BioRec: A Bio Record Retrieving tool for Bioinformatics Research & Application

A THESIS SUBMITTED TO
THE ORISSA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY
BHUBANESWAR
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF SCIENCE IN BIOINFORMATICS

By
Smruti Ranjan Naik

Department of Bioinformatics
Centre for Post Graduate Studies
Orissa University of Agriculture & Technology
Bhubaneswar-751003, Orissa
2006

THESIS ADVISOR DR.P.N.JAGDEV

Dedicated to my Parents & my Brother

Dr. P. N. Jagdev, M.Sc.(Ag.), Ph.D
Head
Department of Bioinformatics,
Centre for Post Graduate Studies
Orissa University of Agriculture & Technology
Bhubaneswar-751003

Bhubaneswar
Dated the 29th Augest, 2006

CERTIFICATE – I

This is to certify that the thesis entitled "BioRec: A Bio Record Retrieving tool for Bioinformatics Research & Application" submitted by Mr. Smruti Ranjan Naik, Adm. No. 11BI/04 in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE IN BIOINFORMATICS is a bonafied record of research work, carried out by him under my guidance and supervision. No part of the thesis has been submitted elsewhere for any other degree or diploma. Help and assistance received are duly acknowledged.

Chairman
Advisory Committee

CERTIFICATE – II

*This is to certify that the thesis entitled "**BioRec: A Bio Record Retrieving tool for Bioinformatics Research & Application**" submitted by **Mr. Smruti Ranjan Naik, Adm. No. 11BI/04** to the Orissa University of Agriculture & Technology, Bhubaneswar in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN BIOINFORMATICS** has been approved by Student's Advisory Committee after an oral examination of the same in collaboration with the external examiner.*

ADVISORY COMMITTEE:

Chairman:

Dr.P.N.Jagdev

Head
Department of Bioinformatics,
Centre for Post Graduate Studies
Orissa University of Agriculture & Technology
Bhubaneswar-751003.

31.10.06

Members:

1. **Mr.Sambit Ranjan Pasupalaka.**
   Asst. Professor
   Department of Bioinformatics,
   Centre for Post Graduate Studies
   Orissa University of Agriculture & Technology
   Bhubaneswar-751003.

2. **Miss Dorey Benjamin.**
   Asst. Professor
   Department of Bioinformatics,
   Centre for Post Graduate Studies
   Orissa University of Agriculture & Technology
   Bhubaneswar-751003.

Dorey

3. **External Examiner**

31.10.06

ACKNOWLEDGEMENT

A pretty long and eventful epoch draws a close. May be one day, I shall be glad to ruminate their immense help acted as invisible lubricant to propel me to steer clear the entrusted task.

His sweet gentility pulsated with undiminished enthusiastic support, I could plunge myself into the trajectory to reach the destination. I deem it my proceed privilege to have worked under the inspiring guidance of my thesis advisor Dr. P. N. Jagdev, who geniously formulated and finalized the programme of work. I wish to express my sincere and deepest sense of gratitude to him from the core of my heart for his enduring guidance, affectionate personal attention and ceaseless encouragement throughout the period of investigation.

I would like to extend my sincere gratitude to Mr. A. Dash, H.O.D; M.C.A for allowing me to use the department facilities. He always there to encourage me in my endeavor.

My profound heartfelt thanks to Surya Bhai who always cared and help through out my project work

My special thanks to Miss.Dorey Benjamin(Asst.Prof in Bioinformatics),Mr.Sambit Ranjan Pasupalak, (Asst.Prof in Bioinformatics) and Mr.Ranjait Dash(Asst.Prof in Bioinformatics), Miss.Lizy Kanungo(Asst.Prof in Bioinformatics), Miss.Rosaly Pattnaik, Miss Sudipta Mohanty, Mahendra Bhai for their tremendous support, care and encouragement throughout the work.

For my dearest friends my entire bioinformatics batch mats, I thank them for their care, love and support during my Bioinformatics Studies. I don't have words to articulate my feelings for them.

Last but not least I thank my parents and my brother for their constant support, encouragement and love. Without their love, support and encouragement it wouldn't have been possible for me to finish my degree. I dedicate my work to them.

Smruti Ranjan Naik

ABSTRACT

Bioinformatics scenario has changed dramatically in the last decade and it will continue to change in future as well. Earlier immediate access to all bioinformatics research papers / literature was a dream of research scientists and now the World Wide Web i.e. search engines have made this dream a reality. Recently bioinformatics related websites like text base research journals are frequently accessed on the Internet for education, research and bioinformatics tool development. Orissa University of Agriculture & Technology is one of the premier research institute of India, involved in various researches like Agriculture, Biotechnology, Veterinary etc. Keeping this on mind this university has started a Bioinformatics program to provide an informatics support to these research areas and the responsibility of this department is also to make records of various research activities of this university. Because of the ever-increasing requirement of current information by the research professionals, the P.G.Department of Bioinformatics, Orissa University of Agriculture & Technology, Bhubaneswar, Orissa has set up a text base database on Bioinformatics research at O.U.A.T and this is known as BioRec. BioRec is a tool for retrieving Bioinformatics research record, comprising the outcome of the sustained bioinformatics research carried out by the students of P.G.Department Bioinformatics, Orissa University of Agriculture & Technology, Bhubaneswar. Interesting records in the field of genomics, proteomics, immuno-informatics, molecular modeling etc research records are incorporated in this Database. The aim of BioRec is to provide access to the end users (students, research persons) for their attempt to enhance indigenous research skills in the field of bioinformatics. BioRec is maintained by the P.G.Department of Bioinformatics, Centre for Post Graduate Studies, Orissa University of Agriculture & Technology, Bhubaneswar, Orissa.

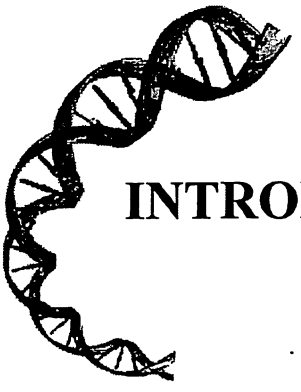
CONTENTS

PAGE NO.

LIST OF FIGURES PAGE

ABBREVIATION USED

ADME	Adsorption, Distribution, Metabolism, Execration.
ADO	Active Data Object
ASP	Active Server Page
CLR	Common Language Runtime.
CTS	Common Type System.
DBMS	Database Management System.
GUI	Graphic User Interface.
HTML	Hyper Text Markup Language
HTTP	Hyper Text Transfer Protocol
IIS	Internet Information Service.
IL	Intermediate Language.
JIT	Just in Time.
MFC	Microsoft Foundation Classes.
MSIL	Microsoft Intermediate Language.
NCBI	National Center for Biotechnology Information.
NIH	National Institute of Health
NLM	National Library of Medicine.
NMR	Nuclear Magnetic Resonance.
OUAT	Orissa University of Agriculture & Technology
QBE	Query by Example.
RAD	Rapid Application Development.
SDK	Software Developer Kits.
SNPs	Single Nucleotide Polymorphisms.
SQL	Structured Query Language.
RDBMS	Relational Database Management System.
WSDL	The Web Service Description Language.
XML	Extensible Markup Language



INTRODUCTION

1. INTRODUCTION

1.1Bioinformatics:

Bioinformatics is the application of computer technology to the management of biological information. Computers are used to gather, store, analyze and integrate biological and genetic information which can then be applied to gene-based drug discovery and development. The need for Bioinformatics capabilities has been precipitated by the explosion of publicly available genomic information resulting from the Human Genome Project. The goal of this project determination of the sequence of the entire human genome (approximately three billion base pairs) will be reached by the year 2002. The science of Bioinformatics, which is the melding of molecular biology with computer science, is essential to the use of genomic information in understanding human diseases and in the identification of new molecular targets for drug discovery. In recognition of this, many universities, government institutions and pharmaceutical firms have formed bioinformatics groups, consisting of computational biologists and bioinformatics computer scientists. Such groups will be key to unraveling the mass of information generated by large scale sequencing efforts underway in laboratories around the world [1].

1.2 Objectives of Bioinformatics:

The fundamental objectives of Bioinformatics are the following[23]:

1. At its simplest and basic level, bioinformatics organizes data that stored in database and allows researchers to access existing information, also to submit new entries. While data curation is an essential task, the information stored in these databases is essentially useless until analyzed. Thus the purpose of Bioinformatics extends far beyond mere volume control of data.
2. The second key objective is to develop tools and resources that aid in the analysis of data. For example, having sequenced a particular protein, it is of interest to compare it with previously characterized sequences. This requires more than just a straight forward database search. As such, programs such as

FASTA and PSI-BLAST must consider what constitutes a biologically significant resemblance. Development of such resources requires extensive knowledge of computational theory, as well as a thorough understanding of biology.

3. The third objective is to use the tools to analyze the data and interpret the results in a biologically meaningful manner. Traditionally, biological studies examined individual systems in detail, and frequently compared those with a few that are related. In bioinformatics, we can also conduct global analyses of all the available data with the aim of uncovering common principles that apply across many systems and highlight features that are unique to some.

1.3 Applications of Bioinformatics:

Some of the applications related to biological information analysis are [23]:

1. Information related to bio-molecules can be mapped (e.g.), the sequences can be parsed to find sites where so-called "restriction enzymes" will cut them.
2. Sequences can be compared, usually by aligning corresponding segments and looking for matching and mismatching letters in their sequences. Genes or proteins that are sufficiently similar are likely to be related and are therefore said to be "homologous" to each other.
3. If a homologous exists then a newly discovered protein may be modeled – that is 3D structure of the gene product can be predicted without doing laboratory experiments.
4. Bioinformatics is used to attempt to predict the function of actual gene products.
5. Information about the similarity, and by implication, the relatedness of proteins is used to trace the "family trees" of different molecules through evolutionary time.
6. Molecular modeling /structural biology is growing field, which can be considered part of bioinformatics. There are, for example, tools which allow making predictions of the secondary structure of proteins arising from given amino acid sequence. These are often based on known

"solved" structures and other sequenced molecules acquired by structural biologists.

7. Structural biologists use bioinformatics to handle the vast and complex data from X-ray crystallography, nuclear magnetic resonance (NMR) and electron microscopy investigations and create the 3-D models of molecules.
8. Information of the research records or the research papers of the journal are stored in text search database for further reference to enhance and improve the research procedure or to get the information for related research techniques as a reference.
9. There are other fields like medical imaging / image analysis that might be considered as a part of the bioinformatics. There is also a whole other discipline of biologically inspired computation: genetic algorithms, AI, neural networks etc. Often these areas interact in unexpected ways. Neural networks, inspired by crude models of the functioning of the nerve cells in the brain, and used in a program called PHD to predict, the secondary structures of proteins from their primary sequences.

1.4 Major areas of Bioinformatics:

Bioinformatics has various applications in research in medicine, biotechnology, agriculture etc. Following areas and research fields has integral component of Bioinformatics [6].

1. Computational Biology: The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems.
2. Genomics: Genomics is any attempt to analyze or compare the entire genetic complement of a species or species (plural). It is, of course possible to compare genomes by comparing more-or-less representative subsets of genes within genomes.
3. Proteomics: Proteomics is the study of proteins - their location, structure and function. It is the identification, characterization and quantification of all

proteins involved in a particular pathway, organelle, cell, tissue, organ or organism that can be studied in concert to provide accurate and comprehensive data about that system.

Proteomics is the study of the function of all expressed proteins. The study of the proteome, called proteomics, now evokes not only all the proteins in any given cell, but also the set of all protein isoforms and modifications, the interactions between them, the structural description of proteins and their higher-order complexes, and for that matter almost everything 'post-genomic'.

4. Pharmaco-genomics : Pharmaco-genomics is the application of genomic approaches and technologies to the identification of drug targets. In Short, pharmaco-genomics is using genetic information to predict whether a drug will help make a patient well or sick. It Studies how genes influence the response of humans to drugs, from the population to the molecular level.
5. Pharmaco-genetics: Pharmaco-genetics is the study of how the actions of and reactions to drugs vary with the patient's genes. All individuals respond differently to drug treatments; some positively, others with little obvious change in their conditions and yet others with side effects or allergic reactions. Much of this variation is known to have a genetic basis. Pharmaco-genetics is a subset of pharmaco-genomics which uses genomic/bioinformatic methods to identify genomic correlates, for example SNPs (Single Nucleotide Polymorphisms), characteristic of particular patient response profiles and use those markers to inform the administration and development of therapies. Strikingly such approaches have been used to "resurrect" drugs thought previously to be ineffective, but subsequently found to work with in subset of patients or in optimizing the doses of chemotherapy for particular patients.
6. Chemo-informatics: 'The mixing of those information resources [information technology and information management] to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the arena of drug lead identification and optimization. [a]. Related terms of cheminformatics are chemi-informatics, chemometrics, computational chemistry, chemical informatics, chemical information management/science, and chemo-informatics.

But we can distinguish chemo-informatics and chemical informatics as follows

Chemical informatics : 'Computer-assisted storage, retri-eval and analysis of chemical information, from data to chemical knowledge.' [b]. This definition is distinct from ' Chemoinformatics ' (and the synonymous cheminformatics and chemo-informatics) which focus on drug design. *chemometrics:* The application of statistics to the analysis of chemical data (from organic, analytical or medicinal chemistry) and design of chemical experiments and simulations. [IUPAC Computational]

computational chemistry : A discipline using mathematical methods for the calculation of molecular properties or for the simulation of molecular behavior. It also includes, e.g., synthesis planning, database searching, combinatorial library manipulation [c]. [IUPAC Computational]

7. Structural genomics or structural bioinformatics refers to the analysis of macromolecular structure particularly proteins , using computational tools and theoretical frameworks. One of the goals of structural genomics is the extension of idea of genomics , to obtain accurate three-dimensional structural models for all known protein families, protein domains or protein folds . Structural alignment is a tool of structural genomics. Comparative genomics: The study of human genetics by comparisons with model organisms such as mice, the fruit fly, and the bacterium E. coli .
8. Biophysics: The British Biophysical Society defines biophysics as: "an interdisciplinary field which applies techniques from the physical sciences to understanding biological structure and function".
9. Biomedical informatics / Medical informatics: "Biomedical Informatics is an emerging discipline that has been defined as the study, invention, and implementation of structures and algorithms to improve communication, understanding and management of medical information."
10. Mathematical Biology: Mathematical biology also tackles biological problems, but the methods it uses to tackle them need not be numerical and need not be implemented in software or hardware. It includes things of theoretical interest which are not necessarily algorithmic, not necessarily molecular in nature, and are not necessarily useful in analyzing collected data.
11. Computational chemistry: Computational chemistry is the branch of theoretical chemistry whose major goals are to create efficient computer programs that calculate the properties of molecules (such as total energy,

dipole moment, vibrational frequencies) and to apply these programs to concrete chemical objects. It is also sometimes used to cover the areas of overlap between computer science and chemistry.

12. Functional genomics: Functional genomics is a field of molecular biology that is attempting to make use of the vast wealth of data produced by genome sequencing projects to describe genome function. Functional genomics uses high-throuput techniques like DNA microarrays, proteomics, metabolomics and mutation analysis to describe the function and interactions of genes.
13. Pharmaco-informatics: Pharmaco-informatics concentrates on the aspects of bioinformatics dealing with drug discovery.
14. In silico ADME-Tox Prediction:(Brief description)- Drug discovery is a complex and risky treasure hunt to find the most efficacious molecule which do not have toxic effects but at the same time have desired pharmacokinetic profile. The hunt starts when the researchers look for the binding affinity of the molecule to its target. Huge amount of research requires to be done to come out with a molecule which has the reliable binding profile. Once the molecules have been identified, as per the traditional methodologies, the molecule is further subjected to optimization with the aim of improving efficacy. The molecules which show better binding is then evaluated for its toxicity and pharmacokinetic profiles. It is at this stage that most of the candidates fail in the race to become a successful drug.
15. Agro-informatics / Agricultural informatics: Agro-informatics concentrates on the aspects of bioinformatics dealing with plant genomes.

1.5 Objectives of the Study:

The main objective of the present study is to know how the Bioinformatics research professionals; students can explore the bioinformatics research database of OUAT, which can be used as a potential tool for the benefits of the bioinformatics education and research.

REVIEW OF LITERATURE

2.REVIEW OF LITERATURE

Retrieval of biological information from biological databases is important for the biotechnological professionals, since it is a scholarly need, an accordance requirement and means of updating oneself professionally. Now a day, information technology, communication networks, Internet and bioinformatics websites offer the biotechnological professionals, the timely and convenient access of the required information from biological databases. The life science librarians have realised that there is a need to organise the information available on the web so that biological research professionals can easily find information to enrich the education process, facilitate the basic research and improve the research skills. The internet has provided the new ways for the research scholars to interact with machine and communicate with the fellow colleagues in their day to day research skills, research procedure and decision making. The researchers irrespective of geographical locations, languages and cost can access the bioinformatics, biotechnology, medical websites. The National Library of Medicine (NLM) USA is playing a significant role in providing the biomedical literature of NLM's MEDLINE database to medical professionals all over the world free of cost. The NCBI's Pub Med also serve same facilities for the research scholar to improve their knowledge sphere to enhanance the indigenous skills for the biological research. In India, biological research professionals are solely dependent on the institute library for current information. With availability of Internet, biological professionals get the required information from many text search database like Pub Med, NLM, MEDLARS databases.

2.1 Biological Database:

A biological database is a large, organized body of persistent data, usually associated with computerized software designed to update, query, and retrieve components of the data stored within the system. A simple database might be a single file containing many records, each of which includes the same set of information. For example, a record associated with a nucleotide sequence database typically contains

information such as contact name; the input sequence with a description of the type of molecule; the scientific name of the source organism from which it was isolated; and, often, literature citations associated with the sequence[3].

For researchers to benefit from the data stored in a database, two additional requirements must be met:

1.Easy access to the information; and
2.A method for extracting only that information needed to answer a specific biological question.

2.2 Types of Database:

(i) Generalized Databases (DNA, Proteins and Carbohydrates, 3D-Structures).
(ii) Specialized Databases (Genomes, Protein Families, Pathways, Microarray data etc.)

2.3 Database Search:

(ii) Text-based database search (SRS,Entrez etc).
(iii) Sequence- based database search (sequence similarity search, BLAST,FASTA etc.).
(iiii) Motif-based database search (ScanProsite,Emotif).
(iiv) Structure-based database search (Structure similarity Search,VAST,DALI etc.).

2.4 Text-base database search:

A wide range of database is present to get the information in text format. Mostly the database of Biological Research Literature are text based search. Some of the text based database search includes[23].
(i) ENTREZ (Integrated database search for nucleotides, proteins, genomes, structures, populations sets and literature at NCBI,USA).

(ii) SRS6 (Search EMBL, SWISS-PROT, TrEMBL, PIR, PDB etc. at EBI, UK, or other SRS5 Servers world wide).

(iii)DBGET/LinkDB (Integrated database search of Link DB at GenomeNet, Kyoto University and University of Tokyo).

(iv)SSS(Sequence and structure searching site at Berkeley,US).

(v)Molecular R US (Full text search of the PDB database at NIH).

(vi)MIA (Molecular Information Agent searching biological databases to find the existing information about a macromolecule at SDSC, USA).

2.5 Literature Search:

This search includes the text-based search of research journals present in database. Some examples of text based search for literature includes:

(i) PubMed (Search public version of full Medline integrated in ENTREZ at NCBI, USA).

(ii) PubMed Central (Free online access to full text of life science research articles at NCBI, USA).

(iii) MEDLINE+ (Search and order full text Medline documents at Healthgate, USA).

(iv) Libri Vision (Search Current Contents, Medline and more at KULeuven Library. Access restricted to KULeuven).

(v) MGAS Online (Access to full text journals at KULeuven library. Access restricted to KULeuven).

(vi) LIBIS (Search in Belgian Library catalogues. Registration Required).

(vii) Pub Crawler (Automated alerting service that scans daily updates to PubMed and Genbank data base at Trinty College Dublin, Ireland).

2.6 Text Search Databases on the Web –PUBMED:

National Library of Medicine located in Bethesda (USA) is the world's largest research library in a single specific and professional field. The NLM provides global access to some of the best on-line sources in the form of searchable database. The

NLM. databases and electronic information sources provide a wide variety of resources related to the bio-medical and health science which has access to 12 million MEDLINE citations record and indexes. It is a text search database which provides relevant information of the research fields. It has around 4500 different titles taken from over 70 countries, which covers back to mid 1960's onwards. The information comes from US National Institute of Health and other federal agencies, professional associations and non-profit health organizations. In addition to health topic MEDLINE plus also connects the users to medical dictionaries, hospitals directories of physicians and dentists, and extensive medical encyclopedia and consumer information about thousands of prescription drugs. NLM provides free access to all its databases and users have been encouraged to look towards the Websites. All users will need to use one of NLM's web-based search engines PubMed (http://www.ncbi.nlm.nih.gov.PubMed) Or Internet Grateful Med (http://igm.nlm.nih.gov).At present PubMed provides access to MEDLINE, PREMEDLINE, and publisher supplied citations. Internet Grateful Med also provides access to the following databases[h].

2.7 MEDLINE/PubMed:

(MEDLARS online) is the largest biomedical bibliographic databases, incorporating the printed Index Medicus. It covers almost 4,500 international journals from 1960 to date and is updated weekly. This journal covers all medical and surgical specialities and preclinical sciences in addition to dentistry, nursing, pharmacology, connectivity, but also facilitates globalization of medical and scientific literature. The use of Internet is growing very fast in the world of health arena. The Internet has become a source of medical information for the medical professionals all over the world. The effective use of the Internet for the professional purpose requires an understanding of the best strategies to search the www as well as compendium of online resources including journals, textbooks medical portals and sites providing high quality information. PubMed Central is a digital archive of life sciences journal literature managed by the National Centre for Biotechnology Information at the NLM. Access to PubMed Central is free and unrestricted[d].

2.8 Internet:

The Internet not only offered global electronic journals, but it provide the latest review of information

2.9 E-Textbooks:

It is believed that unrestricted and free access to scientific knowledge will be the pattern of online education in the years ahead. In context of medical, biological research, such a scenario will have a major impact on research and education. Now a days a many websites are available for free books .

e.g-The recently introduced site FreeBooks4Doctors at

http://www.fb4d.com

http://www.freebooks4doctors.com

is available on the web. It offers links to more than 400 books, sorted in nearly100 specialties and sorted by title alphabetically, with the added promise of new books everyday. This site has been specifically dedicated to the promotion of free access to medical books over the Internet.

2.10 E-Journals:

The www as the most revolutionary development in publishing arena since the printing press. It is evident that electronic version of journals has some inherent advantages over print version such as incorporation of multimedia, liberal use of colours, elimination of delay in manuscript processing, news alerting, instant real time online discussion. Online repositories for articles, such as BioMed Central and PubMed central have embraced the ideal of free access for all the literature. PubMed central provides free access to some print journals already offering their entire contents online in addition to the purely electronic journals in BioMed central. The main advantage of online publication is the speedy dissemination of information. In e-journals, articles can be made available on the web as soon as they are ready. The Web browser and hypertext link provide easy to use tools to view documents and to move between them. Hyperlinks to references and related articles are also available

for full text information about research. With this, research professionals can participate in online continuing education. With the advent of the user friendly World Wide Web and the Internet is no longer a curiosity for the computer savvy. Now a day the lecture notes are published on the Internet, where we can download as and when it is required. In this way students are not dependent upon the library only as various e-books and journals are available on the website. The large amount of biological information demands changes in the teaching techniques used to teach the students. The web-based learning promotes the development of self-directed learning strategies[h].

2.11 Use of Litrature Search Database for Research:

The use of Internet in the research process begins from identifying research issues, through using the web for survey, to pre-publishing and publishing results. The literature searches using different databases such as Medline are obviously an important and integral part of every research process. Material published on the Internet government, biotech organizations, other institutions, medical journals and pharmaceutical companies as well as through medical web directories. The medical professionals are able to retrieve the information from the these websites in order to make a better decision of research procedure. Many research institutions are now having research records on the websites for research and education purposes This help them in review of the research, effective follow ups and in provision of quality new research.

- NLM's resources.
- Lectures by Research Scientests.
- Details of IMC's services (CDROM

databases, training, etc.)

- Downloadable Training Manual (PDF).
- Search Request forms (which can be submitted directly from the webpage) as well as a feedback form.

2.12 Use of Literature Search Websites for Education:

It is believed that unrestricted free access to knowledge will be the trend of online education in the recent times .In reference to education field, such a scenario will have a deep impact on education and research. There are several websites offering online continuing different education. The outcome of this concept is being studied online http://www.cme.center.com may be a valuable resource for researcher, to know about the different fields of research being carried on in the world .The accessibility of knowledge for analysis and the anonymity of the internet allow researchers to analyse text and narrative on websites, to use newsgroups as global focus groups. There are three different research methodologies for qualitative research on the Internet distinguished as Passive analysis, Active analysis and Interviews and surveys

2.13 About Orissa University of Agriculture & Technology:

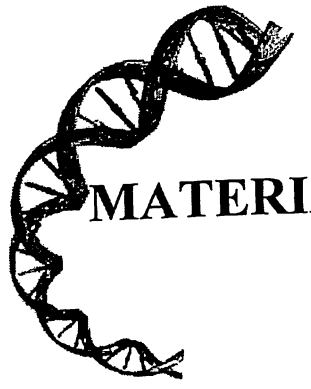
The Orissa University of Agriculture &Technology established in 1962 with only two Colleges of Agriculture and Veterinary Science, has made impressive record of service in the cause of quality education and research. In the pace of decades the University has grown with seven colleges with many research centers. The University inherited a few research stations such as, Economic Botany 1 & 2, Rice Research Station, Jeypore, Agriculture Research Station, Berhampur and Jute Research Station, Kendrapara from the State government. During 1960s the University research was mainly confined to agriculture and a very few research projects were in operation in the College of Veterinary Science and Animal Husbandry. With the launching of Orissa Agriculture Development Project (OADP) in 1978 with the World Bank assistance, four Regional Research Stations were established at Bhubaneswar, Chiplima, Keonjhar and Semiliguda representing the four physiographic regions of the state such as Coastal Plain, Central Table Land, North Central Plateau and Western Ghat High land zones, respectively. With the implementation of National Agriculture Research Projects (NARP) in two phases during the period 1983-1995, the State was divided into ten Agroclimatic Zones and four additional zonal Research Stations and four zonal Substations were established. Thus the research network is functioning in the University across the length and breadth of the state covering all the

10 agro-climatic zones. Research programmes are executed on different aspects of agriculture and its allied sectors through these research centers. The University is credited to be first educational institute in Orissa to develop campus wide Local Area Network with the FO cabling to all the Colleges at Bhubaneswar in 1996. The work was done under Agricultural Research Information System (ARIS) project funded by ICAR. Under the project the University has provided Internet facility to its staff and students through VSAT connected to NICNET. The ARIS project also imparts computer training to its staff from time to time. ARIS also has taken responsibility to develop the Website for University. From Year 2002, the University has migrated to STPI from NIC for improved Internet Connectivity with a bandwidth of 256 Kbps[5].

2.14 About P.G. Department of Bioinformatics:

The Orissa University of Agriculture & Technology is a multi-faculty technical University with an earned reputation of being a leader in teaching, research & extension. OUAT, nurtures the students to face impediments, meet challenges and varying needs of IT career in today's world along with the objective to disseminate advanced knowledge by providing instructions and research facilities.

The Center for Post-Graduate studies forms the indispensable part of this University. This center of learning has been brought out with an offering of career oriented Post Graduate programme like M.Sc Bioinformatics since 2002.The main aim behind the opening of this department is to provide an informatics support to the Agricultural Research at O.U.A.T. Since 1962 this university is engaged with a wide numbers of project works for Agriculture, Biotechnology, and Veterinary Research. Still the research data obtained are not documented for future reference to accelerate the research process. So the department of Bioinformatics has started this work as a database for Bioinformatics research and application at O.U.A.T.



MATERIAL AND METHODS

3.MATERIALS AND METHODS

3.1 System Specification:

(a)Hardware Specification:

This system is developed on the following hardware configuration.

Pentium IV processor

512 MB Ram

Hard disk 80 GB

Microsoft Compatible 101 or more Key Board.

(b)Software Specification:

This system is developed using the following software.

Operating System: Windows Xp Professional

Technology ASP.NET

Data Access Component ADO.NET

Back End: Microsoft Access.

Web browser: Internet Explorer(5.01 or Higher), Mozilla (v 1.7.1).

(c) Web Server i.e. IIS (v 3.0 or Higher).

3.2 Other tools for development:

Photo Impression (Arc Soft): for photo editing.

Scanner: for text data collection.

Pdf995: for conversion of word file to pdf file.

3.3 Collection of Data:

The data were collected from the dissertation records of students (batch 2002,2003). These are the data of research works carried out by students at different laboratories like IGIB, IIT, CCMB, IISC, IICT etc. Currently we have data of 61 records.

3.4 Internet Information Services (IIS):

Internet Information Services (IIS) for Microsoft Windows brings the power of Web computing to Windows. With IIS, we can easily shares files and printers and create applications to securely publish information to improve the way the organization works. IIS is a secure platform for building and deploying e-web solutions. IIS also makes it easy to bring mission-critical bioinformatics applications to the Web [8].

Windows with IIS scales to meet our needs as we can

- Set up a personal Web server.
- Share information within our team.
- Access the Databases.
- Create an enterprise internet.

IIS integrates proven Internet standards with Windows, so that using the web does not mean having to start over and learn new ways to publish, manage, or develop. Windows 2000 with Internet Information Services is the easiest way to share information and run powerful applications on the Web [20].

3.5 Why Microsoft IIS:

- IIS 5.0 is arguably the most powerful Web Server available.
- IIS's component services launch at boot time and are available before we log in to the console.
- In IIS parlance, an application is a collection of Web content (HTML files, ASP scripts and CGI or ISAPI executables) which serves a single purpose.

3.6 Installing Web Server IIS:

Before installation First we insert Windows XP Professional CD into cd-drive. Then after going to control panel, we select Add/Remove Windows Components. Windows Components Wizard window pop-up it's list of components. Then we check which component is not installed. We find Internet Information Service (IIS) as one among them. We select IIS from the list and by clicking OK and Next on main Wizard window it shows a message for finish. Then after finish, we check out the working of server by opening the browser and typing in the address bar as

http://localhost

or

http://127.0.0.1

which is local computer ip address. As everything was successfully set up IIS home page was opened with some information.

3.7 Visual Studio .NET 2003:

Visual Studio .NET 2003 is the comprehensive tool for rapidly building Microsoft .NET connected applications for Microsoft Windows and the Web, dramatically increasing developer productivity, and enabling him or her to satisfy new business and enterprise requirements[8].

3.8 .NET Frame Work:

The .NET Framework is a new computing platform that simplifies application development in the highly distributed environment of the Internet. The .NET Framework is designed to fulfill the following objectives[8]:

- To provide a consistent object-oriented programming environment whether object code is stored and executed locally, executed locally but Internet-distributed, or executed remotely.

- To provide a code-execution environment that minimizes software deployment and versioning conflicts.
- To provide a code-execution environment that guarantees safe execution of code, including code created by an unknown or semi-trusted third party.
- To provide a code-execution environment that eliminates the performance problems of scripted or interpreted environments.
- To make the developer experience consistent across widely varying types of applications, such as Windows-based applications and Web-based applications.
- To build all communication on industry standards to ensure that code based on the .NET Framework can integrate with any other code.

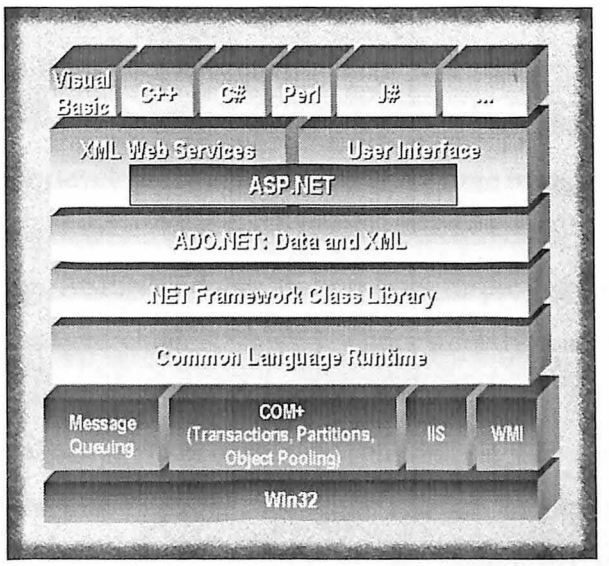


Fig-1:.NET Framework

Benefits of the .NET Framework:

- Based on W3 Standards
- Unified Application Models
- Easy for developers
- Extensible classes

The .NET Framework has two main components: the common language runtime and the .NET Framework class library. The common language runtime is the

foundation of the .NET Framework. We can think of the runtime as an agent that manages code at execution time, providing core services such as memory management, thread management, and remoting, while also enforcing strict type safety and other forms of code accuracy that ensure security and robustness. In fact, the concept of code management is a fundamental principle of the runtime. Code that targets the runtime is known as managed code, while code that does not target the runtime is known as unmanaged code. The class library, the other main component of the .NET Framework, is a comprehensive, object-oriented collection of reusable types that we can use to develop applications ranging from traditional command-line or graphical user interface (GUI) applications to applications based on the latest innovations provided by ASP.NET, such as Web Forms and XML Web services[10].

The .NET Framework can be hosted by unmanaged components that load the common language runtime into their processes and initiate the execution of managed code, thereby creating a software environment that can exploit both managed and unmanaged features. The .NET Framework not only provides several runtime hosts, but also supports the development of third-party runtime hosts.

For example, ASP.NET hosts the runtime to provide a scalable, server-side environment for managed code. ASP.NET works directly with the runtime to enable Web Forms applications and XML Web services. Internet Explorer is an example of an unmanaged application that hosts the runtime (in the form of a MIME type extension). Using Internet Explorer to host the runtime enables us to embed managed components or Windows Forms controls in HTML documents. Hosting the runtime in this way makes managed mobile code (similar to Microsoft ActiveX controls) possible, but with significant improvements that only managed code can offer, such as semi-trusted execution and secure isolated file storage.

3.9 Features of Common Language Runtime (CLR):

The Common Language Runtime (CLR) is a layer between an application and the operating system it executes on. The common language runtime simplifies an application's design and reduces the amount of code developers need to write because it provides a variety of execution services that include memory management, thread execution, code execution, code safety verification, compilation, and other system services. These features are intrinsic to the managed code that runs on the common language runtime[9].

The key benefit of the CLR is that it transparently provides these execution services to all applications, regardless of what programming language they're written in and without any additional effort on the part of the developer. The CLR is also responsible for compiling code just before it executes. Instead of producing a binary representation of our code, as traditional compilers do, .NET compilers produce a representation of our code in a language common to the .NET Framework: Microsoft Intermediate Language (MSIL), often referred as IL. When our code executes for the first time, the CLR invokes a special compiler called a Just In Time (JIT) compiler, which transforms the IL into executable instructions that are specific to the type and model of our system's processor. Because all .NET languages have the same compiled representation, they all have similar performance characteristics. This means that a program written in Visual Basic.NET can perform as well as the same program written in Visual C++.NET.

With regards to security, managed components are awarded varying degrees of trust, depending on a number of factors that include their origin (such as the Internet, enterprise network, or local computer). This means that a managed component might or might not be able to perform file-access operations, registry-access operations, or other sensitive functions, even if it is being used in the same active application. The runtime enforces code access security. For example, users can trust that an executable embedded in a Web page can play an animation on screen or sing a song, but cannot access their personal data, file system, or network. The security features of the runtime thus enable legitimate Internet-deployed software to be exceptionally feature rich.

The runtime also enforces code robustness by implementing a strict type- and code-verification infrastructure called the common type system (CTS). The CTS ensures that all managed code is self-describing. The various Microsoft and third-party language compilers generate managed code that conforms to the CTS. This means that managed code can consume other managed types and instances, while strictly enforcing type fidelity and type safety.

The Common Type System (CTS) is a component of the CLR and provides a common set of data types, each having a common set of behaviors. In Visual Basic, for example, the String data type maps to the CTS System. String class. Therefore, if a Jscript.NET client needs to communicate with a component implemented in VB.NET, the client doesn't have to do any additional work to exchange information because it's using a type common to both Jscript.NET and VB.NET. The CTS eliminates many interoperability problems that exist outside .NET.

.NET programming languages take advantage of the CTS by enabling developers to use their language's built-in data types, the .NET compilers convert the native data types' into their equivalent CTS types at compile time. Developers can also use CTS types at compile time.

In addition, the managed environment of the runtime eliminates many common software issues. For example, the runtime automatically handles object layout and manages references to objects, releasing them when they are no longer being used. This automatic memory management resolves the two most common application errors, memory leaks and invalid memory references.

The runtime also accelerates developer productivity. For example, programmers can write applications in their development language of choice, yet take full advantage of the runtime, the class library, and components written in other languages by other developers. Any compiler vendor who chooses to target the runtime can do so. Language compilers that target the .NET Framework make the features of the .NET Framework available to existing code written in that language, greatly easing the migration process for existing applications.

While the runtime is designed for the software of the future, it also supports software of today and yesterday. Interoperability between managed and unmanaged code enables developers to continue to use necessary COM components and DLLs. The runtime is designed to enhance performance. Although the common language runtime provides many standard runtime services, managed code is never interpreted. A feature called just-in-time (JIT) compiling enables all managed code to run in the native machine language of the system on which it is executing. Meanwhile, the memory manager removes the possibilities of fragmented memory and increases memory locality-of-reference to further increase performance.

The runtime can be hosted by high-performance, server-side applications, such as Internet Information Services (IIS). This infrastructure enables us to use managed code to write our business logic, while still enjoying the superior performance of the industry's best enterprise servers that support runtime hosting[20].

3.10 .NET Framework Class Library:

The .NET Framework class library is a collection of reusable types that tightly integrate with the common language runtime. The class library is object oriented, providing types from which our own managed code can derive functionality. This not only makes the .NET Framework types easy to use, but also reduces the time associated with learning new features of the .NET Framework. In addition, third-party components can integrate seamlessly with classes in the .NET Framework[8].

For example, the .NET Framework collection classes implement a set of interfaces that we can use to develop our own collection classes. Our collection classes will blend seamlessly with the classes in the .NET Framework. To make .NET Class Library easier to work with and understand, it is divided into *namespaces.* The root namespace of the .NET Class Library is called 'System', and it contains core classes and data types, such as Int32, Object, Array and Console. Secondary namespaces reside within the 'System' namespace.
Examples of nested namespaces include the following:

System.Diagnostics: Contains classes for working with the Event Log

System.Data: Makes it easy to work with data from multiple data sources (System.Data.Oledb resides within this namespace and contains the ADO.NET classes)

System.IO: Contains classes for working whit files and data streams.

The below figure (Fig-2) illustrates the relationship between some of the major namespaces in the .NET Class Library.

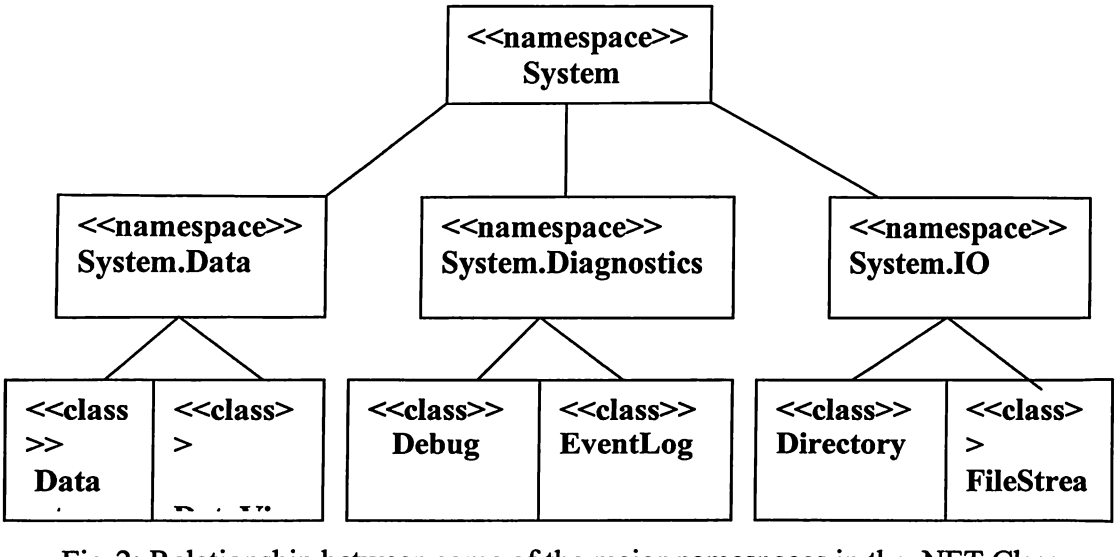


Fig-2: Relationship between some of the major namespaces in the .NET Class Library.

3.11 Organization of the .NET Class Library:

The benefits of using the .NET Class Library include a consistent set of services available to all .NET languages and simplified deployment, because the .NET Class Library is available on all implementations of the .NET Framework.

The .NET Framework types enable us to accomplish a range of common programming tasks, including tasks such as string management, data collection,

database connectivity, and file access. In addition to these common tasks, the class library includes types that support a variety of specialized development scenarios[k]. For example, we can use the .NET Framework to develop the following types of applications and services:

Console applications.
Scripted or hosted applications.
Windows GUI applications (Windows Forms).
ASP.NET applications.
XML Web services.
Windows services.

For example, the Windows Forms classes are a comprehensive set of reusable types that vastly simplify Windows GUI development. If we write an ASP.NET Web Form application, we can use the Web Forms classes.

3.12 Client Application Development:

Client applications are the closest to a traditional style of application in Windows-based programming. These are the types of applications that display windows or forms on the desktop, enabling a user to perform a task. Client applications include applications such as word processors and spreadsheets, as well as custom business applications such as data-entry tools, reporting tools, and so on. Client applications usually employ windows, menus, buttons, and other GUI elements, and they likely access local resources such as the file system and peripherals such as printers. Another kind of client application is the traditional ActiveX control (now replaced by the managed Windows Forms control) deployed over the Internet as a Web page. This application is much like other client applications: it is executed natively, has access to local resources, and includes graphical elements[21].

In the past, developers created such applications using C/C++ in conjunction with the Microsoft Foundation Classes (MFC) or with a rapid application

development (RAD) environment such as Microsoft Visual Basic. The .NET Framework incorporates aspects of these existing products into a single, consistent development environment that drastically simplifies the development of client applications. The Windows Forms classes contained in the .NET Framework are designed to be used for GUI development. We can easily create command windows, buttons, menus, toolbars, and other screen elements with the flexibility necessary to accommodate shifting business needs.

For example, the .NET Framework provides simple properties to adjust visual attributes associated with forms. In some cases the underlying operating system does not support changing these attributes directly, and in these cases the .NET Framework automatically recreates the forms. This is one of many ways in which the .NET Framework integrates the developer interface, making coding simpler and more consistent.

Unlike ActiveX controls, Windows Forms controls have semi-trusted access to a user's computer. This means that binary or natively executing code can access some of the resources on the user's system (such as GUI elements and limited file access) without being able to access or compromise other resources. Because of code access security, many applications that once needed to be installed on a user's system can now be safely deployed through the Web. Our applications can implement the features of a local application while being deployed like a Web page[8].

3.13 Server Application Development

Server-side applications in the managed world are implemented through runtime hosts. Unmanaged applications host the common language runtime, which allows our custom managed code to control the behavior of the server. This model provides us with all the features of the common language runtime and class library while gaining the performance and scalability of the host server.

The following illustration shows a basic network schema with managed code running in different server environments. Servers such as IIS Server can perform standard operations while our application logic executes through the managed code.

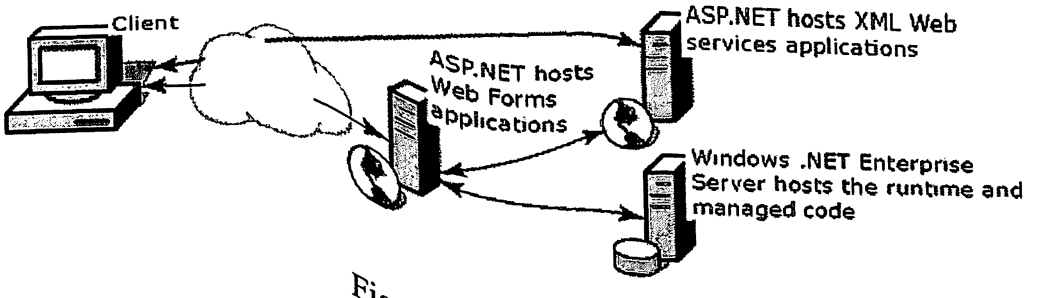


Fig-3: Server-side managed code

ASP.NET is the hosting environment that enables developers to use the .NET Framework to target Web-based applications. However, ASP.NET is more than just a runtime host; it is a complete architecture for developing Web sites and Internet-distributed objects using managed code. Both Web Forms and XML Web services use IIS and ASP.NET as the publishing mechanism for applications, and both have a collection of supporting classes in the .NET Framework. XML Web services, an important evolution in Web-based technology, are distributed, server-side application components similar to common Web sites. However, unlike Web-based applications, XML Web services components have no GUI and are not targeted for browsers such as Internet Explorer and Netscape Navigator. Instead, XML Web services consist of reusable software components designed to be consumed by other applications, such as traditional client applications, Web-based applications, or even other XML Web services. As a result, XML Web services technology is rapidly moving application development and deployment into the highly distributed environment of the Internet[8].

If we have used earlier versions of ASP technology, we will immediately notice the improvements that ASP.NET and Web Forms offers. For example, we can develop Web Forms pages in any language that supports the .NET Framework. In addition, our code no longer needs to share the same file with our HTTP text

(although it can continue to do so if we prefer). Web Forms pages execute in native machine language because, like any other managed application, they take full advantage of the runtime. In contrast, unmanaged ASP pages are always scripted and interpreted. ASP.NET pages are faster, more functional, and easier to develop than unmanaged ASP pages because they interact with the runtime like any managed application.

The .NET Framework also provides a collection of classes and tools to aid in development and consumption of XML Web services applications. XML Web services are built on standards such as SOAP (a remote procedure-call protocol), XML (an extensible data format), and WSDL (the Web Services Description Language). The .NET Framework is built on these standards to promote interoperability with non-Microsoft solutions.

For example, the Web Services Description Language tool included with the .NET Framework SDK can query an XML Web service published on the Web, parse its WSDL description, and produce C# or Visual Basic source code that our application can use to become a client of the XML Web service. The source code can create classes derived from classes in the class library that handle all the underlying communication using SOAP and XML parsing. Although we can use the class library to consume XML Web services directly, the Web Services Description Language tool and the other tools contained in the SDK facilitate our development efforts with the .NET Framework. If we develop and publish our own XML Web service, the .NET Framework provides a set of classes that conform to all the underlying communication standards, such as SOAP, WSDL, and XML. Using those classes enables us to focus on the logic of our service, without concerning our self with the communications infrastructure required by distributed software development. Like Web Forms pages in the managed environment, our XML Web service will run with the speed of native machine language using the scalable communication of IIS.

3.14 Active Server Pages.NET (ASP.NET):

ASP.NET is a unified Web development platform that provides the services necessary for developers to build enterprise-class Web applications. While ASP.NET is largely syntax compatible with ASP, it also provides a new programming model and infrastructure for more secure, scalable, and stable applications. We can feel free to augment our existing ASP applications by incrementally adding ASP.NET functionality to them [14].

ASP.NET is a compiled, .NET-based environment; we can author applications in any .NET compatible language, including Visual Basic .NET, C#, and JScript .NET. Additionally, the entire .NET Framework is available to any ASP.NET application. Developers can easily access the benefits of these technologies, which include the managed common language runtime environment, type safety, inheritance, and so on.

ASP.NET has been designed to work seamlessly with WYSIWYG HTML editors and other programming tools, including Microsoft Visual Studio .NET. Not only does this make Web development easier, but it also provides all the benefits that these tools have to offer, including a GUI that developers can use to drop server controls onto a Web page and fully integrated debugging support.

Developers can choose from the following two features when creating an ASP.NET application, Web Forms and Web services, or combine these in any way they see fit. Each is supported by the same infrastructure that allows us to use authentication schemes; cache frequently used data, or customizes our application's configuration, etc.

Web Forms allows us to build powerful forms-based Web pages. When building these pages, we can use ASP.NET server controls to create common UI elements, and program them for common tasks. These controls allow us to rapidly build a Web Form out of reusable built-in or custom components, simplifying the code of a page. An XML Web service provides the means to access server functionality remotely. Using Web services, businesses can expose programmatic interfaces to their data or business logic, which in turn can be obtained and

manipulated by client and server applications. XML Web services enable the exchange of data in client-server or server-server scenarios, using standards like HTTP and XML messaging to move data across firewalls. XML Web services are not tied to a particular component technology or object-calling convention. As a result, programs written in any language, using any component model, and running on any operating system can access XML Web services. Each of these models can take full advantage of all ASP.NET features, as well as the power of the .NET Framework and .NET Framework common language runtime. These features and how we can use them are outlined as follows: The ASP.NET object model has changed significantly from ASP, making it more structured and object-oriented. Unfortunately this means that ASP.NET is not fully backward compatible; almost all existing ASP pages will have to be modified to some extent in order to run under ASP.NET. In addition, major changes to Visual Basic .NET mean that existing ASP pages written with Visual Basic Scripting Edition typically will not port directly to ASP.NET. In most cases, though, the necessary changes will involve only a few lines of code. Accessing databases from ASP.NET applications is an often-used technique for displaying data to Web site visitors. ASP.NET makes it easier than ever to access databases for this purpose. ASP.NET provides a simple model that enables Web developers to write logic that runs at the application level. Developers can write this code in the globa. asax text file or in a compiled class deployed as an assembly. This logic can include application-level events, but developers can easily extend this model to suit the needs of their Web application. ASP.NET provides easy-to-use application and session-state facilities that are familiar to ASP developers and are readily compatible with all other .NET Framework APIs[13].

For advanced developers who want to use APIs as powerful as the ISAPI programming interfaces that were included with previous versions of ASP, ASP.NET offers the IHttpHandler and IHttpModule interfaces. Implementing the IHttpHandler interface gives us a means of interacting with the low-level request and response services of the IIS Web server and provides functionality much like ISAPI extensions, but with a simpler programming model. Implementing the IHttpModule interface allows us to include custom events that participate in every request made to our application. ASP.NET takes advantage of performance enhancements found in the .NET Framework and common language runtime. Additionally, it has been designed

to offer significant performance improvements over ASP and other Web development platforms. All ASP.NET code is compiled, rather than interpreted, which allows early binding, strong typing, and just-in-time (JIT) compilation to native code, to name only a few of its benefits.

Writing custom debug statements to our Web page can help immensely in troubleshooting our application's code. However, it can cause embarrassment if it is not removed. The problem is that removing the debug statements from our pages when our application is ready to be ported to a production server can require significant effort. ASP.NET offers the Trace Context class, which allows us to write custom debug statements to our pages as we develop them. They appear only when we have enabled tracing for a page or entire application. Enabling tracing also appends details about a request to the page, or, if we so specify, to a custom trace viewer that is stored in the root directory of our application. The .NET Framework and ASP.NET provide default authorization and authentication schemes for Web applications. We can easily remove, add to, or replace these schemes, depending upon the needs of our application. ASP.NET configuration settings are stored in XML-based files, which are human readable and writable. Each of our applications can have a distinct configuration file and we can extend the configuration scheme to suit our requirements[13].

3.15 Overview of DBMS:

An organization must have accurate and reliable data for effective decision - making. To this end, the organization maintains records on various facts of its operation by building appropriate models of diverse class of objects of interest. These models capture the essential properties of the objects and record relationship among them. Such related data is called database. A data base system is an integrated collection of related files along with details about their definition, interpretation, manipulation & maintenance. A data base management system is essentially collection of interrelated data and a set of program to access this data. The objective of DBMS is to provide a convenient and effective method of defining, storing and retrieving the information contained in the database .The DBMS not only makes the

integrated collection of reliable and accurate data available to manipulate applications and users but also exerts centralized control, prevents fraudulent or unauthorized users from accessing the data and ensures privacy. Data base system supports single user and multi - user environment to retrieve and store database information while on one hand DBMS permits only one person to access the database at a given time; on the other hand RDBMS allows many users simultaneous access to the database.

A Database System consists of two parts namely, Database Management System and Database Application. Database Management System is the program that organizes and maintains the information whereas the Database Application is the program that lets us view, retrieve and update information stored in DBMS. DBMS has to protect database against unintentional changes that could be caused by users and applications. In case of multi-user system, it must be capable of notifying and database change to the other user [15].

3.16 Database Management System offers the following services:

- Data Definitions: It is a method of definition and storage.
- Data Maintenance: It checks whether each record has fields containing all information about one particular item.
- Data Manipulation: Allows the data in the database to be inserted, updated, deleted and stored.
- Data Display: This method helps in viewing Data.
- Data Integrity: This ensures the accuracy of data

3.17 Microsoft Access:

Microsoft Access is a relational database management system (DBMS). At the most basic level, a DBMS is a program that facilitates the storage and retrieval of structured information on a computer's hard drive[16].

3.18 The many faces of Access:

Microsoft generally likes to incorporate as many features as possible into its products. For example, the Access package contains the following elements:

- A relational database system that supports two industry standard query languages: Structured Query Language (SQL) and Query by Example (QBE);
- A full-featured procedural programming language— essentially a subset of Visual Basic,
- A simplified procedural macro language unique to Access;
- A rapid application development environment complete with visual form and report development tools;
- A sprinkling of objected-oriented extensions; and,
- Various wizards and builders to make development easier.

For new users, these "multiple personalities" can be a source of enormous frustration. The problem is that each personality is based on a different set of assumptions and a different view of computing. For instance,

- The relational database personality expects you to view your application as sets of data;
- The procedural programming personality expects you to view your application as commands to be executed sequentially;
- The object-oriented personality expects you to view your application as objects which encapsulate state and behavior information. Microsoft makes no effort to provide an overall logical integration of these personalities (indeed, it is unlikely that such integration is possible). Instead, it is up to you as a developer to pick and choose the best approach to implementing your application. Since there are often several vastly different ways to implement a particular feature in Access, recognizing the different personalities and exploiting the best features (and avoiding the pitfalls) of each are important skills for Access developers.

The advantage of these multiple personalities is that it is possible to use Access to learn about an enormous range of information systems concepts without having to interact with a large number of "single-personality" tools, for example:

- Oracle for relational databases

• PowerBuilder for rapid applications development,
• Smalltalk for object-oriented programming.

Keep this advantage in mind as we switch back and forth between personalities and different computing paradigms.

3.19 Access database file:

Although the term "database" typically refers to a collection of related data tables, an Access database includes more than just data. In addition to tables, an Access database file contains several different types of database objects:
• Saved queries for organizing data,
• Forms for interacting with the data on screen,
• Reports for printing results,
• Macros and Visual Basic programs for extending the functionality of database applications. All these database objects are stored in a single file named <filename>.mdb. When we are running Access, a temporary "locking" file named <filename>. ldb is also created. We can safely ignore the *.ldb file; everything of value is in the *.mdb file.

Starting Access:
• To start Access, double click the Access icon (for version 8.0 and 7.0 or for version 2.0) from within Microsoft Windows.

If you are working in the Commerce PC Lab, you will be working with Access version 2.0. If you are working at home, you will able be to tell what version you are using by watching the screen "splash" as the program loads.

Creating a new database:
• Following figure(Fig-4) shows how to create a new database file called myfile.mdb.

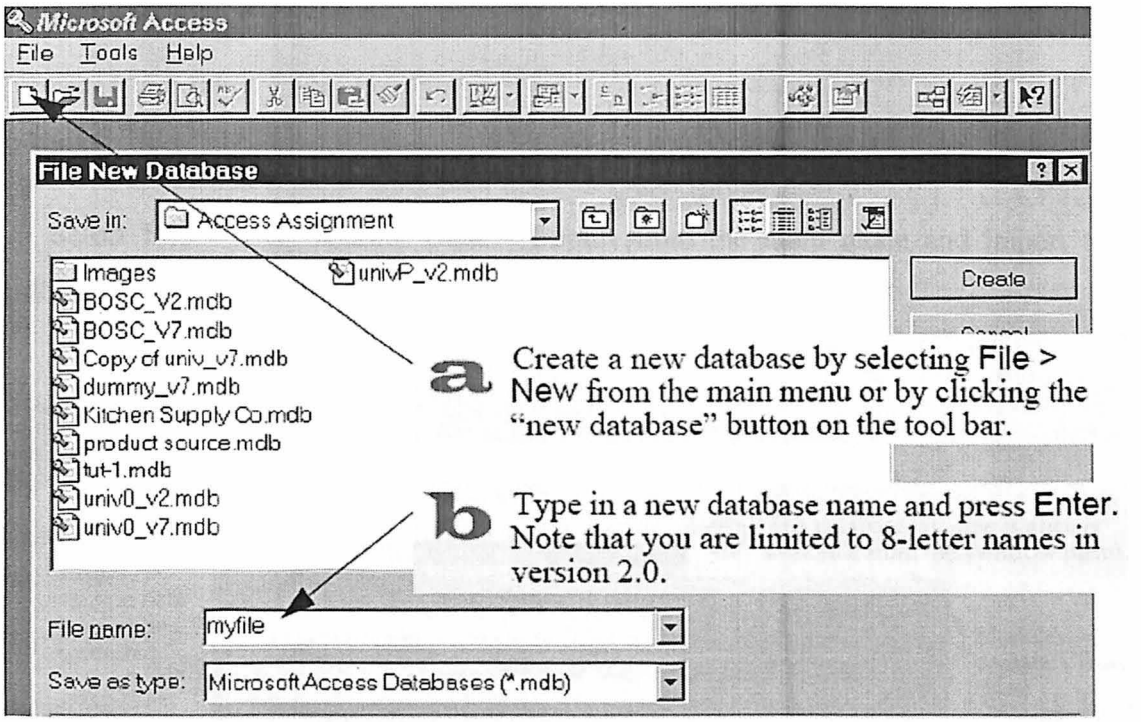


Fig-4: How to create a new database file called myfile.mdb

• Examine the main features of the database window— including the tabs for viewing the different database objects—as shown in (Fig-5).

Opening an existing database:

Although we can open a version 2.0 database with version 7.0, we cannot open a version 7.0 database with version 2.0. Importing and exporting across versions is possible, however.

If you are using version 8.0, you can use either univ0_v2.mdb or univ0_v7.mdb. When you open the file, Access will ask you if we want to convert it to version 8.0. Select yes and provide a new name for the converted file (e.g., univ0_v8.mdb)

• Open the univ0_vx.mdb file and examine the contents of the Sections table, as shown in (Fig-6).

Importing data from other applications:

Access makes it easy to import data from other applications. In this section, we will create a new table using data from an Excel spreadsheet.

• Select File > Get External Data > Import from the main menu and import the depts.xls spread sheet as a new table called Departments.

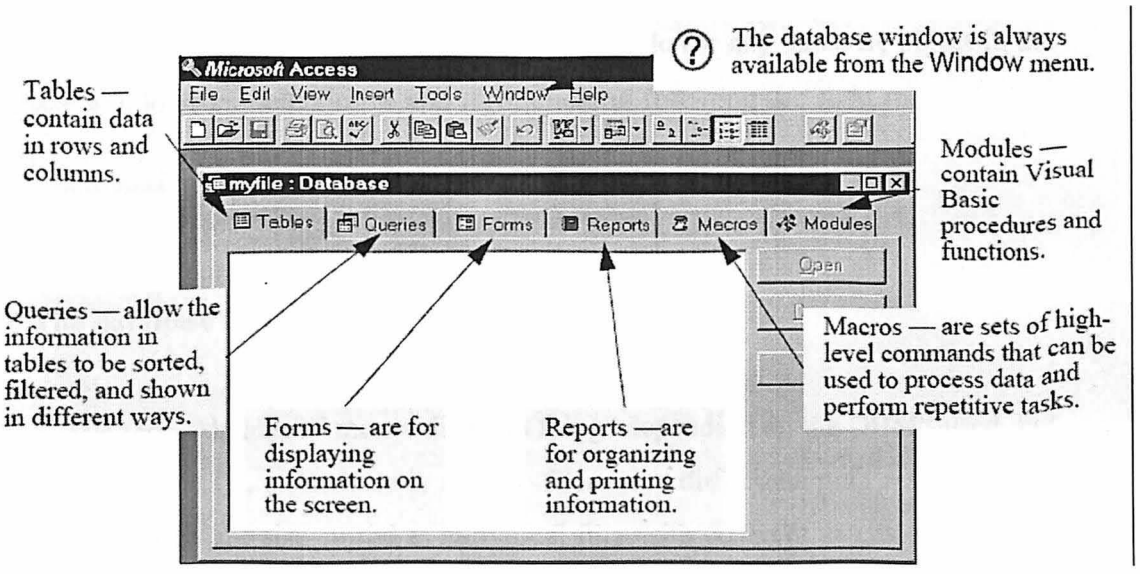


Fig-5:The database window contains all the database objects for a particular application.

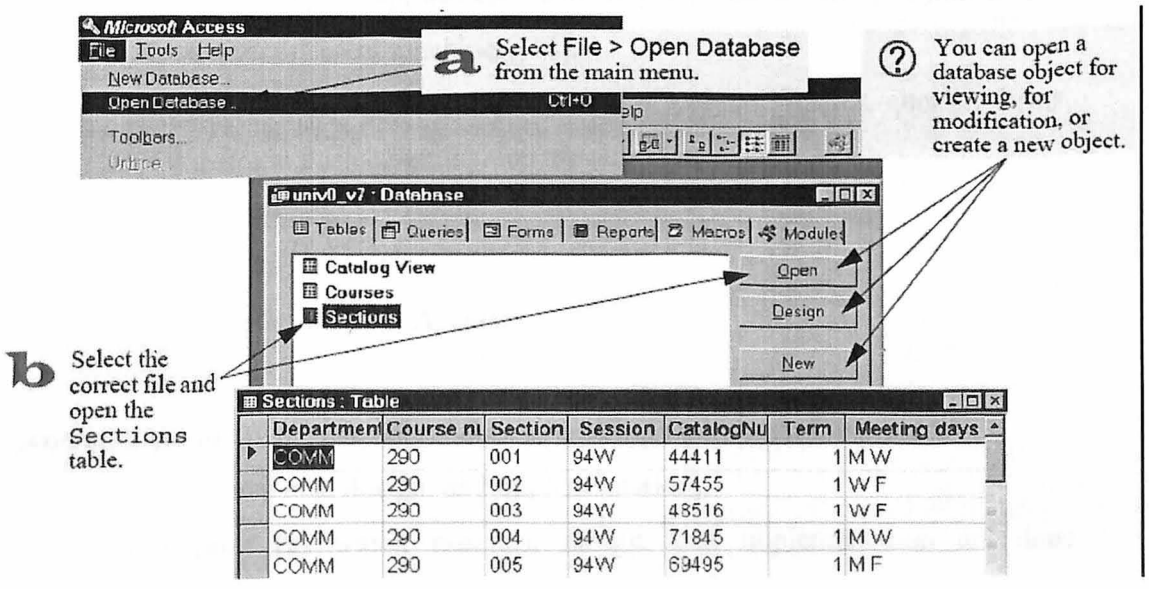


Fig-6: To Open the univ0_vx.mdb file for the version of Access that we are using and then open the Sections table

In version 2.0, the menu structure is slightly different. As such, we must use File > Import.

- Use the import wizard specify the basic import parameters. We should accept all the defaults provided by the wizard.
- Double click the Departments table to ensure it was imported correctly.

If we make a mistake, we can rename or delete a table (or any database object in the database window) by selecting it and right clicking (pressing the right mouse button once).

3.20 The database files in Access:

The term "database" means different things depending on the DBMS used. For example in dBase IV, a database is a file (<filename>.dbf) containing a single table. Forms and reports are also stored as individual files with different extensions. The net result is a clutter of files. In contrast, an Oracle database has virtually no relationship to individual files or individual projects. For instance, a database may contain many tables from different projects/applications and may also be stored split into one or more files (perhaps on different machines)[16].

Access strikes a convenient balance—all the "objects" (tables, queries, forms, reports, etc.) for a single project/application are stored in a single file.

3.21 Developing applications in Access:

In general, there are two basic approaches to developing information systems:

- in-depth systems analysis, design, and implementation,
- rapid prototyping (in which analysis, design, and implementation are done iteratively).

Access provides a number of features (such as graphical design tools, wizards, and a high-level macro language) that facilitate rapid prototyping. Since we are going to build a small system and since time is limited, we will use a rapid prototyping approach to build our application. The recommended sequence for prototyping using Access is the following:

1. Model the information of interest in terms of entities and relationships between the entities (this is covered in the lecture portion of the course).
2. Create a table for each entity.
3. Specify the relationships between the tables.
4. Organize the information in tables using queries.
5. Create forms and reports to support input and output transactions.
6. Enhance the forms with input controls.
7. Create action queries, macros, or Visual Basic programs to perform the transaction processing functions of the application.
8. Create "triggers" (procedures attached to events) to automate certain repetitive tasks.

RESULTS AND DISCUSSION

4. RESULT AND DISCUSSION

BioRec is a bio record retrieving tool for Bioinformatics research record of O.U.A.T. It has integrated, text-based search and retrieval system used for the Research Records of Department of Bioinformatics.

As the software package is deployed to a PC, which is not connected to the server . Hence the IIS (Internet Information Service) is loaded, which act like a local server and like standalone format. We can connect to the home/default page of BioRec through any e-browser e.g.-Internet Explorer 5.01 or higher), Mozila (V 1.7.1) by giving following URL

http://localhost/biorec

Currently BioRec contains records for first two batches(2002 & 2003) of Students. Their details about research works during the dissertation at various research laboratories like IGIB, IIT, CCMB, IISC, ILS, Strand Genomics, IICT etc. are incorporated in this database.

As per the user's requirement, here many options are available in this tool for retrieving the required information. Various Search Selection options are available to retrieve the specific or required record from this database.

4.0 Search Selection:

- Search by any key word of the text of record or browse by subject terms.
- Search by student and guide name by selecting the option available at pull down menu.
- Search by area of work by selecting the option available at pull down menu.
- All the full text records are available in pdf formats and can be accessed by linking to more details option present in output display page.

4.1 BioRec Direct Search Queries:

This search finds citations that correspond to a specific key word category. The search may be either broad and sensitive or narrow and specific. Any word, number present in record can be found out by direct search.

This page provides the following specialized BioRec searches:

- Search by admission number i.e., 01bi/2002.
- Search by title of the dissertation work.
- Search by the Student Name.
- Search by the Scientist (Guide) name.
- Search by the Project Institute name.
- Search by the year of passing.
- Search by the e-mail id.
- Search by any key word of record.

4.2 Out put of the Direct Search Queries:

The output page will display the result of the query with the details about related records. User can get the required record by selecting from the number of related output records. It show the related records with the following details:

- Area of Research Work.
- Student and Guide name.
- Project Title.
- The details about the records as:

1. Admission number of the student.
2. Name of the student.
3. Bachelor Degree.
4. E-mail id.
5. Year of joining and year of pass out.
6. The name of the Scientist (Guide).
7. E-mail of the scientist.
8. Address of the scientist or the research carried out.

9. The website of the research institute.
10. Complete contact numbers like phone, fax numbers.
11. The title of the dissertation work.
12. Abstract of the dissertation work.
13. Summary of the dissertation work.
14. Aim and objective behind the dissertation work.
15. Introduction of work.
16. Final conclusion of the work.

- More details (This is linking to full text of the research record of the student available in pdf format). After clicking the link, the full text pdf file will be down load.

4.3 Search by Student and guide selection category:

This search finds citations that correspond to a specific record category. The search may be either broad and sensitive or narrow and specific. The search filters are more specific because it takes specific record or category. This search finds citations and abstracts related to various topics in Bioinformatics Research Records. For detail about record it has a link to full text record in pdf format of research record.

4.4 Search by Area of Work selection category:

This search is similar search like Student and Guide Search which finds citations that correspond to a specific record category. The search may be either broad and sensitive or narrow and specific. The search filters are more specific because it takes specific record or category. This search finds citations and abstracts related to various topics in Bioinformatics Research Records. For detail about record it has a link to full text record in pdf format of research record.

4.5 Some of the screen shots are below to describe the properties of each page:

Home/Default Page:

fig-7: Home Page/ Default Page

This page is default page, downloaded by putting the URL http://localhost/biorec in address box of internet explorer. This front end form contains some links like:

- OUAT (links to http://www.ouat.ac.in).
- DBI (links to P.G.Department of Bioinformatics OUAT).
- BIOREC (links to details descriptions about BioRec).
- CONTACT(links to Contact information of BioRec).

In this page registration / entry form to database is available, which allow the free registration for first time user but after registration the registered users can enter the database through e.mail id. For first time user some options are mandatory like

- Yourself
- Profession
- Area

- E-mail id

After putting all the above information, the database will stored all log in information along with visitors number, time and date of log in. When same visitor put only e.mail id then this will allow him/her to enter to database but this time no information will be recorded which will save the space of database. As all the information about the registered users are available so the administrator can knows details about the users, whether they are students, research persons etc. Here in default page is properly navigated with home page, about biorec page, contact page, search page, Registration page. The home page is also same as default page.

About biorec:

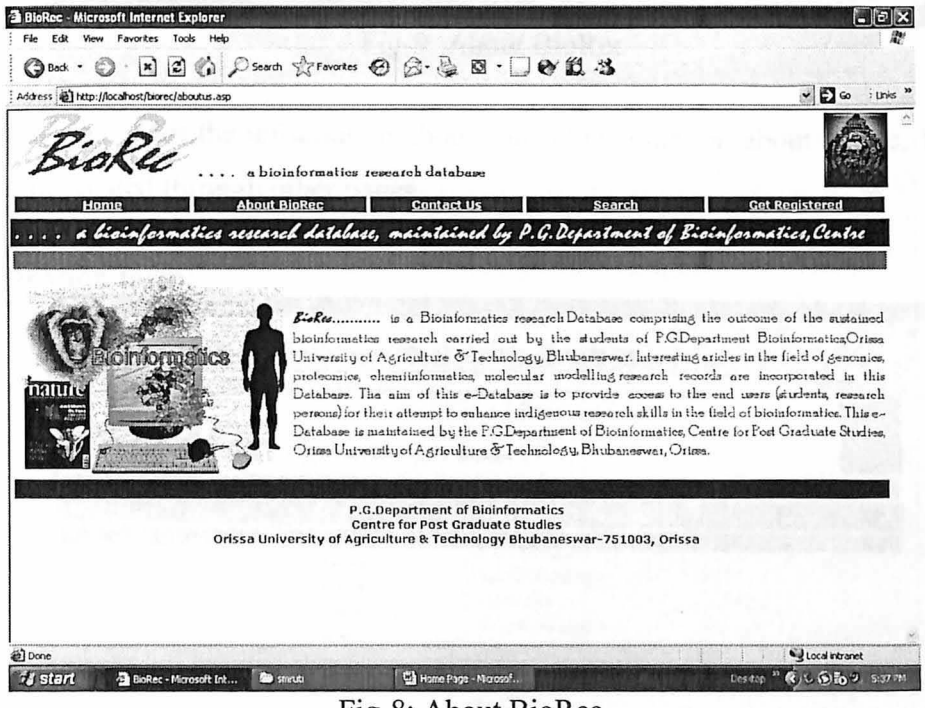


Fig-8: About BioRec

This page provides the information about biorec. Here also proper navigations are available to other pages.

Contact us:

Fig-9: About BioRec

This page provides the information about contact information about biorec. This page is also navigated through other pages.

Registration form:

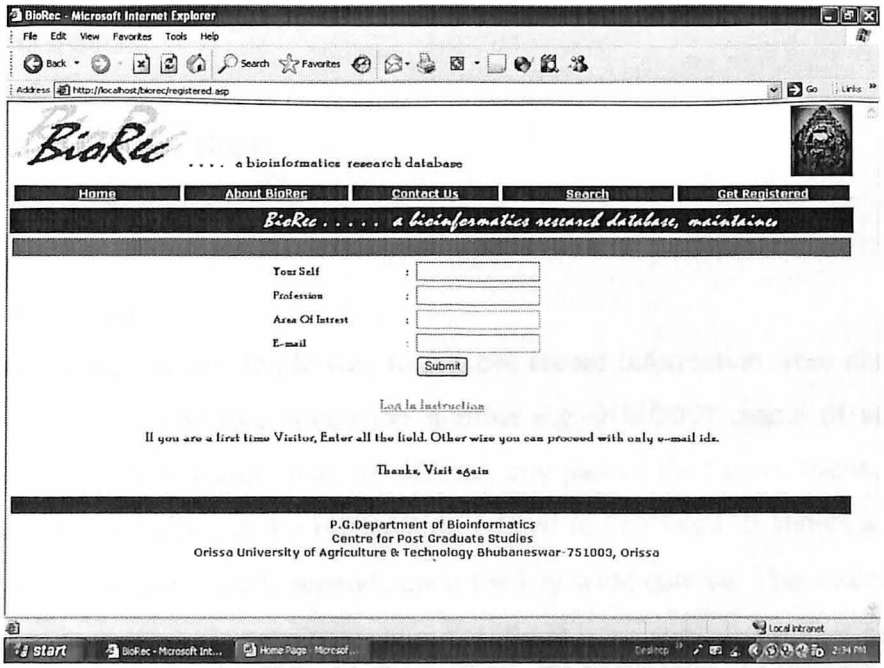


Fig-10: Registration Form

This page has a registration form for the first time user of biorec. This page is also properly navigated through other pages.

Direct search:

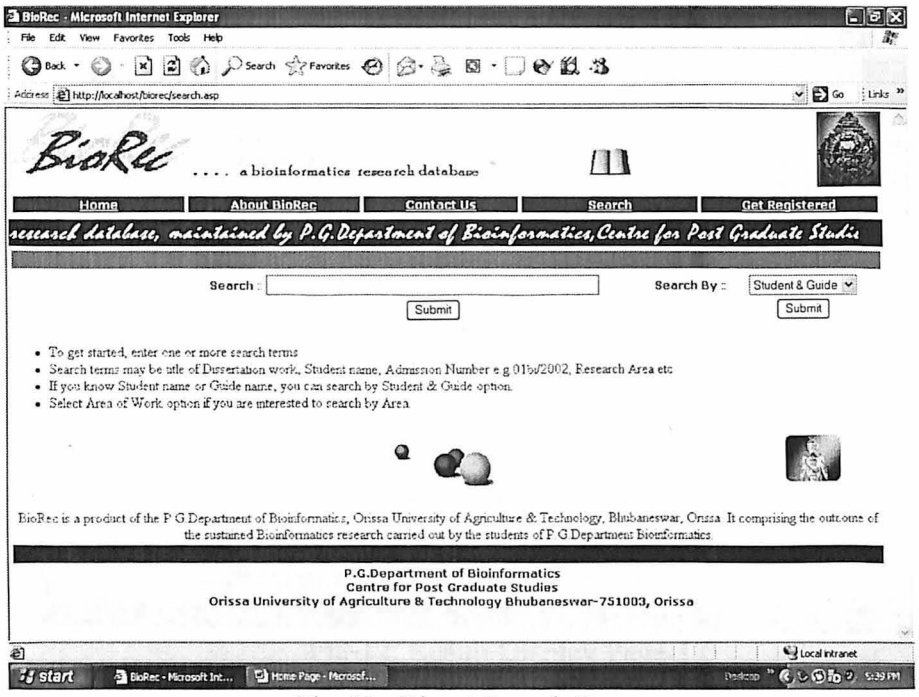


Fig-11: Direct Search Page

After home/default page this page will be downloaded. This page contain two search options.

1. Direct Search.
2. Search by:
    a. Student & Guide.
    b. Area of Work.

1. Direct Search.

   Direct Search is the simple way to retrieve record information from database. By any key word like admission number e.g.-01bi/2002 ,name of student, place of work, research institute address, any part of the title of thesis, guide name etc user can get the information related to key word. It shows a list of records or single record, depends upon the key word entered. This search does not contain any case sensitive character, So it is be very convenient for new users also. Any new user can also get right information according to key word entered.

Result display Page:

Fig-12: Result Display Page-1

This is the output page of direct search. This contains details about the student record information like:

- Area of Research Work.
- Student and Guide name.
- Project Title.
- The details about the records as:

1. Admission number of the student.
2. Name of the student.
3. Bachelor Degree.
4. E-mail id.
5. Year of joining and year of pass out.
6. The name of the Scientist (Guide).
7. E-mail of the scientist.
8. Address of the scientist or the research carried out.
9. The website of the research institute.
10. Complete contact numbers like phone, fax numbers.
11. The title of the dissertation work.

12. Abstract of the dissertation work.
13. Summary of the dissertation work.
14. Aim and objective behind the dissertation work.
15. Introduction of work.
16. Final conclusion of the work.

- More details (This is linking to full text of the research record of the student available in pdf format). After clicking the link, the full text pdf file will be down load and for this installation of Acrobat Reader is essential.

More details/Full text link:

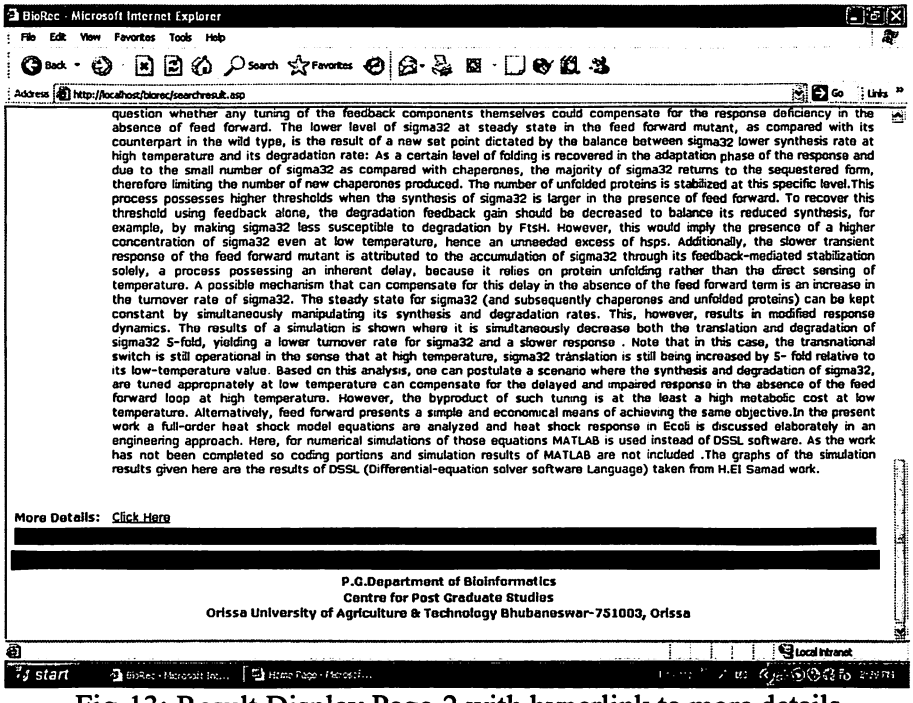


Fig-13: Result Display Page-2 with hyperlink to more details

Search by Student& Guide /Area of Work:

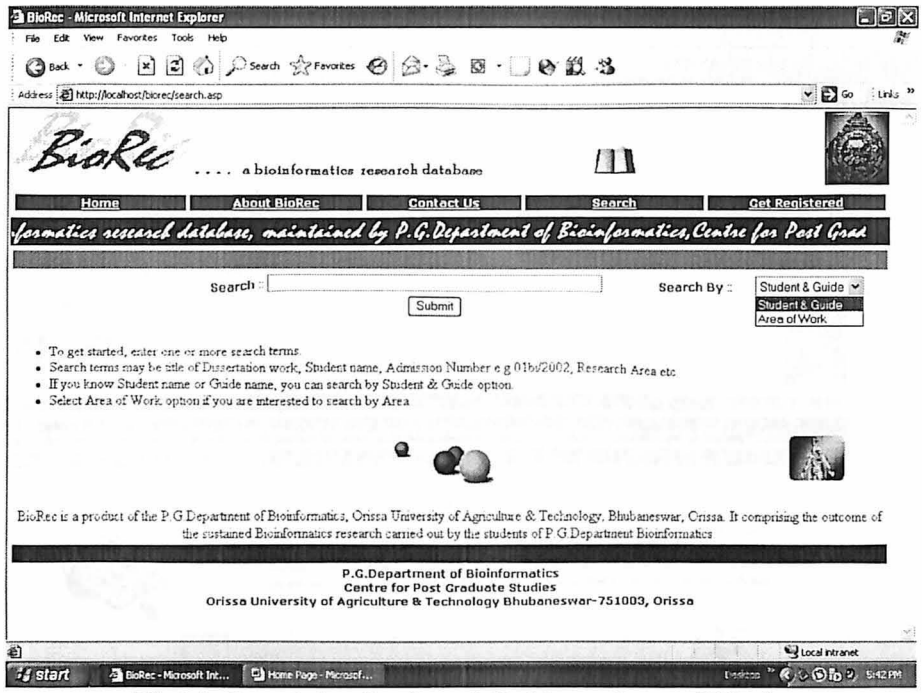


Fig14: Search by Student &Guide or Area of Work.

This page contain one more search option to retrieve the desired record. If the user knows about the student name, guide name or the area of work then he/she can directly retrieve the record to get the required information by this option. At right middle this page contains one combo box that contain option like search by:

a) Student & Guide.
b) Area of Work.

a) Student & Guide.

This is one of the short cut options to get the information. If user knows about the student name or guide name then he/she can directly access with the record. After selecting this option and clicking the submit button, one more page will be down loaded which shows the detail list of the student and guide names. After selecting any one of the name from the list and clicking the submit button, it will take to the specific record. The final output of this selection procedure is same as that of direct search, as it shows the details about the record information like direct search. This option is

very helpful to retrieve the record information, as it saves time along with short cut way to get the information.

Search by student/Guide:

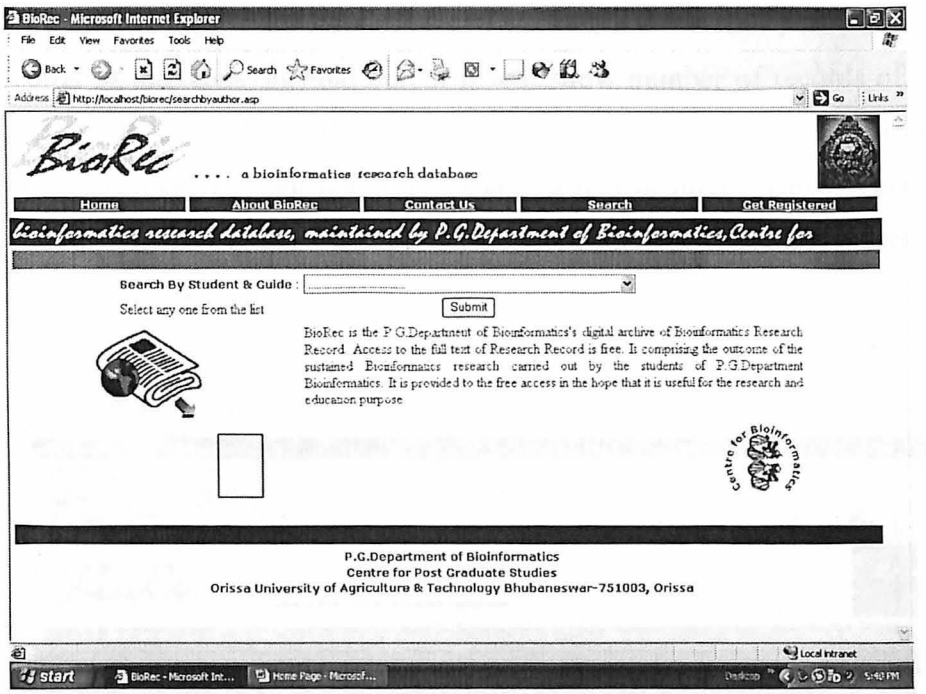


Fig-15: Search by Student & Guide.

Fig-16: Search by Student & Guide.

Search by Area of work:

This option is also same as that of search by Student & Guide option. After selecting this option from the combo box at search page and clicking the submit button, it will take to another page which contain the list of areas in combo box. After selecting any one of the area from list and clicking the submit button, it will take to the records of that area. In final output it will show number of records of that area. User can get the desired record by selecting and choosing any one of the record from the out put records. The out put is also same as that of direct search. This option is important as it is a shortcut way to get the information about a particular area.

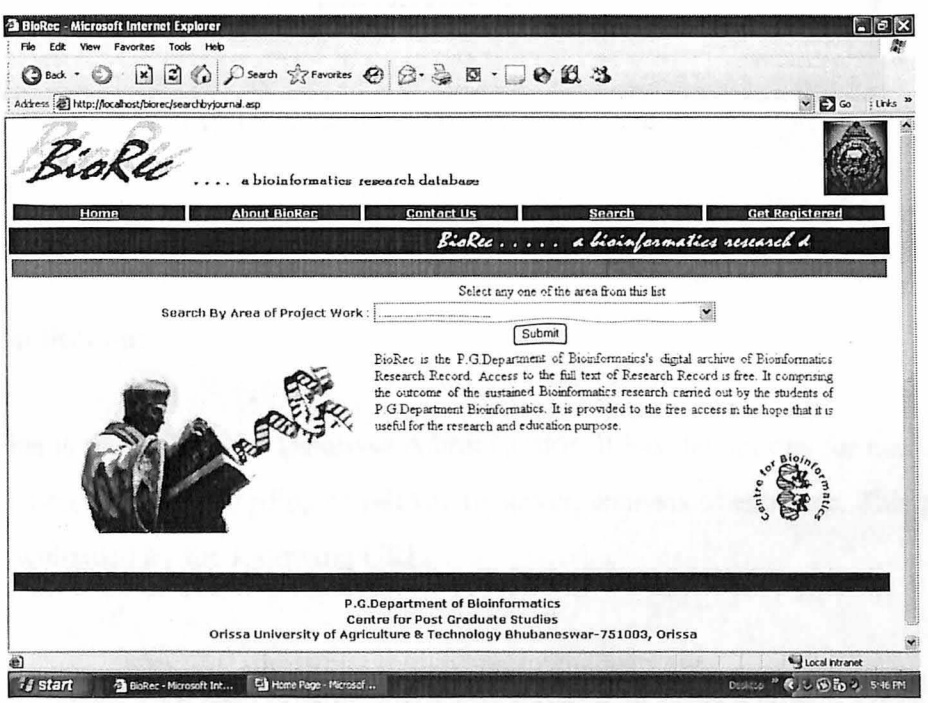


Fig-17: Search by Area of Work Page-1.

Fig-18: Search by Area of Work Page-2.

4.6 Admin Section:

This is the section for Database Administrator. It has the options for new data entry, deletion of data, uploading of pdf file to server, analysis of entry etc. This page can be downloaded by the following URL:

http://localhost/biorec/dynamic/manageart.asp

This page contain options like:

a) Add Record
b) Delete Record
c) Upload file of Record

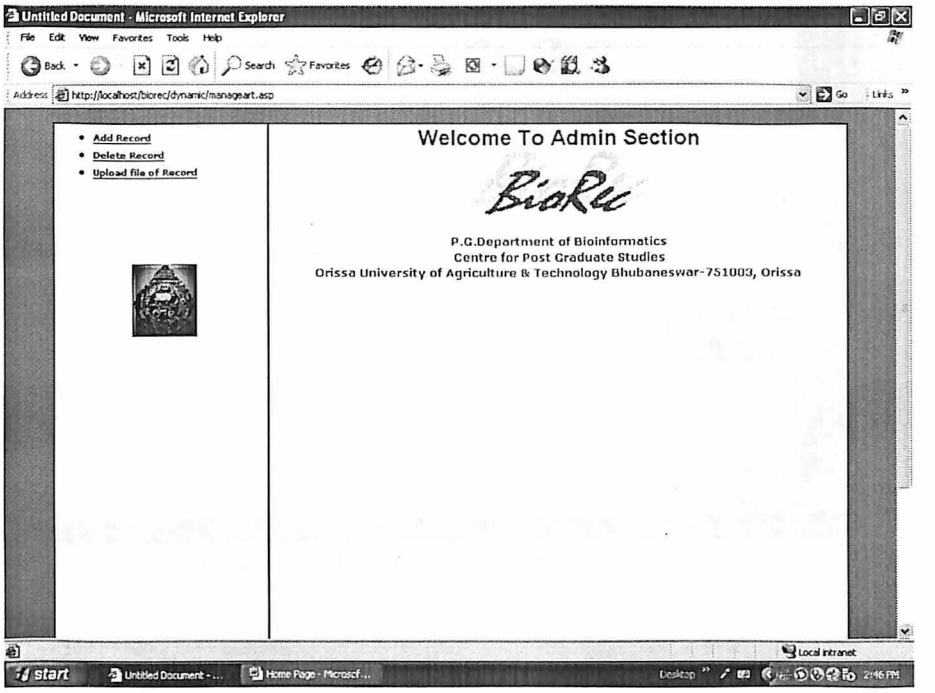


Fig-19: Admin Page.

Add Record:

After selecting add article option from Admin section this page will be downloaded. This is the main page for data entry in specific format, as a number of options are available for data entry for specific format. This form contain number of box for data entry like:

a) Area of Project Work.
b) Project Title.
c) Details of the record.
d) Date of Entry of record.
e) Student & Guide Name.
f) Name of PDF/WORD file to be upload.
g) Put in Home Page.

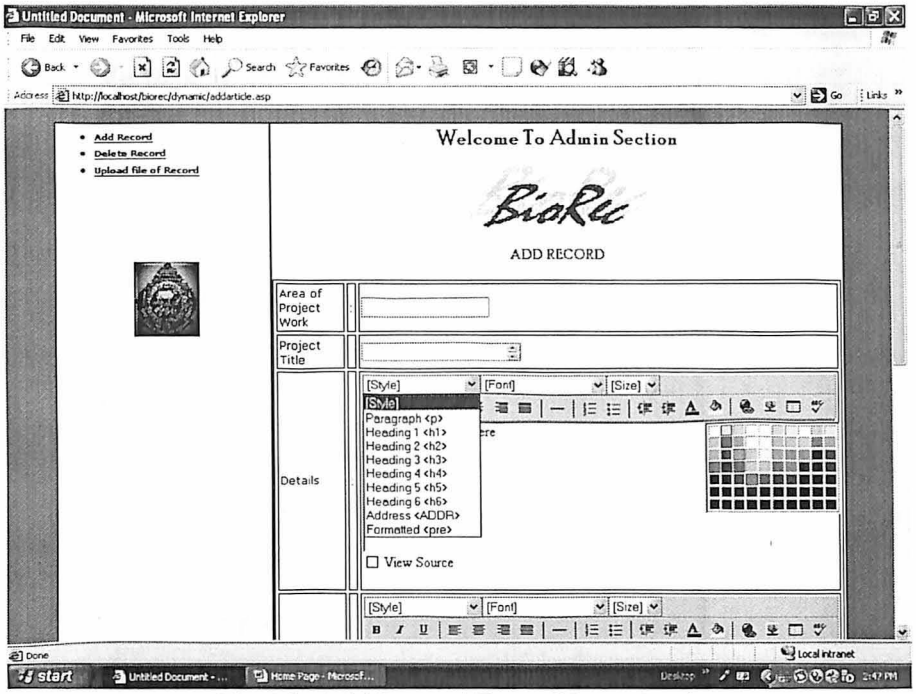


Fig-20: Add Record Form(Upper Part).

a) Area of Project Work.

This box allows to put the area of the project work of a specific record. After putting the record and final submission this new area record will be submitted in drop down menu of search by area.

b) Project Title.

The title of the dissertation work has to enter in this box. So that it will be displayed in output page of search result.

c) Details of the record 1

This is the important box as search conditions are depends on the data of this box. This box contains a wide range of tools for data entry in specific format . The tools are as follows:

1) Style.

It contains some options like Paragraph, Heading1, Heading2, Heading3, Heading4, Heading5, Heading6, Address, Formatted. By selecting any of the options the data can be entered e.g.-Heading, Paragraph etc.

2) Font.

It contain a list of four fonts Arial, Courier New, Times New Roman, Verdana in combo box.

3) Size.

This allows sizing the letter as seven options are present in combo box. We can change the height of letter with these seven options.

4) Bold/Italic/Underline.

By selecting any one or two or three from this selection we can change the letter type.

5) Align left, Align right, Center, Justify.

This allow to align the paragraph in specific format as in Microsoft Word.

6) Horizontal Rule.

This is for putting a line between two sections or paragraph.

7) Ordered List.

It helps to numbered the list with bullet option or number option.

8) Outdent /Indent

This option allow to align the paragraph.

9) Text Colour

Like in Microsoft Word we can change the text in to 63 types of colour by this option.

10) Background Colour

This allow to change the background colour with 63 options of colour.

11) Insert Link

By this option we can link a word or number to other website. Also this is very helpful to create a hyperlink to other website from existing page.

12) Add image

As MS Access database does not allow to put a image file inside the database. But we can store the path to that image in this database. This option help in selecting and putting the image path inside the database. By this option we can display a graph, image in output page

13) Insert Table.

Insert Table option can help to insert a table, with selection type as: Number of Rows, Number of Column, Table Width in percentage, Border Thickness in Pixels, Cell padding, Cell Spacing.

14) Spell Check.

This option allow to check the grammar as in MS Word.

15) View Source.

After entering the text data and proper alignment we can view the HTML code by clicking this option.

d) Details of the record 2

This box allow extra addition of text data.

e) Date of Entry of record.

Date of entry of record can be stored in database by this option. We can stored the date in following format:

mm/dd/yyyy

Also one more option is available. If we select date option a calendar will be downloaded and by selecting any one of the date, it will store the date in same mm/dd/yyyy format.

f) Student & Guide Name.

For searching record through Student & Guide Name, this box has been made. If we enter Student & Guide Name in this box, then it will display that Student & Guide Name in drop down menu of Search by Student & Guide list.

g) Name of PDF/WORD file to be upload.

Here the name of the file is to entered with proper extension of file type.

E.g.- For pdf file .pdf extension and for word file .doc extension is required.

h) Put in Home Page.

If we filled all the boxes and submitting, then the data will be stored inside database. But it will not displayed during search. To display the data in result page we have to select the above option during each record entry.

Reset button will clear the date from each box if any mistake will be found during data entry.

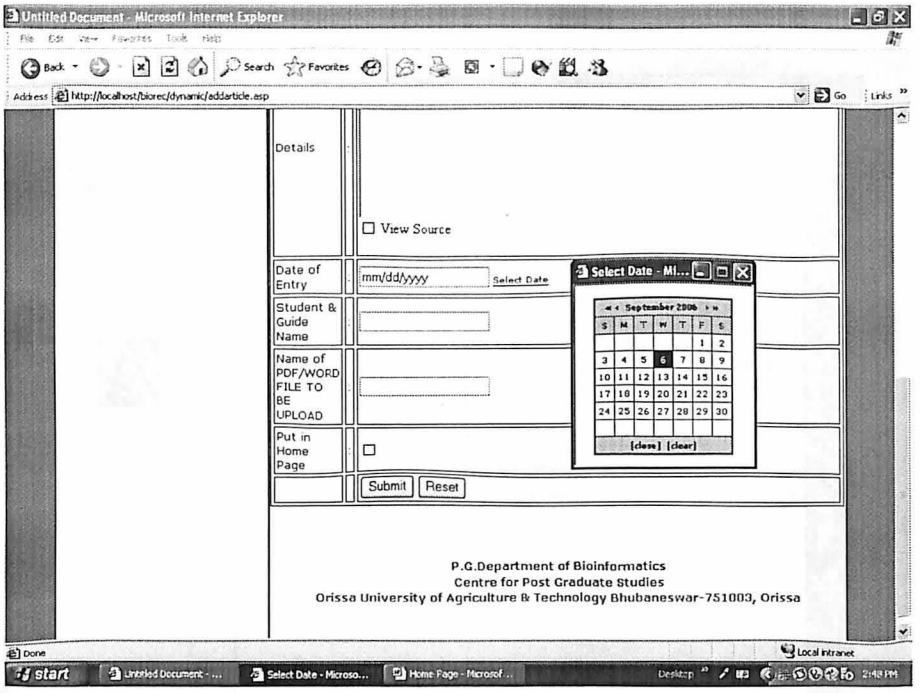


Fig-21: Add Record Form(Bottom Part).

Record deletion form:

This form will display the entire record list in a drop down menu. By selecting any one from this list and clicking Delete button, it delete the record from the list of records.

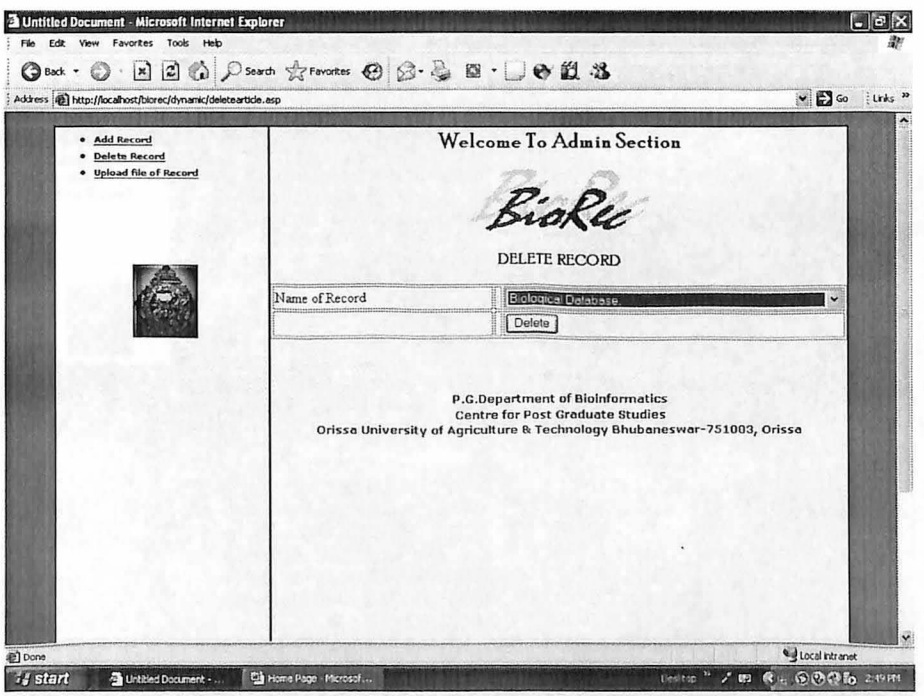


Fig-22: Delete Record Page (1).

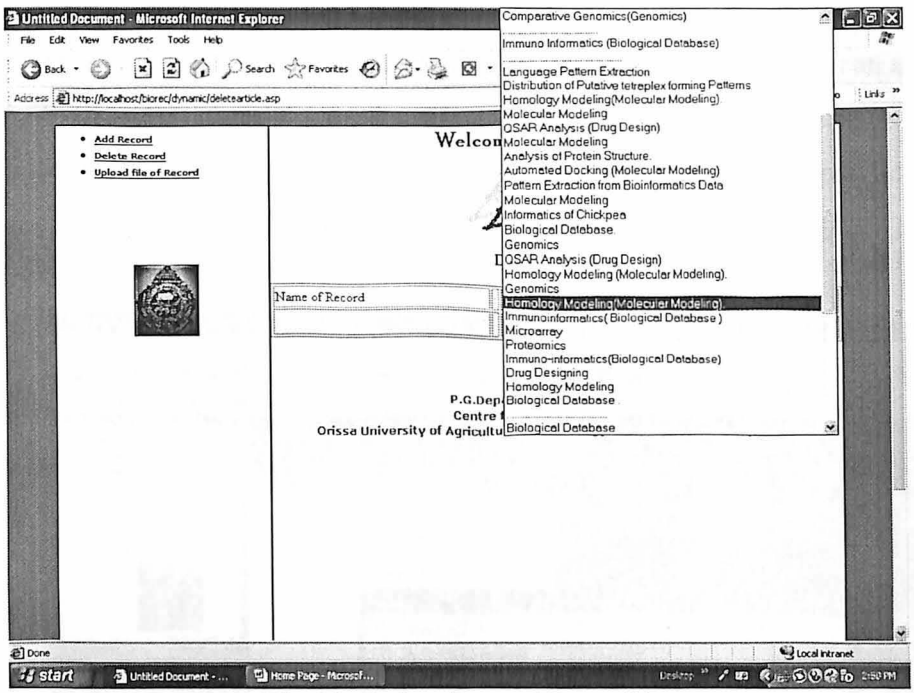


Fig-23: Delete Record Page (2).

File uploading form:

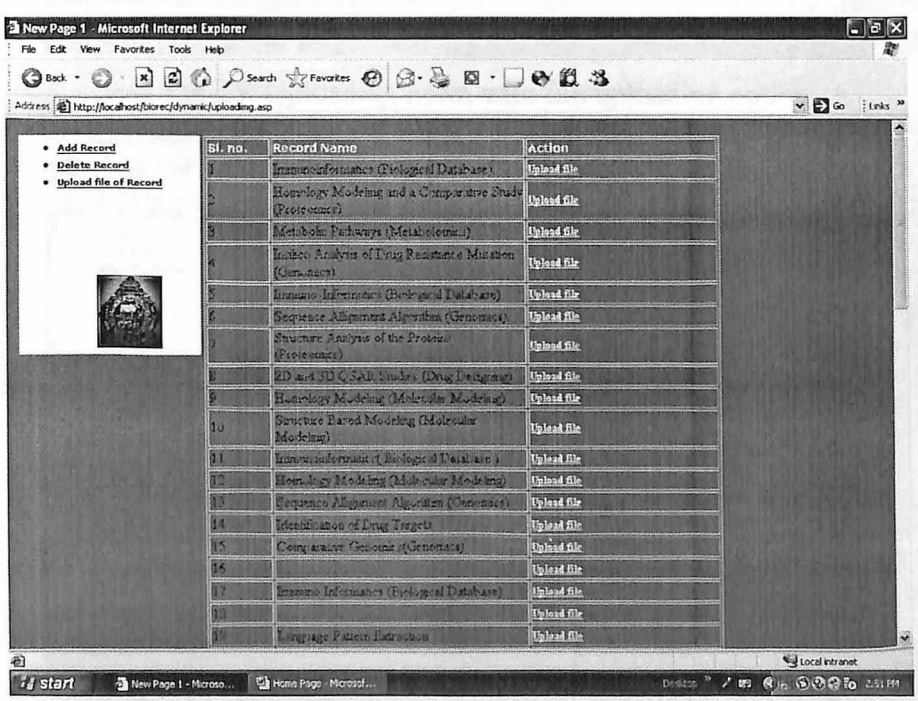


Fig-24: File uploading page with total list of records.

After selecting upload file of the record from Admin Section this page will be down loaded. This page display all the list of records in serial number. Currently this list contains 61 numbers of record for two batches(2002,2003). We can upload pdf/ word file from this page. If we select Upload file from action column, then up loading component will be downloaded. After browsing and selecting specific file from different drives of the computer and clicking upload file, it will upload that specific file into the server.

Fig-25: Uploading Form with file uploading option.

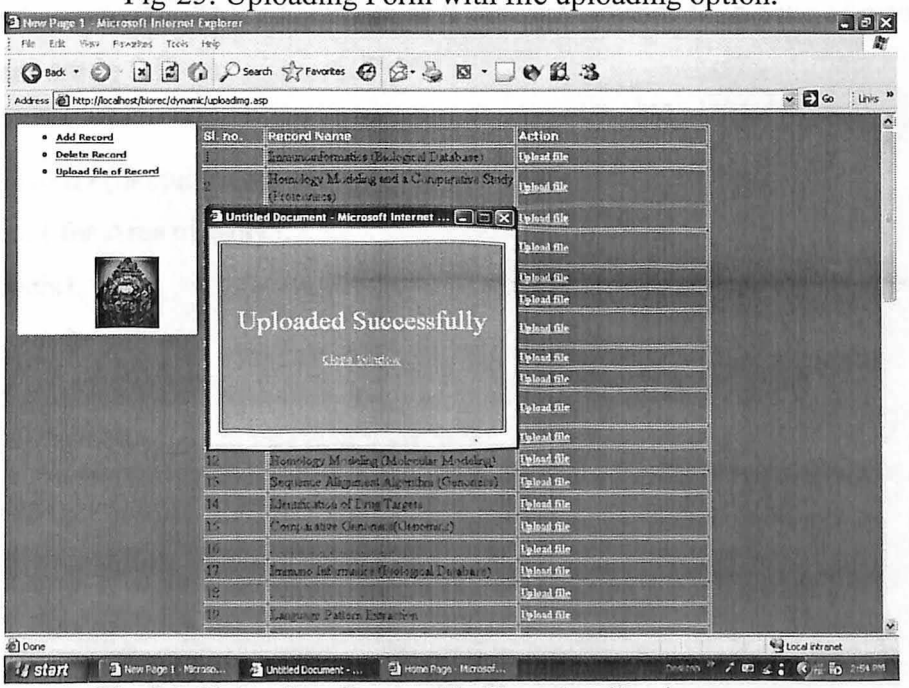


Fig-26: Uploading Form with file uploading in progress.

4.7 The backend database:

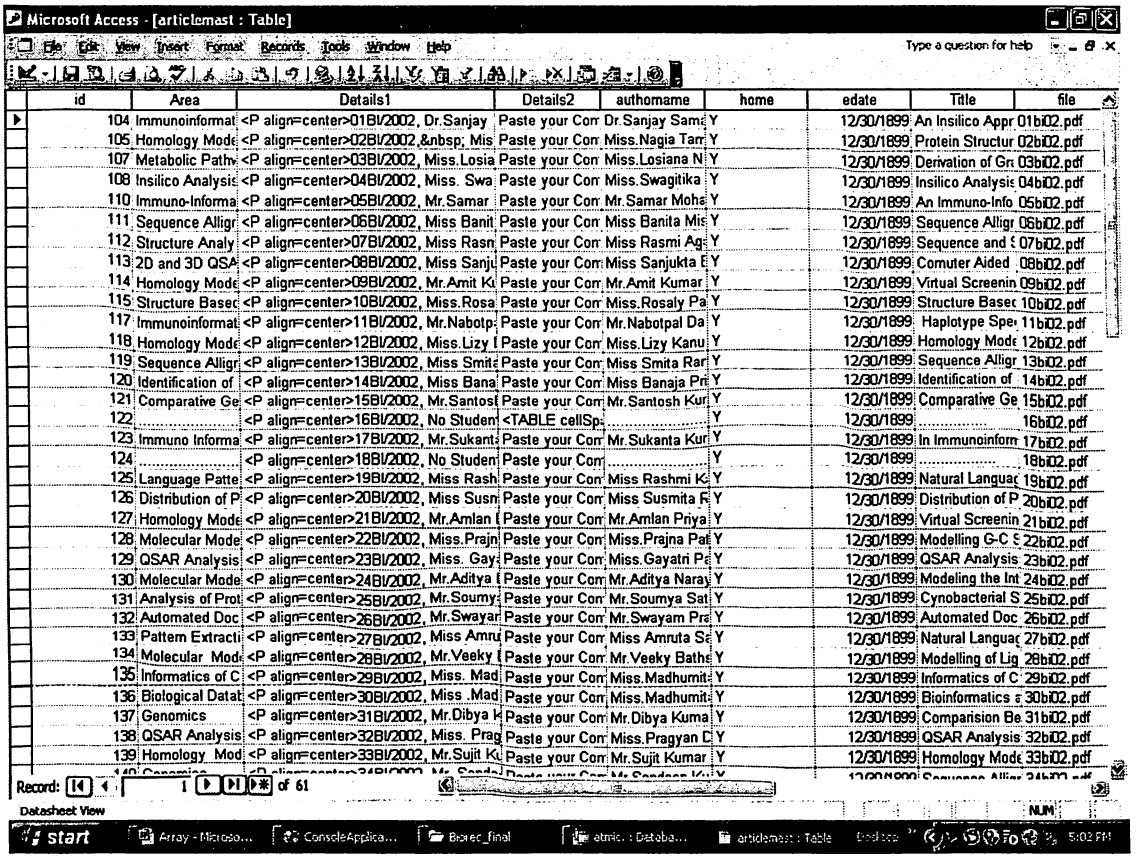


Fig-27: The table in Access.

We have constructed the database with Microsoft Access, as it is very user friendly and easy to use. This table contains 9 separate columns for storing of records. The columns are as follows

1. Biorec ID (default number),
2. Area (for Area of work),
3. Details1,
4. Details2,
5. Author names (for student & guide),
6. home (to put the home page or not),
7. edate (for date of entry),
8. Title (for title of dissertation work),
9. file (for uploaed file name),

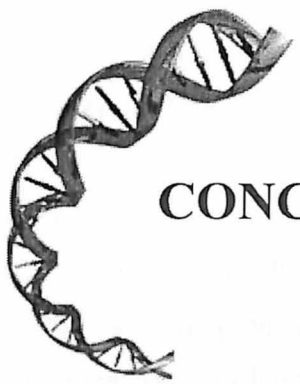
Currently we have 61 records for two batches with detail parameters as explained above.

4.8 Database Connection

A web application can communicate with database is through code running on the web server. So after construction of database and front end form, we have connected the database with declaration of the following string.

Connection string:

```
<%set con=server.createobject("adodb.connection")
con.open "DRIVER=Microsoft Access
Driver(*.mdb);DBQ=c:\inetpub\wwwroot\biorec\dynamic\atmic.mdb"%>
```



CONCLUSION

5. CONCLUSION

The traditional library services are quite affected due to the emergence of new technologies i.e. Internet and bioinformatics text based database for the dissemination of biological information. The biological science librarians and research scientists are well aware of the usefulness of the biological research record for accessibility of timely and required information. Bioinformatics text based websites are global source of information and it revolutionized the way whereby researchers and medical professionals access information. Biorec provides huge amount of information about research activities at different leading laboratories like IGIB, IIT, IISC, CCMB, IICT, Strand Genomics etc. Currently BioRec is not available on web, but it can be accessed through local internet as some of it's development phages are not completed. But any Research professional can get the information in the Bioinformatics Research area with help of easily accessible, user friendly, bio record retrieving tool Biorec. Biorec has the potential of the becoming an important tool for better bioinformatics education, research and maintenance of bioinformatics record in future.

REFERENCES

6. REFERENCES

[a] Frank K Brown 'Chemo-informatics: what is it and how does it impact drug discovery.' Ann. Rep. Med. Chem. 1998, 33 , 375-384.

[b] Chem. Inf. Lett. 2003, 6 , 14.

[c] Hopfinger, 1981; Ugi et al., 1990.

[d] Inderjit IK: Medline. The National Medical Journal of India,2001,14,1pp,50-53.

[e] Lindberg, Donald AB et-al. Use of MEDLINE by physicians for Clinical Problem Solving JAMA 193, 26, 24pp,3124-3129.

[f] Inderjit IK: Biomedical databases. The National Medical Journal of India, 2003, 16, 2pp,100-104.

[g] CPSC 44b – Senior Project Design,Prof. Arosif Mohmood, Hrushikaran Davda, Jignesh Patel, April 7, 2004.

[h] Raj Kumar, Medical Websites: Potential Tools for Education, Research and Patient Care in India, IFLA, 22-27 August 2004.

Website Visited:

[1] http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html

[2] http://www.medline.com

[3] http://www.igm.nlm.nih.gov

[4] http://www.nlm.nih.gov.PubMed

[5] http://www.ouat.ac.in

[6] http://www.geocities.com/bioinformatics web/defination.html

[7] http://www.kbrin.a-bldg.louisville.edu

[8] http://www.microsoft.com

[9] http://www.codeproject.com/dotnet/DotNetWhitePaper.asp

[10] http://www.limas.com/innovation/tech_update.php

[11] http://www.datahousecorp.com/eng/technology/netframework.htm

[12] http://www.msdn2.microsoft.com/en-us/library/zw4w55w.aspx

[13] http://www.msdn.microsoft.com/library/en-us/cpguide/html

[14] http://www.microsoft.com/india/msdn/chat/transcripts/default.aspx

[15] http://www.academicdb.com/microsoft_access.html

[16] http://www.mis.bus.sfu.ca/tutorials/MSAccess/tutorials/intro.pdf

[17] http://www.ncbi.nlm.nih.gov/entrez

[18] http://www.pubmed.gov

[19] http://www.bioinformatics.org

[20] http://www.hfc.nps.gov

Books Referred:

[20] Matthew MacDonald, The Complete Reference ASP.NET, TATA McGRAW-HILL Edition, 2005.

[21] Steven Holzner, Visual Basic .NET Programming, Black Book, 2005 Edition.

[22] Matthew MacDonald and Mario Szpuszta, Pro ASP.NET 2.0 in C# , Apress, 2005 Edition.

[23] S.C.Rastogi, Namita Mendiratta, Parag Rastogi, Bioinformatics Concepts, Skills &Applications, CBS Publication, Edition 2004.