International Journal of Computing Science and Information Technology, **2014, Vol.2(4), 01-05**
ISSN: 2278-9669, July 2014 (http://ijcsit.org)

Some Studies on Relationship between BioTechnology Databases & Bioinformatics

N.Deepak Kumar [1] Email: nagiletideepakvictor@gmail.com,
Dr.A.Ramamohan Reddy [2] Email: ramamohansvu@yahoo.com,
[12]Dept.Of Computer Science &Engineering, SVUniversity,
Tirupati-517502,India,

Abstract

Bioinformatics represents an interdisciplinary compiling of biosciences and biotechnology with mathematics and informatics. Databases, especially factual databases as for example nucleotide and protein sequence databases, are one of the most important tools in bioinformatics. Bioinformatics includes different subject areas, e.g., development of databases and software. The use of sequence and structure information stored in databases opens together with new concepts, such as neural networks, new approaches in molecular modeling for structure prediction and description, theories and methods of 3D-modeling and of optimizing macromolecules, knowledge-bases sequence analysis and prediction of protein folding, development of knowledge-based systems and artificial intelligence methods for applications in genome research, protein design, drug design, etc.
.

Keywords: Bioinformatics,database,protein,entity,Attribute

1. Introduction

The availability of high-quality, up-to-date and comprehensive information is an important requirement in biotechnology and its applications in the ever widening fields of medicine, pharmacy, agriculture, food industry and the environmental sciences. Advances in biotechnology, especially in genome research, depend to an increasing extent on which required information is made available and how it is used. The growing amount of data, especially the genome projects, deliver an enormous number of sequence data, meaning that collecting, processing and disseminating these data is only possible with the help of modern information technology and international co-operation[1].

Modern biotechnology is highly information-dependent and uses a wide variety of information sources and information technologies[2]. The consequences of rapid developments in biotechnology with its tremendous volume of data and in informatics with its potential for processing and using date are:

1. In relation to research –
    - the creation of a new scientific field: Bioinformatics
2. In relation to infrastructures –
    - very large databases and smaller, more highly specialized databases
    - highly sophisticated software for processing and using these databases
    - efficient communication networks for access to databases and information exchange
    - establishment of information centers for collection, processing and distribution of information
    - comprehensive information services.

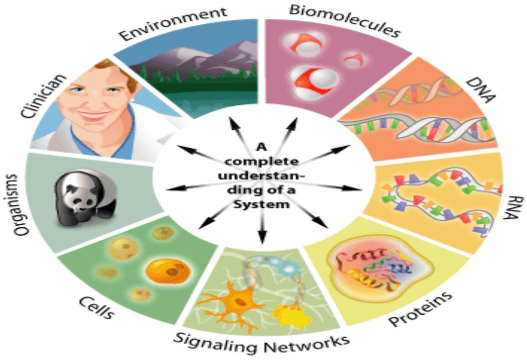


Fig 1.Complete understanding of Bioinformatics

Many databases are available in various types of media: Online via a number of networks and hosts, or on CD-ROM, diskette, magnetic tape, or as a printed version. The overall growth in the online database industry during the past year can be traced through the statistics: 300 online databases were registered in the 1979 edition of Directory of Online Database[5]. In the 1993 edition, the

Gale Directory of Databases registered profiles more than 5200 online databases, among them about 250 with relevance to biotechnology, and more than 3200 database products offered in portable form, among them about 150 with relevance to biotechnology (MAR-CACCIO, 1993). The number of records stored in these databases increased from 52 million (1975) to about 5 billion (1993)[5,7].

system of image segmentation by heuristic approach and Hough transform. This method uses HT transform for identification of curves, lines and circle in coin image. From the acquired image, coin area is identified which can also be used for classification of coin. This method gives 97% accuracy during coin recognition. Fukumi et al [5], tried to achieve 100% accuracy for recognition of coins. They have used 500 yen coin and 500 won coin. In this work they have used Back Propagation (BP) learning and Genetic Algorithm (GA) to design a neural network based coin recognition system. After training the network using BP, GA is used to reduce the size of network by varying the architecture to achieve 99% recognition accuracy rate.

In this paper, we propose a novel approach for coin recognition based on statistical texture features using gray level co-occurrence matrix (GLCM) and sweep line algorithm. The rest of this paper is organized as follows. In section 2, we give a brief introduction of GLCM, Sweep line algorithm and Circle drawing algorithm. Section 3 deals with the proposed coin recognition method. Experimental results are discussed in section 4 and conclusion is presented in section 5.

2. Methodology

With the aim to improve the development of bioinformatics in Europe, in November 1990 the European Chemical Industry Foundation (CEFIC) recommended to the Commission of the European Communities the establishment of European Nucleotide Sequence Center. After long discussion, in 1993 the decision has been made to locate the new European Bioinformatics Institute (EBI) at the Genome Campus near Cambridge in the UK. The EBI will work to:

- ensure that the Data Library continues, develops in response to advancing biological research, and is capable of exploiting advances in informatics;
- reinforce areas which have been neglected, particularly training and user support;
- make the voice of European bioinformatics heard in the global arena;
- increase the effectiveness of dispersed, high-quality European research and service by entering into collaborations with centers of expertise throughout Europe.

The EBI will provide the EMBL Nucleotide Sequence Database and the SWISSPROT Protein Sequence Database; support and distribute other databases in collaboration with European scientist; help to coordinate the EMBnet nodes and molecular biology network services; carry out research and development in the application of information technology, actively tracking advances to explore their utility to biological problems; provide quality user support; and ensure that Europeans are globally competitive in the profession of bioinformatics.

U.S. activities in bioinformatics:

In connection with the U.S. Human Genome Program and the U.S. Plant Genome Research Program may large-scale bioinformatics projects were initialized, especially related to the development and improvement of databases for sequence and map data, and software for sequencing and mapping as well as for the creation and use of databases .

2.1 ER- Diagram

ER diagram is a graphical representation of entities and their relationships to each other, typically used in computing in regard to the organization of data within databases or information systems. An entity is a piece of data-an object or concept about which data is stored.A relationship is how the data is shared between entities. There are three types of relationships between entities:

A. One-to-One

One instance of an entity (A) is associated with one other instance of another entity (B). For example, in a database of employees, each employee name (A) is associated with only one social security number (B).

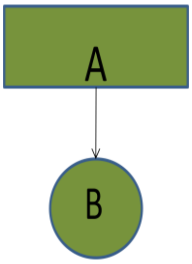


Fig 2.One-to-One

B. *One-to-Many*

One instance of an entity (A) is associated with zero, one or many instances of another entity (B), but for one instance of entity B there is only one instance of entity A. For example, for a company with all employees working in one building, the building name (A) is associated with many different employees (B), but those employees all share the same singular association with entity A.

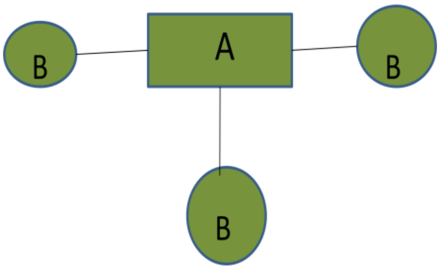


Fig 3.One-to-many

C. *Many-to-Many*

One instance of an entity (A) is associated with one, zero or many instances of another entity (B), and one instance of entity B is associated with one, zero or many instances of entity A. For example, for a company in which all of its employees work on multiple projects, each instance of an employee (A) is associated with many instances of a project (B), and at the same time, each instance of a project (B) has multiple employees (A) associated with it.

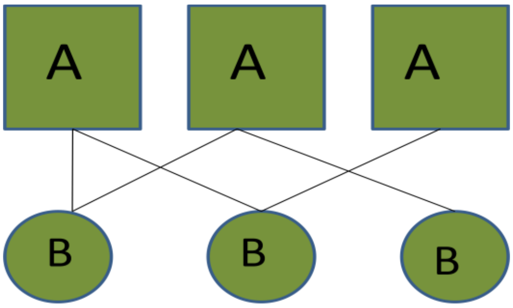


Fig 3.Many-to-Many

2.2 Example for Entities & Attributes

1)are real world objects
ex: Protein
2)contain attributes
ex: Protein-id,Protein-class-name.

below drawn as rectangle boxes that holds the name of the entity and attribute in two different notations as there is no standard!

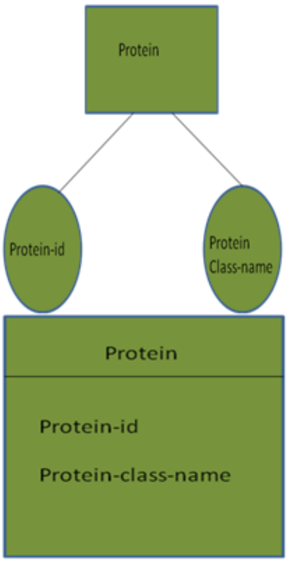


Fig 4.Entity-Attributes Relationship

2.3 Relationship

a)Relationships provide connections between two or more entities
ex: Which protein were used in which attributes
b)When two entities are involved in a relationship, it is known as binary relationship.
c)When three entities are invoved in a relationship, it is called as ternary relationship.
d)When more than three entities are involved in a relationship, it is usually broken in to one or more binary or ternary relationships.
e)are drawn as a line linking the involved entities as.

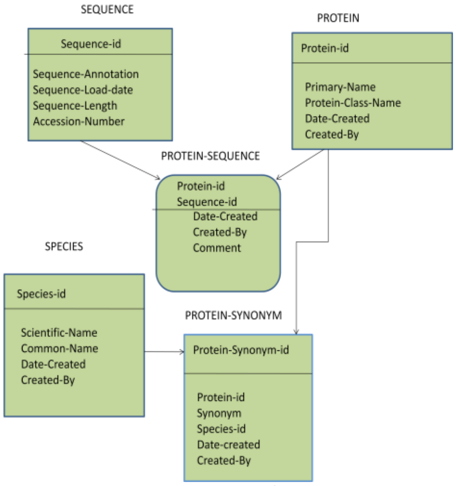


Fig 5. Example ER-Diagram for protein database

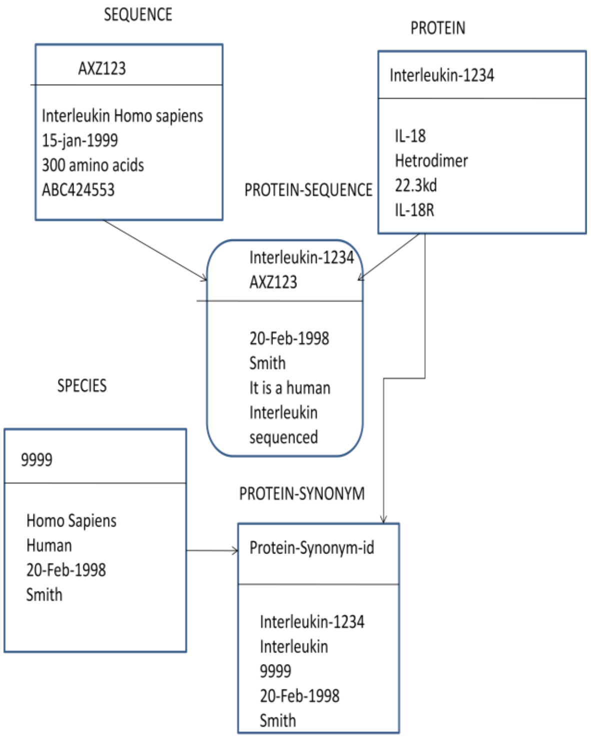
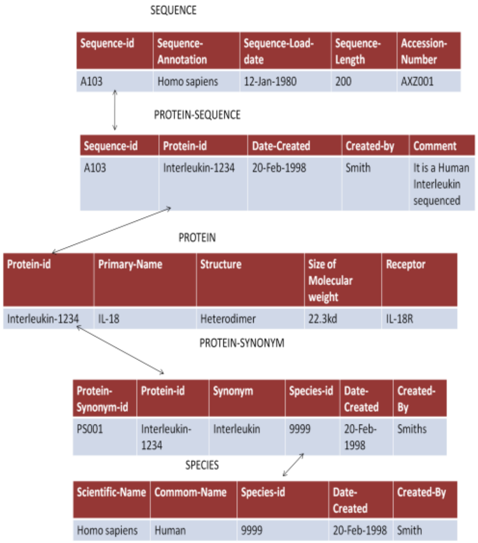


Fig 6. Example values for above Protein Database ER-Diagram


SEQUENCE

Sequence-id	Sequence-Annotation	Sequence-Load-date	Sequence-Length	Accession-Number
A103	Homo sapiens	12-Jan-1980	200	AXZ001

PROTEIN-SEQUENCE

Sequence-id	Protein-id	Date-Created	Created-by	Comment
A103	Interleukin-1234	20-Feb-1998	Smith	It is a Human Interleukin sequenced

PROTEIN

Protein-id	Primary-Name	Structure	Size of Molecular weight	Receptor
Interleukin-1234	IL-18	Heterodimer	22.3kd	IL-18R

PROTEIN-SYNONYM

Protein-Synonym-id	Protein-id	Synonym	Species-id	Date-Created	Created-By
PS001	Interleukin-1234	Interleukin	9999	20-Feb-1998	Smiths

SPECIES

Scientific-Name	Commom-Name	Species-id	Date-Created	Created-By
Homo sapiens	Human	9999	20-Feb-1998	Smith

Fig 7. Design database for above Example ER-Diagram

3. Results

With the large databases we have a huge information potential at our disposal, but until now we are not able to utilize it to a high degree. It is necessary to improve the use of databases which depends on the tasks:

- **In research:**

  Databases serve as a retrieval tool,
  as a research tool (e.g., protein design, drug design),
  as an evaluation tool (e.g., Science citation Index).

- **In industry:**

  Databases serve as a retrieval tool,
  as an aid in decision-making and awareness of competitors,
  as tools for product and technology information and evaluation,
  as tools for sales force productivity.

- **In the public eye:**

  Databases serve to help increase the public perception of biotechnology.

4. Conclusion

1. Creation of a new information consciousness and information behavior

Supporting the development of a new attitude and relationship to information:

- Information is an absolutely necessary component of research
- Information has its price.

2. Interdisciplinary cooperation

Recognizing the interdisciplinary character of biotechnology, increased attention is being devoted to establish

- a close interdisciplinary cooperation among scientists from different fields;
- a broad linkage of biosciences and biotechnology with efficient informatics methods;
- a close interdisciplinary cooperation of all participants in research, industry, government, and in public;
- productive interaction between producers and users of biotechnology information.

References

[1] ALSTON, y., COOMBS, j. (1992), Biosciences, Information Sources and Services, New York: Stockton Press.

[2] CRAFTS-LIGHTLY, A. (1986), Information Sources in Biotechnology, Weinheim: VCH Verlagsgesell-schaft..

[3] GRINDLEY, J.N., BENNETT, D.J. (1993), Public perception and the socio-economic integration of biotechnology, in: Biotechnologia 20, 89-102.

[4] LŰCKE, E.-M., POETZSCH, E. (1993), Biotechnology Directory Eastern Europe, Berlin-New York: de Gruyter.

[5] MARCACCIO, K. Y. (1993), Gale Directory of Databases, Vol. 1: Online Database, Detroit: Gale Research Inc.

[6] MEWES, H.-W. (1990), Workshop Computer Applications in Biosciences, Book of Abstracts, p. 11, Martinsried.

[7] POETZSCH, E. (1986), Faktographische informationsfonds zur Biotechnologie, Berlin: WIZ.

[8] POETZSCH, E. (1988), verzeichnis faktographischer Datenbasen zur Biotechnologie, Berlin: WIZ.

[9]. POETZSCH. E. (1993), Bio Technologie Das Jahrund Adreβbuch 93/94, Berlin: polycom Verlags-gesellschaft.

[10] Chen, P. P. and Li, M., "The Lattice Concept in Entity Set," in: Entity-Relationship Approach, S. Spaccapietra (ed.), North-Holland,Amsterdam, 1987, pp. 311-326.

[11] Chen, P.P., "The Entity-Relationship Model: Toward a Unified View of Data," ACM Trans. on Database Systems, Vol.1, No.1, March 1976, pp. 1-36.

[12] Chen, P. P., "Database Design Using Entities and Relationships," in: S. B. Yao (ed.), Principles of Data Base Design, Prentice-Hall, NJ, 1985, pp. 174-210.