Managing scientific data

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

- Bioinformatics and the management of scientific data is critical step to support life science discovery.

- As computational of protein models of proteins, cells, and organisms become increasingly realistic, much biology researchers will migrate from wet lab to the computer

Problems & scope

- In the last decade biologist have experienced a fundamental revolution from traditional research & development

- This represent a shift from life science's empirical root, in which it was iterative & intuitive

Bioinformatics is the field of science in which computer science and IT merge to form a single discipline

Biological data integration

- Concept identification
- Data transformation
- Concept overloading

Three cahlanges

- Sheer number of available data source & inherent heterogeneity

- Second data formats & data access method (associated interface) change regularly

- Third data & related analysis become increasingly complex

Developing a biological data integration system

System has to overcome difficulties

Specification

- who the user will be
- What functionality is system expected to have
- How this functionality is be viewed by the user
- The performance goal of the system

- System requirement describe the desired system and can be seen as contract greed upon by the target user and the developer
- The user profile is concise description of the target user for a system.
- Once the profile is defined the test the system is supposed to perform must be analyzed

- Finally technical issue must be agreed upon
- Expected to work on (e.g. UNIX, Microsoft)

,it scalability

translating specifications into technical approach

- Once the specifications have been agreed upon, they can be translated into a set of apaches.

- This can be thought be of as an optimization problem n which the hard constrained define a feasibility region

- Each attribute in a system description can be mapped to a dimension

- Existing data management approaches can then be mapped to overlapping regions in this space

- Once the optimal location has identified, these approaches can be used s a starting point for the implementation

Development process

- Code design: describing the various software components/ objects & their respective capabilities

- Implementation: actually writing the code & getting it to execute properly

- Testing: evaluating the implementation

- Deployment: transferring code to set of user(analysis of test & training the user)

Evaluation of system

- Two system may have the same specifications and follow the same approach yet end up with radically different implementation
- Developing feasible system

Question?

Steps of translating specifications?

Reference

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