Allen, R.B., Highly Structured Scientific Publications. ACM/IEEE Joint Conference on Digital Libraries, June 2007, 472.

Highly Structured Scientific Publications

Robert B. Allen
College of Information Science and Technology, Drexel University
Philadelphia, PA 19104
rba@drexel.edu

ABSTRACT

Science is a complex, but highly structured, activity. We propose that reports about scientific research would benefit by reflecting that structure more systematically than is found in current scientific articles. We apply structures for the components of the complete research activity that include the article type, research question, the conceptual model, the research procedure, the data analysis, and the conclusions. We show some complete, though simple, examples and describe issues for extending the approach to collections of complex research reports.

Categories and Subject Descriptors

H.1 MODELS AND PRINCIPLES, H.1.0 General

General Terms

Documentation, Standardization

Keywords

Activity Semantics, Context, Digital Libraries, Modeling, Publication, Research Questions, Science, Workflow, UML

1. INTRODUCTION

While scientific research and scientific publishing has changed dramatically in recent years, the structure of the basic unit of scientific communication – the research article -- is little changed. Although, structured abstracts have now been widely adopted (e.g., [6]), we propose that the entire article should be richly structured.

2. ELEMENTS OF SCIENTIFIC RESEARCH REPORTS

Traditional scientific publications typically include an Introduction, Method, Results, and Discussion (IMRD) [7]. Introductions, for instance, serve several functions: they pose a research question, describe the conceptual model, justify the methodological approach, provide pointers to previous research, and explain the motivation for the work. Even research without a clearly defined hypothesis can be highly structured. In addition, technical publications whose contributions are limited to theory, methodology, or instrumentation, can nonetheless fit this broader framework. Highly structured publications should improve clarity for browsing, visualization (e.g., [1]), and linking across articles. We identify the components of research articles and we show how each can be structured and encompassed in an overall framework.

One key component is the research question and we describe ways to structure research questions (e.g., [9]). Such research questions often refer to a conceptual model (e.g., [8]). Conceptual models themselves can often be described with a workflow language such as UML; for instance, UML is being explored for modeling biological systems (e.g., [4]). When the conceptual model is quantitative, MathML may be used. UML can is also be used for describing the flow of the research procedure, data analyses, and conclusions. Moreover, the specific research paradigm, whether experimental or quasi-experimental, provides a taxonomy of research strategies (e.g., [3]). The data analysis component could be similar to the integration of the R data-analysis system into the Kepler workflow environment. Indeed, meta-models have been proposed as a framework for preservation of scientific data in conjunction with Kepler (e.g., [5]). Here, we go beyond that to develop complete models of the research process and to propose those models as an alternative to traditional scientific publications.

Developing a system for structuring scientific publications should facilitate capturing the discourse claims made in scholarly articles (e.g., [2]). Structure will also facilitate linking articles in digital libraries and should be much richer than traditional article citations. Moreover, we believe that such structure will encourage community annotations by allowing them to pinpoint specific contributions.

JCDL'07, June 17–22, 2007, Vancouver, British Columbia, Canada.
ACM 978-1-59593-644-8/07/0006.

REFERENCES

1 Allen, R.B., Wu, Y.J., and Jun, L. 2005, Interactive Causal Schematics for Qualitative Scientific Explanations, *International Conference on Asian Digital Libraries* (LNCS 3815/2005), 411-415.

2 Buckingham Shum, S.J., Uren, V., Li, G., Sereno, B., and Mancini, C. 2007, Modelling Naturalistic Argumentation in Research Literatures: Representation and Interaction Design Issues. *Int. J. Intelligent Systems*, 22(1), 17-47.

3 Campbell, D.T., and Stanley, J.C. 1966, *Experimental and Quasi-Experimental Designs*, Rand McNally, Chicago.

4 Finney, A.M., and Hucka, M. 2007, *Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model Definitions*.

5 Hunter, J., 2006, Scientific Models – A user-oriented approach to the integration of scientific data and digital libraries, Victorian Association for Library Automation.

6 Mulrow, C.D., Thacker S.B., and Pugh J.A. 1988, A proposal for more informative abstracts of review articles. *Annals of Internal Medicine*, 108(4), 613-615.

7 Swales, J.M. 1990, *Genre Analysis: English in Academic and Research Settings*, Cambridge University Press.

8 Thagard, P. 1992, *Conceptual Revolutions*, Princeton University Press, Princeton, NJ.

9 White, B.Y. 1993, Thinkertools: Causal Models, Conceptual Change, and Science Education. *Cognition and Instruction.* 10(1) 1-100.