

REPORT DOCUMENTATION PAGE					. O 04-01	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0 04-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.						
1. REPORT DATE (- -) 01-06-2011		2. REPORT TYPE Progress			3. DATES COVERED (- -) 15-06-2010 to 15-06-2011	
4. TITLE AND SUBTITLE Categorical Information Theory				5a. CONTRACT NUMBER		
				5b. GRANT NUMBER N 00014-10-1-0841		
				5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Spivak, David				5d. PROJECT NUMBER		
				5e. TASK NUMBER		
				5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Massachusetts Institute of Technology Office of Sponsored Programs Rm E19 750 77 Massachusetts Ave Cambridge, MA 02139-4301					8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)					10. SPONSOR/MONITOR'S ACRONYM(S)	
					11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for Public Release; Distribution is Unlimited.						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT						
15. SUBJECT TERMS						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER ()	

Progress Report for Office of Naval Research Grant N000141010841

David I. Spivak

May 29, 2011

This report will summarize my progress toward fulfilling the goals of ONR grant N000141010841 (“Categorical Information Theory”) during the first year performance period, June 15, 2010 through June 15, 2011. During this time, I have been hosted as a postdoctoral associate in the mathematics department at the Massachusetts Institute of Technology under Professor Haynes Miller. The Technical Proposal for this grant can be found online at: http://math.mit.edu/~dspivak/informatics/technical_proposal2010.pdf.

1 Goals for this performance period

As detailed in Part III of the Technical Proposal for this grant, my goals for this period of the grant were to seek out researchers in neighboring fields, such as mathematics, computer science, and linguistics, with whom I could discuss the subject of information and communication. In particular, I intended to reduce some of my current formulations of databases to practice, i.e. to implement them on a computer. I also planned to write a paper linking ontologies and databases.

I was successful in these ventures. Perhaps most fruitful has been my paper linking ontologies and databases using category theory. These so-called ologs have led to exciting advances in materials science research, as I will discuss below.

2 Progress during this performance period

During this period I have indeed sought researchers in neighboring fields, including computer science, linguistics, and materials science; I will discuss each of these collaborations below. The last of these has been the most successful. I will finish this section with a discussion of the papers I have written and the presentations I have given during this time.

2.1 Computer Science

I worked with Dr. Carlo Curino, a postdoc in the Computer Science and Artificial Intelligence Lab (CSAIL) at MIT, to consider the advantages and challenges associated with my work on categorical databases. Our collaboration was useful for both of us; he now thinks that category theory may well be a useful fundamental framework for databases, and I learned much more about how databases are currently used and designed in practice. We have not completed a paper together because of time constraints and a difficulty finding specific problems for which we could show obvious superiority of our approach. However, we did present a poster together at the New England Database Summit 2011, entitled “Category Theory as a Unifying Database Formalism”, and it can be found online at: <http://math.mit.edu/~dspivak/informatics/talks/NEDB2011.pdf>.

Carlo was also successful at integrating the SQL database definition language and query language with my categorical formulation (discussed in my paper “Functorial Data Migration”), and creating a translating script between the two. (A video of this software being used can be found online at: <http://anonymizedurl.com/014523452/sql2>) This is strong evidence that the theoretical work I have done will be implementable on a computer.

2.2 Linguistics

I worked with Micha Breakstone, a graduate student in the Linguistics department at MIT. Micha has a masters degree in Mathematics from Hebrew University in Jerusalem, where he studied Topological Quantum Field Theories, a strongly category-theoretic field. He was thus equipped to clearly understand what I was working on. We have discussed my work on ologs, and he thinks that they are sufficient to describe most if not all semantic constructs. He further believes that linguistic theory could provide the framework for an automatic translation system of English sentences into ologs. We plan to continue work on this and related areas.

2.3 Materials Science and Engineering

I worked with Dr. Markus Buehler, a professor in the Materials Science and Engineering department at MIT. Buehler’s work focuses on the hierarchical structure of biomaterials and how functionality emerges at different scales. He was looking for a mathematical or linguistic framework to formally express these structures and how they are constructed out of universal building blocks. My work on ologs provided such a framework. Together we produced a paper (joint with his student Elizabeth Wood) called “Category Theoretic Analysis of Hierarchical Protein Materials and Social Networks”, which can be found online at: <http://math.mit.edu/~dspivak/informatics/ProteinSocial-Totalled.pdf>.

We plan to continue to work together in the future, exploring different hierarchical materials to find universal design patterns. This work with Buehler also opens the door to my collaboration with other scientists outside of both mathematics and materials science, because it demonstrates that the olog idea is sufficiently powerful to describe situations like that of proteins and social networks; hence that other disparate research domains may similarly benefit from this formalism.

2.4 Papers and Presentations

During this period, I wrote three papers. The first paper was entitled *Functorial Data Migration* and can be online at: <http://math.mit.edu/~dspivak/informatics/FunctorialDataMigration.pdf>. It has been submitted to the journal *Information and Computation*. In this paper I laid out a simple category-theoretic formulation of database schemas and states, and I showed that schema evolution and data migration can be accomplished functorially using well-known ideas from category theory. I have given presentations on this topic in the following venues:

1. Amgen Inc. 2011/02/17-18
2. New England Database Summit (poster, joint with Carlo Curino) 2011/01/28
3. Boston Haskell 2011/01/20
4. Harvard U. 2010/11/03 (EECS seminar);
5. Galois Inc. 2010/10/22 (Tech talk)

6. MIT 2010/09/20 (Topology seminar);
7. MIT 2010/09/16 (CSAIL seminar).

The second paper I wrote was entitled *Ologs: a categorical framework for knowledge representation* and can be found online at: <http://math.mit.edu/~dspivak/informatics/ologs-basic.pdf>. It has been submitted to the journal *PLoS ONE*. In this paper I explained in very basic terms how category theory can be used as a formalism for knowledge representation using ologs. An olog is a category in which the objects are drawn as text-boxes, the arrows also have text labels, and commutative diagrams (as well as possibly limit and colimit declarations) are recorded. I explained that an olog can serve both as an ontology and as a database schema, and discussed several directions for future research. I gave a few presentations on this topic, namely at the MIT Linguistics semantics seminar (2010/09/15) and briefly in a tech talk at Galois Inc. (2010/10/22).

The third paper I wrote was entitled *Category Theoretic Analysis of hierarchical protein materials and social networks* and can be found online at <http://math.mit.edu/~dspivak/informatics/ProteinSocial-Totalled.pdf>. This paper was a collaboration with M.J. Buehler and E. Wood and was described in Section 2.3. It has been submitted to the *Journal of the Royal Society Interface*. I have not yet given any talks on this subject.

3 Plans for the next performance period

As stated in the Technical Proposal, in the next period I plan to translate the communication protocol that Mathieu Anel and I formulated for pairs of ontologies into a communication protocol for databases. Since I now consider database schemas and ontologies to have identical category-theoretic formulations, this plan becomes somewhat moot. However, I could still work to see how to communicate not only the schema but also the state of the database to another. I hope to submit a paper with Anel on at least the basic communication protocol during the next grant period.

I also plan to extend my notion of database schemas to allow for more complex and hierarchical information storage. While this work is at an early stage, it is my belief that powerful new ideas in the field of ∞ -categories may be applicable to information systems. In fact, this could unify much of my work in this area so far and provide the kind of overarching framework for the study of information that I have been looking for. Of course, this is unclear and will require quite a bit of work. I hope to collaborate with other mathematicians at MIT to accomplish this.