Using the cDnS ontology as upper-level for a Scholarly Debate Ontology

Neil Benn

Abstract

There is an active area of research concerned with designing tools that support quantitative and qualitative analysis of academic knowledge domains. Knowledge Domain Analysis (KDA) research investigates computational support for users who desire to understand and/or participate in the scholarly inquiry of a given academic knowledge domain. KDA technology supports this task by allowing users to identify important features of the knowledge domain such as the predominant research topics, the experts in the domain, and the most influential researchers. However, a key limitation of work to date is its inability to provide machine-readable models of the debate in academic knowledge domains. This paper argues that KDA technology should support users in understanding the features of scholarly debate as a prerequisite for engaging with their chosen domain. To this end, the paper proposes a Scholarly Debate Ontology which specifies the formal vocabulary for representing debate in academic knowledge domains. The ontology is designed with reference to an upper-level ontology that specifies the generic elements of any domain, such as academic domains, where knowledge is collectively constructed and modified by social agents.

Introduction

Knowledge Domain Analysis (KDA) research (e.g. [1][2][3][4][5]) investigates computational support for users who desire to understand and/or participate in the scholarly inquiry of a given academic knowledge domain. KDA technology supports this task by allowing users to identify important features of the knowledge domain such as the predominant research topics, the experts in the domain, and the most influential researchers. However, one limitation of the existing research is its inability to support users in understanding the features of scholarly debate as a prerequisite for engaging with their chosen domain. KDA tools need to be able to reason over formal models of the debate in academic knowledge domains.

To this end, this paper proposes a Scholarly Debate Ontology, which specifies the formal vocabulary for constructing representations of debate in academic knowledge domains. In accordance with best practices for ontology design, the paper presents a design process that involves using an upper-level ontology as the basis for selecting the essential elements of the world to be represented, thus ensuring that the design process adheres to the principle of minimal ontological commitment [6].
The paper begins by characterising knowledge domains as settings for the collective construction of knowledge, thus motivating the reuse of an upper-level constructivist ontology. The paper then defines the Scholarly Debate Ontology, which specifies the essential elements of debate in academic domains. The upper-level ontology acts as a design aid for selecting the essential elements of scholarly debate to be specified in the SDO.

1 The cDnS ontology: codifying social constructivism

The Constructivist Descriptions and Situations (cDnS) ontology “provides the expressivity to talk about the contexts (social, informational, circumstantial, and epistemic), in which collectives make and produce sense” [7]. In other words, cDnS can be characterised as an ontology of collective sensemaking. “Collective sensemaking” or “collective knowledge construction” is a useful way of characterising the key activity of knowledge domains, thus the cDnS ontology provides a suitable vocabulary for describing the portion of reality that is of current interest—in this case, the setting of academic knowledge domains.

Most significantly, the key knowledge domain activities of representing and communicating knowledge constitute semiotic processes. Semiotics is the study of signs and their use in the representation and communication of meaning. In Charles Sanders Peirce’s prominent theory of semiotics [8], the basic structure of signs in semiotic processes consists of three components: the sign-vehicle, which is the entity perceived by the senses; the object referred to by the sign-vehicle, which may include imaginary objects and ideas; and the interpretant, which is the mental representation that links the sign-vehicle to the object in the mind of some conceiving agent. As will be discussed later in the section, these semiotic components correspond to key elements of the cDnS ontology. Indeed, a core configuration of elements within the cDnS ontology can be used to describe any generic semiotic process where an agent conceives some description (or representation) about entities in the world and communicates this description via some object for conveying information.

Although the design approach is to reuse this constructivist ontology, this paper attempts to remain neutral with respect to the perennial philosophical debate about whether we can only construct reality via our subjective and socially-mediated representations of it (a constructivist viewpoint) or whether we can derive true representations of a single objective reality that exists independently of our conceptualisation of it (a realist viewpoint). That philosophical debate is beyond the scope of this paper, which, for the purposes of selecting a suitable upper-level reference ontology, is concerned with characterising knowledge domains as settings for conducting intellectual inquiry through its production of texts as the primary means of representing and communicating knowledge (cf. [9]). This is irrespective of whether or not the “representations of knowledge” that are produced and communicated via published texts correspond to true facts in reality. As the authors of [10] note, a view of knowledge

---

1In the cDnS ontology’s original form as the DOLCE+DnS ontology this configuration of ontological categories was sometimes referred to as the Semiotic Ontology Design Pattern
as socially constructed and socially negotiated does not logically imply an anti-realist position.

The cDnS ontology is depicted in Figure 1. The core\(^2\) cDnS ontology consists of the following main classes: cdns:Entity, cdns:InformationObject, cdns:Description, cdns:Situation, cdns:Concept, cdns:SocialAgent, and cdns:Collection. These, along with the main relations in the ontology, will be described in more detail in the remainder of this section\(^3\). As each class is described, an analysis is made of how, as upper-level classes, they can be interpreted in the context of academic knowledge domains, as a way of demonstrating the suitability of the cDnS ontology for the design task at hand.

1.1 **Entity**

Entity is “the class of everything that is assumed to exist in some domain of interest for any possible world” [11]. In the cDnS ontology, the cdns:Entity class is specified as the uppermost class in the hierarchy from which all other cDnS classes are sub-classed. There are two main categories of Entity: SchematicEntity (typically social entities like organisations and information), and NonSchematicEntity (for example, time intervals and spatial coordinates). However, the main development on the cDnS ontology has focussed on “axiomatising” the former type of Entity—i.e. SchematicEntity.

\(^2\)There are two extended versions of the cDnS ontology that are not treated here.

\(^3\)The description of the cDnS ontology that follows is based on two main publications, [11] and [7]. There are some peripheral modifications made in the cDnS ontology between [11] and [7], which demonstrates that the ontology is still a work in progress. Nonetheless, the core elements of the ontology have become stable enough to be suitable for the purposes of the present work.
1.2 InformationObject

“Information objects” are the vehicles for communicating semantic content or meaning between agents; they are the expression of semantic content, or to use Peircean terminology, they are the sign-vehicle in a given semiotic process, where cdns:Entity plays the role of the “object” in the semiotic process—i.e. an information object in the cDnS sense can be about any other entity. Any unit of information can be treated as an instance of cdns:InformationObject, and this is independently of how the information (as something which is abstract) is realised in a physical medium. However, according to [11], information objects must have a physical realisation so that their informational content is perceivable by some agent. Based on this characterisation a single information object can have multiple physical realisations or modalities—e.g. a newspaper article can have a paper and an electronic realisation.

In the context of academic domains, the most typical examples of information objects are publications, which are the main vehicles for representing and communicating knowledge. A single publication, taken as a whole, can be considered to be an information object. Furthermore, a single publication is composed of clauses and sentences (verbal expressions of knowledge), as well as tables, graphs, and figures (hybrid, i.e. both verbal and non-verbal, expressions of knowledge). Each of these components of a publication can be considered to be an information object in its own right. This corresponds with the view of [12] that academic publications, particularly scientific publications are semiotic hybrids.

1.3 Description

A Description is the abstract, communicable semantic content or meaning that is expressed by an information object. In Peircean terminology, a Description is the “interpretant” that is formed in the mind of some conceiving agent. According to [13], different information objects, possibly even in different languages, can be associated with the same description or semantic content.

In the context of academic domains, just as there are different types of information objects, there are different types of descriptions. For example, a single scholarly publication can be regarded as an information object that expresses a thesis, in much the same manner as a novel can be regarded as an information object that expresses a plot. The thesis of a scholarly publication is therefore an example of a description in the cDnS sense. Furthermore, each clause or sentence that makes up a publication can be characterised as an information object that expresses either some propositional content (as is the case with declarative sentences that may represent some theory conceived by an agent) or some non-propositional content (as is the case with interrogative sentences—i.e. questions). Therefore, the propositional or non-propositional content of individual clauses and sentences are also examples of Descriptions.

---

4 There is an issue here with the ontological status of different modalities or modes of expression. For example, is an orally-delivered speech the same expression as the written speech but just a different modality or are they two entirely different expressions? The intuitive answer seems to be to treat the orally-delivered speech and the written speech as the same abstract conceptual work, in which case the cDnS characterisation is appropriate—i.e. a single information object can have multiple realisations.
1.4 Situation

A situation is said to provide a setting for other entities, including other situations. A situation, in the cDnS sense, represents a state of affairs that is observable by some agent. A situation is said to satisfy a description. Inversely, a description is said to represent an agent’s conceptualisation of a particular situation. The constructivist nature of the cDnS ontology suggests that situations do not exist independently of descriptions—i.e., a state of affairs requires an agent to conceive of it. However, the reverse is not necessarily true—that is descriptions are not dependent on situations, since there exist descriptions that do not describe a particular situation.

1.5 Concept

Concepts are defined by and used by Descriptions. Concepts are used to classify and name other entities. So, for example, in the American constitution, there is an article (i.e. a Description in the cDnS sense) which defines the concept of “US President”. This concept of “US President” can then be said to classify the entity “Barack Obama”. Furthermore, a concept can classify different entities at different times—e.g. the concept “US President” has classified “Bill Clinton” and also “George Bush”—while a concept can also classify different entities at the same time—e.g. “British MP” classifies a number of persons currently sitting in the British Parliament.

In the context of academic domains, concepts correspond to elements of the specialised vocabulary or conceptual system of a particular domain. These domain concepts are typically defined and used by the theories and statements (i.e. the descriptions in a cDnS sense) that are shared and communicated among the agents in the domain.

1.6 SocialAgent

An Agent is required to interpret a given Information Object. When an Agent interprets an Information Object, the agent is said to conceive the Description that is expressed by that particular Information Object.

In the context of academic domains, persons involved in the production of scholarly texts can be characterised as social agents. Also, in the context of knowledge domains, one particularly important feature of agents and their relationship to information objects is that it is possible for two agents (e.g. an author and a reader) to interpret an Information Object differently, thereby conceiving of different descriptions [14]. These different descriptions can even sometimes be contradictory, even though ostensibly they have been derived from the same information object.

1.7 Collection

The collection class captures the intuitive notion of such entities as groups, teams, and associations. A Collection has at least two entities as its members and is said to “emerge” out of its member entities such that, “while retaining their identity, unity, and physical separation, [member entities] are ‘kept together’ in order to form a new entity” [15]. Note however, that the members of collection can change or be substituted
during the life of a collection without affecting the identity of the collection. This is one feature of cDnS collections that helps to differentiate them from mathematical sets. Furthermore, mathematical sets can be empty or singletons, but no empty or singleton collections are allowed in the cDnS “axiomatisation”.

In the context of academic domains, the community of researchers in a given knowledge domain can be thought of as a collection of agents that share one or more descriptions—these descriptions are said to unify the collection [7]. The “collection of agents” characterisation also applies to such entities as organisations, research groups or teams, and conference committees.

With this upper-level semiotic and constructivist framework in place, the paper proceeds to describe the design of the Scholarly Debate Ontology, which will define the vocabulary for representing debate in academic knowledge domains. As will be shown in the next section, the cDnS ontology just described will be used to illuminate some of the design choices in the Scholarly Debate Ontology.

2 The Scholarly Debate Ontology: codifying dialectical exchange

The Scholarly Debate Ontology builds on the debate mapping approach of Robert Horn et al. [16], who, in landmark argumentation modelling work, created seven paper-based, wall-mountable debate maps for analysing the history and current status of scholarly debate about whether computers can think. Horn’s work is directly relevant here because he also recognises that when it comes to the task of analysing knowledge domains, it is important to understand how all the arguments fit together in that knowledge domain. He is particularly interested in representing dialectical exchange between scholars so as to be able to answer the question: “Where can I get an overview of the history of the arguments so I can decide which I want to read?”

What has emerged from Horn’s work is a theory of the structure of scholarly debate, which has subsequently been articulated by his colleague in the creation of the maps, Jeffrey Yoshimi, in what he calls a “logic of debate” [17]. Whereas most argumentation modelling research concentrates on the microstructure of arguments (e.g. modelling the schemes of inference that link premises to conclusions), the concern of a logic of debate is how arguments themselves are “constituents in macro-level dialectical structures” ([17]). This section describes how the basic elements of this logic of debate are implemented as classes and relations in the Scholarly Debate Ontology, all within the semiotic and constructivist framework outlined in the previous section.

2.1 Issue

In the Horn debate mapping approach, argumentative exchange between two or more scholars is depicted as occurring within the context of a particular issue[8]. Figure 2 shows the definition of the Issue class. In the context of academic domains, issues

[7] uses the term “knowledge community” to label such a collection of agents.

[8] Indeed, the approach taken by Horn is sometimes referred to as issue mapping.
typically correspond to the research questions expressed in individual academic publications. Thus, the Issue class is defined as an indirect subclass of the cDnS class cdns:Description via another new class for the Scholarly Debate Ontology, NonPropositionalContent\(^7\). Therefore, as a subclass of cdns:Description, an Issue inherits the cdns:isExpressedBy attribute, and specialises this attribute so that it holds values of type Publication, another class defined in the ontology. The Issue class is specified with another attribute—verbalExpression—which allows an arbitrary text string to be associated with a given Issue instance. The verbalExpression attribute is introduced here primarily because of pragmatic argument modelling, to make the manual process of representing real debates more tractable for a human modeller. The composition\(^8\) of the text string that appears as the value of the verbalExpression attribute has no impact on the reasoning of the system. Indeed, it is the case that an issue could be expressed in a non-textual manner, in which case the verbalExpression attribute might be replaced by a nonVerbalExpression attribute. Note, however, that this paper does not suggest that all verbal and non-verbal forms of expression in scholarly text are directly interchangeable. As discussed previously, scholarly texts consist of both verbal expressions (e.g. sentences and paragraphs) and non-verbal expressions (e.g. graphs and figures). According to [12] “no verbal description can construct the same meaning as a picture”, which suggests that non-verbal expressions cannot be directly reduced to corresponding verbal expressions. However, [12] also explains that we learn to count different abstractions as the same for some restricted purposes. Finally, the listing shows that the Issue class is also specified with one new relation—relatedIssueOf—which allows one issue to be asserted as related to another issue. This is intended currently as all-comprising, generic relation that can be specialised in future research that determines the key relationships between Issues.

Indeed, following the ontology-design principle of minimal ontological commitment, only the essentials of the Issue class for the purposes of representing scholarly debate, have been specified in the ontology. However, it is possible that in future iterations of the ontology design it may be desirable to extend the definition of the Issue class to incorporate explicit constraints on issues such as whether a given issue allows a closed set of answers (e.g. only “Yes” or “No” answers) or an open set of answers to be offered in response to a given issue.

2.2 Proposition and Argument

In addition to issues, the logic of debate consists of propositions and arguments, thus two classes, Proposition and Argument, are introduced into the ontology. As shown in Figure 3, both the Proposition class and the Argument class are defined as subclasses of the class PropositionalContent, which in turn is a specialisation of the cDnS class cdns:Description. As with the Issue class described previously, the verbalExpression

\(^7\)As explained in the previous section, descriptions found in scholarly publications represent either some propositional content (expressed in declarative sentences in the publication) or some non-propositional content (expressed in interrogative sentences—i.e. questions—in the publication).

\(^8\)Even though the text string itself can have an arbitrary composition, it is useful to express the text-string in a grammatically well-formed manner even if it means that the text string no longer corresponds verbatim with the original source from which it has been taken.
attribute is introduced here for the Proposition class primarily because of pragmatic argument modelling reasons. Here the conceptualisation of Argument corresponds to a collection of propositions, one of which is a conclusion and the rest of which are premises. Thus the Argument class is defined with attributes hasPremise and hasConclusion, both of which have values of type Proposition. Note also that the Argument class is defined with the constraint of having at least one (:min-cardinality 1) premise and at most one (:max-cardinality 1) conclusion. In Yoshimi’s logic of debate, the distinction between propositions and arguments is one of a matter of scale—he suggests that it is possible to condense the representation of an entire argument down to a single declarative sentence and that furthermore, for argument visualisation, it is useful to do so.

On Horn’s debate maps, the propositions and arguments depicted as part of an issue region can be said to “address” that particular issue. The addresses relation is one that is implicit in Horn’s representation scheme but which is now made explicit in the Scholarly Debate Ontology.

Horn’s representation scheme then defines two main relations—is supported by and is disputed by—that hold between arguments. [17] offers three examples of types of support between any arguments A1 and A2: (1) logical—i.e. A2 supports A1 if A2 strengthens the conclusion of A1, (2) historical—i.e. A2 supports A1 if A2 is an earlier argument that A1 draws on, and (3) specialization—i.e. A2 supports A1 if A2 is a more specific version of A1. In terms of disputation, according to the logic of

---

In detailing the terminology of his logic of debate, Yoshimi uses the active form of the verbs “supports” and “disputes”. In contrast the terminology of the Horn debate mapping scheme uses the passive form “is supported by” and “is disputed by”, with the reason being that this allows the map reader to visualise the arguments from left to right. With respect to representing the relation in a formal knowledge base, the distinction between the active and passive form is not a fundamental one, and it is typical to have both forms specified in the ontology (as inverses of each other). The choice is then left to the modeller as to which form to use when representing a particular relation instance.
debate, argument A2 disputes argument A1 if the conclusion of A2 is the negation of some statement in A1. In the logic of debate, both supports and disputes are irreflexive, asymmetric, and non-transitive. Figure 4 shows the definition of the addresses, supports, and disputes relations in the Scholarly Debate Ontology. The :sufficient component of the supports relation definition specifies that the relation between premise and conclusion is also a special case of the supports relation.

2.3 Position and ViewpointCluster

The final element of the logic of debate is position, which [17] informally defines as a “family of mutually complementary arguments” relative to a debate. More formally, a position is defined as a collection of arguments related by the supports relation, forming what the author refers to as an “aggregated support path”. The author gives example positions such as Materialism in the Philosophy of Mind and Utilitarianism in Ethics.

According to [17], positions are important elements of the logic of debate because they provide additional information that is essential to understanding the structure of debate. This view about the informational value of positions or philosophical camps\(^\text{10}\) is echoed by [18] and [19], which have identified that one of the difficult aspects of understanding debates is that the protagonists come from quite different worldviews, bringing vastly different assumptions about the nature of reality. So in order to provide support for learners in gaining insight into why particular arguments take place, the debate maps created by [16] include a description of all the major camps from which scholars enter the debate.

The specification of the Position class is shown in Figure 5. This specification shows that the Position class is defined as a subclass of the cdns:Collection class in the cDnS ontology. The specification also includes the attributes hasViewpoint (which links instances of the PropositionalContent class to a given position), associatedPerson (which links instances of Person to a given position) and hasOpposingPosition (which

\(^{10}\)The term “philosophical camp” is used by [18] to describe the same debate phenomenon.
links one position to another when the two positions clash with each other in the context of a particular issue.

Specifying the Position class in the ontology allows for top-down representation of existing intellectual groupings in a scholarly debate. However, it is noted here, and argued further elsewhere [20] that supporting the bottom-up detection of similar intellectual groupings should be a major aim of any technology that purports to enable representing and reasoning about debates in knowledge domains. As described in [20], combining ontology-based analysis with graph-based cluster analysis is one viable approach to enabling bottom-up detection of coherent groups of argument. Thus, Figure 5 also introduces an ontological specification for these debate structures that can be automatically detected. This specification is introduced to be able to distinguish ontologically between what is explicitly represented in a top-down manner and what can be detected in a bottom-up manner. The class ViewpointCluster\textsuperscript{11} is introduced to account for the latter. The ontology specifies that two ViewpointCluster instances can be opposed to each other. The assumption here is that opposition between ViewpointCluster instances can be determined based on the occurrence of disputes relations between individual Argument instances that are part of each ViewpointCluster instance. Two intuitive criteria have been trialled in [20] for detecting opposing ViewpointCluster instances. Using the first criterion, the system infers an opposition relation between two ViewpointCluster instances if at least one viewpoint in one cluster has a disputes relation with at least one viewpoint in the other cluster. This criterion is labelled as

\textsuperscript{11}The rationale behind the name is that “viewpoint” is used in [7] as a synonym for “Description”, and “Cluster” indicates the central role played by cluster analysis to this task.

\begin{verbatim}
(def-relation supports (?p1 ?p2)
 :constraint (and
               (PropositionalContent ?p1)
               (PropositionalContent ?p2)))

(def-relation disputes (?p1 ?p2)
 :constraint (and
               (PropositionalContent ?p1)
               (PropositionalContent ?p2)))

(def-relation addresses (?p ?iss)
 :constraint (and
               (PropositionalContent ?p)
               (Issue ?iss)))
\end{verbatim}

Figure 4: The specification of the “supports”, “disputes”, and “addresses” relations.
"weak opposition". Using the second criterion, the system infers an opposition relation between two clusters if more than half (i.e., the majority) of the viewpoints in one cluster have a disputes relation with the viewpoints in the other cluster. This criterion is labelled as “strong opposition". Weakly and strongly opposed clusters are related to the appropriate ViewpointCluster instance via the hasOpposingClusterWeak and hasOpposingClusterStrong attributes respectively.

### 2.4 Person, Publication, and DomainConcept

These classes do not correspond directly to elements of the logic of debate. However, these classes correspond to what [17] refers to as “argument classifiers”—i.e. additional information within the debate, such as who made a particular argument or in what year and publication was the argument put forward, which is useful for the debate map reader. Yoshimi also recognises the special relevance that such additional information may have for computable representations of the debate, where a user may want to query a system to find (e.g.) all the arguments made by a particular author or in a particular journal. Figure 6 shows the specification of Person (which is defined as a subclass of cdns:SocialAgent), Publication (which is defined as a subclass of cdns:InformationObject), and DomainConcept (which is defined as a subclass of cdns:Concept).

Figure 7 shows a semantic-network-like visualisation of the Scholarly Debate Ontology. The figure shows the relationship between classes in the SDO ontology and upper-level classes of the cDnS ontology.

### Conclusion and Future Work

This paper described the design of an ontology for representing and reasoning about scholarly debate in academic knowledge domains. The ontology has been used to represent and reason about actual scholarly debates as described in more detail elsewhere.
(def-class Person (_cdns:SocialAgent))

(def-class Publication (_cdns:InformationObject)
  ((hasAuthor :type Person)
   (hasTitle :type String)
   (hasYear :type #_time:Year-In-Time)))

(def-class DomainConcept (_cdns:Concept)
  ((_cdns:isDefinedIn :type _cdns:Description)))

Figure 6: The specification of the Person, Publication, and DomainConcept classes as subclasses of the _cdns:SocialAgent, _cdns:InformationObject, and _cdns:Concept classes, respectively.

Figure 7: The Scholarly Debate Ontology
However, what this paper focussed on was the design considerations for the ontology. The paper began by characterising academic domains as settings for the collective construction of knowledge, thus motivating the reuse of the upper-level cDnS ontology. This upper-level ontology acted as a framework for ensuring the design process captured the essential elements of debate in knowledge domains. Finally, the paper specified the classes and relations in the Scholarly Debate Ontology.

For the short-term, pragmatic concerns of building an example application, the ontology-based modelling described in [20] was conducted in a centralised, single-person setting, as this enabled rapid prototyping of the proposed approach to knowledge domain analysis. However, in the longer term, decentralised, mass modelling of scholarly material would be necessary to allow the technology to be widely deployed. The scenario of distributed, mass modelling is likely to involve individual authors submitting representations of their papers. Indeed, it is not overly ambitious to envisage a future scenario where authors submit a formal representation of their paper along with the actual paper itself, in much the same manner that they currently submit abstracts as meta-descriptions of the paper. However, it is likely that other users of such a system would contribute models of literature where they are not the original authors. Thus distributed annotation itself presents a conceptual challenge—i.e. the challenge of determining how to deal with multiple, possibly contradictory representations of the same source material. It is clear that in a scholarly discourse there can be competing conceptualisations (descriptions) of reality AND we can have competing interpretations (descriptions) of a publication that expresses conceptualisations of reality. The ideal solution is to explicitly represent both but make a clear ontological distinction between them. The cDnS ontology, with its formal treatment of situations, provides a comprehensive framework for solving this problem. As mentioned previously, one instance of cdns:Situation can provide the setting or context for another cdns:Situation instance. This means that situations can be layered, with one situation having the other within its scope. Future work will seek to exploit the cDnS ontology in this way.

Finally, future work will seek to extend the Scholarly Debate Ontology with a more thorough treatment of intellectual groupings in knowledge domains. In this regard, one extension of the cDnS ontology that gives an extensive treatment of “collectives” could prove beneficial.

Acknowledgements

The work presented here is part of the PhD thesis of the author [20]. Thus the author would like to thank the supervisors of that PhD—Dr. Simon Buckingham Shum, Prof. John Domingue, and Dr. Clara Mancini—for their tireless support of this research endeavour.

References


