

From Scientific Protocol to Information System: The Development of an Information System Supporting Bioinformatics Analysis

Context of the Work

Background

The advent of bioinformatics, the application of computer technology to the storage, analysis and representation of biological information, profoundly impacts all areas of biomedical research. However, in spite of the critical role that bioinformatics analysis plays in the advancement of biomedical research, and the need to integrate bioinformatics analysis and laboratory research, there is a lack of research into this challenging problem.

The proposed research presented in this poster is part of an ongoing study addressing the problem of linking bioinformatics analysis to scientific practice. To date, this work has produced a systematic protocol describing and detailing the application of bioinformatics analysis to the scientific problem of predicting gene function from sequence data (the functional analysis of a gene sequence) (Bartlett, 2004). The protocol, developed through a combined information behavior and task analysis approach, captures and documents the knowledge and expertise of bioinformatics analysts. It describes twelve individual bioinformatics analysis steps (involving over seventy different bioinformatics information resources) linked in a workflow process with three alternate pathways. The detailed description for each step includes: the rationale for the step, the input/output data, the bioinformatics resources used, and the analysis and interpretation of the results. While the steps are described in a linear sequence, not all steps are appropriate or necessary for every scientific research scenario. Instead, the human judgment of the scientist is needed to determine which steps are appropriate. Likewise, the need for human judgment and decision making applies to the interpretation and application of the results. Automation of the protocol is therefore limited to a semi-automated system.

Objectives

The current research objectives (the subject of this poster) are to develop and evaluate a system to make the bioinformatics analysis protocol accessible and usable to a laboratory scientist. From the perspective of information architecture, the challenge is to develop a system that accesses not one, but many different information resources, and presents them, and the information obtained from them, in an integrated, coordinated and cohesive manner. The system must support the work of scientists who are not familiar with the workflow process underlying the protocol.

Methods

The first step is to develop a prototype information system based on the bioinformatics analysis protocol. The purpose of the system will be to guide and direct a scientist through the process of conducting a functional analysis of a gene sequence.

The system will comprise a hierarchical series of web pages, paralleling the levels of detail in the bioinformatics analysis protocol. At the upper level, the system will

present the protocol at a broad level, providing an overview and allowing orientation within the protocol/system. At the most detailed level there will be a series of web pages, each providing description and instruction on the use and application of a specific bioinformatics analysis step. Information on each step will include:

- what tool(s) may be used (with links to external sites)
- what data is input (including how to format the data)
- what data will be output
- how the output data may be interpreted and applied
- what step(s) to follow next

Intermediate level information will support navigation between the analytical steps, including information about the rationale for each step, and description of the circumstances under which each would be appropriate. This will allow scientists to determine which steps are most suitable to their particular research scenario. The overall flow will parallel the process of conducting a functional analysis of a gene, with scientists being directed from one step to the next in a logical workflow.

The second part of this study involves testing the protocol and system. In a within-subjects experimental design, participants will conduct a functional analysis of a test gene sequence under two conditions, both with and without the system. The results of the two analyses will be evaluated and compared, including consideration of factors such as effectiveness, efficiency, and usability.

Innovative Points

Firstly, the system being designed in this research will not provide access to a single information system. Instead, it must integrate and coordinate the access and use of information from over seventy individual, pre-existing information systems, each of which has its own unique information architecture. The challenge is to provide a framework that logically connects and integrates the use of a diverse variety of resources, and rationally applies them to accomplishing a goal. The system must also support the application of the information in the broader context of the scientific research process.

Secondly, information interaction, the “process that people use in interacting with the content of an information system” (Toms, 2002), is intrinsically influenced by the information architecture of the system. In this work, a detailed analysis of the workflow to be supported, the tasks to be accomplished, and the information to be provided through interaction with the information system has preceded the development of the system itself and will serve as the foundation of the system. In essence, an understanding of the desired information interaction is used as the basis for information architecture.

Finally, within the domain of bioinformatics, this research is innovative in that it is aimed at developing a task-oriented information system. To date, most bioinformatics information systems have focused on individual analytic processes, without reference to the broader context, or other information resources or systems that might also be used. The system proposed here, for the first time, takes a holistic, integrated approach to the architecture of a bioinformatics information system, linking multiple analyses and resources to accomplish a goal.

References

Bartlett, J.C. (2004). *Connecting Bioinformatics Analysis to Scientific Practice: An Integrated Information Behaviour and Task Analysis Approach*. Unpublished Ph.D. Dissertation, University of Toronto.

Toms, E.G. (2002). Information interaction: providing a framework for information architecture. *Journal of the American Society for Information Science and Technology*, 53(10), 855-862.

Poster Content

The poster will present the background to this research, the objectives of developing and testing the information system, the methods, and the anticipated outcomes and implications of the research. It will include a graphic representation of the bioinformatics analysis protocol forming the basis for the information system (Figure 1). It will also describe and illustrate (including screenshots), the various hierarchical levels in the information system (Figure 2).

Poster Mock-up

