Semantic Grid for Biomedical Ontologies

Farhan Shafiq
Department of Computer Science & IT, Federal Urdu University, Karachi, Pakistan

Sheikh Kashif Raffat
Department of Computer Science & IT, Federal Urdu University, Karachi, Pakistan.

Muhammad Shahab Siddiqui
HIIT, Hamdard University, Karachi, Pakistan.

ABSTRACT
The biomedical ontologies contain the complex distributed heterogeneous data, to analyze and process this data is the big challenge for biomedical communities. The common goal of biomedical communities is to annotate this data. These problems generated a need to use the services of grid on semantic web. Semantic Grid is the integration of Grid with the Semantic web, which will play the vital role in future web. The semantic grid architecture provides semantic and knowledge support. In this paper we discuss two biomedical ontologies, Biological Viruses Community Ontology (BVCO) and the most mature Gene Ontology (GO).

General Terms
Biomedical dataset, Future Web.

Keywords
Biomedical Ontology, Grid Computing, Semantic Grid, Semantic Web.

1. INTRODUCTION
Biological sciences are facing the exponential growth of observational, experimental and theoretical data scattered in different laboratories and hospitals. These are distributed heterogeneous datasets because of their source (Imaging device, Sequencer etc.). These laboratories and hospitals are generally not capable to archive, process and analyze these terabyte of dataset. The bioinformatics research community is erg to analyze these distributed heterogeneous dataset in order to extract useful information and knowledge. But these resources which are generating vast amount of biomedical data are widely distributed, highly heterogeneous, may follows heterogeneous protocols and made by different venders so applying appropriate computational technique in this kind of heterogeneous environment is not an easy task. Now a day Grid Computing is playing an important role in obtaining, comparing and analyzing distributed heterogeneous scientific data.

Foster [1] defines the Grid concept as "the controlled and coordinated resource sharing and problem solving in dynamic, multi institutional virtual organizations". This sharing of resource, ranging from simple file transfer to complex and collaborative problem solving, is accomplished with in controlled and well-defined conditions and policies. The dynamic grouping of individuals, groups, or organization that defined the conditions and rules for sharing are called virtual organization (VO) [2].

The grid computing [3] resources include computing power, data storage, hardware instruments, on-demand software and applications. In this context, the real problems involved with resource sharing are resource discovery, event correlation, authentication, authorization and access mechanism. These problems become proportionately more complicated when the grid computing solution is introduced as a solution for utility computing, where industrial applications and resources become available as sharable.

According to Tom Gruber [4], "Ontology defines a set of representational primitives that are typically classes, attributes and relationships". Ontology can be viewed as a controlled vocabulary of well-defined terms with specified relationships between them, capable of interpretation by both human and computers.

Many tools are also available to develop ontology like OBO–Edit (mainly used for Biological Ontologies), Protégé (developed by the Stanford University, USA) and TODE [5] (developed by the National University of Computer and Emerging Sciences, Karachi, Pakistan).

Biomedical research community sequencing more and more genomes day by day and they highly needed processing and analyzing with appropriate algorithms. The integration of grid computing and ontology can play vital role in for biomedical communities [6, 7]. In this paper we are presenting an idea for exploration of huge amount of biomedical datasets by using semantic grids.

2. RELATED WORK
2.1 Grid Computing
Grid computing has the potential to drastically change enterprise computing, as we know today. This advancement of technology play great role in the field of biological sciences, earth science, physics, astronomy and mathematical research. The academia and the industry take great advantage of these significant changes in computing world. The main concept of grid computing is viewing computing as a utility. Grid computing enables resource sharing and dynamic allocation of distributed heterogeneous computational resources while minimizing the associated ownership and operating cost [8]. These resources may be geographically distributed over the network or Internet and follows heterogeneous protocols made by different vendors.

2.2 Semantic Web
"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 co-operation"[9]. It is the idea of having data on the Web [10] defined and linked in a way that it can be used for more effective discovery, automation, integration and reuse across various applications [11], data can be shared and processed by automated tools as well as by people. Tim Berners-Lee proposed layers structure of the Semantic Web as shown in figure 1.
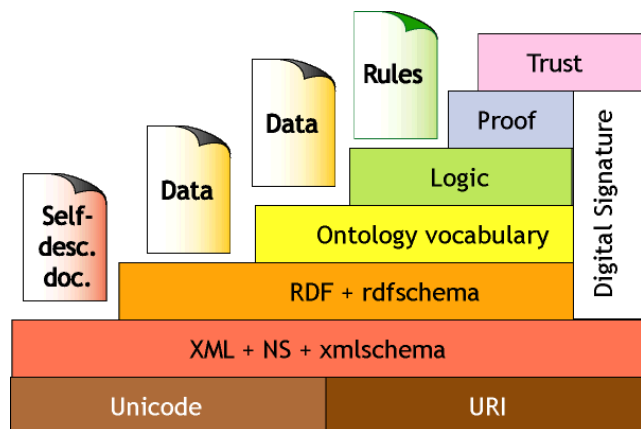
Fig 1: Conceptual layers of the Semantic

2.3 Semantic Grid

Semantic Grid [12] can be described as an extension of the grid in which information and services are given well defined meaning, better enabling computers and people to work in cooperation. It is characterized as an open system in which users, software components and computational resources (all owned by different stakeholders) come and go on a continual basis. It can be seen as a set of services that are offered by entities (which may be software agents). The services [13] are offered under contract, and can be accepted by any of a number of consumers in a marketplace.

According to C. Goble the Semantic Grid [12] is the combination of Grid, Semantic Web and Web Services, which can be seen in figure 2.
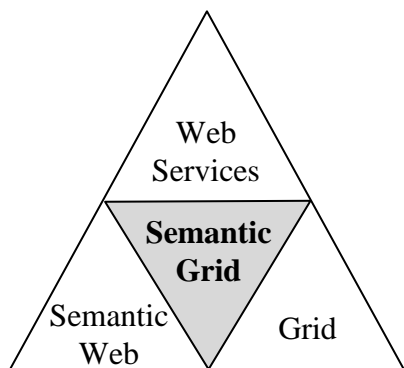

Fig 2: Conceptual Structure of the Semantic Grid

3. SEMANTIC GRID FOR BIOMEDICAL ONTOLOGY

Grid computing provides virtual organization [2] of knowledge and resources that independent of the physical location, it refers to distributed computing infrastructure supporting wide-area networks in world-wide. On the other hand, ontologies provide the structured and controlled vocabulary. Ontological contents are the parts of semantic web [14, 15], in future every thing will be classified in terms of ontological components for semantic web.

Some of the questions raised that:

1. Why we need to use Grid Computing for Biomedical Ontologies?

2. What kind of information should grid nodes contain?

3. Should biomedical data be classified in ontologies?

4. Are there existing ontologies comfortable for Grid?

5. What is the domain of Semantic Grid?

The main challenge facing by biomedical communities is to handle, manipulate and process the huge amount of biomedical data that is growing day by day exponentially. If we link grid to traditional web like "Sealife Project" [16] then the browser must be able to handle ambiguity, which will not happened in ontological data. A grid offers to store these data on different grid nodes, rather then concentrated on larger computing center, where as ontologies help to manage these data in a controlled and manageable way.

Many biomedical ontologies exist in different domains of biology like Gene Ontology (GO), Sequence Ontology (SO), Chemical Entity of Biological Interest (ChEBI), RNA Ontology (RNAO) and Biological Viruses Community Ontology (BVCO), all of these ontologies are defined by using the principles of Open Biomedical Ontology (OBO) [17, 18] and some of these also followed the steps of Ontology Development 101 by Noy and Mcguinness [19]. The ontologies under the umbrella of OBO are comfortable for grid computing as middleware.

We have to integrate the semantic web and grid computing with its many services. Semantic grid allows to use grid services for massive and complex biomedical data more easily accessible. The ontological contents are on the nodes of grid that link to make the semantic grid. Combining a grid services with biomedical ontologies are more beneficial to several aspect to analyze and investigate the complex data. Some of existing e-science project are myGrid, Comb-e-Chem, Geodise and CoAKTing. The table 1 shows the effectiveness of semantic grid over the semantic web and classical grid.

Table 1. Semantic grid over Web, Grid and Semantic Web

Parameters	Classic Web	Classic Grid	Semantic Web	Semantic Grid
Interoperability	Low	Low	High	High
Computation	Low	High	Low	High
Operation Cost	High	Medium	Medium	Low
Performance	Low	Medium	Medium	High
Productivity	Low	High	Medium	High
Efficiency	Low	Medium	Medium	High

3.1 Semantic Grid for Biological Viruses Community Ontology (BVCO)

The Biological Viruses Community Ontology (BVCO) [20] is an ontology that covers the viruses hosted by different species like vertebrate, invertebrate, plant, and unicellular organisms. The impact of biological viruses on human community are drastic, so using this ontology with semantic grid is more useful for end user and as well as virologist. The BVCO is developed by using the principles of Open Biomedical Ontology (OBO) and used the some of existing ontologies to relate nucleic acid.

The BVCO comprises of Core Ontology and Species Ontology, which can be seen in Figure 3. The core ontology contains the information about the viruses listed by the International Committee on Taxonomy of Viruses. Where as the species ontology contains the information about the different communities [21] (species) vertebrate, invertebrate, plant and unicellular organism. The 1,740,330 are the discovered species of earth and still are discovering. This starvation can be efficiently handled by semantic grid. Grid computing provides a suitable middleware infrastructure to enable collaboration and resource sharing.
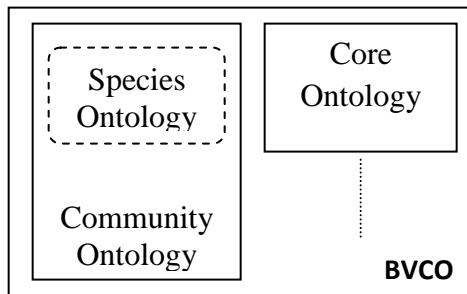

Fig. 3: Block diagram of Biological Viruses Community Ontology (BVCO) presents its major components, which includes Core Ontology and Species Ontology.

3.2 Semantic Grid for Gene Ontology (GO)

The Gene Ontology (GO) [22, 23] is the most mature ontology in bioinformatics, aims to standardize the representation of gene in all the species. The GO consortium is a body that legally responsible for the representation of gene and gene product attributes. GO contains three separate ontologies; molecular functions, cellular component and biological processes.

The GO has reached a substantial size, containing approximately

170,690 terms, with the nodes and leaves, it is linked to a database of more than 120,000 gene products from almost 20 experimental organisms including vertebrates, invertebrates, plants, fungi, bacteria and viruses in which the proteins are tagged with GO IDs. The huge amount of data of gene needed an ontology that is GO, but this also required the services of grid to annotate this data. The middleware of grid is the best option to process the massive information of gene across the biomedical communities. The shared ontological knowledge of gene will be easier for biomedical communities to research in the same direction, the integration of grid with ontology will

share the resources, increases abilities of work with each others and save the computational time.

4. CONCLUSION

There are many reasons to use grid computing with semantic web especially for biomedical data. The some of the reasons are:

1. Handle massive and complex data produced in different location.

2. Use common tools and standard principles to work on theses data in same direction.

3. Provide secure and efficient access of data to biomedical communities.

4. Avoid redundancy in same experiments.

5. Same future direction of the work.

The integration of semantic web and grid computing result a dynamic and intelligent information system in future. We can say that semantic grid is the 'next generation' of the web.

5. REFERENCES

[1] I. Foster and C. Kesselman, "The Grid: Blueprint for a New Computing Infrastructure", Morgan Kaufmann, 1998.

[2] I. Foster, C. Kesselman and S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", International Journal of Supercomputer Applications, **15**(3), 2001.

[3] J. Joseph and C. Fellenstein, "Grid computing", PHI, PTR, Page **48**, Jan 2004,

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

[5] N. Islam, M.S. Siddiqui and Z.A. Shaikh, "TODE–A Dot Net Based Tool for Ontology Development and Editing", Proceedings of 2010 2nd International Conference on Computer Engineering and Technology (ICCET 2010), International Convention Centre of UESTC, Chengdu, China, **6**, Pages 229-233, April 16 - 18, 2010.

[6] G. Napolitano, A. Gonzáles Beltrán, C. Fox, A. Marshall, A. Finkelstein and P. McCarron, "Biomedical Ontologies and Grid Computing as New Resources for Cancer Registries", Proceedings of the International Conference on Health Informatics 2009, Porto, Portugal.

[7] M. Hadzic and E. Chang, "Grid Services Complemented by Domain Ontology Supporting Biomedical Community", Proceedings of SAG'2004. pp.86~98.

[8] J. Khan, F. Shafiq and M. K. Khan, "A Statistical Approach for the Assessment of QOS and Performance in Grid Computing Environment", International Journal of Computer Applications, **19**(3), 2011.

[9] M. Geldof, "The Semantic Grid: will Semantic Web and Grid go hand in hand?", European Commission DG Information Society Unit "Grid technologies" June 2004.

[10] S. Chakraborty and R. Mittal, "Semantic Web Technology: A Strategic Approach to Intelligence", International Journal of Computer Applications, **5**(9), 2010.

[11] H. Stork, "Webs, Grids and Knowledge Spaces-Programmes, Projects and Prospects", Journal of Universal Computer Science, **8**(9), 2002.

[12] C. Goble, D. Kuo, I. Kotsiopoulos, O. Corcho, P. Alper, and S. Bechhofer, "Towards a Semantic Grid Architecture," in CoreGrid workshop on Knowledge and Data Management in Grids, 2005.

[13] S. Adcock, "How does the grid extend the internet, and what is the future vision for this development?", 2000 http://mms.ecs.soton.ac.uk/papers/27.pdf

[14] A. Maedche, and S. Staab, "Ontology Learning for the Semantic Web," The Semantic Web, IEEE Intelligent Systems, 72-79, 2001.

[15] J. Hendler, "Agents and the Semantic Web," The Semantic Web, IEEE Intelligent Systems, 30-37, 2001.

[16] "A Semantic Grid Browser for the Life Sciences Applied to the Study of Infectious Diseases", http://www.nettab.org/2006/Sessions/Burger.pdf

[17] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall et al. "The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration," Nature Biotechnology **25**, 2007, doi:10.1038/nbt1346.

[18] J. Day-Richter, M.A. Harris, M. Haendel, The Gene Ontology OBO-Edit Working Group and S. Lewis, "OBO-Edit—an ontology editor for biologists," Bioinformatics **32**, Pages 2198–2200, 2007, doi:10.1093.

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

[20] S.K. Raffat, M.S. Siddiqui, Z.A. Shaikh and A.R. Memon, "Towards the development of Biological Viruses Community Ontology (BVCO)", Journal of Computing, **3**(4), 2011.

[21] M.S. Siddiqui, Z.A. Shaikh and A.R. Memon, "Towards the development of community ontology," Proceedings of the 12th IEEE International Multi-topic Conference (INMIC) 2008, Bahria University, Karachi, Pakistan, Pages 357-360, December 23-24.

[22] The Gene Ontology Consortium, "The Gene Ontology in 2010," Nucleic Acids Research **38**, 2010, doi: 10.1093/nar/gkp1018, D331–D335.

[23] The Gene Ontology Consortium, "The Gene Ontology project in 2008," Nucleic Acids Research **36**, 2008, doi: 10.1093/nar/gkm883.