

Semantic Models of Research Phenomena

Robert B. Allen
New York, NY, USA
0000-0002-4059-2587

II. IMPLEMENTATION

Abstract—As an example of qualitative modeling of research phenomena, we describe semantic coding of the dynamic Snowball Earth system as an object-oriented program. The program implements a microworld with objects ranging from photons and molecules to glaciers and continents, along with transitionals that make state changes to those objects when specific conditions are met. We show the data structures used to describe objects and their states, and how the program executes state changes. This extends our previous work on highly-structured research reports.

Keywords—explanation, inspectable semantic models, qualitative models, semantic programming

I. HIGHLY-STRUCTURED RESEARCH REPORTS

We have been developing tools to build and provide access to semantic models of scientific research reports [1]. Semantic models are qualitative models at the level of what is typically presented in text. Here we extend our previous work to implement a state-based executable model of the mechanisms and flows that describe research phenomena. Research phenomena are often embedded as part of a complex system. In [2], we demonstrated a qualitative model for experimental research, specifically for Pasteur's swan-neck flask experiment, which partially modeled the underlying research phenomena. Some sciences, such as geology, focus more on developing and reasoning from models of a phenomenon than on experimental evidence. We aim here to make qualitative structured representations of those dynamic systems.

Plausible mechanisms are essential for the credibility of a system model [7]. Because systems are based on the interaction of their components, we emphasize descriptions of transitions and mechanisms. In [3], we implemented a visualization of the chain of transitions that were central for describing evidence supporting the Snowball Earth Theory [9]. We called these chains "Causal Threads". While the Causal Threads provide a general overview of the sequence of events proposed to explain the transitions, they are insufficiently fine-grained and were not formally specified as structured knowledge. Here we supply executable micro-mechanisms that realize those threads.

Our effort is related to qualitative system dynamics, but provides a unified framework for describing the underlying mechanisms. Instead of implementing a quantitative simulation with text annotation or cognitive models, our focus is on representation. Our approach can complement LLMs and provide an alternative because it is inspectable and resistant to hallucinations.

A. Snowball Earth Theory

We chose the Snowball Earth theory [9] as a case study. That theory proposed that the Earth froze over completely during the Neoproterozoic Era. It provides three main lines of evidence, which are reflected in our three Causal Threads: low-latitude glaciation, calcium sedimentation, and iron sedimentation.

B. Knowledge, Control, and Discourse Structures

The Causal Threads in [5] offer a visualization of mechanisms without modeling them in detail. Here, we explore how those mechanisms can be modeled. We build on low-level units of knowledge from ontologies such as BFO [6] and SUMO [10], the latter of which has broad coverage and is axiom-based. Our knowledge structures generally follow those in our previous papers [2, 3]. They include Regions and Layers, which vary in content and transparency. Unlike many traditional ontologies, we allow transitionals, which make state changes so that mechanisms are, essentially, represented by short programs. We draw on linguistic resources such as the Frame Elements of Frame Net [8, 11], which we cast into a state-based programming language. Beyond the semantic representations, there are subjective decisions, such as how the system should be parsed and presented. These are analogous to the choices authors make when writing the text of a research report. In addition, the programs include control structures (e.g., looping).

C. The Program

We use a dynamic, object-oriented program to represent the systems. Object-oriented programs can be considered a type of knowledge graph. They have inheritance and object graphs. In this project, the models were implemented in Python and made extensive use of Python dictionaries.

Descriptions of mechanisms such as those for Snowball Earth interleave descriptions of single objects (e.g., photons of molecules) and descriptions of large collections of those objects. Our qualitative models incorporate both levels. There is a close conceptual relationship between possible transactions for the members of a collection and the entire collection. There are many ways to implement structured descriptions for collections. Some collections, such as the stream of photons, have far too many members to enumerate. To model them, we can pick an exemplar or characterize them by creating bins of some property and then describing the behavior of those bins. In this version, we use the latter. In the case of photons streaming continuously from the Sun, we divide them into collections and model the transit of a

collection as a unit. We would revert to describing individual photons when describing details of their behavior.

As an example of the data structure, we show two selected lines associated with the transitional “to emit,” which is based on the Frame Net’s Emitting Frame. The first line states that a “solarPhoton” and a “solarPhotonCollection” are allowed objects (the theme) to be emitted from the source (cname = “sunObj”).

```
w.objs[self.cname]["trs"]["expected"]["T_emit"]["theme"] |=
    ["solarPhotonSun", "solarPhotonCollectionSun"]
w.objs[self.cname]["trs"]["expected"]["T_emit"]
    ["direction"] |= ["toSpaceRegion"]
```

This structure requires that the emissions are emanating from the sunObj and not the other way around. Because “emitting” is typical of stars, we label it as “expected”. Other frames such as “freezing” or “driving” would be marked as “unlikely”. In the model, the solarPhotonCollection is initially located in the sunRegion, but when the model runs, the location is changed to spaceRegion.

```
self.doEmit("solarPhotonSunStream", "location",
    "sunSurface", "spaceRegion", "toSpaceRegion")
```

Some transitions are more complex. When a stream of photons strikes the Earth’s surface, some photons are reflected and some absorbed. Thus, two new collections are created, and each has its own transition. This applies the “splitting” lexical unit from the Separation frame. Then we apply “reflect” or “absorb” transitions to the new collections. Taken together, these form a compound transition.

```
self.doSplit("solarPhotonAtmosStream",
    "solarPhotonEarthReflectStreamAtmos",
    "solarPhotonEarthAbsorbStream", "albedo")
```

The ratio of the split is based on the albedo of the surface, and the resulting streams must sum to equal the incoming stream. We approximate this with the qualitative representation.

For the stream that is reflected, its direction is changed. The photon stream that is absorbed ends, and its energy is transferred to molecules in the Earth’s surface.

```
self.doAbsorbToEarth("solarPhotonEarthAbsorbStream",
    "earthSurface", "earthSurfaceTemp")
```

The photon absorption would warm the surface, and that warming stimulates blackbody emissions so that over time a temperature equilibrium is maintained.

There are many possible checks for the validity of the model, such as determining whether a transition is legal for a given system. We have implemented some of them, and many more checks are possible.

III. CONCLUSION AND DISCUSSION

We have demonstrated a framework for highly structured qualitative simulation of the transitions proposed in the

Snowball Earth theory. In this brief presentation, high-level issues have been covered. Several other features are included, and there are opportunities for many more, such as:

- The mechanisms described here, along with the graphic interface in [5], could be used to support structured exploration, explanations, evidence, as well as extensions and criticisms of the Snowball Earth theory.
- In many cases, explanations are highly context-specific, but ultimately they should be linked to broad scientific principles. This would reduce the number of special cases. Such explanations can be grounded in different scientific paradigms (e.g., Newtonian, atomic, subatomic).
- In this paper, we focused on handling major cases and avoiding exceptions. Those exceptions can be flagged and, eventually, incorporated into the rule base.
- We have introduced several object-oriented features, but one feature we do not yet support is allowing the objects to behave as independent agents and respond entirely to their context.
- Taken to the limit, our approach suggests that much of natural language can be structured and coded. Toward that end, a new generation of rich semantic resources can be developed. While current semantic resources are useful, still richer resources would be helpful. For instance, we could extend FrameNet by providing coding for all lexical units and, beyond that, by specifying context for applications of those lexical units. While this would be a massive undertaking, modern linguistic tools could automate much of the work.

REFERENCES

- [1] R.B. Allen, “Model-oriented scientific research reports”, *D-Lib Magazine*, May 2011. [DLIB](#), [DOI](#)
- [2] R.B. Allen, “Yoked flows for direct representation of scientific research”, Digital Infrastructures for Scholarly Content Objects (DISCO), Sept. 2021, JCDL Workshop. [CEUR-2976-3](#),
- [3] R.B. Allen, “Using causal threads to explain changes in a dynamic system”, ICADL 2023. LNCS 14458 (open access), [arXiv](#)
- [4] R.B. Allen, Semantic models of flows, ICADL, 2024, pp 110-118, [DOI](#)
- [5] R.B. Allen, Supporting Exploration of Structured Semantic Models, arXiv, 2026, in preparation
- [6] R. Arp, B. Smith, and A.D. Spear, Building Ontologies with Basic Formal Ontology, 2015, MIT Press
- [7] W. Bechtel, Discovering cell mechanisms, 2012, Cambridge University Press
- [8] A. Dolbey, M. Ellsworth, and J. Scheffczyk, “BioFrameNet: A domain-specific FrameNet extension with links to biomedical ontologies”, 2008, KR-Med
- [9] P.F. Hoffman and D.P. Schrag, “Snowball Earth”, 2000, Scientific American, 282(1), pp 68–75. DOI: [10.1038/scientificamerican0100-68](#)
- [10] A. Pease, Ontology: A practical guide, 2011, Articulate Press
- [11] J. Ruppenhoffer, M. Ellsworth. M.R.L. Petruck, C.R. Johnson, and J. Scheffczyk, FrameNet II: Extended Theory and Practice, 2010, available: [https://my.eng.utah.edu/~cs6961/papers/FrameNet_book.pdf](#)