245) and are organized by using different types of relations[13]. The ontology Evolution Explorer (OnEX)[14] is a system developed for exploring ontology changes. OnEX is a three-tier architecture, which includes ontology repository, middleware components and web application. It access to 560+ versions of 16 well-known ontologies. The following are some of the existing OBO ontologies which will be used by BVO and also been held in OnEX.

The *Gene Ontology* (http://geneontology.org/) (GO)[15-16] is the most popular and mature ontology in bioinformatics. The GO Consortium is a body that legally responsible for the representation of gene. GO contains three separate ontologies; molecular functions (8637 terms), cellular components (2432 terms) and biological processes (17069 terms)[15]. By the end of 2006, GO structure was completed because all three GO ontologies became is a relationship complete[16]. GO has also introduced the has part relationship along with other relations used in GO.

The *Sequence Ontology* (SO) (http://sequenceontology.org) is responsible to describe key features for genomic and other structured sequence[17-19]. SO provided information of viral sequence and used cross product terms in OBO[1-2].

The *Chemical Entities of Biological Interest* (ChEBI)[20] (http://www.ebi.ac.uk/chebi/) is defined molecular entities of 'small' chemical compounds. It includes a chemical ontology, which allows the relationships between molecular entities and their parents and/or children to be specified in a structured way.

The *Disease Ontology* (DO/DOID) (http://do-wiki.nubic.northwestern.edu/index.php/Main_Page) describes diseases of human anatomy[21]. It is a controlled vocabulary developed for annotation purpose. It used nucleic acid as parent of order/family and used is a relationship. For nucleic acid they defined their on Identifiers. It is a part of Nugene project at Northwestern University.

The *Ribuo Nucleic Acid Ontology Consortium* (ROC) (http://roc.bgsu.edu/) has created RNA Ontology[22] (RO/RNAO). The aim of RO/RNAO is to describe and characterize RNA sequences, its structures and dynamics[23]. RO/RNAO used some

of CHEBI concepts with relationship of has_functional_parent. It was developed by using the steps of Ontology Development 101 by Noy and McGuinness, RO/RNAO work closely with Gene and Sequence Ontology[19].

The Biological Viruses Community Ontology (BVCO)[24] based on Core Ontology and Species. The core ontology is Biological Viruses Ontology (BVO) contains the viruses listed by the ICTV. BVO contains 6 orders, 87 families, 19 subfamilies and 388 genera and has codes as, Order: BVO:0000001 to BVO:0000020, Family: BVO:0000021 to BVO:0000150, Sub-Family: BVO:0000151 to BVO:0000200,   Genus: BVO:0000201 to BVO:0000650

The species ontology contains the two major living communities of Multicellular Organism Biological Viruses Community Ontology (MBVCO) and Unicellular Organism Biological Viruses Community Ontology (UBVCO). It has code as, Multicellular: BVO:0000652 to BVO:0006000, Unicellular: BVO:0006001 to BVO:0007000

The two major sub communities of MBVCO are Animal Biological Viruses Community Ontology (ABVCO) and Plant Biological Viruses Community Ontology (PBVCO) and have codes as, Animal: BVO:0000652 to BVO:0004000, Plant: BVO:0004001 to BVO:0006000

Proposed plan for developing the HBVO: Some of the queries raised before the development of the HBVO: i. Why we need an ontology for human viruses? ii. Who are the users of HBVO? iii. What is the domain of HBVO? iv. Are there any existing ontologies used in HBVO? v. What kind of information should HBVO contain?

The main components[25] of any ontology are classes, hierarchy (is_a), relations (other then is_a relation as we used has_part) and axioms. We follow the steps described by Noy and McGuinness[26] for developing our proposed ontology of HBVO. The answers of above questions will cover under the following steps given by Noy and McGuinness in Ontology Development 101 method: i. Determine the domain and scope of the ontology, ii. Consider reusing existing ontologies, iii. Enumerate important terms in the ontology, iv. Define the classes and the class hierarchy, v. Define the properties of classes—slots, vi. Define the facets of the slots, vii. Create instances

In first step, we determine the *domain* of HBV Ontology, which is the viruses those hosted by human, classified by ICTV released in 2009. For discussing the scope, we used has_part relationship to relate HBVO with other existing biological ontologies.

In second step, we studied existing ontologies of GO, SO, ChEBI, RNAO, DO/DOID and *reuse* the code of these ontologies in HBVO. In third step, we *enumerate* important

terms relevant to the domain of HBVO (e.g. Order, Family, Sub-family, Genus, Specie, Nucleic acid).

In fourth step, for *classes* and *class hierarchy* we used same hierarchy defined by ICTV.

DO/DOID used nucleic acid as parent (is_a) of order/family but we used nucleic acid as parent of virus species. Also we used relationship has part in HBVO. The classes need *properties* to define them; OBO-Edit has some properties like definition, synonyms (exact, related), comments etc. and these need some source of reference. The figure-1 shows the different level of BVO with is a relationship (arrow with I written on it). Biological_Virus (BVO:0000000) is the base class that contained virus classes with suffixes as order (virales), family (viridae), subfamily (virinae), genus (virus) and species (mumps virus etc. which link with human community). The others ontologies terms are Gene Ontology (biological_process GO:0008150 and cellular_component GO:0005575), Chemical Entities of Biological Interest (chemical entity CHEBI:24431), Disease Ontology (disease DOID:4), Ribuo Nucleic Acid Ontology (molecule RNAO:0000141) and Sequence Ontology (sequence_attribute SO:0000400). that will be link with different hosts to make it BVCO.
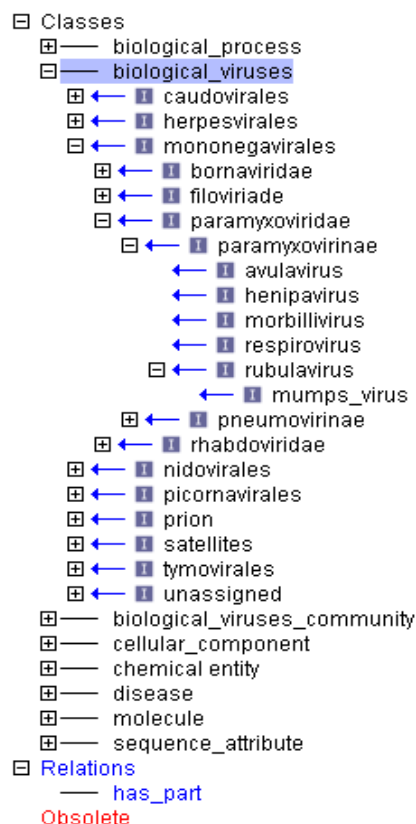

Figure-1
The OBO-Edit screenshot of Biological Viruses Ontology (BVO) tree structure

Results and Discussion

The Human Biological Viruses Ontology (HBVO) is developed as a structured controlled vocabulary for viruses hosted by vertebrate (human). As the research in its initial stage, therefore some changes might be needed in due course of time. Our proposed Biological Viruses Ontology (BVO) will define viruses according to their hosts, which includes Multicellular Biological Viruses Ontology (MBVO) with Animal Biological Viruses Ontology (ABVO), Plant Biological Viruses Ontology (PBVO) and Unicellular Biological Viruses Ontology (UBVO). nucleic acids (DNA/RNA), order, and family as earlier defined by ICTV Classification.

In the first step, we defined HBVO, which is one of the components of ABVO (can be seen in figure-2 and figure-3), which covers the viruses hosted by human. The major components of ABVO are Invertebrate Biological Viruses Ontology (IBVO) and Vertebrate Biological Viruses Ontology (VBVO). VBVO includes Human Biological Viruses Ontology (HBVO).
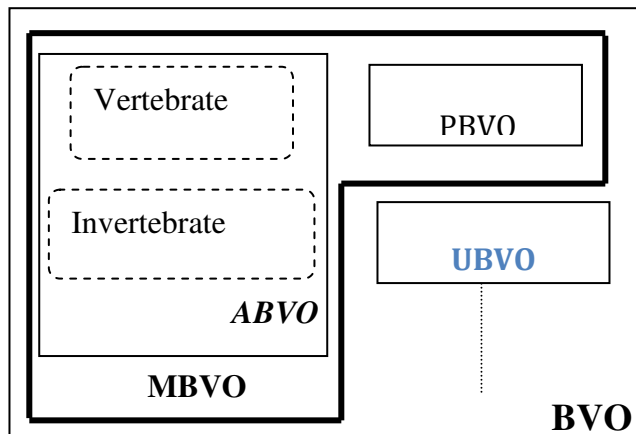

Figure-2
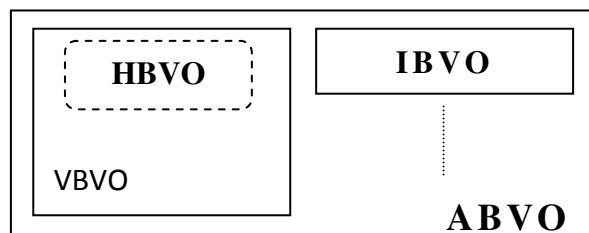**Block diagram of Biological Viruses Ontology (BVO)**

Figure-3
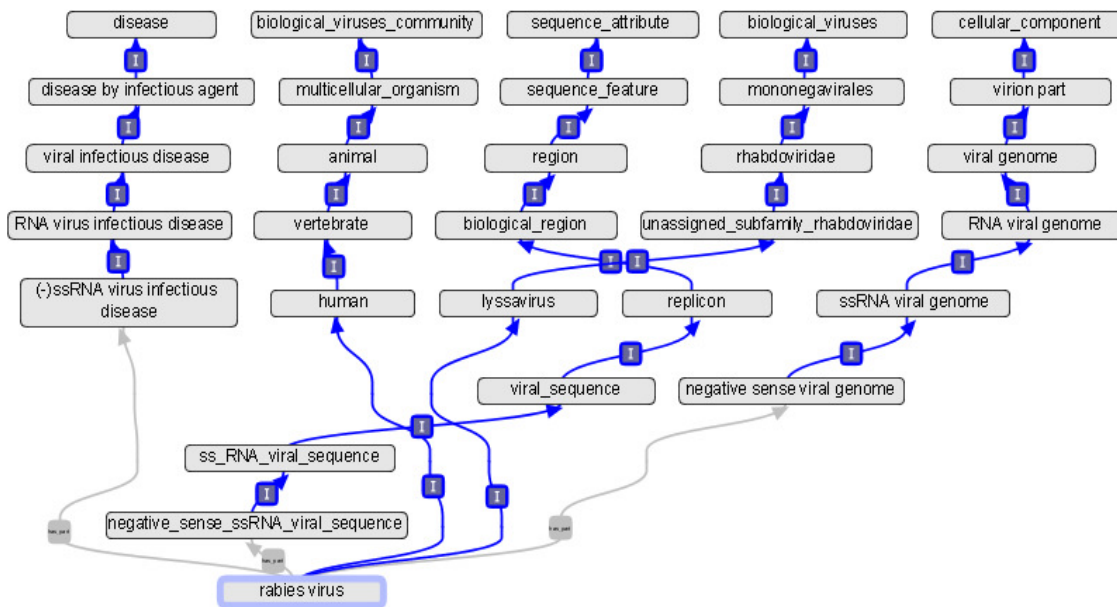**Block diagram of Animal Biological Viruses Ontology (ABVO)**

Figure-4
Graphical structure of Human Biological Viruses Ontology (HBVO)

Our proposed HBVO also follows the basic principles for developing ontology (Ontology Development 101) described by Noy and McGuinnes[26]. For developing HBVO we studied GO, DO/DOID, SO, RO/RNAO, and CHEBI. We developed HBVO as Directed

Acyclic Graph (DAG) by using OBO principles in OBO-Edit[1-2,27]. In HBVO we listed ~58 genera. HBVO terms includes approximately 226 terms (with assigned BVO identifies) and approximately 68 terms include from 5 others ontologies based on MIREOT principles[28].

To develop relationship between virus and nucleic acid, we used has_part relationship as shown in figure-4. Graphical representation, which shows the is_a ('I' is in small square box) relationship and has_part ('has_part' is in small rectangular box) relationship. has_part relationship has be introduced by GO and used in GO[15] and Protein Ontology[29]. It represents a part_whole relationship, as the viruses are the nucleic acid so we use "virus has_part nucleic acid", it means that virus necessarily (always) has nucleic acid as part; i.e., if virus exist then nucleic acid also exists as part of virus. If virus does not exist, nucleic acid may or may not exist. We used has_part relationship as: [Term]: id: BVO:0000706, name: rabies virus, is_a: BVO:0000451 ! lyssavirus, is_a: BVO:0007005 ! human, relationship: has_part DOID:0050503 ! (-)ssRNA virus infectious disease, relationship: has_part GO:0019026 ! negative sense viral genome, relationship: has_part SO:0001200 ! negative_sense_ssRNA_viral_sequence.

Conclusion

We have defined the Human Biological Viruses Ontology (HBVO), which provides a controlled vocabulary of information about viruses hosted by human. The ontology development is an iterative process; in this paper we present the major component HBVO of Biological Virus Ontology (BVO), which is the first objective in our research work. We have selected OBO-Edit tool for its development, as OBO now become unclaimed standard to define any biological ontology. This ontology may be the part of any digital health care networks[30] any where in the world.

Reference

1. Smith B., Ashburner M., Rosse C., Bard J., Bug W., Ceusters W., Goldberg L. J., Eilbeck K., Ireland A., Mungall C.J., et al., The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration, *Nat. Biotech*., **25**, **(2007)**

2. Day-Richter J., Harris M.A. and Haendel M., The Gene Ontology OBO-Edit Working Group and Lewis S., OBO-Edit—an ontology editor for biologists, *Bioinformatics*, **32**, 2198–2200 **(2007)**

3. Siddiqui M.S., Shaikh Z.A. and Memon A.R., Towards the development of human community ontology, *WCSE 2009*, Xiamen, China, **3**, 8-12 **(2009)**

4. Siddiqui M.S., Shaikh Z.A. and Memon A.R., Towards the development of community ontology, *IEEE INMIC 2008*, Bahria University, Karachi, Pakistan, 357-360 **(2008)**

5. Valdivia-Granda W. and Larson F., ORION-VIRCAT: A tool for mapping ICTV and NCBI taxonomies, *Database* (Oxford), **(2009)**

6. Carstens E. B., Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses, *Arch. Virol.*, **155**, 133–146 **(2009)**

7. Carstens E. B. and Ball L. A., Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses, *Arch. Virol.*, **154**, 1181–1188 **(2009)**

8. Altan-Bonnet, A., How RNA Viruses Copy Themselves: Hijack Cellular Enzyme to Create Viral Replication Factories on Cell Membranes, *Science Daily*, May 30, **(2010)**

9. Gruber T., Ontology. Entry in the Encyclopedia of Database Systems, Ling Liu and M. Tomer Ozsu (Eds.), *Springer-Verlag*., **(2008)**

10. Raffat S. K, Siddiqui M. S, Shaikh Z.A, Memon A.R., Ontology: A Scientific Classification Technique, *Sindh Uni. Res. J*., **44(2AB)**, 63-68 **(2012)**

11. Islam N., Siddiqui M. S. and Shaikh Z. A., TODE–A Dot Net Based Tool for Ontology Development and Editing, *ICCET 2010*, International Convention Centre of UESTC, Chengdu, China, **6**, 229-233 **(2010)**

12. Black J., Bio-ontology—fast and furious, *Nat. Biotech.*, **22**, **(2004)**

13. Smith B., Ceusters W., Klagges B., Kohler J., Kumar A., Lomax J., Mungall C., Neuhaus F., Rector A. L. and Rosse, C., Relation in biomedical ontologies, *Gen. Bio.*, **6**, **(2005)**

14. Hartung M., Kirsten T., Gross A. and Rahm E., OnEX: Exploring changes in life science ontologies, *BMC Bioinformatics*, **10**, **(2009)**

15. The Gene Ontology Consortium, The Gene Ontology in 2010, *Nucl. Acids Res.*, **38**, D331–D335 **(2010)**

16. The Gene Ontology Consortium, The Gene Ontology project in 2008, *Nucl. Acids Res.*, **36**, **(2008)**

17. Hartmann S., Kohler H. and Wang J., Ontology consolidation in bioinformatics, *APCCM 2010*, Brisbane, Australia, **(2010)**

18. Moore B., Fan G. and Eilbeck K., SOBA: sequence ontology bioinformatics analysis, *Nucl. Acids Res.*, **38**, W161–W164 **(2010)**

19. Eilbeck K., Lewis S. E., Mungall C. J., Yandell M., Stein L., Durbin R. and Ashburner M., The Sequence Ontology: A tool for the unification of genome annotations, *Gen.Bio.*, **6**, **(2005)**

20. Degtyarenoko K., Matos P., Ennis M., Hastings J., Zbinden M., McNaught A., Alcantara R., Darsow M., Guedj M. and Ashburner, M., ChEBI: a database and ontology for chemical entities of biological interest, *Nucl. Acid Res.*, **36**, D344–D350 **(2007)**

21. Bodenreider O. and Burgun A., Towards desiderata for an ontology of diseases for the annotation of biological datasets, *Int. Conf. on Bio. Onto.*, Buffalo, New York, USA, **(2009)**

22. Leontis N. B., Altman R. B., Berman H. M., Bernner S. E., Brown J. W., Engelke D. R., Harvey S. C., Holbrook S. R., Jossinet F., Lewis S. E., et al., The RNA Ontology Consortium: an open invitation to the RNA community, *RNA*, **12**, 533-541 **(2006)**

23. Batcheor C., Bittner T., Eilbeck K., Mungall C., Richardson J., Knight R., Stombaugh J., Zirbel C., Westhof E. and Leontis, N., The RNA Ontology (RNAO): An ontology for integrating RNA sequence and structure data, *Nat. Precedings: hdl:10101/npre.2009.3561.1*, **(2009)**

24. Raffat S. K, Siddiqui M. S, Shaikh Z.A, Memon A.R., Towards the development of Biological Viruses Community Ontology (BVCO), *J. of Comp.*, **3**, 125–129 **(2011)**

25. Soldatova L.N and King R.D., Are the current ontologies in biology good ontologies?, *Nat. Biotech.*, **23**, **(2005)**

26. Noy N. F. and McGuinness D. L., Ontology Development 101: A guide to creating your first ontology, Stanford Knowledge Systems Laboratory Technical Report, KSL-01-05, Stanford University, USA, **(2001)**

27. Alterovitz G., Xiang M., Hill D. P., Lomax J., Liu J., Cherkassky M., Dreyfuss J., Mungall C., Harris M. A., Dolan M. E., et al., Ontology engineering, *Nat. Biotech.*, **28**, **(2010)**

28. Courtot J. M., Gibson F., Lister A. L., Malone J., Schober D., Brinkman R. R. and Ruttenberg A., MIREOT: Minimum information to reference external ontology terms, *Int. Conf. on Bio. Onto.*, University at Buffalo, NY, USA, **(2009)**

29. Natale D.A., Arighi C. N., Barker W. C., Blake J., Chang T., Hu Z., Liu H., Smith, B. and Wu C. H., Framework for a Protein Ontology, BMC Bioinformatics, **8**, **(2007)**

30. Kalpa S., Health IT in Indian Healthcare System: A New Initiative, *Res.J.Recent Sci.,* **1(6)**, 83-86 **(2012)**