



Interoperable Bioinformatics Web services a Review

P.M.Durai Raj Vincent*

Assistant Professor

School of Information Technology and Engineering

VIT University, India

pmvincent@vit.ac.in

E.Sathiyamoorthy

Associate Professor

School of Information Technology and Engineering

VIT University, India

esathiyamoorthy@vit.ac.in

Abstract: Bio Informatics is a rapidly growing field with large amount of molecular biological data. Over past few years researchers are trying to build service with web-based interfaces. Incompatibility between interfaces and data types prevents users to make use of available services to some extent in web services environment. There are number of service providers and different standards are there in the field of bioinformatics. Any have, it is hard to utilize all those by a non expert developer also it is very difficult to integrate all those services to have a proper workflows. In this paper literature survey performed to identify different problems involved in web services for bioinformatics and it discusses the interoperability problem and how web services are trying to resolve this.

Keywords: Bioinformatics; interoperability; BioMoby; UnitProf; webservices

I. INTRODUCTION

In Recent years web services had great interests over various applications. It could provide better solution to application integration with the help of Web Services Definition Language, Simple Object Access Protocol and Universal Description, Definition and Language. It is becoming very popular in web applications. Web service is an application programming interface and these applications can be accessed through hypertext transfer protocol and it will be executed over a remote system which is hosting the requested services. Web services' applications spread over different disciplines. Among those different applications bioinformatics is one of the important one where it plays vital role.

Bioinformatics is the application of mathematics and computer science to the field of molecular biology. It has developed number of tools to make sense of the rapidly growing pool of molecular biological data.

These biological systems are complex and to understand it, we should link many data sets and we have to use more than one tool. That is why there is a need of experimentation with several techniques to try to integrate these data sets and tools.

II. EVOLUTION OF WEB SERVICES IN BIO INFORMATICS

In the recent past building different services with web based interfaces becomes one of the popular way of sharing the data and tools which are results from many biological projects. SOAP and REST based interfaces have been developed for bioinformatics applications which allows an application running on one computer to use algorithms where data and computing resources on servers will be somewhere else. The important advantage from this is there is no need to deal with software and database maintenance overheads. Basic bioinformatics services are classified as Sequence Search Services, Multiple Sequence

Alignment and Biological Sequence Analysis. In this chapter we are going to see different key things related to web services application in bio informatics. Systems that have been designed to integrate biological data and tools can be divided in two groups:

systems based on a centralisation or data warehousing strategy and systems based on a federated or distributed strategy.

III. INTEROPERABILITY PROBLEM

Always in the biological complex scenario we are having different types of data so there is a need of multiple tools to integrate different types of data. The interoperability problem exists because of the lack of standards between both data and communication between inter application. Systems developed for integrating both data and tools can be divided into two. One is centralized system and the other one is distributed one. Centralized systems can be developed with the help of data base management systems or using indexing system. Either we can use Mysql or oracle for this application even earlier mentioned sequential search services can also be used. If we are having all of the tools as well the data in our local system itself, then we are very comfortable with our research. Because the main factor is the speed because our tools can fetch those data in a faster way from our hard drive than the source somewhere through internet. But with all these advantages if we look after the disadvantages of such system is the maintenance problem. It requires lot of time and money to maintain such a system because those databases and tools used to change frequently due to rapid data growth. The alternative solution for this is to develop a system to access both data and tools over remote access technique.

In this environment the client has to fetch data from one place and it has to reform it for the specific tool and send it to a particular service at another site and find its result and then again reformat the result and so on. Basically web

servers providing such facility to access tools through web browser.

For solving this interoperability we have to develop different grids which can distribute over many servers and our clients have to use special software's to work with these services. Usually these grids will act as a Application programming Interface (API). Such interfaces becoming popular in bioinformatics field. Next to pipette and reading accessing web browser is probably the best option for biologists also it is platform independent. For these XML-based standards has been developed which will enable both data and services and the communication between these two. Some of the standards are

1. Simple Object Access Protocol(SOAP)

It's a protocol to exchange information in a decentralized environment. Here xml is used to create an extensible communicative framework which can exchange data over underlying protocols.

2. Web Services Description Language (WSDL)

It is having xml format to describe network services as set of end points which operates messages. Regardless of communication protocols and message formats this will support like SOAP.

3. Universal Description, Discovery and Integration Protocol(UDDI)

It's a standard which enables dynamic property to all those applications to find its web services.

IV. INTEROPERABLE WEB SERVICES

Several projects have addressed this problem such as Grid[1] and BioMoby[2] which guides discovering bioinformatics web resources from different services. The BioMoby project was started in 2001, MOBY-Services (MOBY-S) and a sub branch known as Simple Semantic Web Architecture and Protocol, however, according to the BioMoby web page[3] these two branches are poised to be combined in the future. In MOBY services we have three different ontologies. Among those first one is service ontology where in

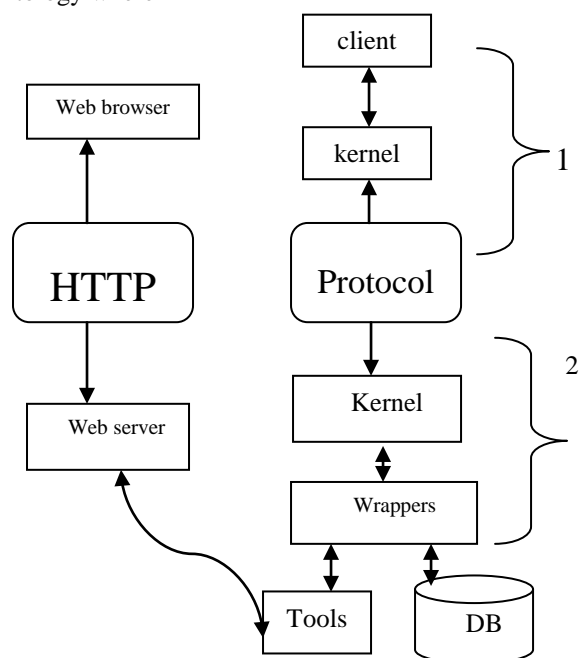


Fig 1 shows the typical web service system in bioinformatics environment. In the above diagram 1 is client side and 2 is server side. tree structure service will be treated as root and alignment as sub types. The main purpose of this is to discover a particular method for a given task. Next one is Name space ontology in this we can have a list of different namespaces to describe data to be produced by web service. It also name different identifiers which will be in a reliable way which overcomes inconsistency in naming [4]. The last ontology is object ontology which is similar to the first one in tree structure where in this the root will be considered as object and its children represents its relationship. This is having a central registry called Moby where all service definitions described. When some new actor wants to register with this BioMoby it has to register their objects, services and its names. The given figure shows the ontology working principles.

V. CASE STUDY

In BioMoby they have integrated three databases like BioModels Databases [5], UniprotKB[6], RCSB[7] and it could be expanded to work with others. Moreover it is platform independent and can be used from any WSDL enabled language and

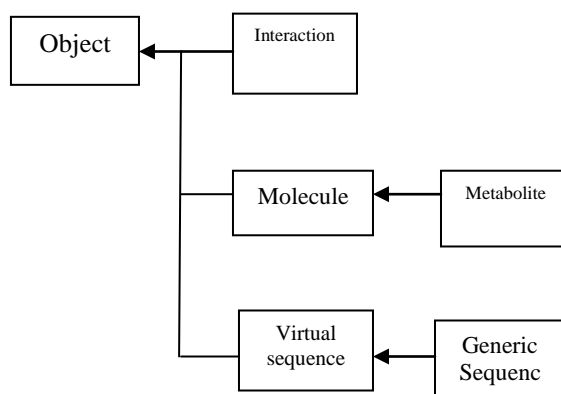


Figure.2 BioMoby

platform. Some of the core methods in this are `getNumberOfSpecies(modelID,n)`, `getPDB(pdb Reference)`, `getNthSpecies(modelID,n)`. For example if we use UniProtKB references from the server to obtain the corresponding xml files sometimes PDB references will be extracted and the same will be displayed to user. So, here PDB becomes the identifying token. In the given figure we can see an example of Taverna a modern web service with Bio application. In the diagram we can see its workflow. In this process list of ID's will be collected for its respective species and for every species we have to determine its UniProt references. Then this software repeatedly calls this method to list out all ID's. The final list will have the UniProt references. One another software called VisTrails which also having the same kind of services but it is more complicated than the previous one.

V. ISSUES

In this section I am summarizing the different techniques which are already discussed, the different issues to be considered for the new web service which can use multiple services for the complex bio scenario. It can be divided into different issues.

The first issue is semantics where the researcher have find the services on its own for some of its particular scientific function or it has find its own service. This automatic finding is called as semantics. For this we should register our service frameworks like MOBY. It will allow its users to find it when it fits at the semantic level.

The second issue is chainability which is related to the previous one. It is able to list for a given type of data. When in some cases it cannot work exactly but it will try to work with subset of the data or it will try to reformat it in other possible way. For example in BioMoby semantics of web service, input output data is given it will understand correct chainability. The reformatting also required means we can use shim services but it should be automatically detected by the framework. In [8], D. Koop et al. presents a method of suggesting suitable services by using predictions.

The next one is granularity always it is good to divide a big problem into small problems. In that view a large problem will be divided into large number of well defined mini problems. Then we can write a function and method each small problem. It may not be a possible solution sometimes because of some complex task of putting every small part together. However, if the blocks which are

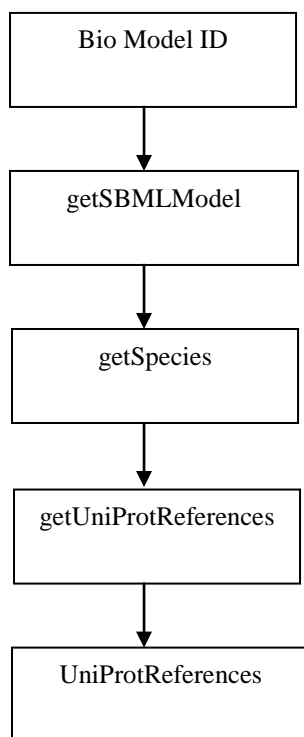


Figure.3 One example service with UniProt

used for building the complete setup is simple then it will be a stronger one at the same time if the function is simple then it will be easier to mention it's semantics which enables good interoperability.

The fourth one is data preparation which plays crucial role when designing a web service. Important thing

in data preparation is the data format for how to represent it when the data items are complex. When we pass the complex objects between different web services it is very important. In this we have chosen the existing SBML and UniProt[3][4] which is using xml standards for bioinformatics. It is a very transparent method to any user because it is already a familiar standard.

The final one is data passing which is doing the finishing work in every service. Selection of data is important because all the time we cannot pass large amount of data so we should avoid such kind of passing large amount of data back and forth when there is no need. If some data is available in the domain which was not generated by the user the particular service should take the responsibility over it. In every case the client can see only the results they intend to see.

VI. CONCLUSION

In this article I have tried to give an overall view about web services and its application in the field of bioinformatics and especially tried to present a literature survey in this application. There is a need of creating complex workflows in this field since various service providers are there in order to integrate all these for any complex scenario this is required. Even they face severe interoperability problems the reason behind this is to find the appropriate services. Always single web service is simple and easy to describe and the methods we are using can semantically well defined. BioMoby and myGrid are such examples which are discussed. This type is usually help the users in discovering the service.

VII. REFERENCES

- [1] K. Wolstencroft, P. Alper, D. Hull, C. Wroe, P. Lord, R. Stevens, and C. Goble, "The mygrid ontology: bioinformatics service discovery," *International Journal of Bioinformatics Research and Applications*, vol. 3, no. 3, pp. 303–325, 2007.
- [2] T. B. Consortium, "Interoperability with moby 1.0-it's better than sharing your toothbrush!" *Briefings in Bioinformatics*, vol. 9, no. 3, pp. 220–231, 2009.
- [3] "Biomoby >>semantic moby," Accessed october 20th 2009. [Online]. Available: <http://biomoby.open-bio.org/index.php/semanticmoby/>
- [4] T. G. O. Consortium, "Go database abbreviations," Accessed october 20th 2009. [Online]. Available: <http://geneontology.org/cgi-bin/xrefs.cgi>
- [5] EMBL-EBI, "Biomodels database," Accessed october 20th 2009. [Online]. Available: <http://www.ebi.ac.uk/biomodels-main/>
- [6] U. Consortium, "Uniprotkb," Accessed october 20th 2009. [Online]. Available: <http://www.uniprot.org/help/uniprotkb>
- [7] RCSB, "Rcsb protein data bank," Accessed october 20th 2009. [Online]. Available: <http://www.rcsb.org/pdb/home/home.do>
- [8] D. Koop, C. E. Scheidegger, S. P. Callahan, J. Friere, and C. T. Silva, "Viscomplete: Automating suggestions for visualization pipelines," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1691–1698, 2008.

- [9] C. E.z Baxevanis, A. D. (2003), 'The Molecular Biology Database Collection: 2003 update', *Nucleic Acids Res.*, Vol. 31, pp. 1–12.d .
- [10] Doelz, R. (1994), 'Hierarchical Access System for Sequence Libraries in Europe (HASSLE): A tool to access sequence databases remotely', *Comput. Appl. Biosci.*, Vol. 10, pp. 31–34.
- [11] T. E. Foundation, "Web tools platform (wtp) project," Accessed october 20th 2009. [Online]. Available: <http://www.eclipse.org/webtools/>
- [12] RCSB, "Rcsb protein data bank," Accessed october 20th 2009. [Online]. Available: <http://www.rcsb.org/pdb/home/home.do>
- [13] EMBL-EBI, "Biomodels database," Accessed october 20th 2009. [Online]. Available: <http://www.ebi.ac.uk/biomodels-main/>
- [14] P. W. Lord, S. Bechhofer, M. D. Wilkinson, G. Schiltz, D. Gessler, D. Hull, C. A. Goble, and L. Stein, "Applying semantic web services to bioinformatics: Experiences gained, lessons learnt," *International Semantic Web Conference*, pp. 350–364, 2004.
- [15] L. Strömbäck, V. Jakoniene, H. Tan, and P. Lambrix, "Representing, storing and accessing molecular interaction data: a review of models and tools," *Briefings in Bioinformatics*, vol. 7, no. 4, pp. 331–338, 2006.
- [16] L. Strömbäck, D. Hall, and P. Lambrix, "A review of standards for data exchange within systems biology," *Proteomics*, vol. 7, no. 6, pp. 857– 867, 2007.