Addressing Entity Change in Procedural Ontologies

Tyler Johnson¹, Mohammed Alliheedi², Yetian Wang³ and Robert E. Mercer¹

¹Department of Computer Science, The University of Western Ontario, London, Canada
²Department of Computer Science, Al-Baha University, Al Bahah 65527, Saudi Arabia
³David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Canada

Keywords: Experimental Procedure, Procedural Steps, Biochemical Ontology, Time series, Forward-Backward Sequencing.

Abstract: Ontologies model a domain by representing the entities, concepts, and the relations between them. The domain of interest in this position paper is the biochemistry experimental procedure. These procedures are composed of procedure steps. These steps represent actions. Actions cause change, a concept being implicitly modelled in this type of ontology. We argue that entities undergoing change need to be properly captured in the ontology. The biochemistry procedure Alkaline Agarose Gel Electrophoresis is used to demonstrate the generality of this procedural ontology.

1 INTRODUCTION

The representational adequacy of an ontology, which provides entities (also known as individuals), concepts, and relations among them, is determined by its ability to model any situation that can occur in the domain being represented. Our focus is the biochemistry experimental procedure domain.

Descriptions of biochemistry experimental procedures exist in scientific writing. These protocols, which typically involve several steps, are described in detail in manuals of standard biochemistry experiment procedures (Boyer, 2012; Sambrook and Russell, 2001).

This position paper builds upon previous work by (Alliheedi et al., 2020) which describes ontologies for two experimental procedures that exist in the manual of standard biochemistry experimental procedures (Sambrook and Russell, 2001): Alkaline Agarose Gel Electrophoresis and Southern Blotting. In this position paper, we provide evidence that to design an ontology that is adequate for this domain, proper labelling of entities that change is required. In addition, we make explicit the connection to a measurement ontology and indicate that the modifications conform to two other ontologies: the geospatial ontology and BFO. To demonstrate these modifications, we show a portion of one example of a biochemistry experimental procedure, Alkaline Agarose Gel Electrophoresis.

2 RELATED WORK

The OBI ontology has been the foundation for several ontologies in the field of biochemistry procedures, including those proposed by (Courtot et al., 2008; Brinkman et al., 2010; Zheng et al., 2013; Soldatova et al., 2013; Dumontier et al., 2014). These ontologies are of great interest to our research, and we provide a brief description of them in this section.

Soldatova and King (Soldatova and King, 2006) proposed EXPO, an ontology of scientific experiments, which offers a detailed description of various aspects of scientific experiments and their relationships. In contrast, OBI describes experimental processes and relationships, and Brinkman et al. (Brinkman et al., 2010) discuss three real-world applications that provide relevant input/output relations for our purpose. Similarly, Zheng et al. (Zheng et al., 2013) introduced the beta cell genomics application ontology (BCGO), which also uses OBI to describe experimental processes but is primarily a descriptive ontology. However, some of the relations in RO, the relation ontology (Smith et al., 2005), that are used in BCGO (e.g., produces, translate_to) have an ordering sense that aligns with our research.
Exact (Soldatova et al., 2013) and the Semantic-science Integrated Ontology (Dumontier et al., 2014) are the two ontologies that are most similar to our work. While both aim to describe sets of actions in scientific protocols, (Alliheedi et al., 2020) represent sequences of actions. Therefore, relations that describe orderings of actions (e.g., ‘precedes’ (Dumontier et al., 2014)) are not applicable to sequences since these relations are transitive. In addition to these ontologies, the Molecular Methods Database (MoMeth) (Klingström et al., 2013) contains scientific protocol ontologies that conform to a set of laboratory protocol standards. Other ontologies that are useful for a biochemistry procedure-oriented ontology include the ontologies for processes, such as (Lenat et al., 1985; Schlenoff et al., 2000), ontology for units of measure (Rijgersberg et al., 2013), classification of scenarios and plans (CLASP) (Devanbu and Litman, 1996), and materials ontology (Ashino, 2010). Foundational theories, such as process calculus and regular grammar, are essential for the formalization of procedure-oriented ontologies. The current paper builds upon previous work by (Alliheedi et al., 2020) and provides a detailed representation of change and an explicit connection to the measurement ontology (Rijgersberg et al., 2013).

3 PROCEDURE-ORIENTED ONTOLOGY

We have proposed in (Alliheedi et al., 2020) a framework for procedure-oriented ontologies. This framework explicitly identifies all steps of an experimental procedure and provides a set of relations to describe the relationships between these steps. This novel approach allows creating a sequence of events (or steps in a procedure) using the ontological concept of “something occurs before”. An ontological concept of “sequence” is used to accomplish this. This new concept is used to model events that happen step by step in some sort of ordering.

This approach will be used to provide ontologies representing experimental procedures for Knowledge Base systems with the required knowledge about the experimental processes used in biochemistry experimental settings. There are manuals of standard procedures in biochemistry (Boyer, 2012; Sambrook and Russell, 2001) which are being used to build extensions of our ontology.

In the next sections, we summarize the current ontological framework found in (Alliheedi et al., 2020). This enables us to connect the focus of this current position paper, the representation of entity change, to this framework.

3.1 Relations

In this section, we describe various properties to satisfy the definition of an experimental procedure. An experimental procedure consists of a series of events, that is, steps in the procedure. These steps occur in either partial or total order. Partial ordering allows steps (more than one step) to precede or follow another step. Total ordering means that a step precedes or follows another step, and this relation is intransitive. These relations are defined for OWL (McGuinness and van Harmelen, 2004).

We represent the choices of a subsequence of steps from more than one possible subsequence. Since the choices among subsequences would be “either” or “or”, the relation ‘optionalStepOf’ needs to be designed based on the different choices of available subsequences in that particular step. To illustrate, the ‘optionalStepOf’ relation is simply an ‘exclusive or’ if there are two choices available, or else it would be a generalization of the exclusive or. We have implemented the aforementioned relations to satisfy the definition of “procedure”.

3.2 Classes and Properties

The ontology framework described in detail in (Alliheedi et al., 2020) consists of three core classes: Step, State, and Action. These classes are described in the following sections. Class names are indicated with capitalized words, e.g., Step. Properties are indicated with single quotes, e.g., ‘subStepOf’. Instance names use typewriter font, e.g., step1. When referring to the actual steps from the protocol, normal font is used, e.g., step 1.1.

3.2.1 Step

Each step in a procedure is represented by instances of the Step class (see Figure 1). We defined object properties such as ‘precedes’, ‘follows’, and ‘parallel’ to represent the ordering relations of each step. Note that the aforementioned object properties are transitive. The properties ‘precedes’ and ‘follows’, inverses of each other, indicate the chronological order between two steps. The property ‘parallel’ is symmetrical which indicates steps may occur simultaneously. Figure 1 illustrates a scenario with parallel

1Available at http://www.ontologydesignpatterns.org/cp/owl/sequence.owl
steps step1 and step2. Intransitive properties 'directlyPrecedes' and 'directlyFollows' are subproperties of 'precedes' and 'follows' respectively. These properties describe the order between steps in which a step immediately precedes or follows another step. Similar to 'precedes' and 'follows', they are also inverses of each other. Therefore, by stating step1.1 'directlyPrecedes' step1.2, and step1.2 'directlyPrecedes' step1.3, a reasoner will automatically infer that step1.1 'precedes' step1.2 as well as step1.3. Also, step1.3 'directlyFollows' step1.2 but only 'follows' step1.1, both being inferable by a reasoner.

Procedures are often formed by an hierarchical structure among steps, that is, a step may consist of a number of sub-steps required to complete. The property 'subStepOf' indicates that some step(s) must be completed for the completion of the parent step. Figure 1 shows an example of a step called step1 which consists of step1.1, step1.2, and step1.3. Then step1.1, step1.2, and step1.3 must be completed in order to state that step1 is complete. The sub-step step1.1 must complete before step1.2 and thus before step1.3. The figure where one and only one of step1.1.1 and step1.1.2 needs to be completed in order to complete step1.1, the property 'optionalStepOf' can be used to indicate that one of the steps (not both) must be completed in order to complete the parent step. Both domain and range of the properties are the class Step.

3.2.2 State

The relations between instances of the class Step outline the structure of a procedure. Each instance of Step is represented as a set of states and are associated to a set of actions. A step involves a transition from state to state via a single or a series of actions, represented by the classes State and Action (see Figure 2).

The State class is connected to the Step class via the property 'hasState'. It has three subclasses, InitialState, MidState, and FinalState. The subclasses are connected via properties that include 'precedes' and 'follows'. InitialState can only precede a MidState or a FinalState. FinalState can only follow an InitialState or a MidState. MidState can 'follow' an InitialState and 'precede' a FinalState, as well as 'precedes' or 'follows' another MidState. The triple (stateX, 'precedes', stateY) implies (stateY, 'follows', stateX), 'follows' being an inverse property of 'precedes'. Figures 1 and 2 omit 'follows' to keep the figures clean. Note that a step has at most one instance of InitialState or FinalState but may have multiple instances of MidState. For example, an instance of Step, step1, may involve two instances of State, i.e., step1.state1 and step1.state2, represented by the following triples: (step1, hasState, state1), (step1, 'hasState', state2), (state1, 'precedes', state2), in which state1 and state2 are instances of InitialState and FinalState respectively. The State class also connects to classes Restriction and Entities, which are discussed in Section 3.2.4 and 3.2.5, respectively.

3.2.3 Action

States are connected to the Action class via 'beforeState' and 'afterState', representing the states before and after an action, respectively. In other words, an instance of Action would transition an instance of State to another. For example, when an action action1 is performed in state1, state1 will be modified and thus transitioned into a new state state2. This would be represented by the triples (action1, 'beforeState', state1) and (action1, 'afterState', state2).
3.2.4 Restriction

A Restriction class was created to represent certain limitations applied to a state in a step. It is linked to the class State via ‘hasRestriction’. For example, (state1,'hasRestriction',restriction1) means that restriction1 will be checked against state1. It may be sufficient to apply a restriction to a FinalState but it is also possible to apply restrictions to all states in a step.

3.2.5 Entities and Biochemistry Domain Knowledge

The State class is connected to the Entities class (see Figure 3) via the property ‘involves’. Domain knowledge of biochemistry can be described by extending the Entities class with subclasses such as Instruments, Materials, and Devices involved in a specific state. For demonstration purposes, we have included only selected general concepts related to experimental procedures described in Section 4. The class Instrument includes Container and Device where Container ‘contains’ Material which is a class for Chemical and Non-Chemical materials used in biochemistry experiment procedures. Compound materials and assembled instruments are represented using the property ‘consistsOf’. Instrument and Material can be connected to the class Measure which is a combination of numerical values and Unit_of_Measure, e.g., 10m is a measure where the value is 10 with a unit of measure of meter (Rijgersberg et al., 2013). The Measure class was extended with subclasses to represent absolute measures (e.g., 10m), range values (e.g., 5m-10m), and ratio (e.g., 1/2).

4 ITEM LABELLING

The main contribution of this position paper is to provide direction in how to represent the changes in entities during the experimental procedure in the ontology. We use the Alkaline Agarose Gel Electrophoresis procedure (see Figure 4) to demonstrate our ideas. We have generated a timeline to follow the changes in all of the entities found in this procedure. A portion of this timeline is given in Figure 5. We will now detail the new items in this timeline.

4.1 Details of the Timeline

The timeline in Figure 5 shows our proposed method for tracking changes to entities in the procedure ontology. First, entities have “identifiers” (e.g., item1). The relation ‘alias’ is used when an entity changes its identifier. The relation ‘becomes’ is used to indicate a change in some property of an entity. Examples of these properties are temperature (e.g., the slurry in the flask is heated) or a change in what constitutes the entity (e.g., the empty flask is filled with slurry). We have adopted a naming scheme for the entities. The name contains the item name, the step in the procedure, and a short description of the property change. An example of this naming scheme is item1.step1.1.mixture which becomes item1.step1.1.mixtureheated.

For entity A to “become” entity B, the following must be satisfied:

- A and B must have the same identity (we follow Sider’s view of identity, that is, an entity maintains its identity throughout time (Sider, 2003)).
- A measurable change has occurred with respect to entity A and thus is transitioned to entity B.
- If entity B is a compound material or assembled instrument based on entity A the ‘becomes’ relation also has the effect of a ‘consistsOf’ relation in the reverse direction.

Another relation in the timeline which helps to track the changes in the entities is ‘consistsOf’. We take the definition of this relation from the work summarized in Section 3.2.5: this relation connects the entity which is a compound material or assembled instrument with the components that comprise it. It is important to note that the use of these relations can be subjective (a matter of choice) or objective (given by the steps in the procedure description). For example, there is an ‘alias’ relation between glassbottle1 and Container1 and another between flask1 and Container1. These relations are objectively given in the procedure since
Alkaline Agarose Gel Electrophoresis

1. Prepare the agarose solution
   1.1. Adding the appropriate amount of powdered agarose to a measured quantity of H2O
       in either:
       1.1.1. An Erlenmeyer flask (Container 1)
       1.1.1.1. Loosely plug the neck of the Erlenmeyer flask with Kimwipes
       1.1.2. OR a glass bottle (Container 1)
       1.1.2.1. Make sure that the cap is loose
       1.2. Heat the slurry (Item 1) in (Container 1) for the minimum time required to allow all
           of the grains of agarose to dissolve using either:
           1.2.1. A boiling-water bath
           1.2.1.1. Check that the volume of the solution (Item 1) has not been decreased by
                    evaporation during boiling in (Container 1):
                   1.2.1.1.1. if yes: replenish with H2O in (Container 1)
                   1.2.1.1.2. If no: do not add H2O in (Container 1)
           1.2.2. OR a microwave oven
           1.2.2.1. Check that the volume of the solution (Item 1) has not been decreased by
                    evaporation during boiling in (Container 1):
                   1.2.2.1.1. if yes: replenish with H2O in (Container 1)
                   1.2.2.1.2. If no: do not add H2O in (Container 1)
           1.3. Cool the clear solution (Item 1) to 55°C.
           1.3.1. Add 0.1 volume of 10x alkaline agarose gel electrophoresis buffer in (Container 1)
           1.3.2. And immediately pour the gel (Item 1) into mold (Container 2)
           1.4. After the gel (Item 1) is completely set
           1.4.1. Mount it (Item 1) in the electrophoresis tank (Container 3)
           1.4.2. Add freshly made 1x alkaline electrophoresis buffer until the gel (Item 1) is just covered.

2. Prepare DNA samples
   2.1. Collect the DNA samples (Item 2) by standard precipitation with ethanol2
   2.2. Dissolve the damp precipitates of DNA (Item 2) in 10-20 µl of 1x gel buffer. (Item 3)
   2.3. Add 0.2 volume of 6x alkaline gel-loading buffer
   2.3.1. It is important to chelate all Mg2+ with EDTA before adjusting the electrophoresis samples to alkaline conditions

3. Initiate the electrophoresis
   3.1. Load the DNA samples dissolved in 6x alkaline gel-loading buffer into the wells of the gel (Container 3)
   3.2. Start the electrophoresis at < 3.5 V/cm when the bromocresol green has migrated into the gel approx. 0.5-1 cm; Turn off the power supply, and place a glass plate on top of the gel in (Container 3) and then continue electrophoresis until the bromocresol green has migrated approximately two thirds of the length of the gel in (Container 3).

4. Finalize the experiment
   4.1. Process the gel according to one of the procedures either Southern hybridization by:
       4.1.1. Transfer the DNA either:
           4.1.1.1. Directly (without soaking the gel) from the alkaline agarose gel to a charged
                    nylon membrane. Please see Southern Blotting: Capillary Transfer of DNA to
                    Membranes
           4.1.1.2. OR after soaking the gel in neutralizing solution for 45 minutes at room
                    temperature to either:
           4.1.1.2.1. An uncharged nitrocellulose as described in Southern Blotting: Capillary
                    Transfer of DNA to Membranes
           4.1.1.2.2. OR nylon membrane as described in Southern Blotting: Capillary
                    Transfer of DNA to Membranes
           4.1.2. Detect the target sequences in the immobilized DNA by hybridization to an appropriate labeled probe. Please see Southern Hybridization of Radiolabeled Probes to Nucleic Acids Immobilized on Membranes
           4.2. OR Staining
           4.2.1. Soak the gel in neutralizing solutions for 45 minutes at room temperature
           4.2.1.1. Stain the neutralized gel with 0.5 µg/ml ethidium bromide in 1x TAE or with SYBR Gold
           4.2.1.1.1. A band of interest can be sliced from the gel and subsequently eluted by one of the procedures described Recovery of DNA from Agarose Gels.

Figure 4: The steps of Alkaline Agarose Gel Electrophoresis.

once the type of container is chosen, the procedure continues and performs the same steps using the container labelled as Container 1. On occasion the choice to use a ‘becomes’ relation is subjective. For example, item1-step1.3_cooled ‘becomes’ item1-step1.3_cool withA0xAgarose and ‘consistsOf’ 10xAgarose and item1-step1.3_cooled, by virtue of the ‘becomes’ relation.

4.2 Measurement Ontology

The ontology presented in Section 3 uses the concept of measurement. Here, we have made this precise by incorporating the Ontology of Units of Measure and Related Concepts (OM) (Rijgersberg et al., 2013). In Figure 5 connections with the Measurement Ontology are indicated by green relations and are labelled with a label prefixed with “OM:”. With these explicit measurements, the ontology is able to compare measurements of the changing entities. These comparisons are indicated by red relations. Some of the measurements are quantitative (e.g., 55°C C) and some are qualitative (e.g., heated).

To reduce the number of lines in Figure 5, we assume that measurements are transitive with respect to the ‘becomes’ property. To add, properties will be transitive unless they are overridden by new measurements of the same type; however, in practice the measurements are not transitive and thus need to be connected to each one that they apply to. For example, if the temperature does not change in an entity, it will
Figure 5: Proposed Timeline for the Alkaline Agarose Gel Electrophoresis Procedural Ontology.
need to be linked back up to the previous measurement of the temperature.

At times, the ontology does not specify a specific quantity for a measure but allows the use of the ontology to determine that quantity. For instance, in Step 1.1 “appropriate amount of powdered agarose” and “measured quantity of H2O” will give the desired concentration to be used in the preparation of the gel which is determined by the size of the protein being studied (Lee et al., 2012).

4.3 Other Ontologies

There are other ontologies which seek to map out a series of events. One ontology is from (Kauppinen and Hyvönen, 2007; Kauppinen et al., 2008) where they seek to describe changes in an arbitrary geospatial region. In particular, they create an ontology out of different areas and their changes of Finland since the start of the 20th century. Using what are called “change bridges”, they connect different regions as they grow and change throughout time. In contrast to our work which uses object properties to link changes in our entities, (Kauppinen and Hyvönen, 2007; Kauppinen et al., 2008) uses individuals of a certain type of “change bridge” to represent what kind of change has occurred in the region(s).

The object properties “before” and “after” are connected from the bridge to the regions to give direction to the flow of time. This works analogously to the Action class in Section 3.2.3 with each state in time linked to an instance of Action by ‘beforeState’ and ‘afterState’.

Another ontology which aims to represent series of events is the Basic Formal Ontology (BFO) (Arp et al., 2015). Unlike the previously mentioned ontology, this is a top-level ontology which aims to support information retrieval and analysis of other scientific ontologies across a myriad of individual domains. In particular, we look at the classes of Continuant and Occurant in BFO. The former represents the concept of continuants. These entities will continue or persist throughout time. In contrast, occurants are entities which occur or happen. With respect to the ontology in Section 3, there are continuants such as a container or an instrument which can be labelled with the appropriate subclasses of Continuant. Whereas, there are occurrants such as the instances of the Action class, which can be labelled with the appropriate subclasses of Occurant.

5 CONCLUSIONS

In this position paper, evidence is provided showing that to design an ontology that is adequate for the biochemistry experimental procedure domain, entities that change must be properly labelled. We also make explicit the connection to a measurement ontology (Rijgersberg et al., 2013) and indicate that the modifications conform to two other ontologies: time (Kauppinen and Hyvönen, 2007; Kauppinen et al., 2008) and BFO (Arp et al., 2015). We demonstrate these modifications in Figure 5 with a portion of one example of a timeline of entity changes in a biochemistry experimental procedure, Alkaline Agarose Gel Electrophoresis in Figure 4, and how the measurement ontology is integrated. We have also done a timeline for another experimental procedure: Southern Blotting.

Going forward, we will incorporate the timeline into the ontology classes that have been described in Section 3. The class hierarchies provided in Section 4 indicate that the presented ontology will accept this modification. Details how to accomplish this will be developed in the near future.

ACKNOWLEDGEMENTS

We would like to thank all reviewers for their comments, which improved this paper. This research is funded by an Undergraduate Student Research Internship (USRI) from the University of Western Ontario to T. Johnson, and from The Natural Sciences and Engineering Research Council of Canada (NSERC) through a Discovery Grant to R. E. Mercer.

REFERENCES


