

BIOINFORMATICS AND INTERNET: NEW PARADIGM TO DISCIPLINES AND INFORMATION TECHNOLOGY

*Shiva Kanaujia**

The paper deals with the Bioinformatics and Internet. It introduces Bioinformatics and leads to Internet and its relation with Bioinformatics. Further discusses the various facilities, services provided by Internet for Bioinformatics. It further mentions required resources. Explains the links available on Internet, various products and services, involvement of private firms. It highlights some aspects related with techniques and tools for Bioinformatics on Internet. Explains some searching methods and role of library and information centres regarding Bioinformatics and Internet. Concludes with the idea that shows the importance of Bioinformatics on Internet

KEYWORDS/DESCRIPTORS: Bioinformatics; Internet; Internet Links; Searching Methods

1 INTRODUCTION

Bioinformatics is a new and emerging branch of Biotechnology that has come up very recently. It mainly involves the use of software to utilize information from vast biological database that is developed by experienced Biotechnologists. Gene sequencing is a part of Bioinformatics wherein a lot of data related to biotechnology is processed. This brings biotechnology within the ambit of information technology, and hence the label, Bioinformatics. In fact Bioinformatics is a new discipline that involves molecular biology and computer science. Presently genomic research, sequencing of human genome and advances in disease related issues have required and helped in developing this at fast level. So we can say that in Bioinformatics, computers are required to store, retrieve, analyze or predict the composition or the structure of biomolecules. It is the fascinating hybrid of computer science and biology.

* Jr. Research Fellow, Dept. of Library and Information Science, University of Lucknow, Lucknow, INDIA.

The National Center for Biotechnology Information (NCBI 2001) defines Bioinformatics as [1]:

"Bioinformatics is the field of science in which biology, computer science, and information technology merge into a single discipline...There are three important sub-disciplines within Bioinformatics: the development of new algorithms and statistics with which to assess relationships among members of large data sets; the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structures; and the development and implementation of tools that enable efficient access and management of different types of information."

Bioinformatics is a particularly international subject, with a notably high degree of information sharing among researchers in different countries. It is also known as computational biology e.g., USC Computational Biology, NCSA Computational Biology.

2 USE OF INTERNET IN BIOINFORMATICS

When we talk about sources of biological information and computers for providing it, we can not ignore the role and impact of information superhighway i.e., Internet. Internet is the most potential tool of this information age and it is serving as a platform for Bioinformatics tool. It provides the opportunity to search that information, which was available only by reaching to the information centre.

2.1 Areas of Services

The Internet provides various facilities for Bioinformatics, such as;

- Bioinformatics research
- Courses
- Resources
- Biological databases
- Construction tools
- Software resources
- WWW search tools
- Courses of Bioinformatics

- Advanced topics in Bioinformatics
- Scientific databases
- Electronic journals
- Asking queries from the librarian in online manner
- News events and activities such as; announcement for Bioinformatics interest group, meetings on federated databases, molecular biosciences and technology seminars.

2.2 World Wide Web Virtual Library: Biotechnology

This directory, provided by Cato Research Ltd., contains over 1000 URLs specific to biotechnology, pharmaceutical development, and related fields. The emphasis is on product development and the delivery of products and services [2].

2.3 Subject Specific Sites

These sites are more likely to concentrate on a particular area of Bioinformatics. These sites are further divided into the various areas of Bioinformatics e.g., Codon usage, and Genome analysis/Genomic comparisons, Phylogenetics etc.

2.4 General Bioinformatics Web Sites

Many of the sites are offering the same sorts of links and many to other Bioinformatics sites; many have links to a Sequence Retrieval System or other facilities for sequence retrieval. These are categorized as under:

- Academic Sites
- Corporate/Government Sites

2.5 Access to Journals

Providing access to journals such as; Nature, Science, Molecular biology and Evolution, Nucleic Acids Research, Bioinformatics, The Journal of Molecular Biology, Genetics, New Scientist, Online Journal of Bioinformatics, Internet Science Journal.

2.6 As a Centre for Biotechnology Information

One can explore extensive sites of resources and including newsletters, Bioinformatics databases, and links to the major medical bibliographic databases. It not only connects to textual databases but also to Protein Structure Servers. These include 3DB browser, biomolecular modeling and structural classification of proteins etc. Biotechnologists can reach any Bioinformatics Centers on Internet. From DNA Databank Japan to European Bioinformatics Centre can be reached by using Internet. One can search databases on protein, nucleotides, and protein structure.

2.7 News Services

Various news services can be categorized as following:

- 2.7.1. *Via e-mail*: It can serve as an important tool for being aware of new developments. The Scientist is providing this service with frequent coverage of Bioinformatics news. For this facility registration is free. It enables news automatically be sent via e-mail the tables of contents for each bi-weekly issue [3].
- 2.7.2. *Directories of news sites*: Some directories of news sites are focusing in Bioinformatics. **Southwest Biotechnology and Informatics Center (SWBIC)** is a good "launch pad" to news sites [4].
- 2.7.3. *Headline news service*: **Genomics Today** is a daily headline news service that provides links to genomics news in other sites. It culls the relevant headlines from a wide variety of sources including wire services, newspapers, Yahoo, selected web sites, and university news sites. The Pharmaceutical Research and Manufacturers of America sponsors it [5].
- 2.7.4. *News Network*: **Genome News Network** is a good source for news on scientific, as opposed to business, aspects of Bioinformatics. Bioinformatics news is clearly marked. The short news summaries are to be commended for giving the full citation to the original scientific article at the end of each news piece. In addition to news there are also featured articles and a few educational links [6].
- 2.7.5. *Conferences etc.*: Internet also informs about conferences and other platforms where designers and architects of high-performance computer chips, software and systems can share their vies and ideas for developing

more sophisticated and useful tools for Internet for Bioinformatics. IT provides detailed information about conferences, symposiums like Intelligent Systems for Molecular Biology [7] and Pacific Symposium on Biocomputing [8].

3 RESOURCES

For running any kind of service resources are required by Internet to be efficient in Bioinformatics. The resources include, Bioinformatics resources hardware, seminars, databases, agencies and societies, archives, centres, genome centres, genome databases and protein structure servers. It works with the help of Profile Search, Profile Scan, Manual, User's Guide, Help Page. These resources are also the part of provided information to users about.

4 BIOINFORMATICS, INTERNET AND LINKS

Internet facilitates linking to current programs and initiatives utilizing the Internet to form clearing-houses and distributed networks of biological information [9]. Some are integrated system for agricultural genome analysis, including databases, conferences, publications, courses, and a particularly good plant genome online database tutorial [10]. Links to many tools and programs are available from the National Institutes of Health for sequence analysis and molecular biology, including databases, protocols and tutorials [11]. Many a pages include links to a number of model organism databases, banks and tables, and to a number of genetic databases. Dept. of Molecular and Cellular Biology, Harvard University is maintaining such kind of service [12]. Presently "Molecules R US", which is a World Wide Web (WWW) Forms interface, facilitates access (browsing, searching and retrieval) to the molecular structure data contained within the Brookhaven Protein Data Bank (PDB). Through the Molecules R US request FORM, one may request a PDB data file to be returned in one of several different formats [13].

Bioinformatics online is providing "Tables of Contents" and abstracts for *Bioinformatics* issues published since 1996 [14]. Center for Genomics and Bioinformatics at Indiana University is doing the service including links to major Internet resources in Bioinformatics and also links to granting agencies, meetings, various genome projects and a large list of sequence analysis tools and databases and all this is possible due to Internet [15].

This is the Internet that is providing access to the site of the University of California, Santa Cruz containing working drafts for the human genome and the mouse genome. The human genome is over 90% complete. Approximately 85% of the sequence is in a highly accurate 'finished' state. The other 15% is merely 'draft' quality. Some care must be taken when interpreting draft regions, but these are still very useful to the working scientist [16]. One can access to SRS, EMBL, SWISS-PROT, and TrEMBL, RHD and MSD databases of European Bioinformatics Centre [17].

5 BIOINFORMATICS: PRODUCTS AND SERVICES

Like information products and services, Bioinformatics is also offering these “products and services”. We can be benefited for various products and services like American Type Culture Collection [18], Biotechnology Industry Organization [19], Professional directories [20], National Biotechnology Information Facility of New Mexico State University [21], Proteome Inc. Products and Services [22], Protein Data Bank [23], Online resources for biotechnology professionals [24].

6 INVOLVEMENT OF PRIVATE FIRMS

Various firms are maintaining Bioinformatics tools, for example- Federal Biotechnology Transfer Directory, a database of over 4,300 records of federal research projects and patents is maintained by a private consulting firm [25]. We can find non-profit, academe-based organizations, which are committed to opening access to Bioinformatics research projects, providing Open Source software for Bioinformatics by hosting its development, and keeping biological information freely available. For example, Bioinformatics.Org is a community focused on the freedom of information as it pertains to the bioscience [26]. Yahoo Directory of Medicine: Bioinformatics can be mentioned here which provides links to pages covering a wide variety of Bioinformatics categories, including: Bioinformatic Servers, Conferences, Human Genetics, Institutes, Professional Organizations, Schools, Departments and Programs, and Telemedicine [27].

7 INTERNET EDUCATIONAL RESOURCES FOR BIO-INFORMATICS

Biotechnology organizations are making their biological information available on the Internet. Some Indian universities are doing it also. We can mention here Centre for Biotechnology, Delhi University [28] and Bioinformatics Centre, Pune.

The Bioinformatics Centre of Pune provides upto-date and accurate information in the area of Biotechnology. It has emphasis on virology, proteins and nucleic acid sequences and structures, microbial strain data. It also provides an access to the other related areas through networks. Apart from answering the bibliographic queries and supplying the sequence data, the Centre also offers unique facilities for data analysis. The hardware, software and the expertise are available at this Centre. This availability facilitates the users in every possible way. The Center is equipped with the state-of-the-art services and facilities in terms of hardware, software, and information storage and retrieval [29]. Another one, Biotechindia.com provides information of facts. Their catalogue requests are solved. The services on Internet provide connection with R&D Systems Inc., Super Array Biosciences Corporation, Santa Cruz Biotechnology Inc., Cambrex Bioscience Rockland Inc., Cambrex Bioscience Walkersville Inc., Peninsula Laboratories Inc., Bellco Biotechnology Inc. of USA, and Bachem AG, Switzerland

8 COURSES FOR INTERNET RESOURCES AND BIOINFORMATICS TOOLS

Programming courses are being run on Internet for Bioinformatics and Internet, which requires basic Unix commands, file editing on Unix, telnet, FTP, E-mail, WWW and HTML. Online tutorials are available for starters. For example Stanford University has various courses for Bioinformatics in computing syllabus like “Stanford Algorithms in Computational Biology (MIS 214) Online, “Stanford Molecular Biology (Biochem 203), “Stanford Computational Biology Biochem (218 Online) [30].

Arranging workshops can provide an overview of the Bioinformatics/computational molecular biology tools available for use over the Internet. The workshops can enable participants for:

- Becoming familiar with databases and analysis tools on the Internet.
- Participating in hands-on computer labs to practice the use of these tools.
- Learning skills that allow exploring and exploiting more specific tools that may be useful.

9 SEARCHING

Various ways are available to search; a text-based query can be submitted through the system like Entrez system [31]. A sequence query can be submitted

through the program [32]. To search a particular type of database or item, it should be selected from the menu. Various lists include a number of databases. The list can be accessed by category/type of database or alphabetically by title. By clicking on the short description of each database, a paragraph-long description can be accessed from the database. Another system, Sequence Retrieval System interface provides links to sites that allow web based searching and retrieval of nucleotide and protein sequence. With the facility to query most of the major Bioinformatics databases and retrieve textual information but it is not the complete list.

10 ROLE AND USE BY LIBRARY AND INFORMATION CENTRES

As the library and information centres are playing a great role in Bioinformatics. Separate “Bioinformatics” have been established and many organizations are planning for it. As the field is vast and holds a wide potentiality for future prospects so some aspects should always be kept in mind. These are as following:

1. Researchers should be well informed about the services on Internet regarding Bioinformatics.
2. The user community must be taught and learnt the tools and techniques to explore Bioinformatics available on Internet.
3. Some customized services can be started, which can be provided in any of the manner i.e., in anticipation and on demand.
4. These customized services can include searching of databases, graphical information, and recent developments in the requested subject-discipline, providing news facilities by e-mail service.
5. The “Bioinformatics” of the organization can offer documentation services regarding availability of biological information on Internet and available other Bioinformatics services on Internet.
6. By using available resources of Bioinformatics on Internet, users can be informed about their availability.
7. Internet can help in a wide manner as it can help them in searching that information which is available far from their organization and country.
8. As Internet is spreading the “Bioinformatics”, it should be vital to publicize the role of Internet. Organization should provide facilities for data analysis of Bioinformatics from Internet.

9. As hardware, software and the expertise are required for various activities and services so these should be facilitated for the users.

11 CONCLUSION

Like other fields of knowledge, Bioinformatics has grown and seeing the researches going on in the genetic scenario, nobody can ignore the increasing impact of Bioinformatics. We can relate it with the increasing impact of Internet in information scenario. The future holds the ever increasing dependability on information and information technology so it can be said that Internet is an efficient tool for accessing the Bio-information in the form of Bioinformatics. The need is to know, explore and exploiting it.

REFERENCES

1. www.cats.ucsc.edu
2. www.cato.com/biotech/
3. www.the-scientist.com
4. www.nbif.org/links/1.20.php
5. www.genomics.phrma.org/today/
6. www.gnn.tigr.org/main.shtml
7. www.ebi.ac.uk/ismb-97/papers2.html
8. www.cgl.ucsf.edu/psb/
9. www.Biosis.org.zrdocs/zoolinfo/biol_inf.htm
10. www.ars-genome.cornell.edu/
11. www.molbio.info.nih.gov/molbio/
12. www.golgi.harvard.edu/Biolinks.html
13. www.molbio.info.nih.gov/cgi-bin/pdb
14. www.oup.co.uk/bioinformatics/contents/
15. www.cgb.indiana.edu/bioinformatics/resources
16. www.genome.ucsc.edu/
17. www.ebi.ac.uk/index.html
18. www.atcc.org
19. www.bio.org/welcome.html
20. www.biotech-register.com/
21. www.nbif.org

22. www.proteome.com/services/index.html
23. www.rcsb.org/pdb/
24. www.sciewb.com
25. www.bioinfo.com/fbdhome.html
26. www.bioinformatics.org
27. www.yahoo.com
28. www.cbt.org
29. www.unipune.ernet.in
30. www.mcrcr0.med.nyu.edu/rcr/molbio/syllabus-98.html
31. www.ncbi.nlm.nih.gov/Entrez/index.html
32. www.ncbi.nlm.nih.gov/BLAST

Appendix-1

Internet educational resources for Bioinformatics:

- **NCBI:** sequence data repository, US Bioinformatics center. <http://www.ncbi.nlm.nih.gov/>
- **EBI:** sequence data repository, European Bioinformatics center. <http://www.ebi.ac.uk/>
- **Pasteur:** France bioinfo. center, Bio Netbook is an excellent database of Internet information for biosciences, Bioinformatics. <http://bioweb.pasteur.fr/intro-uk.html> Bio Netbook
- **ExPASy/SWISSPROT:** protein sequence data center. <http://www.expasy.ch/> good list of bioinfo resources
- **Sanger:** European sequencing, Bioinformatics center. <http://www.sanger.ac.uk/>
- **Weizmann:** Israel Bioinformatics center. <http://bioinformatics.weizmann.ac.il/>
- **GenomeWeb:** Bioinformatics resources. <http://www.hgmp.mrc.ac.uk/GenomeWeb/>

- **CSHL**: US sequencing, Bioinformatics center. <http://www.cshl.org/>
- **WUSTL**: US sequencing, Bioinformatics center. <http://www.ibr.wustl.edu/>
- **Stanford genome center** US sequencing, Bioinformatics center. <http://genome-www.stanford.edu/>
- **TIGR**: US sequencing, Bioinformatics center. <http://www.tigr.org/>
- **Celera**: US commercial sequencing, Bioinformatics center. <http://www.celera.com/>
- **GenomeNet**: Japan Bioinformatics center <http://www.genome.ad.jp/>
- **Bionet**: Usenet network news for biology. <http://www.bio.net/>
- **BioMedNet**: Bioinformatics resources including HMS Beagle, <http://www.bmn.com/>
- **BioInform** mostly commercial news, services - good list of companies in Bioinformatics. <http://www.bioinform.com/>

Appendix- 2

Web Sites in the News:

- *Bioinformatics via D'Trends*: A good place to begin your research on the field, because it provides all of the background information by the person who coined the term himself --Dr. Hwa A. Lim (aka HAL)
- *Human Genome Project Information Web Site*: A government source and a great place to go for most of the information relating to current genome research. Sponsored by the U.S. Department of Energy Human Genome Program.
- *The National Human Genome Research Institute (NHGRI)*: Grant information, intramural research, ethical, legal and social implications, genomic and genetic resources. Also includes links to the policy and public affairs, workshops and conferences, and "The Genome Hub."

- *Bioinformatics at the NIH*: Contents include what's new, general information, impact to agency, and funding links. Also has a link to BIST (Biomedical Information Science Technology Initiative).
- *Center for the Study of Comparative Functional Genomics/Faculty*: Did you know the University at Albany has a local site? This site is well-equipped with mission statement, news, research, faculty descriptions and recent publications, seminars, employment, and pre-doctoral graduate program general and contact information.
- *Buffalo Center of Excellence in Bioinformatics*: In addition to having a link to news, this site also provides a wonderful overview with valuable information on informatics, bioinformatics, functional informatics and the future, with particular attention associated events in Western New York.
- *Bioinformatics*: A site that provided basic knowledge on the subject as well as the goals of the site. Explains why there should be interest in the field and had given relevant obtainable links for further research. Site contains links to scholarly journals, dictionary and FAQs.
- *Bioinform*: The Global Bioinformatics News Service Explore a New York-based journal, published every other Monday, focused exclusively on current events in the global Bioinformatics community.
- *Bioinformatics.org*: The site contains links to news (and news archives), repositories, mailing lists, and bulleting board. New additions include FAQ and Banner Exchange pages.
- *Bioinformatics News*: Self-described as a "Yahoo-like hierarchical Web directory." A Weblog that provides daily updates and links to news items and stories of interest to Bioinformatics community.
- *Bioinformatics homepage*: Internet Resources for Bioinformatics Coordinator: Dr. George Michaels. The aim of this page is to provide CSI students and researchers of Bioinformatics and Computational Biology an internet resource for their research as well as news, events, activities.

Other Sites of Interest:

- *NASA Bioinformatics*: A government site via NASA (National Aeronautics and Space Administration) which branches the field into two

subareas: Center for Informatics and Information Technology. These two areas include 3D construction, simulation, virtual reality, and neurotechnology. Site has links to virtual live science, space biology, image gallery, publications and people. There are also staff and contact information for students interested in their lab.

- *Bioinformatics at Latrobe University*: In addition to course description, this site describes bioinformatics, DNA and protein structure and molecular modeling, and the human genome. Discusses the field's popularity and has a very nice link to a powerpoint presentation with extensive details on concepts and terminology.
- *Gerstein Group-Yale Bioinformatics*: With much to die for regarding the field, Mark Gerstein, Assistant Professor of Molecular Biophysics and Biochemistry and Computer Science has a premiere Web source. It includes research on genomics, structure and expression with links to lectures, papers, people, jobs and calendar searches for events.
- *Masters of Science in Bioinformatics*: Through Georgia Institute of Technology, this site is tailored for students in pursuit of graduate degrees. Despite its promotion of the graduate program, it does piece together useful and brief information for those doing some background research on the field as a career. Also provides application and FAQ.