Full Length Research Paper

Development of a prototype hybrid-grid-based computing framework for accessing bioinformatics databases and resources

Oluwagbemi, Olugbenga Oluseun

[1]Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ogun State, Nigeria.
[2]Rochester Institute of Technology, 28 Lomb Memorial Drive, Rochester NY 14623, New York, USA. E-mail: gbemiseun@yahoo.com, ooluwagbemi@acm.org. Tel: +2348066533717.

Bioinformatics has already entered into its post genomic era, where research has advanced from data collection to data analysis using advanced computational and analytical tools. Due to current high demands on bioinformatics data, the various shortcomings in the computing infrastructure associated with the handling and processing of such biological data has constituted a great challenge. In this paper, effort was made at developing and describing a prototype of a new-generation computing framework known as the hybrid-grid-based computing framework for bioinformatics (HGCFB), with the aim of maintaining, sharing, discovering, and expanding bioinformatics knowledge in geographically distributed environments. This paper proposed the system architecture of a prototype hybrid-grid-based computing framework for bioinformatics (HGCFB), and described its corresponding functionalities. Attempts were also made at implementing some aspects of this framework with an event driven programming language. This framework will be very useful in facilitating the effective, efficient use and management of bioinformatics databases and resources.

Key words: Bioinformatics, biological databases and resources, computing infrastructure, hybrid -grid-based framework.

INTRODUCTION

The emergence of bioinformatics as a novel field in the last few decades has brought about a major revolution, coupled with great problems and opportunities for researchers in other fields such as computer science, mathematics, public health, biomedicine, biotechnology, agriculture and the pharmaceutical industry in the history of biological research.

The production of genomic data is currently being executed in large volumes and at increased rates by series of experimentations and biological analyses (Louie et al., 2007; Huttenhower and Hofmann, 2010; Schadt et al., 2010; Bell et al., 2011; Schmidberger et al., 2011; Swedlow et al., 2011). Thus, numerous experimental techniques are responsible for producing these new types of data. These techniques are capable of generating data from the scientific experimental analysis of cells, tissues, organisms and even populations (Walter et al., 2010).

The major force behind the expansion of computational biology research is the emergence of novel, reliable, and efficient experimental methods, among which are high-thoroughput sequencing, which has led to an unprecedented growth (Alberts, 2002; Chance et al., 2004).

Advancement in bioinformatics research has gradually shifted from data gathering and accumulation to analysis and implementation of data. The correct analysis and interpretation of this large volume of data depends on the ability to adequately handle and analyze these data (Morgan and Sonquist, 1963). One of the daunting tasks in bioinformatics is managing the vast amount of genomic data generated from large scale experiments (Barrett et

al., 2007; Kann, 2009). The challenges enormous biological data presents are in different levels of variations and complexities (Louie et al., 2007); this constitutes some of the challenges of our generation. The problems range from difficulty associated with understanding the human genome, the detailed functions of gene encoding proteins, and sourcing for useful information for drug design among others. These problems are intensive (Zomaya, 2005). Consumers search, identify, and make use of archived data generated by many biological and bioinformatics laboratories. Thus, bioinformatics is an essential field that requires the use of Grid computing for a wide variety of distributed applications (Manjula and Raju, 2010). Presently, the current types of grid computing infrastructures are not adequate to meet all the needs of bioinformatics researchers worldwide. The main reasons for their inabilities are:

1. They were not developed taking into consideration all the necessary requirements for bioinformatics research.
2. There are no extensive and comprehensive formal biological frameworks for the grid computing and its application to bioinformatics and biological research domain.

Disparities between bioinformatics and non-biological data were identified. In biology, data have great effect on the system in which they are integral part of. There is therefore, the need to identify features of data and corresponding information about relationships existing within such data in a biological system context (Ahsan and Shah, 2008; Bornberg and Paton, 2002).

FEATURES OF BIOINFORMATICS DATA, DATABASES AND RESOURCES

The features of bioinformatics data are as follows:

1. There are close associations, similarities, and functional requirements among bioinformatics data, it is essential to analyze and study such data simultaneously (Ahsan, 2001).
2. There exist heterogeneity in the structures of homogeneous bioinformatics data in varying environment (Baldi and Brunak, 2001; Shah, 2009).
3. Bioinformatics data are semi-structured (Alberts, 2002; Lord et al., 2004).
4. Environment plays a major role in determining the behavioral variability of objects within a biological system (Alberts, 2002).
5. There exists broad usage of bioinformatics data based on varying research perceptions.
6. Bioinformatics and biological data are evolutionary in nature (Gentleman et al., 2004; Andreas et al., 2005).
7. Data curation, misinterpretation and experimentation can lead to erroneous results in a biological system (Rhee et al., 2006; Shah, 2009).

8. Due to the impreciseness and incompleteness of bioinformatics and biological data, it is often a challenge determining their functional specifications (Ahsan and Shah, 2008; Alcalde, 2011).
9. Biological and bioinformatics data have iterative characteristics (Shah, 2009).

As a result of the characteristics listed in the previously, a single grid computing infrastructure may not be sufficient to adequately handle the requirements of bioinformatics data, databases and resources. Thus, it is necessary to develop a prototype of a hybrid-grid computing framework, which if fully implemented, will greatly facilitate solving most of the problems associated with bioinformatics, thus ensuring effective, efficient and prudent management of bioinformatics data, databases and resources.

Types of grids in existence

The following types of grids form the most prominent types currently in existence. They are:

1. Autonomic grids
2. Semantic grids
3. Utility grids
4. Cognitive grids

Autonomic grids models and related technologies have been designed, developed and deployed in some aspects of bioinformatics research over the years (Dai et al., 2006; Rahman et al., 2011; Deb et al., 2011; Folino and Mastroianni, 2011; Pandey et al., 2012; Visan et al., 2011; Hunter et al., 2008).

Semantic grids represent the integration of grids with the semantic web. Similarly, semantic grids, related frameworks and technologies have been developed and applied to bioinformatics and related fields over the years (Shafiq et al., 2011; Ramirez et al., 2011; Xu et al., 2011; Hadzic and Chang, 2004; Geldof, 2004; Chakraborty and Mittal, 2010; Goble et al., 2005; Hendler, 2001; Smith et al., 2007; Napolitano et al., 2009; Vega-Gorgojo et al., 2006; de Roure et al., 2003).

Utility grids, cognitive grids and related technologies have been developed and also applied to computational and biological related researches over the years (Tan and Hunter, 2004; Jiang et al., 2011; Niu et al., 2007; Foran et al., 2011; Castellano et al., 2011; Estrada et al., 2009).

None of these grid infrastructures can single-handedly meet all the demands of diverse bioinformatics databases. As a result, the purpose of this research is to propose and partly develop a prototype of a hybrid-grid-based computing framework for effectively accessing, managing bioinformatics databases and possibly sharing knowledge among bioinformatics' researchers among different research clusters.

Furthermore, part of this newly proposed framework will

be implemented by using an event-driven programming language to show how it will perform in the real-life hybrid-grid computing infrastructure. This architecture will be very useful in accessing and analyzing the vast amount of data that usually characterize bioinformatics databases, and so help to access, process and disseminate them to the appropriate quarters efficiently.

MATERIALS AND METHODS

Many grid frameworks, applications, systems and infrastructures, have been developed and deployed over the years. These have found relevance in bioinformatics research and other health-related sectors. Teo et al. (2004) proposed a system for developing and deploying large-scale bioinformatics-grid. In their study, they proposed a grid-programming toolkit for bioinformatics research.

In bioinformatics research, modeling and simulation of cellular processes and important pathways in humans, organisms, plants and animals play an prominent role towards novel discoveries. The first grid-enabled tool for modeling and simulating cellular processes (Grid-Cellware) was developed by Dhar et al. (2004). Konagaya (2006) presented a review of the trends in life science grid, stating the importance of data and knowledge grid to the advancement of bioinformatics research. Shabab et al. (2005) developed a grid portal interface for interacting with and monitoring high-throughput proteomic annotations.

Grid computing has found expression in bioinformatics research especially in the area of driving data-intensive genetic research (Andrade et al., 2007). Other studies applied the use of web services to integrate heterogeneous bioinformatics applications (de Knikker et al., 2004; Martín-Requena et al., 2009; Ramirez et al., 2010). Stankovski et al. (2007) developed a data mining grid system that is scalable, efficient, flexible, extensible and easy to use.

In a recent study, the adoption of a web portal enabled by a robot certificate to enable bioinformaticians to perform large-scale Bayesian Phylogenetic analyses on the grid was implemented (Barbera et al., 2011). A computational grid framework for immunological applications that enables scientists to surf through a single interface local, national and international resource was recently developed by a group of researchers (Halling-Brown et al., 2009). In another study, a simpler grid framework was developed for biological sequence alignment in bioinformatics on a computational grid (YarKhan and Dongarra, 2005). Pandey et al. (2011) presented a survey of data intensive bioinformatics applications on a variety of parallel computer architecture for accelerating the processing of large volume of biological data set.

Other bioinformatics and health-related grid computing researches include: development of a grid-based solution for executing blast on computational grids (Yang et al., 2009; Tang et al., 2005; Manjula and Raju, 2010; Jiang et al., 2011; Mareuil et al., 2011; Villamizar et al., 2011; Taha and Elmasri, 2011; Mirto et al., 2009; Daramola et al., 2008; Oluwagbemi, 2008).

However, in our work, we are proposing a prototype of a hybrid-grid computing framework which will be useful for accessing bioinformatics databases and resources for computational biology and bioinformatics researchers. Attempt was made at implementing the semantic web application aspect of this prototype.

The newly proposed framework (HGCFB)

The prototype of this new framework was proposed based on insight gained from studies on the integration of the four types of grids earlier mentioned (Figure 1).

The proposed hybrid-grid framework for bioinformatics (HGCFB)

The autonomic concept was inspired by knowledge gained from the human body's autonomic nervous system (Hariri et al., 2006). Incorporating this concept into the hybrid-grid framework has a good tendency of enhancing the performance.

The newly proposed architecture of a prototype hybrid-grid-based computing framework for bioinformatics (HGCFB), as shown in Figure 2, is as described below:

(i) Hybrid-grid fabric layer: This layer provides a platform for managing local bioinformatics grid resources. Such resources include different operating systems, servers, clusters etc.
(ii) Middleware layer: This layer provides storage access, programming frameworks and security.
Grid applications layer: This layer enables end-users to utilize bioinformatics grid services.
(iii) Integrated layer: This layer consists of an integrated layer of the collaborative functionalities of utility, semantic, cognitive and autonomic grid facilities to help constitute a hybrid-grid framework for bioinformatics research.

Implementation of a bioinformatics hybrid-grid web application and interface

Implementation was done with the Visual Basic 6 programming language within the Visual Studio environment. This implementation was achieved by coding in visual basic 6.0. This was implemented in visual basic programming environment- an event driven programming language. This graphical user interface describes the basic control buttons for navigating different bioinformatics databases. It also has a web address space box.

RESULTS

This was implemented in visual basic programming environment- an event driven programming language. This graphical user interface describes how the hybrid-grid web application tool navigates both the micro array databases of the EMBL-EBL bioinformatics and the Protein database for EBI (European Bioinformatics Institute).

DISCUSSION

The proposed framework introduces the possibility of having a hybrid-grid for bioinformatics research which has the tendency to integrate the capabilities of semantic web applications, autonomic grids, cognitive functionalities and utility grids. These combined functionalities has the advantage of enhancing the efficiency of conducting bioinformatics research as it applies to databases and other related bioinformatics resources. This proposed hybrid-grid framework if fully implemented will help promote research in bioinformatics and computational biology. The positive impact of the preliminary results obtained in this research as it relates
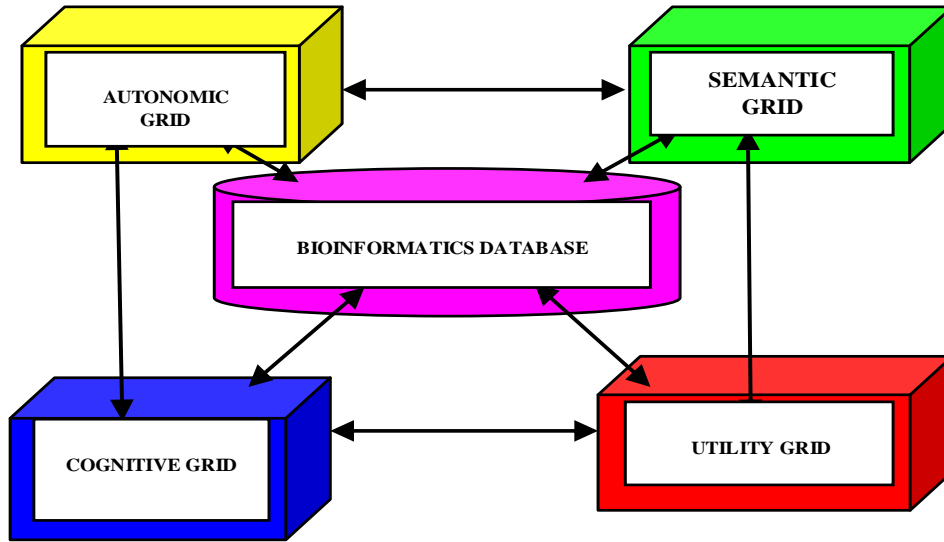
Figure 1. Interconnectivity and collaboration of different grid platforms to form the newly proposed hybrid-grid framework for bioinformatics databases and resources.
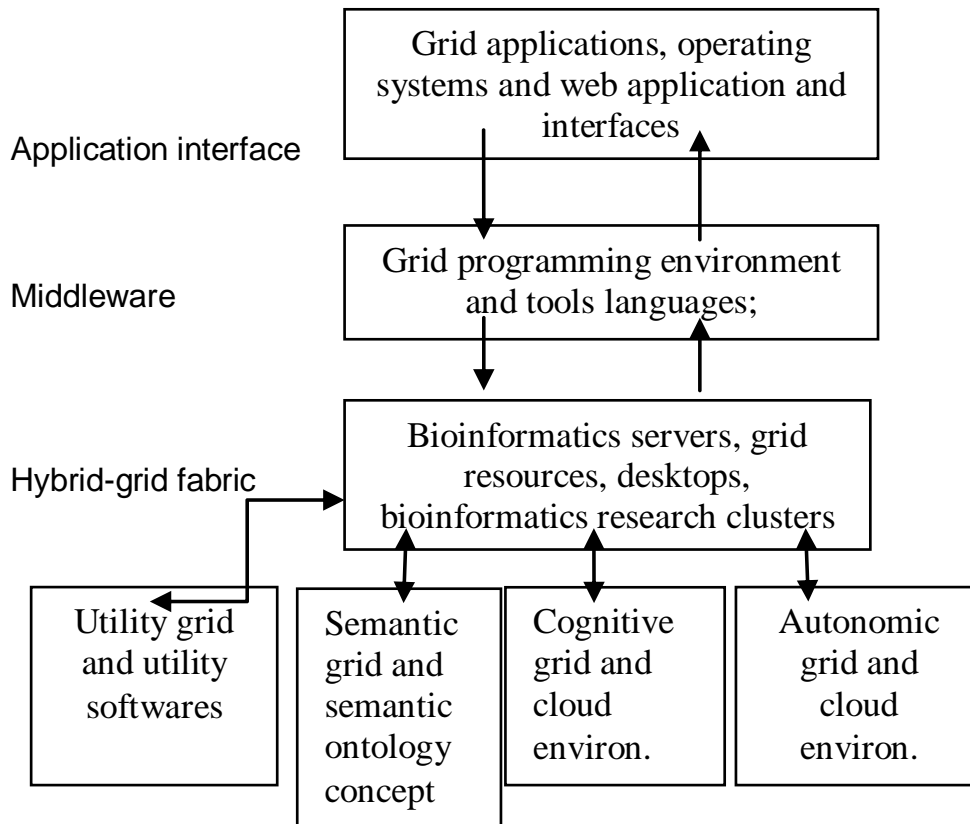

Figure 2. Proposed hybrid-grid framework for bioinformatics databases and resources.

to the advancement of knowledge in the fields of bioinformatics and computational biology cannot be thus

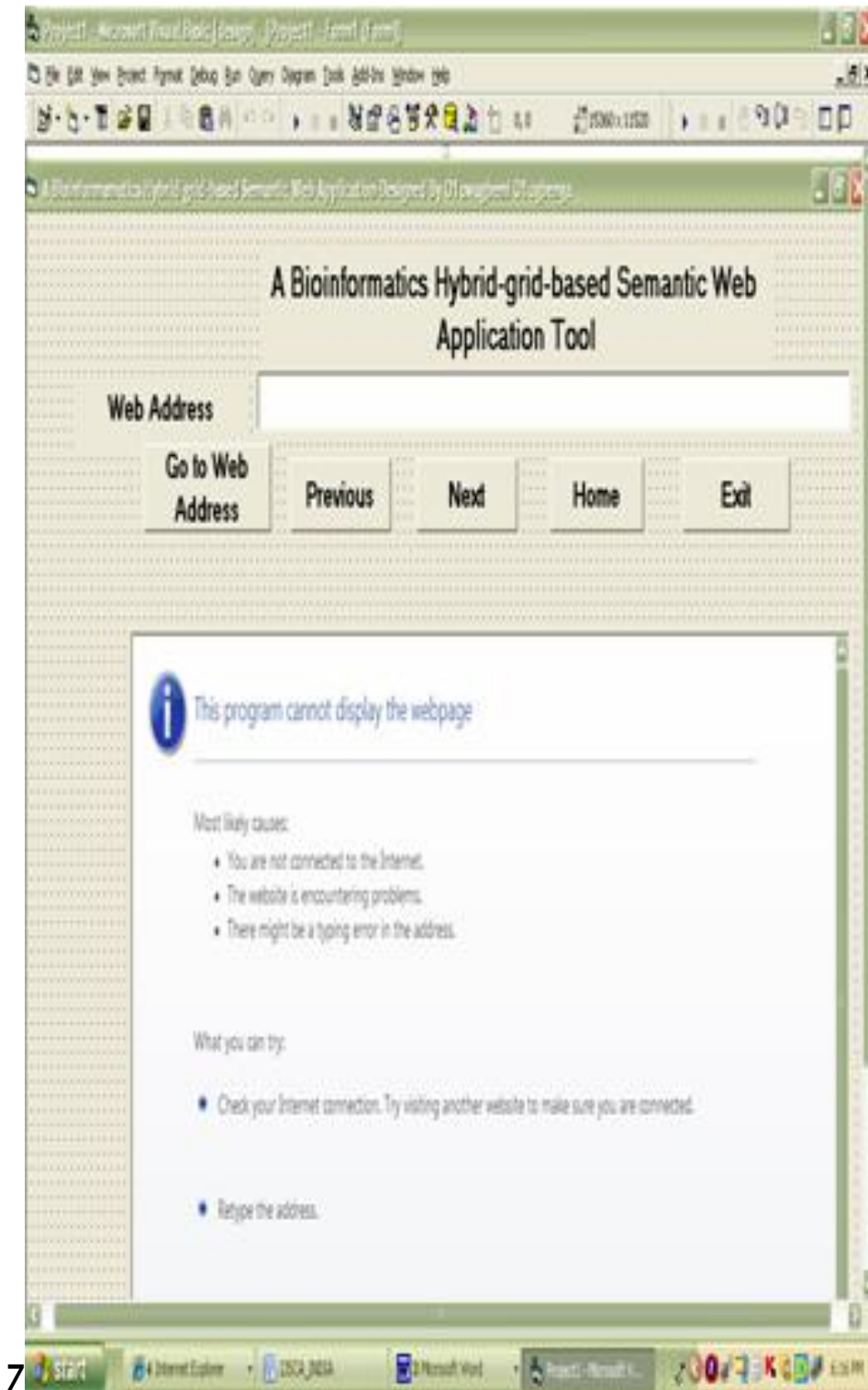underestimated. In this paper, only some aspect (hybrid-grid semantic web application and interface) of this

Figure 3. The design phase of the bioinformatics hybrid-grid web application too1

proposed framework was actually implemented, since this is just a framework depicting part of the overall model. The results obtained in Figure 3 showed the design phase of the bioinformatics hybrid-grid web application tool in the visual basic programming environment. This is the initial stage in the development
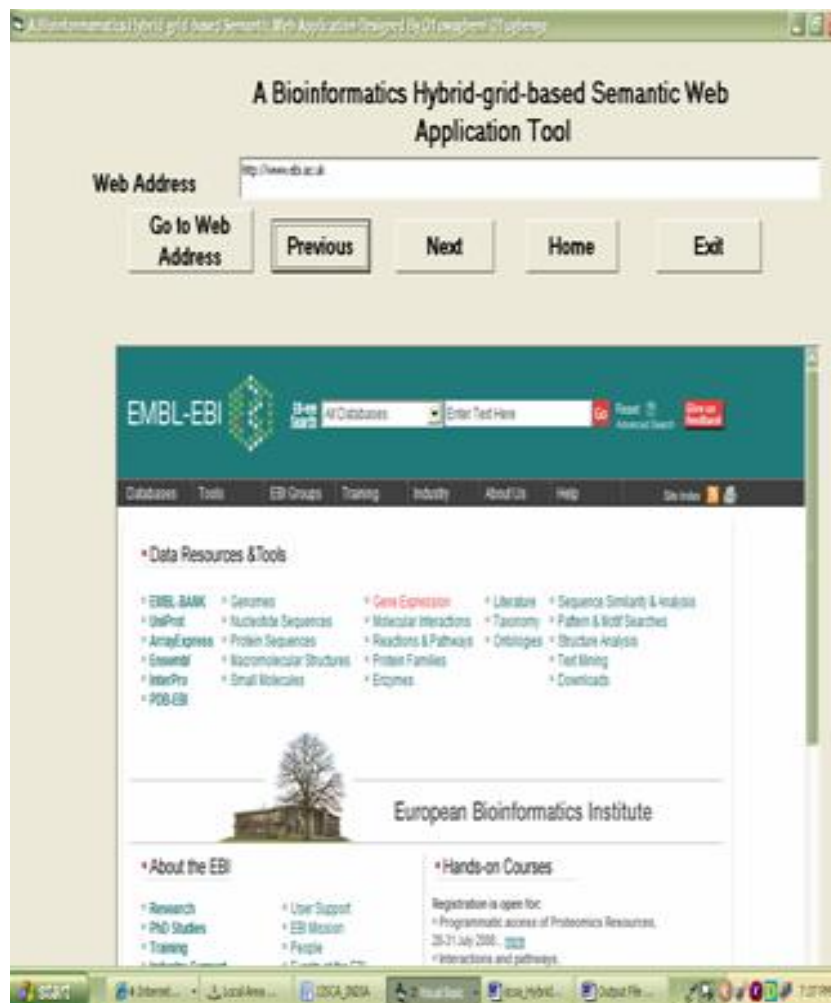
Figure 4. The implementation phase of the of the bioinformatics hybrid-grid web application tool.

of a bioinformatics web application tool. Figure 4 shows the implementation of the web application tool which was able to successfully connect to a Bioinformatics Database EBI (European Bioinformatics Institute).

The results obtained in Figure 5 showed that the web application tool was able to successfully access EBI's micro-array database. Figure 6 shows how an effective access was established to the protein database for EBI's Database using the hybrid-grid-based semantic web tool. Figure 7 shows how the developed bioinformatics web tool was able to access the databases for molecular interactions of the EBI (European Bioinformatics Institute). This shows the flexibility of the tool to navigate and access different bioinformatics databases and resources. This tool was implemented in visual basic programming environment- an event driven programming language. This graphical user interface describes the how the hybrid-grid web application tool navigates the bioinformatics databases for molecular interactions

Conclusion

In this research paper, we have proposed a hybrid-grid-based framework for analyzing bioinformatics databases and resources. A subset of the hybrid-grid framework (a web application for accessing and connecting bioinformatics databases and resources) was also developed and implemented for accessing useful bioinformatics data and resources for research purpose.

However, other features and improvements to the existing framework can be accomplished in further work, by implementing all other aspects (the cognitive, autonomic, and utility grids) of the proposed framework.
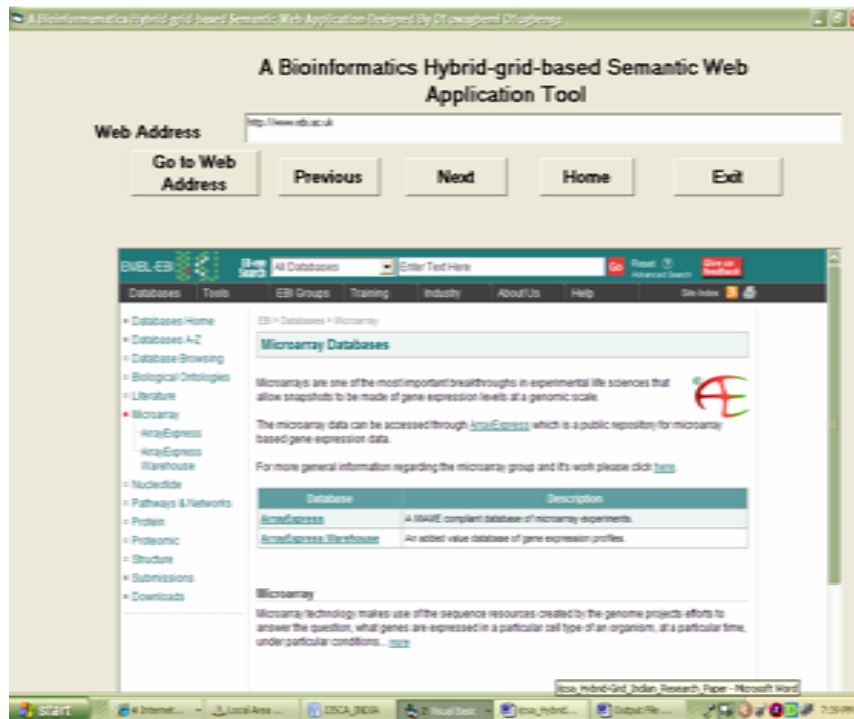
ACKNOWLEDGEMENT

Figure 5. The implementation phase of the of the bioinformatics hybrid-grid web application tool navigating the micro array database.
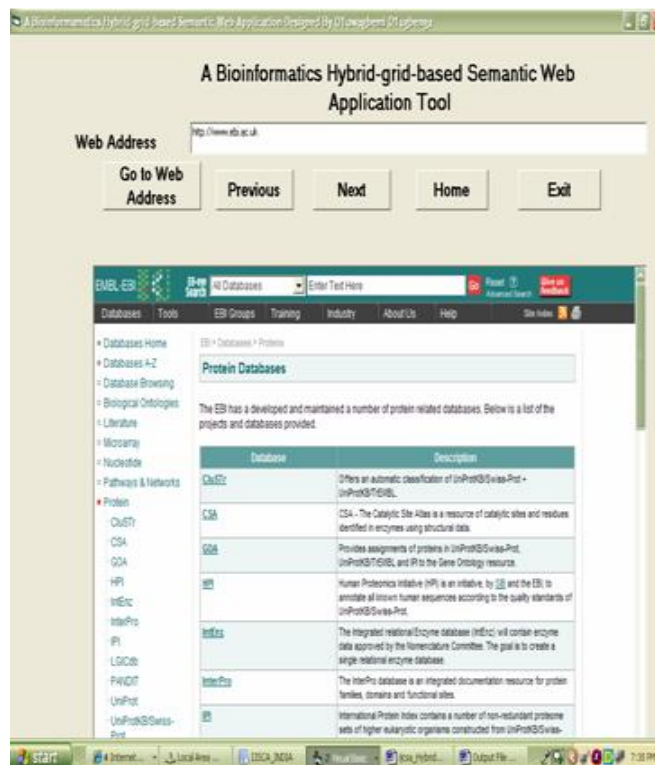

Figure 6. The implementation phase of the of the bioinformatics hybrid-grid web application too1 navigating the protein database.
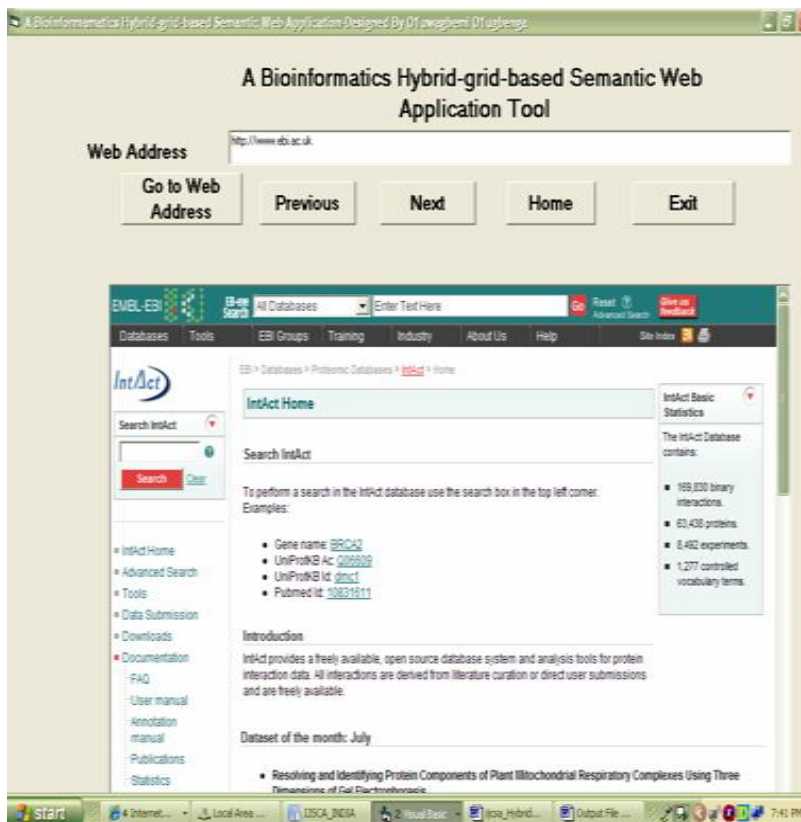
Figure 7. Output connecting to and displaying bioinformatics databases for molecular interactions.

REFERENCES

Ahsan S, Shah A (2008). An Agile Development Methodology for the Development of Bioinformatics. J. Am. Sci., 4(1): 1545-1003.

Ahsan S (2001). A Framework for Life-Cycle of the Prototype-Based Software Development Methodologies, J. King Saud Univ.,13: 105-125

Alberts B (2002). Molecular Biology of the Cell, 4th Edition.. New York, NY. Garland Publishing.

Alcalde FG (2011), Fuzzy Technology for Gene Ontology and Transcription Factor Binding Sites, A PhD Dissertation.

Andrade J, Andersen M, Sillén A, Graff C, Odeberg J (2007). The use of grid computing to drive data-intensive genetic research, Eur. J. Hum. Genet., 15: 694-702. doi:10.1038/sj.ejhg.5201815.

Baldi P, Brunak S (2001). Bioinformatics: The machine learning approach, 2nd ed, MIT Press, Cambridge.

Barbera R, Andronico G, Donvito G, Falzone A, Keijser JJ, La Rocca G, Milanesi L, Maggi GP, Vicario S (2011). A grid portal with robot certificates for bioinformatics phylogenetic analyses, Concurrency and Computation: Practice and Experience Special Issue: IWPLS 2009, 23(3): 246-255.

Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R (2007). NCBI GEO: mining tens of millions of expression profiles--database and tools update, Nucleic Acids Res. 2007 Jan;35(Database issue):D760-5. Epub 2006 Nov 11.

Bell G, Hey T, Szalay A (2011). Beyond data deluge, Science 6 March 2009, pp.1297-1298.

Bornberg-Bauer E, Paton NW (2002). Conceptual Data Modeling for Bioinformatics, Briefings in Bioinformatics.

Chakraborty S, Mittal R (2010). "Semantic Web Technology: A Strategic Approach to Intelligence", Int. J. Comput. Appl., 5(9).

Chance MR, Fiser A, Sali A, Pieper U, Eswar N, Xu G, Fajardo JE, Radhakannan T, and Marinkovic N (2004). High-Throughput Computational and Experimental Techniques in Structural Genomics, Genome Res., 14(10b): 2145-2154: doi: 10.1101/gr.2537904.

Dai YS, Palakal M, Hartanto S, Hartanto S, Wang X, Guo Y (2006). A Grid-Based Pseudo-Cache solution for MISD biomedical problems with high confidentiality and efficiency, Int. J. Bioinformat. Res. Appl., 2(3): 259-281

Daramola JO, Osamor VC, Oluwagbemi OO (2008) A Grid-Based Framework for Pervasive Healthcare using Wireless Sensor Networks: A Case for Developing Nations. Asian J. Inf. Technol., 7(6): 260-267. ISSN 1682-3915.

Deb D, Fuad MM, Oudshoorn MJ (2011). Achieving self-managed deployment in a distributed environment, J. Comput. Methods Sci. Eng., 11(1): 115-125.

de Roure D, Jennings N, Shadbot N (2003). The semantic grid: a future e-science infrastructure, F. Berman, G. Fox, A. Hey, Editors , Grid Computing: Making the Global Infrastructure a Reality, (first ed.), John Wiley and Sons, Chichester, United Kingdom, pp. 437-469.

Dhar PK, Meng TC, Somani S, Ye L, Sakharkar K, Krishnan A, Ridwan ABM, Wah SHK, Chitre M, Hao Z (2004). Grid Cellware: the first grid-enabled tool for modeling and simulating cellular processes. Bioinformatics, 21(7): 1284-1287.

Estrada K, Abuseiris A, Grosveld FG, André GU, Tobias AK, Fernando R (2009). GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data, Bioinformatics, 25(20): 2750-2752.

Folino G, Mastroianni C (2011). Preface - Special Issue: Bio-Inspired Optimization Techniques for High Performance Computing, New Gen. Comput., 29(2): 125-128, DOI: 10.1007/s00354-011-0101-8.

Foran DJ, Yang L, Chen W, Hu J, Lauri A Goodell, Reiss M, Wang F, Kurc T, Pan T, Sharma A, Saltz JH (2011). ImageMiner: a software system for comparative analysis of tissue microarrays using content-based image retrieval, high-performance computing, and grid technology, J. Am. Med. Inf. Assoc.,18:403-415, doi:10.1136/amiajnl-2011-000170.

Geldof M (2004). The Semantic Grid: will Semantic Web and Grid go hand in hand? European Commission DG Information Society Unit "Grid technologies".

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J (2004). Bioconductor: open software development for computational biology and bioinformatics, Genome Biol., 5(10):R80. Epub 2004 Sep, 15.

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

Hadzic M, Chang E (2004). "Grid Services Complemented by Domain Ontology Supporting Biomedical Community", Proceedings of SAG', pp. 86-98.

Halling-Brown MD., Moss DS, Sansom CE, Shepherd AJ (2009). A computational Grid framework for immunological applications, Philosoph. Trans. Royal Soc. A, 367(1898): 2705-2716.

Hariri S, Khargharia B, Chen H, Yang J, Zhang Y, Parashar M, Liu H (2006). The Autonomic Computing Paradigm, Cluster Comput., 9: 5-17.

Hendler J (2001). Agents the Semantic Web, The Semantic Web, IEEE Intell. Syst., pp. 30-37,

Hunter A, Schibeci D, Hiew HL, Bellgard M (2004). GRID-enabled bioinformatics applications for comparative genomic analysis at the CBBC, Cluster Computing, 2004 IEEE International Conference Proceedings 486 held between 20-23 Sept. 2004; 10.1109/CLUSTR.2004.1392652.

Hunter PJ, Crampin EJ, Nielsen PMF (2008). Bioinformatics, multiscale modeling and the IUPS Physiome Project, Briefings Bioinf., 9(4): 333-343.

Huttenhower C, Hofmann O (2010). A Quick Guide to Large-Scale Genomic Data Mining, PLoS Computational Biology 6(5): e1000779. doi:10.1371/journal.pcbi.1000779.

Jiang W, Baumgarten M, Dai Q, Zhou Y (2011). The deployment and evaluation of a bioinformatics grid platform – The HUST_Bio_Grid, Computers and Electrical Engineering (in press, corrected proof).

Kann MG (2009). Advances in translational bioinformatics: computational approaches for the hunting of disease genes, Briefings in Bioinformatics, 11(1): 96 -110.

Konagaya A (2006). Trends in life science grid: from computing grid to knowledge grid, BMC Bioinformatics, 7(Suppl 5):S10 DOI: 10.1186/1471-2105-7-S5-S10.

Lord P, Bechhofer S, Wilkinson MD, Schiltz G, Gessler D, Hull D, Goble C and Stein L (2004). Applying Semantic Web Services to Bioinformatics: Experiences Gained, Lessons Learnt, Lecture Notes in Computer Science, 2004, Volume 3298/2004: 350-364, DOI: 10.1007/978-3-540-30475-3_25.

Louie B, Mork P, Martin-Sanchez F, Halevy A, Tarczy-Hornoch P (2007). Data integration and genomic medicine, J. Biomed. Inf., 40(2007): 5-16.

Manjula KA, Raju G (2010). A Study on Applications of Grid Computing in Bioinformatics, Int. J. Comput. Appl. CASCT, (2): 69-72.

Mareuil F, Blanchet C, Malliavin TE, Nilges M (2011). Grid computing for improving conformational sampling in NMR structure

calculation, Bioinformatics, 27(12): 1713-1714.

Martín-Requena V, Ríos J, García M, Ramírez S, Trelles O (2009). jORCA: easily integrating bioinformatics Web Services. Bioinformatics, 26(4): 553-559.

Mirto M, Epicoco I, Cafaro M, Fiore S (2009). ProGenGrid: A Grid Problem Solving for Bioinformatics. In M. Cannataro (Ed.), Handbook of Research on Computational Grid Technologies for Life Sciences, Biomedicine, and Healthcare, pp. 269-291. doi:10.4018/978-1-60566-374-6.ch014.

Morgan JN, Sonquist JA (1963). Problems in the Analysis of Survey Data, and a Proposal, J. Am. Stat. Assoc., 58(302): 415-434.

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

Niu BF, Lang XY, Lu ZH, Chi XB (2007). ScBioGrid: a commodity supercomputing environment supporting bioinformatics research, Int. J. Comput. Math., 84(2).

Oluwagbemi OO(2008). Development and implementation of a bioinformatics online distance education learning tool for Africa, Int. J. Nat. Appl. Sci., 4(3): 256-262.

Pandey BK, Pandey SK, Pandey D (2011). A Survey of Bioinformatics Applications on Parallel Architectures, Int. J. Comput. Appl., 23(4): 21-25.

Pandey S, Voorsluys W, Niu S, Khandoker A, Buyy R (2012). An autonomic cloud environment for hosting ECG data analysis services, Future Gen. Comput. Syst., 28(1): 147-154.

Rahman M, Ranjan R, Buyya R, Benatallah B (2011). A taxonomy and survey on autonomic management of applications in grid computing environments, Concurrency and Computation: Pract. Exp., 23(16): 1990-2019.

Ramirez S, Muñoz-Mérida A, Karlsson J, García M, Pérez-Pulido AJ, Claros MG, Trelles O (2010). MOWServ: a web client for integration of bioinformatic resources, Nucleic Acids Res., 38(2): W671-W676.

Ramirez S, Karlsson J, Trelles O (2011). MAPI: towards the integrated exploitation of bioinformatics Web Services, BMC Bioinformatics, 12:419, DOI: 10.1186/1471-2105-12-419.

Rhee SY, Dickerson J, Dong Xu (2006). Bioinformatics and Its Applications in Plant Biology, Ann. Rev. Plant Biol., 57: 335-360.

Schadt EE, Linderman MD, Sorenson J, Lee L, Nolan GP (2010). Computational solutions to large-scale data management and analysis, Nature Rev. Gen., 11: 647-657. (September, 2010) |, DOI:10.1038/nrg2857.

Schmidberger M, Lennert S, Mansmann U (2011). Conceptual Aspects of Large Meta-Analyses with Publicly Available Microarray Data: A Case Study in Oncology, Bioinf. Biol. Insights, 5: 13-39.

Shafiq F, Raffat SK, Siddiqui MS (2011). Semantic Grid for Biomedical Ontologies. Int. J. Comput. Appl., 23(5): 1-4. Shahab A, Chuon D, Suzumura T, Li W, Byrnes R, Tanaka K, Ang L, Matsuoka S, Bourne P, Miller M, Arzberger P (2005). Grid Portal Interface for Interactive Use and Monitoring of High-Throughput Proteome Annotation. In Grid Computing in Life Science (LSGRID2004). Edited by Konagaya A, Satou K. Berlin Heidelberg New York: Springer; 2005: Lecture Notes in Bioinformatics, 3370: 53-67.

Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ (2007) "The OBO Foundry: Coordinated evolution of ontologies to support bi omedical data integration," Nat. Biotechnol., 25 DOI:10.1038/nbt1346.

Stankovski V, Swain M, Kravtsovc V, Niessen T, Wegener D, Kindermann J, Dubitzky W (2007). Grid-enabling data mining applications with DataMiningGrid: An architectural perspective, Future Gen. Comput. Syst., 24(4): 259-279.

Swedlow JR, Zanetti G, Best C (2011). Channeling the data deluge, Nature Methods, 8: 463-465. DOI:doi:10.1038/nmeth.1616.

Taha K, Elmasri R (2011). IRTG: A Grid Middleware for Bioinformatics, Services Computing (SCC), 2011 IEEE International Conference on 4-9 July 2011 in Washington DC, Conf. Proceed., pp. 440-447. 10.1109/SCC.2011.52.

Tan FB and Hunter MG (2004). Cognitive Research in Information Systems Using the Repertory Grid Technique: The Handbook Inf. Syst. Res., 1-30.

Tang F, Chua CL, Ho L, Lim YP, Issac P, Krishnan A (2005). Wildfire:

distributed, Grid-enabled workflow construction and execution, BMC Bioinformatics, 6: 69. DOI: 10.1186/1471-2105-6-69.

Teo Y, Wang X, Ng Y (2004). GLAD: a system for developing and deploying large-scale bioinformatics grid, Bioinformatics, 21(6): 794-802.

Vega-Gorgojo G, Miguel LB, Eduardo G, Yannis AD, Juan IA (2006). A semantic approach to discovering learning services in grid-based collaborative systems, Future Gen. Comput. Syst., 22(6): 709-719.

Villamizar M, Castro H, Mendez D, Restrepo, S, Rodriguez L (2011). Bio-UnaGrid: Easing Bioinformatics Workflow Execution Using LONI Pipeline and a Virtual Desktop Grid, ThinkMind // BIOTECHNO 2011, The Third International Conference on Bioinformatics, Biocomput.Syst. Biotechnol. Conf. Proceed., pp. 12-19.

Visan A, Istin M, Pop F, Cristea V (2011). Bio-Inspired Techniques for Resources State Prediction in Large Scale Distributed Systems, Int. J. Distr. Syst. Technol., 2(3): 1-18.

Walter T , Shattuck DW, Baldock R , Bastin ME , Carpenter AE , Duce S, Ellenberg J , Fraser A , Hamilton N ,Pieper S , Ragan MA, Schneider JE , Tomancak P, Jean-Karim H (2010). Visualization of image data from cells to organisms, Nature Methods 7, S26–S41 (1 March, 2010) | DOI:10.1038/nmeth.1431.

Xu K, Yu Q, Liu Q, Zhang J, Bouguettaya A (2011). Web Service management system for bioinformatics research: a case study, Service Oriented Comput. Appl., 5(1): 1-15. DOI: 10.1007/s11761-011-0076-9.

Yang, C., Han, T., Kan, H. (2009). G-BLAST: a Grid-based solution for mpiBLAST on computational Grids, Concurrency and Computation: Pract. Exp., 21(2): 225-255.

YarKhan A, Dongarra JJ (2005). Biological sequence alignment on the computational grid using the GrADS framework, Future Gen. Comput. Syst., 21(6): 980-986.

Zomaya A (2005). Grid computing: opportunities for bioinformatics research, Third Int. Conf. Inf. Technol. Appl., ICITA, 1(3): 4-7, DOI: 10.1109/ICITA.2005.167.