





Abstract

Bioinformatics tools built for isolated tasks do not adequately support translational researchers' analytical needs for hypothesis formulation. We address the design of integrated systems to support exploratory translational research. We examine user cognition and analysis patterns, their implications for system requirements, and biological knowledge representations for causal analysis.

Introduction

- Bioinformatics tools are generally developed to perform an isolated task without full understanding of how a translational researcher would apply the tool in the context of their work
- Efforts to build integrated systems of tools at NCIBI make it necessary to understand how users approach their research and how they try to employ these tools
- Observations of users of integrated tools at NCIBI are helping develop understanding of requirements for integrated systems of bioinformatics tools
- This effort meshes with ongoing work on biological modeling exemplified by the BioWorkflow project

User Observations and Interviews

Methodology

Method: Field observations and interviews over time

Sample: 15 biomedical researchers

Unit of Analysis: Software-supported problem solving session aimed at formulating an hypothesis about disease mechanisms

Data Collection: Think-alouds, task duration, actions, outcomes, impasses, goals,

intentions, reasoning, judgments

Data Analysis: Qualitative uncovering of patterns related to stages, tasks, behaviors, ways of knowing and reasoning

[ctnb1_human] -- [tf7l2_human]

eport a problem with this entr

 Tools and Their Uses in Scientists' Exploratory Analysis

Tool: Michigan Molecular Interaction (MiMI) web interface and protein interaction database **Used to:** Query on candidate genes or concepts; retrieve data about relevant protein interactions, literature, GO annotations, and literature-mined descriptions

Tool: MiMI plugin to Cytoscape.

Used to: Interactively explore and manipulate networks of protein-protein interactions in Cytoscape, complete with conceptual and quantitative data carried over from MiMI.

Findings

► All scientists flowed through 4 stages, each with a dominant mode of reasoning: 3. Mentally Modeling Causes & Effect 1. Confirming

- Verification reasoning
- Causal mental modeling/inference 2. Separating the Wheat from Chaff Classification



4 / beta-catenin complex, unspecified role, unspecified role, neutral agents unspecified role, unspecified role, neutral component neutral PubMed

- 4. Building a Biological Story Narrative reasoning
- Validation occurs throughout inquiry but in different forms in each stage.
- Classification reasoning is far better supported by bioinformatics tools than causal mental modeling and narrative reasoning.
- Novel insights into credible mechanisms of a disease depend on supporting users in integrating all modes of reasoning and inquiry stages of analysis.
- Support for novel insights requires improvements in tools, as follows:
- **Provide ample interactivity** and control over manipulating data as needed for turning data to knowledge – i.e. based on observed patterns

Provide cues, not just data read-offs, to guide inferences – e.g. emphasis, default groupings, "analysis in a keystroke" that embodies domain knowledge

Provide surrogates for validation acceptable to scientists when test statistics are not available or when biological knowledge is incomplete.

Provide domain-based contexualizing information: Types of interactions, types of molecules, ability to import one's own data, overlays of protein interactions and pathways, overlays of protein interactions and disease-ome.

Design of Integrated Translational Bioinformatics Systems

Barbara Mirel, DArts¹; Benjamin J. Keller, PhD^{1,2}; Mor Peleg, PhD^{3,4}; Russ Altman, MD, PhD⁴ ¹University of Michigan, Ann Arbor, MI; ²Eastern Michigan University, Ypsilanti, MI; ³University of Haifa, Israel; ⁴Stanford University, Stanford, CA

User Model

- Representation of scientists' ways of knowing and reasoning:
- Scientists iteratively integrate 3 modes of reasoning
- They continuously intermix creative exploration and rigorous validation
- They require coherent transitions between ways of knowing to succeed
- Allows for arbitrary transitions between kinds of reasoning with explicit verification after each transition, and validation at any point



Supporting Modeling

- Many bioinformatics tools (including most NCIBI tools) are focused on classification
- The BioWorkflow approach provides an insight into how modeling can be supported

Example: BioWorkflow Process Model of C-like Nucleosides

- Model shows alternative processes involved in the cellular processing of C-like nucleosides, including the native C nucleoside and the drug Gemcitabine – a C-analog (dFdC)
- Four alternative processes (green ovals) can transport the C-nucleosides into the cytoplasm, each using a different enabling transporter (blue rectangle): hENT1, hENT2, hCNT1, or hCNT3
- ➤ C-like nucleoside (blue rectangle) is shown to be participating with a substrate link on all of the alternative processes, which converge at the XOR/XOR router.
- ▶ Two alternative processes follow: Deactivation of C-like nucleoside or its phosphorylation.
- Phosphorylation step is followed by 3 alternative processes: Dephosphorlyation, Deactivation, or a second phosphorylation.



Matching Requirements to the User Model

- The user studies provide a basis for deriving a set of requirements based on the types of information inspected and tasks performed
- Basic requirements relate to need to support
- All three kinds of reasoning
- Ability to verify outcomes from queries and previous stages
- Ability to validate current information, and
- Different approaches to moving among types of reasoning and verification/validation activities
- Within modes of reasoning requirements are more complex
- Example: Framing requirements for scientists' different types of classification reasoning during the Separating the Wheat from Chaff stage
- Scope: Mechanisms for identifying
- Groupings by various forms of similarity distinguishing between sharing of single versus multiple attributes
- Patterns: temporal, localization, regulation, compatible annotation, graph theoretic • Subsystems: internal structures, relationships, normal vs abnormal behavior
- ➤ Visualizations of multiple relationships crossing biological scale and attributes e.g., a protein-protein interaction network painted by GO Biological Process
- User controls for selecting; aggregating and abstracting selections; filtering
- Display of information that allows the user to verify query results based on personal knowledge or literature provenance

Structural Part of the Workflow

- Molecules that are part of the Gemcitabine/C-nucleoside processing model are shown as blue squares. The roles that they play are shown as pink ovals.
- ◆ The insert on the right shows the details of the Gemcitabine chemical. The cellular location is specified as "Nucleus" – a term taken from the biomedical vocabularies that include TAMBIS and UMLS.



Petri Net corresponding to the BioWorkflow Process Model of C-like Nucleosides

+ --++++D<mark>F</mark> - + + + + + + +









Moving Forward

- Consideration of a cognitive view of user tasks leads to important insights in how to build integrated systems of tools
- Now possible to state more precise requirements for classification reasoning
- Tools and scientists reasoning are dynamically related, and so improved tools will lead to cycle of identifying better targeted requirements followed by tool improvements
- Few open source tools support contextualized causal reasoning for mental modeling and narrating putative biological events and abnormalities
- Well-targeted requirements will evolve in modeling and story-telling with user experience with prototypes based on best guesses such as the BioWorkflow project
- Remaining challenges include the design of visualizations for classification, and supporting the transition from classification to modeling

Acknowledgements

This work was partially supported by National Institutes of Health: Grant #U54 DA021519

References

- Cohen AM and Hersh W: A survey of current work in biomedical text mining, Briefings in Bioinformatics 2005; 6: 57-71.
- Saraiya P, North C, Duca K. Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. Information Visualization 2005; 4:191–205.
- Mirel B. Supporting cognition in systems biology analysis; findings on users' processes and design implications. Journal of Biomedical discovery and Collaboration 2009; 4:1-17
- Peleg M, Yeh I, Altman RB. Modeling biological processes using Workflow and Petri Net models. Bioinformatics 2002;18(6):825-837.
- Peleg M, Rubin D, Altman RB. Using Petri Net Tools to Study Properties and Dynamics of Biological Systems. J Am Med Inform Assoc 2005;12(2):181-199.