INTRODUCTION TO BIOINFORMATICS

مقدمة في علم المعلومات الحياتية

W1P. Biological Databases

كلية التقنية الطبية / قسم الأدلة الجنائية

Dr. Ragheed Hussam Yousif

Ragheed.Hussam@alfarahidiuc.edu.iq

OUTLINE

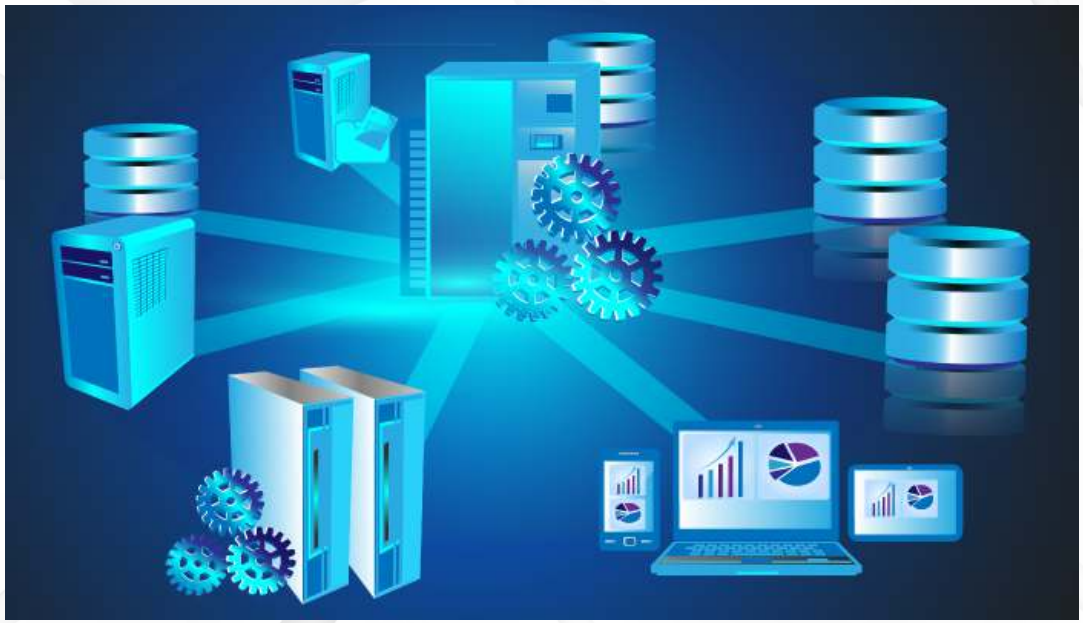
1. What is Database?
2. Biological Database.
3. Sequence Database?
4. Nucleotide Database.
5. Standard contents of a sequence database.

OUTLINE

6. NCBI.
7. Standard contents of a sequence database.
8. Structural databases

What is Database?

A database is an organized collection of structured information, or data, typically stored electronically in a computer system.



What is Database?

DATA VS INFORMATION ?

Biological databases

Biological data are complex, vast and incomplete. Therefore, several databases has been created and interpreted to ensure unmistakable results. A collection of biological data arranged in computer readable form that enhances the speed of search and retrieval and convenient to use is called biological database. A good database must have updated information.

What we expect from a database?

- Sequence, functional, structural information, related bibliography.
- Well Structured and Indexed.
- Well cross-referenced (with other databases).
- From time to time updated.
- Tools for analysis and visualization.

Biological Databases

- Sequence databases
- Structure databases

Sequence databases

- Nucleotide databases
- Protein databases

Nucleotide databases

- International Nucleotide Sequence Database Collaboration (INSDC)
  - NCBI
  - EMBL
  - DDBJ

Standard contents of a sequence database

- Sequences
- Accession number
- References
- Taxonomic data
- Annotation/curation
- Keywords
- Cross-references
- Documentation

NCBI National Center for Biotechnology Information

- Very comprehensive biological database
- GENBANK: The nucleotide sequence database
- Provides 42 different resource
- Provides a simple and easy to use web interface

http://www.ncbi.nlm.nih.gov/

- Sequence submission.
- Search Engine for data retrieval: Entrez
- Retrieves information across all the resources under NCBI

  Example: PubMed, taxonomy, SNP, PubChem etc.

Tools for analysis

- BLAST
- Primer-BLAST
- B-Link
- ORF finder
- Genome workbench

Protein Sequence databases

- UniProt
- PFAM
- Gene Index project

UniProt

- Universal Protein Resource
- Formed through the merger of :
  - SIB
  - EBI-SwissProt
  - TrEMBL
  - PIR-PSD

Structural databases

PDB – Protein Data Bank

- Contains information about experimentally determined structures of proteins, nucleic acids, and complex assemblies
- RCSB-PDB, PDBe, PDBj, BMRB – repositories of protein structure data
- Files in PDB, mmCIF, PDBML/XML formats

- Advanced search – provides comprehensive information about a protein.
- Sequence info, domain info, sequence similarity, literature, apart from the details of the structure.

Thank You!

جامعة الفراهيدي
2012 - 1433
Al-Farahidi University