Issues to Address While Designing a Biological Information System

Issues to Address While Designing a Biological Information System

MBI301-DataMining & Data Analytics

Mamta Sagar
Department of Bioinformatics
University Institute of Engineering & Technology, CSJM University, Kanpur

- Life scientist has experienced a fundamental revolution from traditional in vivo discovery methods
- To cope with this dramatic revolution, life scientist need tool that enable them to access, integrate ,mine, analyze, interpret, , simulate & visualize the wealth of complex & diverse electronic biological data.

Engineering vs. Experimental Science

Design Implementation

This loop aim to correct implementation failure to perform effectively the requirement & to extend significantly, the implementation to new requirements

Generic System vs. Query-Driven Approaches

Computer scientist aim to build systems. a system is implementation of an approach that is generic to many applications having similar characteristics

Life scientist are motivated by an hypothesis they wish to validate. The validation process typically involve some data sets extracted from identified data sources & follows a predefined manipulation of data the collected.

Legacy

- Biological Data
- Biological Tools and Workflows

A Domain in Constant Evolution

- Traditional Database Management and Changes
- Data Fusion
- Fully Structured vs. Semi-Structured
- Scientific Object Identity
- Concepts and Ontologies

Biological Queries

- Searching and Mining
- Browsing
- Semantics of Queries
- Tool-Driven vs. Data-Driven Integration

Query Processing

- Biological Resources

1. Coverage of Information Sources
2. Source Capabilities:-capture the type of queries supported by the sources
3. Statistical Pattern:-include the description of information clusters
4. Delivery Pattern:-include the response time, that is needed to the receive the first block of pattern, the size of the these blocks so on

Query Planning:-

- Query planning consist in considering the many potential combination of accesses to evaluate to the query. Each combination is query evaluations plan.
- There exist many plan evaluate query.

Query Optimization

- Query optimization is the science and the art of applying equivalence rules to rewrite the tree of operator evoked in a query and proceed an optimal plan.
- A plan is optimal if it return the answer in the least time or using the least space.
- The overall cost is typically defined as the total time needed to evaluate the query and obtain all of the answer

Visualization

- Multimedia database
- Browsing scientific Objects

Question?

What is data fusion?

References

- **Bioinformatics: Managing Scientific Data** Z. Lacroix, Lacroix Zoe Critchlow Terence, T. Critchlow Published 2013 Computer Science