

Biomedical Data Bases

Saul Pierotti

November 19, 2019

Introduction

- A database is an organized collection of data
- A DBMS is a piece of software that allows DB implementation and data mining
 - It allows data storing, retrieval, to perform backups, to maintain security
- Curation is a synonym of manual annotation
- In the last few years there has been a big effort to standardize DBs
- Even in the best DBs there are errors (!)
- Integration among DBs happens with corss-linking and with merging of DBs
 - Uniprot is an example
- Not all the structures available are on PDB (!)
 - Pharmaceutical companies have their own provate DBs
- Some DB terms
 - A record is a DB entry containing different fields
 - An accession key is a unique identifier of a record
 - A table is a DB file containing many records
 - * Different tables can be connected by the same accession key for some records
 - A query is a data request submitted to a DB
- A query can be organized with boolean operator, in order to retrieve the desired result
- The schema of a DB is the logical structure of the data
 - A schema can be written in a flatfile, XML, ecc.
- The instance is the set of actual data
- The schema allows the interpretation of the instance
- Pubmed is based in Betsheda, at the National Library of Medicine
- Many DBs have built-in a way to retrieve data from the command line in a structured way
 - Entrex Direct allows to retrieve results from Entrez in the Unix command line

For upload