Technical Proposal

BAA number: #10-001

Title of Proposal: Categorical Information Theory

Identity of prime Offeror:
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

Technical contact:
David I. Spivak
MIT Math Department
77 Massachusetts Ave, Bldg 2, Rm 236
Cambridge MA 02139
`dspivak@gmail.com`
(510) 684-6425

Administrative/business contact: Stacey Sullaway
Massachusetts Institute of Technology
Office of Sponsored Programs
77 Massachusetts Avenue
Cambridge, MA 02139-4307
ph: (617) 324-7210
fx: (617) 253-4734
`sullaway@MIT.EDU`

Duration of Effort: 3 years: From 2010/06/01 to 2013/05/31.

Contents

Part I
Technical Approach and Justification

1 Topic of study: Information and communication

The same frustration is facing institutions and individuals throughout modern society. The information that one entity gathers cannot be readily converted into a format that another entity can understand and use. Databases are incompatible, vocabularies are mismatched, rule-systems are vague and imprecise. The symptoms of this disease include the "data glut" in the natural sciences, the failure of communication between researchers in even neighboring academic disciplines, the database merging problem in commerce, and the difficulty in implementing the semantic web. Broadly, we suffer from an inability to create effective communication between foreign entities.

I propose that mathematics has the capacity to offer a solution to this problem. Throughout history, mathematics has been invented to formulate, discuss, and ultimately solve the issues of the day. Counting was created to found chronology and economics; geometry was created to aid surveying, construction, and astronomy; calculus was created to understand relationships between physical variables. If we consider mathematics not as its diverse collection of subfields but instead as a commitment to formal expression and rigorous methodology, then its relevance to the current issue of communication becomes more clear.

Mathematics is already being broadly used within the information and communication sciences. In computer science, databases are founded on the theory of relations in mathematical logic, and functional programming languages are steeped in category theory. In linguistics, one may study syntax using categorical grammars and study semantics using the lambda-calculus. Finally, our overview would not be complete without mentioning "Information Theory" in the sense of Shannon, in which one uses mathematics to describe how to efficiently send data across a noisy channel.

Unfortunately, the above disciplines are not well-integrated. They each study a particular segment of the issue of information and communication, but are not formulated in a way that can be readily related to the others. It is interesting that these theories are each designed to describe information and communication, and yet communication between them is woefully inadequate. It is imperative that we find a coherent overarching framework for this whole field of study.

Information is inherently a combinatorial affair: facts combine with other facts to create new facts. Like atoms combine in set ways to create molecules and words combine in set ways to create sentences, there is some underlying structure governing the behavior of information. However, the combinatorics of information is not numerical, it is conceptual. In order to approach it, we must use a mathematical theory that is equipped to organize layers of abstraction. Such a theory already exists within mathematics, namely the theory of categories and functors.

In the early 1940s, Saunders Mac Lane and Samuel Eilenberg developed Category Theory in order to discuss the rich relationships that exist between geometric shapes and equational algebra. Category theory gained prominence throughout the 20th century because of its ability to connect diverse mathematical fields and translate results between them. Today, most papers in algebra and geometry could not even be written without categorical terminology and proof technique because the concepts would be too complex to present coherently.

As alluded to above, category theory has been broadly applied outside of mathematics as well, e.g. in computer science, linguistics, and physics. In some sense category theory is the essence of structure. In philosophy, a category consists of a collection of objects that are related in some way, such as the set of people in a family or the set of idioms in a language. Similarly, a category in mathematics consists of a set of objects, a notion of "morphism" or structure-preserving transformation, and an ability to compose these morphisms: if one can relate A to B and relate B to C, then one can relate A to C. It is remarkable that such a simple axiom system can lead to such a profound ability to precisely formulate relationships between disparate fields.

In this technical proposal, I will explain how category theory may be applied to the situation at hand. In the remainder of this section I will discuss the frameworks of information that are currently in use, how they manage the issue of communication, and an overview of how category theory may be applied. In Section 2, I will proceed to discuss my current progress in this direction, much of which was accomplished during the first one-and-a-half years of a previous grant from the Office of Naval Research (`N000140910466`: "Geometric networks: A higher-dimensional approach to networks and databases"). Section 3 will be devoted to explaining what I hope to accomplish in the next three-year period.

1.1 Frameworks of information

Computers were invented to store and process information. In order to know what information is, one should begin by taking stock of how it is used in practice.

The most common device for storing information is a database. A database is more than a spreadsheet or a table. It consists of a system of tables whose rows represent things in the world, and whose columns represent attributes of various sorts. Attributes from one table may match attributes from another; it is this fact that leads to the complexity of the subject. Fundamental to understanding a database is a systematic account of the connections that exist between columns of different tables. I will discuss my research on the category of databases in Section 2.1.

Recently researchers have developed another systematic device for storing and processing information, called an ontology. Ontologies are not as structured as databases are: ontologies are designed to be flexible whereas databases are designed to be regimented.

There are currently several competing ideas of precisely what ontologies are and how they should be modeled. Often ontologies take the form of "controlled vocabularies" or taxonomies. For example, one could input that parrots and robins are both birds. More powerful ontological systems may allow for more interesting facts like "the weight of an adult scarlet parrot in grams is a number between 900 and 1100." The issue is not simply to record these facts, but to organize them in such a way that a computer can make inferences (e.g. to answer questions like "which type of parrot weighs the most?"). I will discuss my current progress and future plans to categorify ontologies in Section 3.2.

There are many other frameworks for information currently in existence. Any combination of syntax and semantics could be considered a framework for information – in particular a language such as English can be considered a framework for information. The theory of programming languages is another framework for information: given input, it is a matter of rigorous information-processing to determine what comes out.

In order to formulate all of these information frameworks into one coherent picture, one must find what is common to all of them. The primary function of information in all its forms is to be communicated; that is, information only exists to be queried and transfered. Thus communication and information are inextricably linked.

1.2 Communication

Communication between foreign entities is inherently challenging. The reason is that each entity has already developed internal language and jargon to reference its information, but this jargon does not extend beyond its borders. Just as the human brain stores and accesses information via an intricate system of neurons that is totally inexplicable, each agency or business accesses its information in a way that is difficult to relate. Furthermore, each entity makes assumptions about the structure of its world that become ingrained and taken for granted.

And yet communication does occur! In order to communicate with a foreign entity, assumptions should be made explicit and mutually acceptable conventions established. The necessary negotiation process is simplified if there is some accepted standard for information and its transfer.

There are several distinct varieties of information transfer. The simplest is that which occurs within a single entity; for example, this includes querying, manipulating, or restructuring a database. A harder problem occurs between two entities that are using the same framework, e.g. both using the same model of databases. This is not an easy problem in practice, but it is manageable; I will discuss my progress in this direction in Section 2.1. The hardest problem occurs when two entities are using different frameworks, for example if one uses a database and the other uses an ontology.

Tackling these communication issues demands a coherent and broad theory of information. One can think of each piece of information as an intricately-structured object and each message as a structure-preserving transformation. This viewpoint is precisely what category theory is designed to describe.

1.3 Applicability of category theory

To some degree one needs to have experience working directly with category theory in order to see its breadth and explanatory power. Adopting the arabic number system may have seemed arbitrary and strange to societies using roman numerals; however, once one uses the arabic system its superiority in matters of arithmetic becomes obvious. The arabic number system was designed with the relevant properties of numbers, namely the associative law and the distributive law, in mind.

The point is that not all systems are of equal merit, and category theory is the best system in which to describe the formal aspects of information and communication. As mentioned above, it is designed to rigorously enforce a fixed structure within a fixed category and yet allow for deliberate changes in structure by way of functors between these categories.

As an example, consider the notion of a context-free grammar. A context-free grammar is a system of rules for establishing the syntax of a language. For example, one could declare a rule designating every proper name to be of type <subject>, every present-simple verb (e.g. "eats" or "works") to be of type <verb>, and every noun to be of type <object>. Finally one could declare that a <sentence> be of type <subject><verb><object>. While some expressions within that grammar will be nonsensical, the grammar sets up a system for parsing sentences.

Context-free grammars are broadly used in computer and information science. However, it is rare that two entities would use precisely the same grammar without convening ahead of time. Thus, in order to communicate there needs to be an established notion of "morphism of grammars." There may be many good notions of morphism between grammars; each one would form a category and there would presumably be functors connecting these categories. To illustrate such a category, I will take from the presentation in [Hermida, Makkai, Power. "Higher dimensional multigraphs."]

Definition 1.1. A *grammar* consists of a sequence (C_0, C_1, S, dom, cod), such that C_0 and C_1 are sets, $S \in C_0$ is an element, and $dom\colon C_1 \to C_0$ and $cod\colon C_1 \to (C_0)^*$ are functions, where $(C_0)^*$ denotes the set of ordered lists of elements in C_0. Elements of C_0 are called *symbols*, S is called *the start symbol*, and elements of C_1 are called *rules*. An element $t \in C_0$ that is not in the image of *dom* is called a *terminal symbol*.

A *morphism of grammars*, denoted $f\colon (C_0, C_1, S, dom_C, cod_C) \to (C_0', C_1', S', dom', cod')$, consists of functions $f_0\colon C_0 \to C_0'$ and $f_1\colon C_1 \to C_1'$ such that

1. $dom_D \circ f_1 = f_0 \circ dom_C$,
2. $cod_D \circ f_1 = f_0 \circ cod_C$,
3. $f_0(S) = S'$, and
4. if $t \in C_0$ is a terminal symbol then so is $f_0(t) \in C_0'$.

Let us note a consequence of this definition. Suppose every entity has a grammar and speaks in the language associated to that grammar. If two entities establish a morphism $f\colon G \to G'$ between their respective grammars, it is not hard to prove that this will induce a function from sentences in the language of G to sentences in the language of G'. In other words, any sentence that the G-entity can parse can be automatically converted into a sentence that the G'-entity can parse. Whether or not meaning is preserved is another issue (indeed, it is an issue of great concern), but the point is that a morphism of grammars provides a ground for communication.

My first goal in the present project is to find categorical representations of a broad spectrum of information systems. These formulations will already ease communication between entities using a given system. Moreover, finding functors between distinct systems will promote communication between foreign entities. My final goal is to note the similarities and differences between various systems and, to the extent possible, give an overarching framework for the entire affair. These goals will be discussed more thoroughly in Section 3.

2 Progress: 2009 – present

In this section I will briefly review the work I have already done toward accomplishing the goals laid out in Section 1.

2.1 Simplicial databases

As mentioned in Section 1.1, a database consists of a system of tables, such that the columns in various tables overlap. A table T represents a type of thing by laying out the set of attributes that such a thing has: each row in T is a specific example of that type and each column in T is an attribute. The overlaps between columns of different tables signify connections between distinct types of thing.

Given this intuition, one seeks to create a category that makes it rigorous. This was done in the paper [Spivak. "Simplicial databases."] Roughly, a simplicial database consists of a schema and a sheaf of data. The schema is a simplicial set; in other words it looks like a diagram of vertices, edges, triangles, and higher-dimensional tetrahedra glued together in a precise way. Each simplex in this diagram represents a table and each vertex represents a column in that table; a connection between simplices represents columns shared between the tables. Finally, the sheaf of data provides the set of rows of each table.

This formulation has many advantages over the typical ER-diagram model. Most notably is the geometric intuition it provides. One can imagine tables as labeled triangles, or "tiles," which can be put together as long as their labels match. One can thus mix and match these tiles to create a custom database. To query such a database, one simply draws a curve through the simplicial set signifying the type of data to be input and the type to be output. All of this was made precise in [Spivak. "Table manipulation in simplicial databases."]

2.2 Networks

Common to many areas of academia is the notion of a network. In neuroscience one studies the network of neurons in the brain; in economics one studies the network of individuals trading goods; in computer science one studies the network of web pages on the internet; in mathematics one studies the network of objects of a given type (a category).

In my view, all of these are examples of the same phenomenon. A network is a collection of nodes and connections between these nodes, such that information is passing along these connections. The connections may not simply be 1-dimensional edges – they may be simplices such as triangles and tetrahedra, representing 3-way or 4-way conversations. This was described in the previous grant proposal `N000140910466`.

There are many different models of network (a graph is one and a simplicial set is another) but they all have certain aspects in common, as was described in [Spivak. "Higher dimensional models of networks"]. In a collaborative work-in-progress, myself and Mathieu Anel (currently at U. Quebec at Montreal) will describe a communication protocol between nodes in a network, given that the language of each node is encoded as an ontology.

The strength of the connection that binds any group of nodes in a network is measured as some type of quantity. I believe that this "quantity" should be an object in some information category such as simplicial databases. But suppose that one wishes to express the strength of each connection as simply a real number between 0 and 1; for example the probability that a message transfered along that connection will arrive in tact. The result is called a fuzzy simplicial set, which was defined in [Spivak. “Metric realization of fuzzy simplicial sets.”] In that paper, I showed that there is a “metric realization” functor which sends fuzzy simplicial sets to metric spaces; under this functor nodes that are strongly connected will be sent to points that are close together.

2.3 Presentations, professional connections, and projects in progress

In the last few years I have spent time traveling and explaining my ideas to a wide variety of experts in the field. I have given talks on this subject at the “Topological Methods in Computer Science 2008” conference in Paris and the “Agent-based complex systems” conference at the Institute of Pure and Applied Mathematics, as well as in colloquia and seminars at U. Oregon, UC Riverside, McGill, Carnegie Mellon, and Reed. I will give another talk on the subject at U. Chicago in May and at the “Foundational Methods in Computer Science 2010” conference in Calgary in June 2010. I have also discussed the issue in person with well-known academic researchers such as Bob Harper, André Joyal, and Andre Scedrov.

My work on simplicial databases was featured on the popular programming language blog `Lambda the Ultimate`. Soon after, it was picked up by David Balaban, Vice President at Amgen corporation. This has led to a good deal of collaboration with Balaban and his group and others. In particular, I spent two days in February 2010 relating the idea of simplicial databases to the informatics group at Amgen, as well as to Allen Brown at Microsoft Research. I have a weekly conference call to discuss these issues with Peter Gates at Johnson and Johnson and Eric Prud’hommeaux at the World Wide Web Consortium (W3C).

The point of all this collaboration is to disseminate what I have already learned, to continue to challenge myself to address additional topics in the field, and to create connections with and learn from the experience of active researchers. This project thrives when existing models are held up against new challenges and varying viewpoints.

3 Proposed project: 2010 – 2013

In the next period of this grant, there are many directions in which the research can be taken. In the following sections I will lay out a number of them.

3.1 Databases

There are many directions to take the simplicial model of databases. These can be roughly divided into pure directions and applied directions.

3.1.1 Pure theory

The pure directions have to do with studying the category of simplicial databases and related constructions. First, I want to generalize the construction of simplicial databases somewhat. Currently I have in mind something called categorical databases which are much more descriptive than simplicial databases are. I plan to find a diverse array of applications of these databases to show that they constitute a worthwhile generalization.

Second, I look forward to studying the topological nature of simplicial databases. The schema of a simplicial database has an underlying topological space, and curves through that space have real-world meaning. In fact, curves through a database form a non-trivial 2-category, which one might call the *fundamental 2-category* associated to the database. It is the 2-category of queries.

There may be classes of simplicial databases whose behavior mimics interesting categories from algebraic topology, such as that of bundles. Simplicial databases seem to have a strong relationship to the theory of operads, which I hope to make explicit.

Finally, one can import ideas and terminology from algebraic geometry to simplicial databases. Algebraic geometry sets up a correspondence between algebraic equations (e.g. $x^2 + y^2 = 4$) and the sets of points that satisfy them. A scheme is a union of such solution-spaces. Similarly one can write down propositions about rows in a table. There is a correspondence between propositions and the sets of rows that satisfy them. A database is a union of such solution-spaces.

One can make a precise notion of prime spectra for tables, and thus have the beginnings of a dictionary between algebraic geometry and simplicial database theory. For example, consider the field $k = \mathbb{F}_7$. The database Spec $k[x]$ would consist of a 1-column, 7-row table (of elements of k). The database Spec $k[x,y]/(xy = 1)$ consists of a 2-column, 6-row table (of all pairs of elements whose product is 1, e.g. $3 \times 5 \equiv 1 \pmod 7$), together with its projections to Spec $k[x]$ and Spec $k[y]$. One can obtain a "projective space" database $\mathbb{P}^1_k$ as the colimit of the diagram

$$\text{Spec } k[x] \leftarrow \text{Spec } k[x,y]/(xy = 1) \rightarrow \text{Spec } k[y]$$

of affine databases, just like in algebraic geometry. I plan to explore this connection.

3.1.2 Applied theory

Conversations with David Balaban and Allen Brown (see Section 2.3) have led to several ideas about how to proceed with simplicial databases. The most important is simply finding a way to implement them on a computer. The theory acts as a kind of proto-code that can be used to guide the programmer. However, to actually bring it to fruition will take more time and work than I had originally envisioned. In order to do so, Balaban and I have plans to sit down with a working database and make a refined dictionary between the ideas and terminology used in my paper and those used by working database administrators. This negotiation process will surely lead to new insights for all involved.

While the work discussed in the above paragraph is necessary, it is not quite academic research. Academic research possibilities include applications to constraint programming and to linear logic and workflow, and exploring the relationship with formal grammars.

In constraint programming, one forces a set of variables to conform to a set of propositions. For example, one may say “for variables a, b, c and n, all of which are positive integers, find solutions to the equation $a^n + b^n = c^n$. ” As easy as it is to phrase, this constraint question took 350 years to conclusively solve. Most constraint problems are much more complex to phrase and simpler to solve. These are precisely the kinds of questions that computers are most capable of attacking.

Constraint problems can be phrased in terms of simplicial databases. The variables are typed and often grouped together in various ways. Each constraint, phrased as a proposition in the language of these variables, constitutes an “ideal” in the sense of Section 3.1.1, and thus yields a table of local solutions. These tables of local solutions are joined together along common variables to form a database: solutions to the constraint problem are global sections of the data-sheaf.

Brown suggested that the query-paths in a simplicial database (see 2.1) are akin to workflows. In business, a workflow is a specification of a location from which certain materials or data will be taken, instructions on how those materials will be combined and processed, and a location at which to deliver them. Given a simplicial database, a query-path is precisely this information. The only difference is that the data at the start location is not moved or deleted in the process, whereas materials in a workflow are moved. This brings up the subject of linear logic, in which data cannot be freely duplicated.

There is surely a way to incorporate linear logic and workflow into the simplicial database model. In order to do so, however, one has to decide how to deal with information that is shared. Such decisions must not only satisfy the minds of those with expertise in these areas, they must satisfy the rigors of category theory to ensure that nothing is being forgotten. Once this is done, a nice model will emerge.

Finally, there seems to be a relationship between presentations of database schema and context-free grammars. Balaban suggested that I consider the mathematics of collapsing portions of database schemas into complex types. This process is akin to making a rule in a context-free grammar. I hope to finding some sort of functorial relationship between these two worlds, which would be very fruitful in my pursuit of an overarching framework.

3.2 Ontologies

There is a strong connection between ontologies and toposes. In some sense, a topos is a utopian version of an ontology, in which all concepts expressed in the ontology have been explored to their fullest degree. In actuality, however, the project of converting an ontology into its associated topos can never be completed. Thus toposes are the wrong model for ontologies – we need something more like “incomplete toposes” or finite pre-toposes.

To that end, I have begun work with Mathieu Anel on finding an adequate model. We have ideas which need to be written up in a form that is coherent to working ontologists with at least some interest in category theory. Once that is done, I will look for functors that connect ontologies to databases, and then attempt to convert the “communication protocol” we developed for ontologies into a communication protocol for databases.

In a totally different direction, it would be interesting to apply concepts from Bayesian probability theory to the theory of ontologies. Suppose a type (e.g. playing card) has several subtypes (e.g. queen or heart). A subtype relationship may be annotated with its "likelihood" (so that, given a playing card one can evaluate how likely it is to be a queen or a heart). Without intending to do so, humans often annotate subtypes of a given type with likelihoods. In its positive formulation this is called judgment or familiarity; in its negative formulation it is called stereotyping or bias. It would be interesting to consider ontologies in this light, because the added information is certainly relevant to the way ontologies are used in practice.

3.3 Mathematical referencing project

In mathematics research, one provides rigorous definitions, theorems, and examples about precise topics. In order to prove results, one references a result or definition either within his or her own paper, or within a different paper. In so doing, the notation may change or the theorem be applied in a way that is perhaps unexpected. Changes of notation and indexing can be annoying or even stifling to research.

One way to phrase this issue is that mathematics documents present structured information but not in a structured way. Reading mathematics would be much more efficient if the documents had some internal configuration by which references to other theorems could be established. For example, suppose one proves in theorem A that if x is even then xy is even, and suppose in theorem B we wish to apply that fact to the case of n^2 where $n = 6$. To do so, we first point theorem B to theorem A, we then say that x and y are both pointing to n, and we finally say that $n = 6$ satisfies the hypothesis for A.

The issue here is only that of substituting variables. This process may sound trivial, but when one attempts to connect notations between two distinct documents with different notations and types that are more complex than even integers (e.g. categories of sheaves on a site), this kind of thing can take much more time for the reader than it does for the original researcher. Furthermore, composition of even trivial changes-of-variables can be quite painful. The process of converting the results of paper A into a notation that is more convenient for researcher B can be automated, given an underlying grammar structure. Usage of this automated process could save days or weeks in some cases.

3.4 Applications to other fields

Information and communication are part of every human affair. Learning from researchers in other academic fields such as linguistics, philosophy, communications, education, and law would certainly lead to new insights in the study of information.

I hope to form active collaborations with people in these fields. I believe that the rigorous aspects of these fields can be formulated in terms of category theory and that doing so will lead to new insights for both parties. For example in creating laws, a legislature may attempt to say something that is quite precise. Doing so may be beyond the capacity of the English language, and yet that is all the legislators have to use. Perhaps a good and comprehensible category of information would be useful to them. If the links between different laws were formed within a rigorous structure (like the one suggested in Section 3.3), an average citizen might be able to navigate the legal system more easily.

At a “meta” level, forging communication between mathematics and other disciplines (like linguistics or law) is like forging communication between foreign entities (see Section 1). The terminology and structure of the two entities differs, and yet to some degree they are talking about the same things. Learning to engage other disciplines and make sense of their world-views within the language of mathematics will surely be beneficial to the community on many levels.

3.5 Overarching framework

The long-range goal of this project is to understand what information and communication are. Such an understanding probably could not be phrased in the language of information itself (as something like Tarski’s “undefinability theorem” probably would apply). However, having such a goal in mind gives a sense of direction within the chaos presently existing within the world of information.

It is obvious, but also exemplified in this document, that information and communication make up a vast and intricate field. If information is one complex thing, then in order to know what it is one must know how it is used in all its forms, and then integrate these diverse forms into a single framework. To the extent possible, I will seek to provide such a framework.

Part II

Future Naval Relevance

Complex systems have a pressing need to process information efficiently, and the Navy is no exception. A group only functions as a unit when all the parts are in good communication. Data from one part of the structure needs to be transferable to and understandable by other parts. Short-term solutions, such as creating links between data sets in an ad hoc manner, will inevitably fail as the system evolves and becomes more complex.

Needed is a more encompassing and foundational viewpoint about how information should be modeled. Category theory not only provides this foundation, it is a powerful tool for organizing information in a way that is flexible, transferable, and scalable.

Part III

Project Schedule and Milestones

I will begin by seeking out researchers in neighboring fields with whom to discuss the subject of information and communication. Most of my efforts will be spent in mathematics, computer science, and linguistics. Within computer science, I hope to interest people in collaborative efforts to reduce my current ideas, especially simplicial databases, to practice.

By the end of 2010, I plan to have written a paper that links ontologies to databases. I will then work to translate the communication protocol that Anel and I formulated for pairs of ontologies (see Section 3.2) into a communication protocol for databases. Around the same time, I will work to formulate a connection between grammars and databases. I will write up my findings in these areas by the end of 2011. For the rest of the grant period, I will continue to work with my collaborators to consolidate and extend the theory so that it begins to take on the format of an overarching framework for the study of information.

Part IV

Reports

The following data deliverables will be provided:

- Annual Technical and Financial Progress Reports
- Final report

Part V
Management approach

No personnel apart from the PI are planned under this research. MIT provides basic office space and library services. The campus at the Massachusetts Institute of Technology is networked by both a wired 100/1000Mbps Ethernet LAN and campus-wide 802.11b/g wireless networks.

Part VI
Current and Pending Project and Proposal Submissions

I currently have a grant from the ONR, detailed below; it is set to end 2010/06/15. I have no other current or pending projects or proposals.

1 Current grant: N000140910466

Title of Proposal: "Geometric networks: A higher-dimensional approach to networks and databases").

Summary: A category-theoretic study of databases, networks, and learning.

Source and amount of funding: Office of Naval Research. Annual direct cost: \$76,804. Percentage effort devoted to this project: 66%.

Identity of prime Offeror: David I. Spivak

List of subcontractors: None.

Technical contact:

David I. Spivak
MIT Math Department
77 Massachusetts Ave, Bldg 2, Rm 236
Cambridge MA 02139
`dspivak@gmail.com`
(510) 684-6425

Administrative/business contact :

Chingling Reed
Office Manager, Mathematics department
218 Fenton Hall
University of Oregon

Eugene OR, 97403
Phone: (541) 346-0989
Fax: (541) 346-0987

Period of performance: Period 1: 2009/01/01 – 2009/09/30.
Period 2: 2009/10/01 – 2010/09/30. Note: this grant will be de-obligated as of 2010/06/15 at the University of Oregon.

Total award amount and person-months : Total obligated on award: $175,000. Grant total: $300,000.
Person-months: 9 months (PI).

Relation to present proposal: The present proposal is a renewal of this award.

Part VII
Qualifications

The PI received a Ph.D. in Mathematics from the University of California, Berkeley in 2007. The PI was then employed as a Paul Olum Visiting Assistant Professor by the University of Oregon from 2007 to 2010. He is well-trained in category theory and has studied information and communication for several years.

The PI's curriculum vitae is attached.