Practical Semiotics: A Formal Theory

Adam Pease, Ian Niles
Teknowledge
1810 Embarcadero
Palo Alto, CA 94303 USA
[apease | iniles]@teknowledge.com

Abstract
We address the development of a practical theory of semiotics within the context of the IEEE Standard Upper Ontology effort. The theory has the form of a set of predicates and other terms that are axiomatized in first order logic. We argue that this theory is sufficient for representing many of the subtleties of semantic representation that are likely to arise in the context of interoperability of information systems.

Keywords: ontology, semiotics, logic

1. SUO Scope And Purpose

Recognizing both the need for large ontologies and the need for an open process leading to a free, public standard, a diverse group of collaborators from the fields of engineering, philosophy, and information science have come together to work on a Standard Upper Ontology (SUO). The SUO email list was created in May of 2000 and, shortly after that date, had over 150 subscribers. The Project Authorization Request (PAR), which details the scope and purpose for the SUO effort was submitted to the IEEE in October and was approved as working group P1600.1 in December.

The Standard Upper Ontology (SUO) will provide definitions for general-purpose terms, and it will act as a foundation for more specific domain ontologies. It is estimated that it will eventually contain between 1000 and 2500 terms and roughly ten definitional statements for each term. The SUO will have the following purposes:

- Design of new knowledge bases and databases. Developers can craft new knowledge and define new data elements in terms of a common ontology, and thereby gain some degree of interoperability with other compliant systems.
- Reuse/integration of legacy databases. Data elements from existing systems can be mapped just once to a common ontology.
- Integration of domain-specific ontologies. Such ontologies (if they are compliant with the SUO) will be able to interoperate (to some degree) by virtue of shared terms and definitions.

There are two official "starter documents" approved by the SUO email list. The formal semiotic theory to be addressed in this paper is contained in one of those documents, the Suggested Upper Merged Ontology (SUMO) [1]. The SUMO uses a simplified version of KIF [2] called SUO-KIF [3] as the language for specifying the ontology.

2. Semiotics Background

Semiotics is the theory of signs, and it has aroused considerable interest in a wide range of fields, including linguistics, anthropology, literary criticism, political science, and philosophy. Two of the original thinkers in semiotics were Charles Sanders Peirce [4] and Ferdinand de Saussure [5]. Each of these thinkers posited a basic theory of signs and their meanings. According to De Saussure, a sign is made up of two elements, the signified and the signifier. The signifier is the sound or image that stands for something else, e.g. the word "bachelor", and the signified is the concept represented by the signifier, e.g. the concept of being an unmarried, adult male. Peirce complicated this picture somewhat by introducing an additional element. According to him, a sign consists of a Representamen (corresponding to De Saussure's signifier), an Interpretant (corresponding to De Saussure's signified), and an

Object, the thing or set of things in the actual world to which the Representamen refers.

In this paper, we will flesh out these basic pictures of semiotics in the context of ontological engineering. In particular, we will propose a set of formally specified predicates, which we believe can represent that part of semiotics that is relevant to semantic interoperability between heterogeneous information systems. It is important to point out that this concern leaves out much that conventionally goes under the heading of semiotics. Another helpful reference in this area is [8]. However, we are not concerned with branches of semiotics that address the interpretation of signs, pictorial semiotics [6], discourse analysis, nor the broader application of semiotics in cultural or political contexts [7].

3. SUMO Semiotics

The SUO email list has created a well-defined set of semiotics predicates that are not, as far as the authors of this paper are aware, available in any other publicly available ontology. The members of the list were able to agree on this stock of predicates (and their corresponding formal definitions) after working through some concrete examples relating to intellectual property issues. These examples concerned the appropriate representation of the following items:

- The fictional character Hamlet
- An edition of the play Hamlet
- A copy of the play Hamlet
- A performance of Hamlet
- A performance of Hamlet captured on video
- The timeless informational content of the play Hamlet

Before we discuss these examples and their representation, we'll define all of the predicates that were developed on their basis. Note that these definitions will be somewhat informal. The complete, formal definitions of these predicates and their supporting concepts are available in text and hyperlinked, browsable format at http://ontology.teknowledge.com .

The central semiotic predicate in the SUMO is **represents**. **(represents THING ENTITY)** means that THING in some way expresses, connotes, pictures, describes, etc. ENTITY. The two subrelations of **represents** are **containsInformation** and **realization**. The former predicate relates a **ContentBearingObject** to the **Proposition** that is expressed by the **ContentBearingObject**. Examples include the relationships between a physical novel and its story and between a printed score and its musical content. As for the other subrelation of **represents**, **(realization PROCESS PROP)** means that PROCESS is a **Process** that expresses the content of the **Proposition**. Examples include a particular musical performance, which realizes the content of a musical score, or the reading of a poem. Note that the only difference between **containsInformation** and **realization** concerns the type of instance of **Physical** that is permitted to occupy the first argument position. This is a small difference, but it is reflected prominently both in natural languages and in semiotics discussions, and so we decided to find a place for it in the SUMO.

It is important to note that, in some cases, the binary relation **represents** is too general in scope. On the basis of suggestions from SUO list member Phil Jackson, we have added the ternary predicates **representsForAgent** and **representsInLanguage**. These predicates are useful when one wants to relativize the representation relationship to a particular **Agent** or **Language**.

Although **represents** can be regarded as the central semiotics predicate in the SUMO, it is a subrelation of another semiotics predicate **refers**. **(refers OBJECT1 OBJECT2)** means that OBJECT1 mentions or includes a reference to OBJECT2. The predicate **refers** is more general in meaning than **represents**, because presumably something can represent something else only if it refers to this other thing. For example, an article whose topic is a recent change in the price of oil may refer to many other things, e.g. the general state of the economy, the weather in California, the prospect of global warming, the options for alternative energy sources, the stock prices of various oil companies, etc.

The predicate **refers** and its subrelations all relate a **ContentBearingObject** to something denoted by the object. Another class of predicates in the SUMO relates two **ContentBearingObjects** or classes of **ContentBearingObjects** to one another. These predicates are **subsumesContentInstance**,

equivalentContentInstance, **subsumesContentClass**, and **equivalentContentClass**.

(subsumesContentInstance OBJECT1 OBJECT2) means that the content expressed by OBJECT2 is part of the content expressed by OBJECT1. An example is the relationship between a handwritten poem and one of its stanzas. This is a relation between instances, rather than classes. If one wants to assert a content relationship between classes, e.g. between a version of an intellectual work and a part of that work, the relation **subsumesContentClass** is used. Note that this latter predicate is needed in many cases. Consider, for example, the relation between the King James edition of the Bible and its Book of Genesis. This relation holds for every copy of this edition and not just for a single instance.

The predicate **equivalentContentInstance** means that the two **ContentBearingObjects** related express exactly the same content. An example would be the relationship between a handwritten draft of a letter to one's lawyer and a typed copy of the same letter. The predicate **equivalentContentClass**, on the other hand, means that the content expressed by a subclass of **ContentBearingObject** is identical with the content expressed by another subclass of **ContentBearingObject**. An example of this sort of relationship holds between English and Russian editions of Agatha Christie's 'Murder on the Orient Express'. Note that **(equivalentContent OBJECT1 OBJECT2)** is equivalent to **(subsumesContent OBJECT1 OBJECT2)** and **(subsumesContent OBJECT2 OBJECT2)**, whether one is talking about instances or subclasses of **ContentBearingObject**.

Aside from the predicates discussed in the preceding paragraphs, it is important to understand the SUMO concepts of **Proposition** and **ContentBearingObject**. Propositions are abstract entities that express a complete thought or a set of such thoughts. As an example, the formula '(instance Yojo Cat)' expresses the Proposition that the entity named Yojo is an element of the class of cats. Note that propositions are not restricted to the content expressed by individual sentences of a language. They may encompass the content expressed by theories, books, and even whole libraries.

ContentBearingObjects, on the other hand, are objects that express a **Proposition**. It is important to distinguish **Propositions** from the **ContentBearingObjects** that express them. A **Proposition** is a piece of information, e.g. that the cat is on the mat, but a **ContentBearingObject** is an object that represents this information. A **Proposition** is an abstraction that may have multiple representations: strings, sounds, icons, etc. For example, the **Proposition** that the cat is on the mat is represented here as a string of graphical characters displayed on a monitor and/or printed on paper, but it can be represented by a sequence of sounds or by a non-latin alphabet or by a cryptographic form.

It may be useful to see how all of these semiotics-related terms are organized into type hierarchies.

```
refers
  represents
    containsInformation
    realization

Entity
  Physical
    Object
      ContentBearingObject

subsumesContentInstance
  equivalentContentInstance

subsumesContentClass
  equivalentContentClass
```

Note that indentation indicates a subsumption relation between the indented term and the preceding term.

Now that we've presented the semiotics predicates in the SUMO, let's return to the representational challenge problems involving Hamlet. Returning to our "Hamlet" example,

Hamlet the fictional character

```
(instance Hamlet Human)
```

Hamlet an edition of the printed play

```
(subclass
  Hamlet-FolgerEdition
  ContentBearingObject)
(subclass
  Hamlet-ScribnerEdition
  Hamlet-FolgerEdition)
(equivalentContentClass
  Hamlet-ScribnerEdition
  Hamlet-FolgerEdition)
```

A copy of the printed play

```
(instance
  Hamlet-ScribnerEditionOnMyBookshelf
  Hamlet-ScribnerEdition)
```

A performance of Hamlet

```
(instance
  HamletPerformanceByRoyalSPCoOnJune18-99
  IntentionalProcess)
 (realization
  HamletPerformanceByRoyalSPCoOnJune18-99
  Hamlet-ThePlay)
```

A performance of Hamlet captured on video

```
(instance
  VideoOfHamletPerfByRoyalSPCoOnJune18-99
  ContentBearingObject)
(refers
  VideoOfHamletPerfByRoyalSPCoOnJune18-99
  HamletPerformanceByRoyalSPCoOnJune18-99)
```

The timeless informational content of the play

```
(instance Hamlet-ThePlay Proposition)
(containsInformation
  Hamlet-ScribnerEditionOnMyBookshelf
Hamlet-ThePlay)
```

4. Formal Details

In this final section, we'll present the axioms that provide the formal, semantic underpinning to the predicates that were considered in the previous section. Because the semiotics section of the SUMO is only one small part of this large, upper-level ontology, there are additional, relevant axioms that are too numerous to reproduce here. Please consult the on-line browser for complete information.

In the discussion below it may also be helpful to refer to the on-line specification of SUO-KIF [3]. In particular, terms that begin with a question mark '?' denote variables, and 'exists', 'forall', 'and', '=>', and '⇔' stand for the existential quantifier, the universal quantifier, and the truth-functional operators of conjunction, material implication, and material equivalence, respectively.

Realizing some proposition in a process requires that there be an object that represents the information in that proposition.

```
(=>
  (realization ?PROCESS ?PROP)
  (exists (?OBJ)
    (and
      (instance ?OBJ ContentBearingObject)
      (containsInformation ?OBJ ?PROP))))
```

If one class subsumes the informational content of another, instances of the second class express all of the information represented by instances of the first class.

```
(<=>
  (subsumesContentClass ?CLASS1 ?CLASS2)
  (forall (?INFO ?OBJ1 ?OBJ2)
    (=>
      (and
        (instance ?OBJ1 ?CLASS1)
        (instance ?OBJ2 ?CLASS2)
        (containsInformation ?OBJ1 ?INFO))
      (containsInformation ?OBJ2 ?INFO))))
```

If one instance subsumes the informational content of another, then the first object contains all the information contained in the second object.

```
(<=>
  (subsumesContentInstance ?OBJ1 ?OBJ2)
  (forall (?INFO)
    (=>
      (containsInformation ?OBJ1 ?INFO)
      (containsInformation ?OBJ2 ?INFO))))
```

Content bearing objects contain some sort of information

```
(=>
  (instance ?OBJECT ContentBearingObject)
  (exists (?PROP)
    (containsInformation ?OBJECT ?PROP)))
```

If two instances of `ContentBearingObject` subsume each other's content, then these instances must express the same content.

```
(<=>
  (and
    (subsumesContentInstance ?OBJ1 ?OBJ2)
    (subsumesContentInstance ?OBJ2 ?OBJ1))
  (equivalentContentInstance ?OBJ1 ?OBJ2))
```

If two subclasses of `ContentBearingObject` subsume each other's content, then each subclass must express the same content.

```
(<=>
  (and
    (subsumesContentClass ?CLASS1 ?CLASS2)
    (subsumesContentClass ?CLASS2 ?CLASS1))
  (equivalentContentClass ?CLASS1 ?CLASS2))
```

If two sentences have equivalent content, they imply one another.

```
(=>
  (and
    (equivalentContentInstance ?SENT1
?SENT2)
    (instance ?SENT1 Sentence)
    (instance ?SENT2 Sentence))
  (<=> ?SENT1 ?SENT2))
```

One sentence subsumes the content of another sentence only if it implies the other sentence.

```
(=>
  (and
    (subsumesContentInstance ?SENT1 ?SENT2)
    (instance ?SENT1 Sentence)
    (instance ?SENT2 Sentence))
  (=> ?SENT1 ?SENT2))
```

Acknowledgments

We wish to acknowledge our Army sponsors for funding and guiding this work, and we would like to thank members of the SUO email list for their many helpful suggestions and criticisms.

References

[1] Niles, I., & Pease, A. (2001) "Toward a Standard Upper Ontology", in *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, Ogunquit, Maine, Oct. 17-19.

[2] Genesereth, M. (1991). "Knowledge Interchange Format," Proceedings of the Second International Conference on the Principles of Knowledge Representation and Reasoning (KR-91), J. Allen et al., (eds), Morgan Kaufman Publishers, 1991, 238-249. See also http://logic.stanford.edu/kif/kif.html.

[3] Pease, A., (2000), Standard Upper Ontology Knowledge Interchange Format, unpublished web document edited from (Genesereth, 1991), http://suo.ieee.org/suo-kif.html

[4] Peirce, C. S. (1931-1958) *Collected Papers of C. S. Peirce*, ed. by C. Hartshorne, P. Weiss, & A. Burks, 8 vols., Harvard University Press, Cambridge, MA.

[5] De Saussure, F. (1916) *Cours de Linguistique Générale*, translated by W. Baskin as *Course in General Linguistics* (1959), Philosophical Library, New York.

[6] Barthes, R. (1964) "Rhétorique de l´image", *Communications*, 4, 40-51. Also in Barthes, R., L´Obvie et l´Obtus (1982), Seuil, Paris, 25-42.

[7] Lotman, M., & Uspenskij, B. A. (1978) "On the semiotic mechanism of culture", *New Literary History*, IX, 2, 211-232.

[8] Eco, U. (1976) *A Theory of Semiotics*, Indiana University Press, Bloomington.