Representing Roles in a Biology Textbook

Vinay K. Chaudhri, Nikhil Dinesh, and Stijn Heymans

Artificial Intelligence Center, SRI International, Menlo Park, CA, 94025

Abstract. A biology textbook contains sentences such as "role of cytoskeleton is support and motility of a cell" and "NADH is an electron carrier that transfers electrons". The first sentence explicitly uses the word *role* while in the second sentence the role of electron carrier is implicit. Role concepts have been defined in the context of upper ontologies for some time, but the word *role* has also been used in linguistics and description logics to denote related but different notions. In this paper, we consider the *roles* in a biology textbook, adapt a definition of roles based on upper ontologies, and relate it to the corresponding notions in linguistics and description logics. We consider our solutions to several conceptual modeling problems for roles as we have used them in representing the knowledge contained in a biology textbook. The primary contribution of this work is in applying the representation of roles to a complex and challenging domain, in giving definitions that can be applied by biologists, and in collecting several concrete examples that require such representation.

Introduction

For Project Halo, we recently completed a knowledge engineering effort that resulted in a knowledge base (KB) called `KB_Bio_101` [9, 11, 13] which represents a significant fraction of an introductory college-level biology textbook [23]. We have used `KB_Bio_101` as part of a prototype of an intelligent textbook called Inquire which is designed to help students to learn better [8]. Inquire answers questions [11], gives explanations and engages in dialog through natural language generation [3].

The textbook contains concepts that suggest roles as illustrated in the following examples:

S1 Microfilament is an elegant structure of a eukaryotic plasma membrane protein that plays a crucial role in nerve cell signaling.
S2 Role of Cytoskeleton: Support and motility. The most obvious function of the cytoskeleton as a whole is to give mechanical support to the cell and maintain its shape.
S3 Considering the overall result of glycolysis, would you expect ATP to inhibit or stimulate activity of this enzyme? Make sure you consider the role of ATP as an allosteric regulator, not as a substrate of the enzyme.

The word *role* explicitly appears in the above example sentences. In sentence S2, the word *role* refers to the functions of the cytoskeleton. There are, however, many example sentences in the textbook that do not explicitly use the word *role* but still implicitly suggest a role concept. Here are a few examples (the words that suggest roles are emphasized).

S4 Some evidence suggests that during the Cambrian period, *predators* acquired novel adaptations, such as forms of locomotion that helped them catch *prey*, while *prey* species acquired new defenses, such as protective shells.
S5 NADH and a similar *electron carrier*, a coenzyme called FADH2, transfer electrons derived from glucose to electron transport chains, which are built into the inner mitochondrial membrane.
S6 A *nutrient* is a chemical that is used in an organism's metabolism and is obtained from the environment. Different organisms have different nutrient requirements.

The goal of representing the knowledge contained in the textbook was to answer questions about that knowledge. Though some questions are purely about roles, roles can also appear as answers to questions that are not necessarily about roles. Here are some example questions that we were interested in answering.

Q1 What is the difference between an oxidizing agent and a reducing agent?
Q2 In pyruvate oxidation, what acts as an oxidizing agent?
Q3 What is the relationship between an electron carrier and photosynthesis?
Q4 What makes a human host for one organism but not for another?

To answer the above questions correctly and in a uniform manner, we needed to adopt a clear set of modeling guidelines for representing roles followed by appropriate computational and presentation techniques for answers. Overall contribution of our work is to show how the role concepts developed in formal ontologies can be applied to a large scale knowledge engineering project. At a more specific level, we show how we have adapted the results from formal ontologies by (1) developing an easy to understand definition of roles that can be applied by biologists (2) clarifying trade-offs between representing the roles as objects vs relationships (3) identifying that when roles are represented as a class, they need to be disjoint from entities (4) clarifying the relationship between roles and functions.

We will begin this paper by a review of prior definitions, by explaining the difficulties that arise in directly applying them, by presenting our solutions, and finally, by giving example answers to some of the questions that use roles.

Defining Roles

We will consider how roles are defined in upper ontologies, linguistics. and description logics. We will then discuss how biologists think of roles and the definition that we adapted for our project.

Roles in Foundational Ontologies

Our starting point for the definition of a role is an upper ontology called the Component Library [5] or CLIB. CLIB distinguishes between an Entity (i.e., things that are) and an Event (i.e., things that happen). The distinction between an Entity and an Event is comparable to a similar distinction in many other common upper ontologies such as DOLCE [15] that distinguishes between a Perdurant and an Endurant). A Role can be thought of as a temporally unstable entity. It is what an entity is in the context of some event. For example, PERSON is an entity while EMPLOYEE is a role. A PERSON remains a PERSON independent of the events in which she participates. Conversely, someone is an EMPLOYEE only by virtue of participation in an EMPLOY event.

There exists a rich body of literature on ontological analysis of roles [25, 21, 22]. An excellent survey of the usage and definition of roles from many different perspectives from computer science, conceptual modeling, linguistics and philosophy can be found elsewhere [21]. We will only focus on relating our usage of roles with a few key alternatives available in the literature.

Based on an extensive survey of the prior literature, the following commonalities were observed when roles are reified [21]: (1) each role is an individual (2) each role has exactly one player (3) there is at least one context for a role (4) no role is a context of another role. For example, Student could be a role whose player is some Human, and whose context is determined by some relation to a University. Roles have been classified into three kinds: relational roles, processual roles and social roles. The notion of role in CLIB, and the one considered in our work here, is the closest to the notion of processual roles, i.e., roles where the context is an Event.

In Basic Formal Ontology or BFO, a role is defined as follows [1]. A role is associated with an entity in an event when that event is not typical of that entity's physical structure. The role is played by an entity because this entity is in some special natural, social, or institutional set of circumstances. For example, a chemical compound can play the role of a catalyst. A person can play the role of a student. Roles are optional, and they often involve social ascription. This is why a person can play the role of being a lawyer or a surrogate to a patient, but it is not necessary for persons that they be lawyers or surrogates. BFO also distinguishes the notion of *disposition* as distinct from role. A disposition invariably leads to a certain result given certain circumstances. For example, the disposition of a car windshield is to break if struck with a sledgehammer moving at 100 feet per second.

Even though the definition given by BFO is most widely used in bio-medical ontologies, it is unsatisfactory for several reasons. First, it is open to debate whether being a student is typical of a person's physical structure. Of course, to be a student, the entity has to have the structural properties of a person that are necessary to be a student. Second, in the context of a biology textbook, it is unclear how one should determine social or institutional circumstances. Finally, even though many roles are temporary, some of the roles such as a daughter, are taken for the life of the entity, and the BFO definition does not address such cases.

Roles in Natural Language Processing

Case roles are generalizations of the semantic relationships between an activity (a process or event) and the participants in, and circumstances around the activity [4]. The activity is usually represented by a verb and the participants and circumstances by the verb's syntactic arguments. Natural language processing systems that use cases reduce much of the meaning of a sentence to instances of case relationships. Some examples of case roles are Agent, Object, Instrument, Experiencer, etc. For example, in the sentence: Wilma printed a paper, the Agent role is Wilma, and the Object role is paper. The relation vocabulary in CLIB is inspired by case roles.

Roles in Description Logics

In the early literature on description logic systems, the word *role* was used to describe potential relationships between instances of a Concept and those of other closely associated Concepts [7]. Although the usage of roles in this sense is equivalent to binary relations, modern description logics have continued to use this term [2]. Furthermore, the deescription logics and modern knowledge representation languages do not distinguish amongst different kinds of relations such as relational roles, qualities and parts [16]. In the Web Ontology Language, which is based in description logics, the word *role* is, however, replaced by the word *property*.

Biologist Definition of Roles

Because the textbook frequently uses the word *role*, we interviewed three biology teachers to gain insight into how they understand and use this word. The following consensus definition emerged: A role is a temporary duty or job of an entity, and it does not make sense to classify entities permanently under that class. For example, electron donor, electron carrier, nutrient, etc. are considered roles by biologists.

Discussion

Clearly, the intuition of the biologists is the closest to the notion of roles adopted in the upper ontologies. Case *role*s as used in linguistics for natural language processing applications have an interesting parallel to the notion of *role* in an upper ontology: case roles also capture the relationships between participants and verbs that is temporary only in the context of a sentence of interest. Not surprisingly, a case role such as Agent can also be reified as a role concept Agent-Role in an upper ontology. The usage of *role* in description logics to refer to relations is for a completely different purpose, and is simply ambiguous when it is used in the same context as an upper ontology. For our work, we adopted the definition of roles as provided by the upper ontologies as described in Section 1. More specifically, we consider here how to distinguish between entities and roles.

We say that a concept represents an entity if that entity has a structural description, and it can be measured and drawn [24]. We say that a concept is a role if it is defined in terms of an entity's participation in an event, it is temporary, and it can be acquired and/or released during an entity's lifetime. We consider a few exceptions to this general definition. Some roles exist which once acquired persist for a life time. By default, we treat such roles as entities (for example, an Autotroph and Heterotroph are treated as entities). They are roles because the entities play them only in the context of certain processes, but an entity is heterotroph or autotroph throughout its lifetime for those processes. A second exception is when one role of an entity persists for lifetime, but another role of the same entity satisfies the temporary nature of a role, we do not treat such life time role as an entity. For example, Daughter is a role that persists for the lifetime of an entity, but the same entity can play a role such as Parent which is an acquired role. In this case, we treat both Parent and Daughter as a role.

Using Roles in Representation

In this section, we consider the issues and challenges that arose during the process of using roles to represent the biology textbook. This involves making decisions such as should we reify roles, where should we place them in a class taxonomy, how should we use them in writing rules, and what relations should we associate with them. Some of these challenges may not be important if one is representing isolated examples. But, in a project like ours, where substantial amount of knowledge on a diverse set of topics is to be represented, a uniform and consistent modeling approach is essential for two reasons: (a) knowledge encoders, who themselves are not ontologists, must have clear and unambiguous guidelines on how to model different kinds of knowledge, and (b) the representation must integrate with reasoning and natural language processing, and it must be as clearly specified as possible for the engineering of the system.

Should we reify roles?

We need to decide whether a role should be represented using a binary relation or as an object in the ontology. For example, consider a process in which a virus infects a cell, and we need to specify the Agent role of the virus in this process. We can either use a relation such as agent or create an object such as Agent-Role. When we reify a role as an object, we associate two relations with it: plays and $\mathrm{in\text{-}event}$. By $\mathrm{plays}(x, y)$ we mean that an entity x plays role y. By $\mathrm{in\text{-}event}(x, y)$ we mean that the role x is played in the event y. The relationship plays is commonly used in reified representation of roles, and the relationship $\mathrm{in\text{-}event}$ is similar to the context relationship [21]. Alternative representations for these two choices are shown below.

$$\forall x : \mathsf{Virus\text{-}Infection}(x) \Rightarrow \mathrm{agent}(x, f_1(x)) \wedge \mathsf{Virus}(f_1(x)) \tag{1}$$

$$\begin{aligned} \forall x : \mathsf{Virus\text{-}Infection}(x) \Rightarrow\ & \mathrm{in\text{-}event}(x, f_2(x)) \wedge \mathsf{Agent\text{-}Role}(f_2(x)) \\ & \mathrm{plays}(f_3(x), f_2(x)) \wedge \mathsf{Virus}(f_3(x)) \end{aligned} \tag{2}$$

The above axioms could have also been written using existential quantifiers instead of using skolem functions. For the purpose of the material presented in this paper these two alternative forms do not make any difference. In the broader context of our project, the use of skolem functions plays an important role because there are many situations in which we need to author axioms that refer to objects used in other axioms which is only possible if one uses skolem functions [10].

Axioms 1 and 2 are equivalent except for the following differences: Axiom 1 is more compact as it introduces only one new relation compared to axiom 2 which introduces two new relations and one extra object. Such compactness is particularly advantageous when interfacing to a natural language processing system which can generate the representation of axiom 1 in a straightforward manner by parsing the text. A compact representation is also advantageous if a complex set of axioms are to be visualized in a user interface. If we wish to state additional properties about the relationship between a participant such as Virus and the process such as Virus-Infection, that can be done in the reified representation by introducing additional relationships. If we wish to add additional detail when representing a role using a relation, we will need to introduce n-ary relationships. AURA supports n-ary relationships, but as a matter of modeling principle, most of the relationships in our system are binary relationships. In our knowledge engineering work so far, we have not found any examples of situations where such additional detail is needed to be specified to justify the reified representation.

The knowledge authoring system AURA [20], which uses this representation, was designed on the principle that the knowledge engineers should control the relation vocabulary and the domain experts should be allowed to create new concepts that use that vocabulary [12]. From that perspective, the representation of axiom 2 is especially advantageous because using this representation, the biologists will be able to create new domain-specific roles. This ability will not be possible if we were using the representation in axiom 1, in which case the roles must be pre-built into the system or require intervention from a knowledge engineer.

In a project such as ours with multiple goals, the choice between the two alternatives was not clear-cut. Therefore, we used the following pragmatic solution. For the relations motivated by case roles, which are also referred as participant relations, we use both representations (i.e., we represent them using both a relation and reify them as object as well). In the knowledge authoring interface, we only expose the participant relations, and hide the reified object representation from the end-user. The KB contains axioms stating the equivalence of the two representations. For example, for the Agent role, we have the following axiom in the KB:

$$\begin{aligned}\forall x : \text{agent}(x, f_1(x)) \wedge \mathsf{Virus}(f_1(x)) \Leftrightarrow\ & \text{in-event}(x, f_2(x)) \wedge \mathsf{Agent\text{-}Role}(f_2(x)) \\ & \text{plays}(f_1(x), f_2(x)) \wedge \mathsf{Virus}(f_1(x))\end{aligned} \tag{3}$$

For the other cases, we use the following approach. We prefer a reified role representation when we anticipate that biologists will need to create sub-roles of a role and compare one role to the other. For the cases requiring no further specialization, a biologist makes the decision. They can either reify the role by creating a new class in the KB, or request that the knowledge engineers to add a new relation. This solution is, of course, directly motivated by the knowledge authoring choice that we have made for AURA. For projects, where end-users can freely create new relationships, we expect that the roles will never need to be further specialized and we do not need to state any additional properties about the role, roles can be represented using a relationships.

Where to place the roles in the taxonomy?

In prior work, several researchers have concluded that roles should not be in a taxonomic relationship with entities (i.e., they should neither be a superclass nor a subclass of a class representing entities) [25, 14]. We briefly review the reasons cited in the literature. Making a role a subclass of an entity that plays it causes problems when a role could be played by multiple entities. For example, both companies and people can play the customer role, and if Customer is made a subclass of both Person and Company, its instances can only be the intersection of these two classes which does not make sense. A Customer cannot be made a superclass of Person as it has more properties than a person does. Further, such an arrangement leaves no way to separate the entities from the roles that they play, because now every person is also a customer. The above conclusion is also consistent with the Ontoclean methodology [17], because roles are anti-rigid, and they cannot be in a taxonomic relationship with classes that are rigid.

There has also been a proposal to use role mixin classes to solve this problem [18]. The use of role mixin classes is unsatisfactory for the following reasons: (1) It forces the duplication of the taxonomy. For example, for every Company, we need a class such as CompanyCustomer; (2) The temporal extent of classes such as CompanyCustomer cannot be captured in the class taxonomy.

Assuming that the roles are represented in the class taxonomy in a way that they are not in direct taxonomic relationship to entities, the prior work gives us no clear guideline on whether roles should be disjoint from entities. For example, energy foods are processed into energy through the process of Cellular-Respiration, and that fact needs to be represented. We have two ways of representing this information as shown below. In the axiom 4, we state that for every instance of Cellular-Respiration, there exists an Organic-Molecule that is its $\mathrm{raw\text{-}material}$, and further, that Organic-Molecule plays the role of Energy-Food in Cellular-Respiration. Next, in axiom 5, we state that for every instance of Cellular-Respiration, there exists an Organic-Molecule that is its $\mathrm{raw\text{-}material}$, and further, that Organic-Molecule is an instance of the role class Energy-Food.

$$\forall x : \mathsf{Cellular\text{-}Respiration}(x) \Rightarrow \begin{array}{l} \mathrm{raw\text{-}material}(x, f_4(x)) \wedge \mathsf{Organic\text{-}Molecule}(f_4(x)) \wedge \\ \mathrm{plays}(f_4(x), f_5(x)) \wedge \mathsf{Energy\text{-}Food}(f_5(x)) \wedge \\ \mathrm{in\text{-}event}(f_5(x), x) \end{array} \tag{4}$$

$$\forall x : \mathsf{Cellular\text{-}Respiration}(x) \Rightarrow \begin{array}{l} \mathrm{raw\text{-}material}(x, f_6(x)) \wedge \mathsf{Organic\text{-}Molecule}(f_6(x)) \wedge \\ \mathsf{Energy\text{-}Food}(f_6(x)) \end{array} \tag{5}$$

In the axioms above, Energy-Food is a role class. We use Organic-Molecule as the $\mathrm{raw\text{-}material}$, which is a superclass of Fat, Carbohydrate, and Protein. We expect that in more specific versions of Cellular-Respiration, Organic-Molecule will be specialized to Fat, Carbohydrate, or Protein. In axiom 4, we use the reified representation of roles that we have discussed previously. In axiom 5, entity $f_6(x)$ which is an instance of Organic-Molecule, is also an instance of the role class Energy-Food. Even though such relationship violates the principle that the roles and entities cannot be in a direct relationship, it is allowed unless we also require the roles and entities to be disjoint. Suppose, we are now asked the question: "What are the energy foods?" Axiom 5 facilitates answering the question about energy foods: we can simply look for terms in the KB that are subsumed by Energy-Food, and return them in the answer. If we use axiom 4, we must deal with an extra indirection. The approach used in Axiom 5 breaks down in some situations as we illustrate in the next example.

Suppose, we wish to represent the following predator and prey relationship: a bear eats a seal that eats a fish. If roles and entities are not disjoint, the following will be allowed:

$$\begin{aligned}\forall x : \textsf{Ecological-Region}(x) \Rightarrow\ & \textsf{Consume}(f_7(x)) \wedge \mathrm{site}(f_7(x), x) \wedge \\ & \mathrm{object}(f_7(x), f_8(x)) \wedge \textsf{Fish}(f_8(x)) \wedge \\ & \textsf{Prey}(f_8(x)) \wedge \text{in-event}(f_8(x), f_7(x)) \\ & \mathrm{agent}(f_7(x), f_9(x)) \wedge \textsf{Seal}(f_9(x)) \wedge \\ & \textsf{Predator}(f_9(x)) \wedge \text{in-event}(f_9(x), f_7(x)) \\ & \textsf{Prey}(f_9(x)) \wedge \text{in-event}(f_9(x), f_{10}(x)) \wedge \\ & \textsf{Consume}(f_{10}(x)) \wedge \mathrm{site}(f_{10}(x), x) \wedge \\ & \mathrm{object}(f_{10}(x), f_9(x)) \wedge \mathrm{agent}(f_{10}(x), f_{11}(x)) \wedge \\ & \textsf{Predator}(f_{11}(x)) \wedge \text{in-event}(f_{11}(x), f_{10}(x)) \\ & \textsf{Bear}(f_{11}(x))\end{aligned} \tag{6}$$

Axiom 6 is problematic, because $f_9(x)$ which is a Seal, is both a Predator and a Prey, but we cannot tell in which particular Consume event each of those roles is played. One way to address this problem is to introduce the following axiom:

$$\forall x : \textsf{Predator}(x) \Leftrightarrow \exists e : \text{predator-in}(x, e) \tag{7}$$

This solution is not satisfactory because it introduces specialized versions of the in-event relationship for every role in the ontology.

Although allowing an individual to be an instance of both role and an entity may lead to a simpler representation and question answering for some examples, it is not a general solution. Allowing it under some situations and not in other situations is also problematic, as this approach leads to confusion and a lack of consistency. A desired representation for this example is shown below:

$$\begin{aligned}\forall x : \textsf{Ecological-Region}(x) \Rightarrow\ & \textsf{Consume}(g_1(x)) \wedge \mathrm{site}(g_1(x), x) \wedge \\ & \textsf{Prey}(g_2(x)) \wedge \text{in-event}(g_2(x), g_1(x)) \wedge \\ & \textsf{Fish}(g_3(x)) \wedge plays(g_3(x), g_2(x)) \wedge \\ & object(g_1(x), g_3(x)) \wedge agent(g_1(x), g_5(x)) \wedge \\ & \textsf{Predator}(g_4(x)) \wedge \text{in-event}(g_4(x), g_1(x)) \wedge \\ & \textsf{Seal}(g_5(x)) \wedge plays(g_5(x), g_4(x)) \wedge \\ & \textsf{Consume}(g_6(x)) \wedge \mathrm{site}(g_6(x), x) \wedge \\ & \textsf{Prey}(g_7(x)) \wedge \text{in-event}(g_7(x), g_6(x)) \wedge \\ & plays(g_5(x), g_7(x)) \wedge \\ & object(g_7(x), g_5(x)) \wedge agent(g_7(x), g_8(x)) \wedge \\ & \textsf{Bear}(g_8(x)) \wedge plays(g_8(x), g_9(x)) \wedge \\ & \textsf{Predator}(g_9(x)) \wedge \text{in-event}(g_9(x), g_7(x))\end{aligned} \tag{8}$$

In axiom 8, we introduce separate roles for each of the two Consume events, and it is made clear which role is being played by each entity in a particular event. This axiom is more verbose than axiom 6, but this approach is necessary to ensure its correctness. The example also illustrates that we cannot, in general, allow entity terms to be instances of roles. To achieve this discipline, we must also declare Role and Entity as disjoint classes in our ontology. Doing so will disallow axioms 5 and 6.

Relationship between roles and functions

In sentence S2 considered earlier, we noted that support and motility are noted as roles as well as functions of cytoskeleton. This raises a natural question regarding whether roles and functions refer to the same notion, or whether they are separate? This question has been extensively discussed previously [1], where roles and functions are distinguished by asking the following questions: Is the realizable entity such that its typical manifestations are based upon its physical structure? If the answer to this question is yes, it should be considered a function. Roles are optional, and they often involve social ascription. For example, a person can play the role of a lawyer or a surrogate to a patient, but a person need not necessarily be a lawyer or a surrogate.

In our analysis of several examples where roles and functions were used interchangeably in the biology textbook, we found them to be overlapping notions. Some roles are also indicative of function of an entity. For example, an electron carrier is a role. NADPH plays the role of an electron carrier, and its function is to be an electron carrier. But, not all roles indicate function. For example, solute, waste and debris are roles, but they are rarely ever the functions of the entity playing them.

Answering Questions about Roles

Answering question involves three distinct phases: (1) interpreting questions stated in English, (2) reasoning to produce a logic level answer, and (3) presenting the answer to the user. Questions Q1—Q3 that were shown earlier require the same computation as is needed for entities. For example, a version of question Q1 for entities is: what is the difference between a ribosome and a centrosome? The main difference in answering this question for roles vs entities arises in presenting the answer. Some of the answers contain natural language sentences, and any sentence that contains a role requires different handling than a sentence that contains only entities. In our current system, we produce the following four forms of sentences involving roles: (a) NADH is an electron carrier during photosynthesis (b) NADH carries electrons during photosynthesis (c) Reduction — an electron transfers to P700+. Here, P700+ is an electron recipient (d) Organic molecules, such as carbohydrates, proteins, and fats, are fuels in cellular respiration. In sentence (a), emphasis is clearly on indicating the role of an entity. In sentence (b), the role is converted to a verb phrase and the role itself becomes implicit. Sentence (c) also explicitly calls out the entity that plays the role. Sentence (d) contains multiple roles that are aggregated using the phrase "such as". Next, we consider a few example questions and answers.

In response to a question such as Q1, the system produces a table comparing the two roles that contain their definitions, their super classes, and examples of their usage in the KB. In response to Q2, the system responds by saying that NAD+ is an oxidizing agent.

In response to Q3, the system produces a graphical representation showing the relationship between photosynthesis and an electron carrier, which can be read as follows: photosynthesis has a step called Calvin cycle that consumes NADPH, which plays the role of an electron carrier. Our current system does not support Q4 as it requires further representation and reasoning design.

Another important type of question from an application standpoint is "What plays the role of R in the context of C?". There are two parts to this question. First, we need to determine what x plays the role r and in which event y which can be done by the following query:

$$\exists z : \text{plays}(x, r) \wedge R(r) \wedge \text{in-event}(r, y)$$

Next, if a context C is given, we use that context to limit the aforementioned results x (the players of the role R) and y (the corresponding event). We will distinguish between two cases depending on the nature of the context:

- the context C is an Entity. In this case, the context C is assumed to be the entity that has the function y (the event the role is played in). Therefore, we look for x and y such that

$$\exists r, c : \text{plays}(x, r) \wedge R(r) \wedge \text{in-event}(r, y) \wedge \text{has-function}(c, y) \wedge \mathsf{C}(c)$$

  For example, for a question "What plays the role of Electron-Recipient in the context of a Photosystem", we would have Chlorophyll-A as the answer as it plays the role of an electron recipient in a reduction, and reduction (of electrons) is a function of the photosystem. In many situations it is also interesting to report *co-occurring* roles, i.e., what other players play what other roles in that same event (Reduction). In this example, we have that Chlorophyll-B plays the role of reducing agent in the same reduction event. Including such co-occurring roles in an answer gives the user a feel for the role landscape in a concept of interest.
- the context C is an Event. In this case, the context C is assumed to be an event that is related to the event the role is played in; usually the event in which the role is played is a sub event of the context event. The corresponding query would be:

$$\exists r, c : \text{plays}(x, r) \wedge \text{in-event}(r, y) \wedge \text{sub-event}(y, c)$$

  For example, "What plays the role of Electron Donor?". Note that the context does not have to be given by the user; we can determine that a suitable context is Calvin Cycle: NADPH plays the role of Electron Donor in the event of an Oxidation. This indeed happens in the context of a Calvin cycle, as a Calvin Cycle has a subevent *Reduction in Calvin Cycle*, which has a subevent *Redox reaction*, which finally has a subevent *Oxidation*. While this chain is not necessary if one adopts a strict logical reading of the original question, but users demand such additional information. Therefore, whenever we answer role questions, the system proactively computes the context and reports it as part of the answer. Finally, for events we also show the co-occurring roles (in the example, PGAP plays the role of Oxidizing agent in the same Oxidation event).

Our work is a unique attempt in using the representation of roles for answering questions that are of direct interest to a practical application. In the formal analysis of roles, researchers have used a question based on *counting problem* [19] which may be interesting from a philosophical perspective, but have little practical relevance.

Summary and Conclusions

In this paper, we consider the challenges and solutions to the problem of representing roles in a biology textbook. Although we presented several solutions, many problems remain open. In our current representation, the role concepts can have only two relations: plays and in-event. Whether additional relations might be required, and what those relations might be are unclear. In the literature some discussion exists on transfer axioms (i.e., axioms that define properties which should be transferred to an entity only when it is playing that role). We have not considered such axioms in our work so far. Our work on answering questions has been preliminary as we answer only rudimentary forms of questions. Answering more complex questions that involve roles and leveraging suitable formalisms for this purpose remains open for future work [6].

The representation work was undertaken in the context of an intelligent textbook that is aimed at helping students learn better. We showed how we were able to take the results from ontology research and apply them to this novel and complex domain. Our contributions lie primarily in the application of ideas on role representation, and addition of much needed practical perspective to the formal analysis of roles. This application required adapting the definition of roles so that biologists could easily use it, making pragmatic knowledge engineering decisions about whether to reify the roles, showing the need for keeping them disjoint from entities, and clarifying their relationship to functions.

Acknowledgments

Vulcan Inc. and SRI International funded this work. We thank Debbie Frazier for her contributions to this work.

References

1. Robert Arp and Barry Smith. Function, Role, and Disposition in Basic Formal Ontology. *Nature Preceedings*, pages 1–4, 2008.
2. Franz Baader, Ian Horrocks, and Ulrike Sattler. Description Logics. *Foundations of Artificial Intelligence*, 3:135–179, 2008.
3. Eva Banik, Eric Kow, and Vinay K. Chaudhri. User-Controlled, Robust Natural Language Generation from an Evolving Knowledge Base. In *ENLG 2013 : 14th European Workshop on Natural Language Generation*, 2013.
4. K. Barker, T. Copeck, S. Delisle, and S. Szpakowicz. Systematic Construction of a Versatile Case System. *Journal of Natural Language Engineering*, 3(4):279–315, 1997.
5. K. Barker, B. Porter, and P. Clark. A Library of Generic Concepts for Composing Knowledge Bases. In *First International Conference on Knowledge Capture*, 2001.

6. Patrick Barlatier and Richard Dapoigny. A type-theoretical approach for ontologies: The case of roles. *Applied Ontology*, 7(3):311–356, 2012.
7. Ronald J Brachman and James G Schmolze. An Overview of the KL-ONE Knowledge Representation System. *Cognitive science*, 9(2):171–216, 1985.
8. Vinay K Chaudhri, Britte Cheng, Adam Overholtzer, Jeremy Roschelle, Aaron Spaulding, Peter Clark, Mark Greaves, and Dave Gunning. Inquire Biology: A Textbook that Answers Questions. *AI Magazine*, 34(3), September 2013.
9. Vinay K. Chaudhri, Daniel Elenius, Sue Hinojoza, and Michael Wessel. Kb bio 101: Content and challenges. In *In the Proceedings of International Conference on Formal Ontologies in Information Systems*, 2014.
10. Vinay K. Chaudhri, Stijn Heymans, and Son Cao Tran Michael Wessel. Object-Oriented Knowledge Bases in Logic Programming. In *Technical Communication of International Conference in Logic Programming*, 2013.
11. Vinay K. Chaudhri, Stijn Heymans, Michael Wessel, and Son Cao Tran. Query Answering in Object Oriented Knowledge Bases in Logic Programming. In *Workshop on ASP and Other Computing Paradigms*, 2013.
12. Vinay K Chaudhri, Bonnie E John, Sunil Mishra, John Pacheco, Bruce Porter, and Aaron Spaulding. Enabling Experts to Build Knowledge Bases from Science Textbooks. In *Proceedings of the 4th International Conference on Knowledge Capture*, pages 159–166. ACM, 2007.
13. Vinay K. Chaudhri, Michael A. Wessel, and Stijn Heymans. KB_Bio_101: A challenge for OWL Reasoners. In *The OWL Reasoner Evaluation Workshop*, 2013.
14. James Fan, Ken Barker, Bruce Porter, and Peter Clark. Representing roles and purpose. In *Proceedings of the first international conference on Knowledge capture*, pages 38–43. ACM, 2001.
15. Aldo Gangemi, Nicola Guarino, Claudio Masolo, and Alessandro Oltramari. Sweetening WordNet with DOLCE. *AI Magazine*, 24(3):13–24, 2003.
16. Nicola Guarino. The ontological level: Revisiting 30 years of knowledge representation. In *Conceptual Modeling: Foundations and Applications*, pages 52–67. Springer, 2009.
17. Nicola Guarino and Christopher Welty. Evaluating Ontological Decisions with OntoClean. *Communications of the ACM*, 45(2):61–65, 2002.
18. Giancarlo Guizzardi. *Ontological foundations for structural conceptual models*. CTIT, Centre for Telematics and Information Technology, 2005.
19. Giancarlo Guizzardi. Agent roles, qua individuals and the counting problem. In *Software Engineering for Multi-Agent Systems IV*, pages 143–160. Springer, 2006.
20. David Gunning, Vinay K. Chaudhri, Peter Clark, Ken Barker, Shaw-Yi Chaw, Mark Greaves, Benjamin Grosof, Alice Leung, David McDonald, Sunil Mishra, John Pacheco, Bruce Porter, Aaron Spaulding, Dan Tecuci, , and Jing Tien. Project Halo Update: Progress Toward Digital Aristotle. *AI Magazine*, Fall 2010.
21. Frank Loebe. An Analysis of Roles: Towards Ontology-Based Modelling, 2003.
22. Claudio Masolo, Laure Vieu, Emanuele Bottazzi, Carola Catenacci, Roberta Ferrario, Aldo Gangemi, and Nicola Guarino. Social roles and their descriptions. In *KR*, pages 267–277, 2004.
23. Jane B. Reece, Lisa A. Urry, Michael L. Cain, Steven A. Wasserman, Peter V. Minorsky, and Robert B. Jackson. *Campbell Biology*. Benjamin Cummings imprint of Pearson, Boston, 2011.
24. A. D. Spear. Ontology for the Twenty First Century: An Introduction with Recommendations. `http://www.ifomis.org/bfo/documents/manual.pdf`, 2006.
25. Friedrich Steimann. On the Representation of Roles in Object-Oriented and Conceptual Modelling. *Data & Knowledge Engineering*, 35(1):83–106, 2000.