Issues for the Direct Representation of History

Robert B. Allen(✉)

Yonsei University, Seoul, South Korea
rballen@yonsei.ac.kr

Abstract. We propose that representations for structured models of human and social history need to go beyond traditional ontologies to the combination of rich semantic ontologies with programming languages. We base our approach on the Basic Formal Ontology (BFO) and then consider how to extend it beyond traditional approaches to ontology with higher-level structures. For instance, we propose the need for composite entities that allow transitions in the configuration of component entities. We then explore the relationship of these composite entities to notion of systems and consider how they may provide a definition of "causal unity" and be related to models of social systems. We identify some challenges in defining the nature of social entities. Finally, we introduce structured applied epistemology as a framework for managing historical evidence, analysis, and argumentation.

Keywords: Causal unity · Conditionals · Community models · Composite reference entity · Historical newspapers · Late binding · Scenarios · Systems

1 Introduction

Ontologies can be applied to the description of history [12]. However, histories go beyond typical applications of ontologies, in having extended interactions of complex instances across many different contexts. In addition, consider the difficulty that traditional search engines have in indexing the broad range of stories in historical newspapers. Individuals, institutions, and locations may be closely coordinated in some articles but are found in very different settings in other articles. We propose that higher-level structures and semantic search are needed. Such higher-level structures could support a wide range of new interactive services ranging from games to educational and scholarly applications. Semantic search may greatly enhance performance for searches of the full range of historical newspaper articles. We have termed the use of rich semantics to organize content from digitized historical newspapers – and by extension all historical material – as developing semantic "community models" [5].

We adopt the Basic Formal Ontology (BFO) [8] as the basis for modeling history. BFO is an upper ontology that is widely used in biology. BFO is an appropriate choice because it supports "realist" modeling and there should be value in working toward a unified framework that has the potential to be integrated across different domains. While BFO is relatively well developed with respect to the description of natural science entities, it is still a work in progress and important aspects of it are still open. Here, we

A. Morishima et al. (Eds.): ICADL 2016, LNCS 10075, pp. 218–224, 2016.
DOI: 10.1007/978-3-319-49304-6_26

discuss some high-level structures that might be developed to facilitate the representation of history. Several of these issues have been discussed in our earlier work such as developing complex entities and basing representations on object-oriented programming languages (e.g., [4]).

2 Rich Semantics

2.1 Continuants and Composite Entities

As an upper ontology, BFO specifies different types of entities. These different types of entities are coordinated through different types of relationships to form descriptions of complex entities. For instance, an ontology would likely specify that Continuant Material Entities (e.g., Objects) would have the Dependent Continuant Quality of Mass. If Mass were not specified in a given ontology of physical properties, then it cannot be used for descriptions based in that ontology.

It would be useful to have greater ability to examine and evaluate configurations of composite Entities. Such configurations could be developed as schemas. Indeed, there could be several types of schemas. A Composite Reference Entity would be a schema of all the relevant parts of a non-instantiated composite entity. There could be inheritance from the Composite Reference Entity. The Composite Reference Entity could be one part of a Reference Ontology [8, p39]. The Composite Reference Entities could also be instantiated, what we term Instantiated Composites. In the example above, the Reference Entity would have a Mass and, possibly, a range of acceptable values for that Mass, but only Instantiated Composites will have would have a specific value for Mass.

An Instantiated Composite could have a long existence. Consider if the Instantiated Composite represented a person. In the course of a lifetime, that individual could have a vast range of experiences. One hand the potential for those experiences may have been present in the original person Reference Entity. On the other hand, there is such a broad range of outcomes for an individual; it does not seem plausible associate every possible outcome with the original Reference Entity. Rather, some outcomes (e.g., Roles associated with a person) would have to be attached to the Instantiated Composite across time. In in terminology of object-oriented programming, this is known as late binding [4].

2.2 Occurrents and Transitions

A classic paradox for ontologies is to distinguish between the sense in which a river is a fixed entity and the sense in which it is continually being renewed. In BFO, this paradox is addressed by allowing the river to be both a Continuant and an Occurrent. They are entities that change through time in the sense that a river is always changing. Occurrents are composed of the "present participle" form of verbs (e.g., the running done by a Continuant across a time region) [8, p. 121]. In BFO, Occurrents include Processes, Process Boundaries, and Temporal Regions (i.e., intervals and periods). One specific type of Process is a History, which according to BFO is composed of the processes that occur in the Temporal Region covered by that Material Entity [8, p122]. While such

definitions are clearly useful for describing different types of biological activities, the BFO definition of a History might more accurately be called a chronology. We also note the similarity of a Temporal Region to the historical notion of an era or events that extend across time.

We believe it is also useful to identify Transitions. A Transition is any change in the structure of an Composite Entity. We can distinguish between transitions of individual Material Entities (BFO:Object) and Transitions across Reference Composites or Instantiated Composite. Transitions are due to Processes. Because the Processes are related to Transitions, they should specify a Domain and Range. Moreover, the relationship between Processes and Material Entities is said to be analogous to the relationship of a Specifically Dependent Continuant to a Material Entity [8, p. 121]. This linkage is notably similar to the linkage of classes and methods in Java. It implies that Material Entities can have both quality-like processes and "realizable" Processes. Thus Processes like breathing could be considered as part of the definition of an Independent Continuant. This is a very low level of modeling and is consistent with our call for a programming-language approach to rich semantics. In the example above, it considers both the river and the changes in the river's molecules. Another high-level structure is a Scenario [2, 4]. The Scenario is the relevant context for a collection of entities. We can say that, a Transition between parts of a Process is caused by other Entities in the Scenario (cf. [7]).

A Procedure is a prescription whose execution can cause Transitions in a Process. A Procedure is different from a Process in having gates or contingencies. It is usually controlled by a third party. For instance, in Gibbon's "The Decline and Fall of the Roman Empire" we find:

> Those princes, whom the ostentation of gratitude or generosity permitted for a while to hold a precarious sceptre, were dismissed from their thrones, as soon as they had performed their appointed task of fashioning to the yoke the vanquished nations. The free states and cities which had embraced the cause of Rome were rewarded with a nominal alliance, and insensibly sunk into real servitude.[1]

According to the passage, the princes were permitted by Rome to hold their thrones until "they had fashioned a yoke". Computer programs are artifacts that specify Procedures.

2.3 Structure and Function

Structure is the relationships among the parts of a composite entity. For instance, the chapters of a book form the structure of the book. Structure is often contrasted with function. Across many academic disciplines, insights are gained by considering complexity in terms of both structure and function. Consider anatomy (structure) versus physiology (behavior and function).

In BFO, Functions are Dispositions that are Realizable Entities. Functions specify potential Transactions [14]. For example, a hammer has the function "to drive in nails"

[1] From E. Gibbon, Decline and Fall of the Roman Empire, Chap. 1: The Extent of the Empire In The Age Of The Antonines.—Part III. http://www.gutenberg.org/files/731/731-h/731-h.htm#link22HCH0002.

[8, p 103] and a Continuant with a Function can have Agency (though not intention). A structured set of interacting entities forms a System (e.g., [4]) and a full description would include the entire set of sub-systems that constitute the overall system. We believe that such systems provide a definition of "causal unity".

Both structure and function may provide useful perspectives to societal models. A range of sociological "grand theories" has been developed applying the different approaches. There are structuralists such as Levi-Strauss, functionalists such as Malinowski and structural-functionalists such as Parsons. The different approaches to societal models employ different assumptions and tools. Our approach, which is grounded in object-oriented modeling, seems most compatible with structure-functionalism as an approach to social systems.

Many aspects of historical models involve social entities. However, social entities are contentious and difficult to model. They are based on an unpredictable and even chaotic foundation – the needs of individuals. Social systems may be self-organizing and different cultures at different times have found their own ways of balancing these factors. We consider some outlines that appear useful to explore further. We might develop an ontology of human Needs based on the work of Maslow [11]. Presumably, Needs are grounded in biological processes that could be linked back into BFO. Even without being able to model the mechanisms in detail, we can see that relatively stable social structures (e.g., religions, governments) may develop. Work on Social Ontologies which is related to BFO has explored deontic entities (i.e., rights and obligations). Particular attention is paid to document acts (laws, regulations, records, etc.) [13]. One of government's major activities is establishing frameworks for document acts and enforcing their commitments.

3 Example Application

We have been exploring community models to support access to digitized historical newspapers (e.g., [5]). As a specific example to illustrate some of the points explored here, we selected an individual essentially at random and found a set of news reports about him: D.C. O'Connor, who appeared in the newspapers of Norfolk, Nebraska over several years primarily in his role as Superintendent of Schools of Norfolk, Nebraska. He moved from Superintendent of Schools for the town of Madison to Norfolk. O'Connor worked in Norfolk until 1905 when he resigned and became Superintendent of Schools for the Panama Canal Zone. During the years he was in Norfolk, there are numerous reports in the newspapers of his activities for the School Board such as making reports and participating in professional organizations. There are also reports of his participation in local social activities. In one news item,[2] a teacher (McCoy) resigned to

[2] *The Norfolk Weekly News Journal* (Norfolk, Neb.), 29 March 1900. *Chronicling America: Historic American Newspapers*. Library of Congress. http://chroniclingamerica.loc.gov/lccn/sn95070060/1900-03-29/ed-1/seq-7/

The Norfolk Weekly News-Journal. (Norfolk, Neb.), 04 Sept. 1903. *Chronicling America: Historic American Newspapers*. Library of Congress. http://chroniclingamerica.loc.gov/lccn/sn95070058/1903-09-04/ed-1/seq-6/.

take a position in a different school district and the Superintendent needed to replace him. We can model this particular development with Entities such as the School District and the Roles of the School Superintendent and Teacher (cf., [9]) as well as implementing Transitions for hiring, resigning, and traveling. We interpreted employment as a type of Role attachment. Thus, hiring is implemented as a transition in which the Role (e.g., Teacher or School Superintendent) becomes linked to the Person. Constraints could be included such as requiring an employee to be of a certain age. Instances could be created for the school entities of the local towns, the individuals involved (O'Connor and McCoy) and their initial roles. In terms of narrative structures, this could thought of as defining the Setting.

As a first transition, O'Connor resigned from the Madison School District. A second transition was his hiring by the Norfolk School District. These transitions are directly connected and might be thought of as forming a thread, episode, or scenario. A second set of related transitions starts with McCoy's resignation as a Teacher. The vacancy created by McCoy's resignation triggers an activity associated with O'Connor's Superintendent Role, a recruiting trip. Unlike the other transitions in these passages, the transition between McCoy's resignation and O'Connor's trip is best characterized as a causal relationship. In response to the vacancy, O'Connor's obligation to find a replacement is triggered and to fulfill that obligation, he makes a trip to find a replacement teacher. We could implement that trip with the Motion frame.

The news item ends with the description of O'Connor on the train without reporting the results of the trip. We may believe that he succeeded in recruiting a new teacher and thus filled the vacancy but we need to mark that as a speculation. We can also be confident that the positions of Superintendent and Teacher include specific activities in a job description though we can probably never know the details of those job descriptions. A common challenge for these implementations is gaps in knowledge. The use of plausible inference is common to compensate. The formal structure highlights the number of such inferences that are required and the need for explicit placeholders to show that some details are unknown.

4 Coordinated Collections of Historical Resources and Structured Applied Epistemology

[3] called for coordinated, large-scale repositories of historical resources. We renew that call and add that material could be enriched in several ways. We could create rich social and historical domain ontologies for descriptions. For that, the approach of the OBO Foundry for BFO-based biological domain ontologies could be extended to the development of social and historical domain ontologies. A related project could develop collections of standard procedures and policies (e.g., job descriptions), geographic locations, train schedules, as well as census and economic data. These could be linked into collected texts. Moreover, the texts, reviews about them as well as facts about their authors, and their publication histories could be positioned on a broad temporal map (see [6]).

The texts could also be enriched as analysis or argumentation structures. The texts may provide evidence for claims about history. The evidence could be a quotation from

another text. The links into the texts could be thought of as comparable to footnotes in narrative and analytical histories [1, 3].

Epistemology is usually distinguished from ontology in focusing on criteria for how we can be said to know something. Structured applied epistemology is concerned with providing support for belief and knowledge about claims based on semantic models. This could also be considered structured semantic abduction where abduction is defined as determining the best explanation given the evidence. These analyses could consider the plausibility alternative semantic models of scenarios. The scenarios could be evaluated at several levels starting with whether they meet basic physical constraints. For instance, two Material Entities cannot to occupy exactly the same Spatio-Temporal Region. The scenarios should also be supported by evidence linked to specific aspects of the scenario with structures reflecting the traditional criteria for the evaluation of historical claims [10]. For instance, the reliability of the source of a quotation used as evidence could be critiqued. If there were disagreement about the evidence and warrants, the differing viewpoints could be explored with argumentation structures.

References

1. Allen, R.B.: Model-oriented information organization: part 2, discourse relationships. D-Lib Mag. (2013). doi:10.1045/july2013-allen-pt2
2. Allen, R.B.: Repositories with Direct Representation, NKOS Workshop, Seoul (2015) arXiv:1512.09070
3. Allen, Robert, B., Chu, Y.: Towards a full-text historical digital library. In: Tuamsuk, K., Jatowt, A., Rasmussen, E. (eds.) ICADL 2014. LNCS, vol. 8839, pp. 218–226. Springer, Heidelberg (2014). doi:10.1007/978-3-319-12823-8_22
4. Allen, R.B., Chu, Y.M.: Architectures for complex semantic models. IEEE Big Data Smart Comput. (2015). doi:10.1109/35021BIGCOMP.2015.7072809
5. Allen, R.B., Japzon, A., Achananuparp, P., Lee, K.-J.: A framework for text processing and supporting access to collections of digitized historical newspapers. HCI Int. Conf. (2007). doi:10.1007/978-3-540-73354-6_26
6. Allen, R.B., Song, H., Lee, B.E., Lee, J.Y.: Describing scholarly information resources with a unified temporal map. In: Morishima, A., et al. (eds.) ICADL 2016. LNCS, vol. 10075, pp. 212–217. Springer, Heidelberg (2016)
7. Allen, Robert, B., Wu, Y., Luo, J.: Interactive causal schematics for qualitative scientific explanations. In: Fox, Edward, A., Neuhold, Erich, J., Premsmit, P., Wuwongse, V. (eds.) ICADL 2005. LNCS, vol. 3815, pp. 411–415. Springer, Heidelberg (2005). doi:10.1007/11599517_50
8. Arp, R., Smith, B., Spear, A.D.: Building Ontologies with Basic Formal Ontology. MIT Press, Cambridge (2015)
9. Chu, Y., Allen, Robert, B.: Formal representation of socio-legal roles and functions for the description of history. In: Fuhr, N., Kovács, L., Risse, T., Nejdl, W. (eds.) TPDL 2016. LNCS, vol. 9819, pp. 379–385. Springer, Heidelberg (2016). doi:10.1007/978-3-319-43997-6_30
10. Howell, M.C., Prevenier, W.: From Reliable Sources: An Introduction to Historical Methods. Cornell U. Press, Ithaca (2001)
11. Maslow, A.: Theory of human motivation. Psychol. Rev. **50**, 370–396 (1943)

12. Meroño-Peñuela, A., Ashkpour, A., van Erp, M., Mandemakers, K., Breure, L., Scharnhorst, A., Schlobach, S., van Harmelen, F.: Semantic technologies for historical research: a survey. Semant. Web **6**, 539–564 (2015)
13. Smith, B.: Document Acts, Conference on Collective Intentionality, Basel (2010). http://ontology.buffalo.edu/smith/articles/document-acts.pdf
14. Spear, A.D., Ceusters, W., Smith, B.: Functions in Basic Formal Ontology, Applied Ontology, 11 (2016, in press)