Meaning and the Semantic Web*

Bijan Parsia
University of Maryland
College Park, MD, USA
bparsia@isr.umd.edu

Peter F. Patel-Schneider†
Bell Labs Research
Murray Hill, NJ, USA
pfps@research.bell-labs.com

ABSTRACT

The meaning of names (URI references) is a contentious issue in the Semantic Web. Numerous proposals as to how to provide meaning for names in the Semantic Web, ranging from a strict localized model-theoretic semantics to proposals for a unified single meaning. We argue that a slight expansion of the standard model-theoretic semantics for names is sufficient for the present, and can easily be augmented where necessary to allow communities of interest to strengthen this spartan theory of meaning.

1. INTRODUCTION

The Semantic Web [2] is an extension of the World Wide Web. The major philosophical difference between the Semantic Web and the World Wide Web is that the Semantic Web is supposed to provide machine accessible meaning for its constructs whereas in the World Wide Web this meaning is provided by external mechanisms. (For example, the meaning an HTML document is whatever humans glean from its presentation in a browser, and the meaning of an XML document, even with a schema, is determined by the document or schema designer and is not present in the document or schema itself.) Meaning in the Semantic Web is largely based on the meaning of names which, in the Semantic Web, are URI references with optional fragment identifiers [5].

The initial view of the meaning of names in the Semantic Web was that the meaning of a name was determined by the owner of the name, if there is an owner. For names that use schemes based on authorities, such as the http scheme, this owner can be easily discovered by stripping off the fragment identifier, if any, and using the standard World Wide Web mechanisms to determine the owner of the resulting URI. On this view, good practice requires that the URI's owner to supply documents, accessible from the corresponding URI, which more or less express a definition of that URI. That definition is determinative, at least, in that a third party which discovered that definition through normal Web mechanisms and made use of it in reasoning with documents using that URI has, *certeris paribus*, exercised due diligence with respect to the URI owner's definitorial authority. On this view, it is unclear whether publishing documents at the relevant URI *defines* that URI, or simply provides a respectable default for random Web agents. This view of meaning led to Section 4.3, on the authoritative definition of terms, of the 23 January 2003 version of *Resource Description Framework (RDF): Concepts and Abstract Syntax* [6].

During the last call period of the above document, considerable pressure was applied against this view of meaning. As a result, the current version of the document [5] does not have a section on the meaning of RDF. This leaves RDF [7] with only the sparse meaning provided by the model-theoretic semantics of RDF [4]. In this account, the meaning of a name in RDF is relative to a particular RDF graph (which roughly corresponds to an RDF/XML document in the World Wide Web or perhaps a collection of such documents). Furthermore, the relativization does not strongly constrain the possible consistent interpretations of the name (and, thus, of the graph).

However, there is still a need to provide a stronger meaning for names than that provided by this model-theoretic semantics, in particular, to allow for or require the use of the meaning available from other documents. This paper examines the possibilities for meaning in the Semantic Web in general, and the meaning of names in the Semantic Web in particular.

The basic thesis of this paper is that it is sufficient to provide a rather sparse, and formal, notion of meaning, and that going beyond this spare notion as part of the core machinery of the Semantic Web is not warranted, at least, at this time. This notion of meaning will be able to utilize the machinery of the World Wide Web to access bodies of information which can provide common meanings, but does not require a common meaning and thus allows for divergences of meaning between different systems.

If a stronger meaning is required than that which can be provided by the formal machinery of the Semantic Web, then communities can define this meaning outside of the Semantic Web and agree to abide by this stronger meaning in their applications. In this way the expressive and computational limitations of the Semantic Web and the systems that work within it can be sidestepped when necessary.

Thus, we argue that the Semantic Web should not be too aggressive in providing standard machine accessible meaning for its constructs. The Semantic Web is an *extension* of the World Wide Web and, we hypothesize, will also depend heavily on external mechanisms to fix the meaning of its constructs. Growing the Semantic Web requires a delicate balance between interoperability, homogeneity, shared understanding, misunderstanding, and divergent needs.

2. A SPECTRUM OF MEANINGS

One of the problems when discussing meaning in the Semantic Web is how to specify meaning in the Semantic Web. There are many possibilities here, ranging from very sparse to very rich. How we choose to specify meaning has enormous effect on how we subsequently determine meaning in the Semantic Web. If we are too permissive in what we count as a specification of meaning, then determining that meaning will be very hard or impossible, and not just for programs. Some ways of making the meaning easy to determine put too great a burden on the specifier, thus inhibiting

*This is an extended version of a poster from WWW-2004.

†Work on this paper was partly supported under DARPA contract HR0011-05-C-0094.

the creation of Semantic Web data. The tradeoff is not unlike the tradeoff between the expressivity and computational complexity of logics.

At the rich end of the spectrum of meaning specification, one could say that the meaning of a document, or a piece of a document, is the meaning that was intended by whoever wrote the document. This version of meaning, often called "intended meaning", has some very useful properties. In particular, it provides a strong sense of cohesion.

However, there are several problems with this very rich sense of meaning. If the meaning of a piece of a document is always the meaning provided by the creator of that document then there is no possibility for other systems to use a different meaning. This prohibition prevents different systems from disagreeing about the validity of information, as in, for example, disputing an invoice.

(In some formalisms, it is possible to explicitly represent agreement and disagreement even of the meaning of terms, or to explicitly indicate which meaning is "in play" at any point in the document. In such formalisms there is not *this* problem in requiring global agreement by default, as that default agreement can be sidestepped as necessary. Of course, global agreement by default might produce other infelicities.)

Further, there is no real possibility of systems actually determining this intended meaning. In almost all cases part of intended meaning remains hidden in the internals of the creator of the intended meaning, and cannot be transmitted to another system, even if the other system is a person. Indeed, the original author might be unsure or ambivalent about the intentions that were to ground the meaning, or simply have no way to distinguish what was actually intended at the moment of authorship, and what is desirable at any subsequent moment. There needs to be some fairly strict constraints on how intentions ground meaning and what counts as evidence for the grounding intentions (and, thus, for the meaning). For example, we would find it troubling if an author of a commonly used ontology had the right and ability to come along and say: "Oh, I had *intended* to publish a completely different ontology. Everyone who had hard coded reasoning for that ontology got it wrong and *always did*."

At the sparse end of the spectrum, one could say that the meaning of a document, or a piece of a document, is only the meaning that is provided by a formal semantic account of the language in which the document is written. For example, the meaning of a URI reference or a triple in an RDF/XML document would simply be the meaning provided to that URI reference or triple by the RDF Semantics [4].

This formal meaning of meaning has very different characteristics than the intended meaning variant above. Meaning here is local to a document, and thus allows divergence between various systems. Meaning here is also completely determinable; there is no portion of meaning that is inaccessible to systems that process documents. However, this determinability is achieved by leaving the meaning to be much less constrained than people may be used to expect. In most cases, a software agent which acts in response to statements in a document (except to draw conclusions sanctioned by the formal meaning of that document) is imposing a specific interpretation on that document. On a plausible intended meaning accounts the agent will respect the meaning of a document just in case the agent acts in a way consonant with the rational expectations of the author of the document. (Whether this is the case may only be discernible, if at all, by examining the code of the agent.) This is not be so on the formal meaning account.

The main contribution of this paper is to discuss which possible meaning for meaning is suitable as the core, standardized meaning for the Semantic Web.

3. MEANING IN THE WORLD WIDE WEB

Let us first examine how meaning is determined in the other parts of the World Wide Web.

In what is sometimes called the "Visual Web", i.e., that portion of the World Wide Web that consists of documents that are meant to be rendered by web browsers, meaning is truely in the eye of the beholder. The only real meaning of documents in the Visual Web is that meaning that is given to the documents when they are viewed by a human user, after rendering by a browser. In fact, many important commercial sites in the Visual Web attempt to defeat any other possible way of determining meaning by changing the format of their web pages in ways that are difficult to automatically parse but that produce similar visual renderings. Similarly, many users avoid putting their email addresses on their web pages in textual form, instead using a visual rendering that is, again, difficult or impossible to automatically parse but that is easy for humans to read.

In what is sometimes called the "Syntactic Web", i. e., that portion of the World Wide Web that consists of XML documents that contain data not just meant to be rendered by web broswers, the situation is very different. In the Syntactic Web meaning is determined by agreements reached between communities of users, generally written down in standardization documents. These agreements are then used in the specification of software systems that act on the data. (In practice, of course, it is often the case that meaning in the Syntactic Web is really determined by the behaviour of the software systems and authoritative documents are only produced afterwards if at all.) Often part of this meaning is accessible to software by means of DTDs, XML Schemas, or other ways of specifying the form of XML documents, but only a small part of the meaning is available in this way [9].

For example, XHTML documents are a core part of both the Visual and the Syntactic Web, and are especially impoverished with regard to the interesting meaning of their contents. In the Visual Web, tables, images, and CSS are all used to encode, for example, the very same navigation bar for a presentation with almost equal successful use by the average Web surfer using a reasonably modern and capable browser. None of these techniques, however, allow Web browsers to use the navigation bar to provide keyboard shortcuts for the next and previous slide in the presentation. In the Syntactic Web, while there was no standard way to distinguish an HTML table of strings from a table of integers, it is relatively straightforward to do that in XHTML given the appropriate XML Schema. It is much more difficult to distinguish a table of calories burned for a given activity and a table of calories ingested with certain foods.

4. MEANING IN THE SEMANTIC WEB

The Semantic Web aims to make more of the meaning of Web documents accessible to software systems, by writing down more of this meaning in Web-accessible forms whose meaning is standardised and thus can be reliably processed by software systems. In this way the importance of extra-Web agreements and *ad hoc* software behaviour as specifiers of Web meaning can be reduced.

This vision of the Semantic Web relies heavily on a theory of meaning for documents in the Semantic Web. Such a theory of meaning has to be such that it both is reasonably and usefully processed by programs we currently know how to write, rich enough to allow us to connect the meaning of our documents to the behaviour, both actual and desired, of those programs, and intelligible to people who produce and use both the documents themselves and the programs which process them. How should this meaning be determined however? There are many options here, even leaving aside issues having to do with the expressive power of Semantic Web languages.

One important aspect of meaning in the Semantic Web arises from the fact that the World Wide Web is a (single) web of documents. We would thus expect that part of the meaning of any document (or part of a document) in the Semantic Web must depend, to some degree on other documents and parts thereof published on the Semantic Web.

At one end of the spectrum of solutions to this aspect of meaning in the Semantic Web is for meaning in the Semantic Web to be global. That is, the meaning of any document in the Semantic Web cannot be considered in isolation but must take into account the meaning of all documents anywhere in the Semantic Web. At the other end of this spectrum, another possible way of determining the meaning of a document in the Semantic Web is determined by only that document itself, and no other documents whatsoever contribute to the meaning of the document. At this extreme end of the spectrum, constructs like the explicit importing mechanism in OWL [3] would be forbidden.

There are several obvious intermediate points in this spectrum. One intermediate point near the idea of global meaning is to incorporate into the meaning of a document the meaning of all documents that are mentioned in that document. A document is mentioned by another document if there is a URI reference with optional fragment identifier (i. e., a Semantic Web name) in the second document that results in a URI for the first document when the fragment identifier, if any, is removed.

A variation on this method for determining which other information to incorporate would be to somehow select a portion of documents to utilize. Methods for this selection would, of course, have to be determined and would result in variations of this solution.

(In RDF/XML, there are already two mechanisms that identify pieces of an RDF/XML document that do not carry meaning, and potentially a third. First, since the meaning of an RDF/XML document is determined by the meaning of the RDF *graph* that results from the parsing (and other syntactic processing) of the document, any URI used in the document which does not make it into the graph is not significant to the meaning of that document. This includes many of the W3C defined URIs prefixed with http://www.w3.org/1999/02/22-rdf-syntax-ns#, in particular, those concerned with syntax. Second, URIs in literals do contribute to the meaning of the document, but in a way that is strongly disanalogous to the way URIs used directly in the graph do. It is entirely unlikely, for example, for the use of URIs in an XML literal to change the number of explicit assertions in the RDF graph in which that literal appears. Finally, RDF reification is intended to suppress certain aspects of the meaning of a URI reference, although the current underspecification of the meaning of reification in RDF's semantics leaves quite a bit open on how the intended suppression would affect the relation between URIs in "reified triples," the documents retrievable with those URIs, and the meaning of the document containing the reified triples.)

Yet another intermediate point, and a variation on the previous two solutions, would be to only incorporate external meaning under certain circumstances. Again different circumstances would result in variations of this solution; one variant that has been proposed would only consult external documents for names that are used as predicates, and not for other purposes.

A different intermediate point would be to consult only other documents (or portions of documents) that are explicitly mentioned as to be consulted, as in, for example, the OWL imports construct [8]. In this solution the use of a name is mostly divorced from the use of any document related to the name.

Aside from the question of which documents or document parts contribute to meaning, however, there is the question of how to determine the Semantic Web meaning of a collection of documents or document parts, once this collection has been determined. (It is possible, of course, for systems to go beyond this sanctioned meaning but they would then be "on their own", so to speak.) There is, again, a spectrum of solutions to this portion of meaning in the Semantic Web.

At one end of the spectrum the meaning of such collections is whatever meaning can be gleaned from them by any means whatsoever. This need not even be limited to effective means, but could include divining the intent of whoever (or whatever) created a document in the first place, even if that person (or, indeed, any agent) is no longer available for interrogation.

A less inclusive account of meaning admits only information that can be gleaned using resources available in the Semantic Web. This solution is still very expansive, as it could, in principle, incorporate meaning contained in documents or document fragments that are written in arbitrary formal languages, such as, for example, Montague logic, or even natural languages, such as, for example, Sanskrit.

A still less inclusive account restricts the kinds of documents to documents written in a formal language that has been standardized as part of a Semantic Web standardization effort and only allows the meaning provided by the formal specification of the language. Under this solution, then, the only meaning that is currently part of the Semantic Web is the meaning that comes from the model theories of RDF(S) [4] and OWL [8].

There is an even less inclusive account yet. In this account the meaning of a Semantic Web document is restricted to that provided by a particular, given set of standards, not all the Semantic Web standards. Under this account, there would be several different meanings for Semantic Web documents that use OWL constructs, one that incorporates only the RDF (or RDFS) meanings of these constructs and one that incorporates the additional meaning given to these constructs in OWL.

All of these possibilities seem reasonable to *some* application or community. The question here, however, is what should be the meaning sanctioned by the Semantic Web, that will serve as the base meaning for all systems that work within the Semantic Web. Making this base meaning too weak will mean that there is little benefit to be gained from working within the Semantic Web, as every community must define its own extended meaning and thus reduce the likelihood of interoperability and sharing between those communities. Making this base meaning too strong will make the Semantic Web too rigid and confining, again requiring communities to define their own, certainly incompatible, alternatives. Both cases will reduce the utility and impact of the Semantic Web as a unifying force in the World Wide Web.

5. AN EXAMPLE

The above discussion is all rather high-level and informal. While a formal account of each of the possibilities would permit the precise determination of their characteristics and effects, it is extremely unclear how to even begin to do so, particularly in a way that fits in with the current model-theoretic accounts of RDF and OWL. For our current, polemical purposes, we will illustrate some of the effects of the various possibilities by means of a simple example.

This example takes the form of a fairly facile abstraction of an electronic commerce domain.[1] The example is of necessity quite sketchy, because a large portion of it alludes to some future extension of the current Semantic Web and thus needs to be independent of the particulars of any Semantic Web language. The only semantic relationship that will be used is the notion that some information follows from some other information. Readers with a formal background may think of this relationship as some sort of entailment relationship in a powerful formal framework; readers with a less formal bent may think of it as determining what information is implicit in what is being considered.

This example will, mostly, ignore any aspect having to do with justification of this relationship. Although this would be a very important notion in some of the possibilities it is mostly independent of the points made here. The example will also ignore most aspects concerned with reasoning about what (other) agents believe, at least inside the Semantic Web. Again this will be an important part of some of the possibilities, but is also mostly independent of the points made here.

In this simplified example, sellers publish[2] catalogues of the products they sell, including, among other things, prices for the products. Buyers can then publish orders, which indicate which products they wish to buy from a particular seller. This seller then ships the product and publishes an invoice indicating how much the buyer owes the seller. The buyer pays the seller and publishes a payment notice. Finally the seller publishes a receipt.

To support reasoning there are published documents (ontologies) that provide information about catalogues, orders, invoices, payments, and receipts. This background theory is sufficient to make the appropriate information follow from the combination of the background theory and information about a particular invoice, payment, etc.

5.1 The Easy Case

Under the normal course of events no one publishes incorrect information and the steps above happen in the appropriate sequence and with the appropriate timing. We would like a theory of meaning in the Semantic Web to support reasoning appropriate to this normal case of events. For example, there should be ways to set up documents such that it follows from an invoice document that the buyer in the document needs to pay the seller in the document.

This is, in fact, quite easy to arrange for, just by ensuring that each document contains all the information needed to make the conclusions, perhaps by copying the contents of all the background documents into the document. However, as this destroys just about all of the sharing and community aspects of the Semantic Web, we will thus exclude solutions that require this amount of copying from now on.[3]

Even with this restriction just about all the solutions described above are adequate. Obviously solutions that collect the meaning of all documents in the Semantic Web or all documents implicitly mentioned will gather sufficient information from the background documents so that the appropriate information follows. Similarly solutions that allow the intent of document creators to be utilized will have sufficient information available in the formalism to make the appropriate inferences.

Whether software systems can really access and process this information is a separate matter, however. It is possible to "hard-code" meaning that is not accessible on the Web into special-purpose software, but we prefer solutions where meaning in particular domains can be determined by software systems that have not been written solely for that particular domain, and are, in fact, quite generic. This indicates that notions of meaning that require more direct access to the intent of human designers than the public meaning of their explicit statements are to be avoided in Semantic Web meaning. The same considerations indicate against notions of meaning that involve access to documents that are not written in languages that can be processed by current software.

However, this still leaves quite a few of the above possibilities as viable candidates for Semantic Web meaning.

5.2 The Hard Cases

So the easy case, where everyone agrees as to what is going on, does not provide much guidance as to which notion of meaning to use. What about when there is disagreement? This disagreement can have many sources (from honest error to exceptions to outright fraud) and many manifestations (from disagreement about particular facts all the way to disagreement about the fundamental assumptions concerning electronic commerce). The problem is how to handle disagreements in a flexible manner without reverting to total anarchy, which in this context means that no one can reliably reuse any information external to their documents. It is important not only to *mark* disagreements as such, but to allow disagreeing parties to be able to, so to speak, agree to disagree without necessarily having to use almost entirely disjoint vocabularies. After all, disagreement requires *some* commonality if it is to be disagreement, and not mere difference.

Disagreement makes it essentially impossible to "hard-code" intended meaning into software, as new software would have to be written to handle these divergent intended meanings. It is not a viable solution to require new programs to be written just because of arbitrary disagreements between any two systems in the Semantic Web.

Disagreement also makes notions of meaning that involve the entire Semantic Web unworkable, as the inconsistencies involved when, for example, a seller and a buyer disagree on whether a payment has actually been sent. In one example of this situation the buyer would publish a page stating that the payment has been sent. In an account of meaning in the Semantic Web that combines the meanings of each page, the seller would then have no effective way of disputing this information because any disputing information would simply cause a contradiction to follow, because of the global nature of this version of Semantic Web meaning.

Well, at least this is what would happen in simple accounts of global meaning in the Semantic Web. It would be possible to relativize all information in the Semantic Web. In this account each bit of information is only a claim by its publisher, and not a global truth. This account is actually quite useful and perhaps is the best account of meaning in the long run. However, it does require quite sophisticated machinery to make work correctly, and it would be useful to have a simpler account of meaning, at least for now.

It would also be possible to use some sort of paraconsistent theory of meaning, perhaps a form of relevance logic [1], where local contradictions can be part of a globally consistent view. However, paraconsistent logics are difficult to work with, and generally have very weak inference, too weak for most purposes.

[1] Yes, this is a very tired example, but it does serve to illustrate the main points here.

[2] Some of this "publishing" would be public, i.e., readable by others, and some would not. This example does not consider any effects of this difference.

[3] Note, however, that this solution *is* an improvement over the situation where there are multiple, incompatible languages and formats. While this solution eliminates much of Web like aspects of the Semantic Web, it is not entirely devoid of value.

Disagreement need not be limited to information about objects, such as particular invoices or payments, but can also extend to disagreements about concepts, such as disagreements about how invoices are to be structured or even the meaning of the properties of invoices. Disagreement can thus concern any of the information in question, including factual properties of objects, definitions of classes, and definitions of properties. Under a global theory of meaning in the Semantic Web, these sorts of disagreements would also lead to contradictions.

Theories of meaning that bring in the meaning of a document when a name from that document is mentioned also have problems when disagreements are possible. The above example illustrates this problem as well. The buyer publishes a document written in some formal Semantic Web language (currently RDF or OWL) that defines an object (using a particular name) that the buyer states is a valid payment record. The seller is again unable to dispute this, as any reference to the object will bring in the information from its defining document that it is a valid payment record, causing a contradiction to follow from the disputation.

Instead, what the seller needs to do is to refer to the object without necessarily bringing in any of the information that the buyer associates with the object. Then the seller can provide its own information about the object, perhaps to say that the account paid into is not the account of the seller, and thus that this is not a payment of the initial invoice. This can be done simply by not bringing in any of the information provided by the buyer.

How then can commonalities of information be achieved? The simplest way is to have a mechanism for explicitly bringing in information. Communities can publish general information (ontologies) and members of the community, and others, can explicitly include that ontology in their documents. If there are no disagreements, then one system (like the buyer) can explicitly include information that another has published (such as information generated by the seller about the invoice).

If there is a disagreement, then the buyer just does not include information from the seller, but can still use the same names and even copy some of the information from the seller's document. This results in two, contradictory versions, one from the seller and one from the buyer, but does not cause any contradiction within the Semantic Web as a whole, as there is no global meaning in which to place the contradiction.

6. THE PROPOSAL

We therefore propose that meaning in the Semantic Web be defined in a local sense, from a particular document or collection of documents. For determining which documents to consider when determining the meaning of a document, we propose that only documents explicitly mentioned in constructs like the OWL importing mechanism contribute to the meaning of that document.

We would like to be able to include portions of documents, as well as entire documents. However, there are currently few mechanisms for syntactically delineating portions of Semantic Web documents written in RDF or OWL. The only such mechanism of note involves Named Graphs (http://www.w3.org/2004/03/trix/), but it is not a standard part of the Semantic Web. Any mechanism for importing a portion of a standard Semantic Web document would thus have to use some external way of determining which portion of the document to import. Unfortunately, there is no good way of performing this determination. Any scheme that pulls in only information that mentions some name will be too weak (consider the problem of pulling in an OWL description) or impossible to implement (just what information in a document is relevant).

For determining the meaning of a collection of documents, we propose to use only that meaning determined by the formal language specifications of the Semantic Web, currently the RDF model theory [4] and the OWL model theory [8]. So the "follows from" relationship, which was only informally defined above, is really OWL entailment.

6.1 Discussion

Our proposal allows for divergences of meaning between different documents. A document that does not explicitly import a well-known ontology document, or, indeed any commonly-used document, can easily diverge from any portion, or indeed all, of the common meaning of any name. For example, a document could ignore the common meaning of an invoice and instead use one that has the consequence that the seller owes the buyer money.

One might think that our theory of meaning thus results in complete anarchy in the Semantic Web. Not so, however, or, if so, then we have embraced only those portions of anarchy that are necessary to prevent totalitarianism, for any proposal for Semantic Web meaning that cuts off easy access to disagreements will inevitably end up stultifying the Semantic Web.

Of course we really do not have a *solution* to handling disagreement in the Semantic Web. Our proposal just makes disagreement possible. A full solution to disagreement requires much more formal machinery than we feel is appropriate at this juncture in the development of the Semantic Web. Futher, no strictly formal means will be completely adequate to handle disagreement, short of a full solution to the AI problem, as determining which one of a collection of contradictory claims to believe inevitably brings in matters of trust and judgement. Our point is that it is necessary to *allow* unqualified disagreement even, or especially, at this stage of the Semantic Web.

Our proposal does, however, allow for consensus to be achieved, and in an easy fashion. All that is required to achieve consensus concerning meaning is to have the same background theory (and same external intuitions, but this is outside the Semantic Web) and our proposal makes it easy to support such consensus by placing a representation of the consensus in commonly-used documents.

Communities of interest that want to mandate a shared meaning can require the use of such consensus documents. These communities would have, of course, cut themselves off from potentially valuable dissent, but there are many cases, including electronic commerce, where requiring the use of some common meaning is useful or even required for progress, particularly with our poor understanding of how to build truely cognitive software systems. The need in our scheme for explicit importing of these consensus documents provides signals that a document adheres to this meaning, and these signals can be read both within and without the community.

Our proposal also makes it easy to determine (most of) at least the formal part of consensus meaning that is required by a document. The explicitly imported documents in a document (and the documents that these imported document import, and so on) provide an excellent indication of just which consensus meanings a document uses. Further the Semantic Web meaning of these consensus documents is just their formal meaning, which is easy to determine.

We freely admit that this notion of Semantic Web meaning is insufficient to capture the entirely of the meaning intended by document writers, and likewise insufficient to capture the entirely of the meaning which the behaviour of many effective software agents will act upon. There is nothing, however, in our proposal that prevents software systems from augmenting, or even replacing, the Semantic Web meaning with their own notions of meaning. Semantic Web meaning only serves as a core, common meaning for Semantic Web documents, to be used or abused as desired. As the Semantic Web evolves, some of these augmented notions of the meaning of the document may become common enough, and well understood enough, to augment or replace the core, standardized meaning. But this is an juncture where we feel strongly that (further) standardization should follow (future) practice.

6.2 Vs. the Semantic Web

As its name indicates, the Semantic Web is all about *meaning*. It is part of the vision of the Semantic Web that it will abound with meaning and that this meaning will be rich and interdependent in interesting ways allowing for the collaborative, largely uncoordinated development of a global knowledge oriented system. The Semantic Web, as an extension of the Web, is to be in key respects *like* the Web itself. One difficulty in this vision is that it is very unclear which structural features of the Web are necessary for achieving Web-like virtures, and which were accidental. Furthermore, the techno-social landscape is vastly different now than before the Web. For one, we have the Web itself to contend with.

Most Semantic Web advocates agree that URIs are somehow a key part both of realizing the Semantic Web and for integrating it with the World Wide Web. With an http URI, there is a strong expectation that using them with the HTTP protocol should do something useful and to fail to make good use of HTTP would be a significant missed opportunity. This distinguishes between views that merely take URIs to be, by default, univocal, from those which take the semantics of different URI schemes to be interesting, significant, fruitfully exploitable, and perhaps essential. There is a reason why few, if any, Semantic Web proponents advocate wholesale adoption of URNs as RDF and OWL terms. Indeed, the fact that the core RDF and OWL vocabulary entirely consist of HTTP URIs sets a very strong example.

In HTML documents, there are several contexts where URIs are used with an expectation that their dereferenced document will be part of the meaning of the HTML document. URIs in the `src` attribute of `img` elements, for example, clearly have inclusion semantics (though whether the included images are part of the expository content of the document, or "merely" navigational is not determinable). URIs in the `href` attributes of anchor elements sometimes affect the content of the document (for some entities) and other times do not. For example, there is quite a difference between the standard HTML idioms:

```
Click <a href=".....">here</a> to
accept the license.
```

and:

```
Sometimes one just adds a
<a href="http://dictionary.reference.com/-
                         search?q=sprinkle">
sprinkle</a> of links one
<a href="http://www.kentgallery.com/as/-
                         rodthe.htm">
thinks</a> are interesting.
```

for a human surfer.[4] For an HTML rendering engine, a search engine bot, a link checker or some mirroring program, all the URI uses above have the same status. URIs in textual content in HTML might still have link semantics for a human, and some browsers or other rendering engines might add missing anchors, but they do not alter the *standard* machine readable, hypertextual semantics of the document.

So, even in HTML, it seems useful to distinguish, for both programs and people, different ways of treating URIs which appear in the document. It is perfectly reasonable for a link harvesting program to grab every URI it can—from `hrefs`, from text content, from wherever—but it would be quite wrong for a link checking program to do so. Of course, both link checkers and link harvesters rely on a very thin sense of the semantics of used URIs, and have a very low barrier to success. Even so, they do require some machinery (e.g., HTTP response codes) to function properly. They also tend to have varyingly weal levels of respect for the obvious intentions of the document author. Human surfers, on the contrary, tend to be much more discriminating in their interpretations of how URIs in documents affect the content of that document. In the HTML fragments above, we have one URI which most human readers will interpret following as indicating an agreement to a licence, one URI which explicates the meaning of the term, and one which is just a fun, random annotation.

On the Semantic Web, we might expect programs to be more like the human surfer and less like the link checker. We might therefore expect that programs should be free to be rather judicious in what URIs they follow and incorporate into their understanding of the meaning of the document. We contend that our thin notion of Semantic Web meaning better supports such behaviour. As with HTML, you want to have some URIs to directly and "dumbly" influence the meaning of the document (e.g., inline images or external style sheets), and it seems best, given our current knowledge, to require those to be explicitly syntactically marked. Thus, constructs akin to the OWL imports mechanisms are the right idea.

It should be noted that even such an apparently mundane mechanism as OWL's imports construct arouses considerable controversy. It is hard to see how the Semantic Web community can arrive and the requisite confident consensus on anything substantially thicker in the near term.

One observation: Semantic Web documents tend to be substantially denser, regardless of their domain, than HTML or even arbitrary XML formats. Substantive HTML documents intended for human consumption, aside from bookmark files, have a fairly high text (or multimedia content) to link ratio. In XML, the use of QNames to name many distinct nodes tends to reduce the number of distinct URIs used. In RDF and RDF based languages, on the other hand, URIs are *everywhere*. This makes it much more difficult for human authors to understand the implications of using an URI, if mere use tends to bring in a lot of other assertions. Given such a high cost for using a URI, authors will be inclined to only use URIs under their control. We believe that to be unfortunately inhibitory. It seems more fruitful to let people use URIs with possibly divergent, but easily determinable meaning, and allow reconciliation to be demand driven, on a case by case basis. If there is a set of URIs that many document authors which to use with a shared meaning, that will tend to produce documents that exactly describe the desired meanings for those URIs, and authors can signal their use of the common meaning with an explicit imports statement.

6.3 A Refinement of the Proposal

To allow for software systems of differing sophistication and differing needs we augment our basic proposal above to allow Semantic Web meaning to be contingent on a set of Semantic Web languages that the system understands. Semantic Web meaning for software systems that do not implement OWL entailment would

[4] Yes, the first example goes against certain principles of the World Wide Web (see http://www.w3.org/TR/webarch/#safe-interaction), but this kind of "overloading" is not unknown.

then be less powerful than Semantic Web meaning for software systems that do.

There are mechanisms in the World Wide Web that can be used to support these variations on Semantic Web meaning. We would augment the importing mechanism to use content negotiation, allowing OWL-aware systems to request the OWL version of a resource identified by a URI and RDF-aware systems to request the RDF version of the same resource.

This does require considerable care on the part of designers of documents, so that, for example, OWL and RDF documents at the same resource have compatible meanings. However, this is not really different in spirit from issues having to do with the relationship between JPEG and GIF documents at the same URI.

One advantage of this refinement is that it allows for growth in the Semantic Web. New languages, such as a rules extension for OWL, can be added to the Semantic Web and retrofitted to previous ontologies, augmenting their Semantic Web meaning for software systems that can process the new language while still retaining the old behaviour for existing systems. It is even possible to have Semantic Web languages that are not compatible with OWL or even RDF.

6.4 Future Refinements

This view of meaning in the Semantic Web is certainly not the one that we would like to have for all time. When the Semantic Web becomes more widespread, when information in it becomes more sophisticated, and when more powerful software systems for the Semantic Web become available our simple version of Semantic Web meaning will be inadequate.

For example, it would be useful to reason about contradictory information in different documents. In the example above, it would be useful to reason *within the Semantic Web* that the buyer and the seller do have mutually contractory views. This can be done in certain kinds of modal logics, which would thus be useful as the foundations of a more powerful theory of meaning in the Semantic Web.

In this context it would also be useful to be able to reason, again within the theory of meaning of the Semantic Web, about the rights and obligations of agents within the Semantic Web. This sort of reasoning can perhaps be supported by a theory of meaning that includes these sorts of concepts.

It would also be useful to import only specific portions of documents. A future extension of Semantic Web languages that allow portions of documents to be identified would permit this more fine grained version of importing.

Finally, it does seem a bit odd, if somewhat harmless, that the OWL importing mechanism is an *OWL* importing mechanism, rather than an RDF one. It seems harmless as this is not a large extension to RDF and, if it proves popular, may become a *de facto* standard extension to RDF, eventually to be incorporated into RDF itself. (This possible course of events echoes the migration of the DAML+OIL collections constructs into RDF.) While the need for richer imports mechanisms becomes more acute for higher layers of the Semantic Web language stack, such as rules, it seems unlikely that any would conflict in a systematic way with OWL's imports construct. So we get exactly what we want: *some* mechanism that we can use *now*, that is plausibly forward compatible with future mechanisms.

7. CONCLUSION

We have argued that a formal account of meaning is appropriate for the Semantic Web and that this formal account need not require a common, universal notion of meaning in the Semantic Web. The only sharing mechanism needed, at least for now, is an explicit importation mechanism, similar to that provided in OWL. In this way information can be shared as appropriate, thus preventing total anarchy, but need not be, thus allowing for differences of opinion, which are needed to prevent totalitarianism and its resultant stultification and ossification.

8. ACKNOWLEDGMENTS

Many of the options for Semantic Web meaning in this paper are revisions and rehashes of previous proposals by others, in particular pioneers of the Semantic Web, including Tim Berners-Lee, and participants of the Tech Plenary session and the WWW2003 Birds of a Feather meeting on social meaning, as well as the Technical Architecture Group's task force on Semantic Web meaning. We have not attributed most of the options, partly because even with the strong record keeping activities of W3C it is difficult to determine just who originated an idea and partly because we have modified some of the options for expository reasons and the authors might not want to be associated with the modified option.

9. REFERENCES

[1] A. R. Anderson and N. D. Belnap, Jr. *Entailment: The Logic of Relevance and Necessity*, volume I. Princeton University Press, Princeton, New Jersey, 1975.

[2] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, May 2001.

[3] M. Dean, G. Schreiber, S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein. OWL web ontology language reference. W3C Recommendation, 10 Feb. 2004. http://www.w3.org/TR/owl-ref.

[4] P. Hayes. RDF model theory. W3C Recommendation, 10 Feb. 2004. http://www.w3.org/TR/rdf-mt.

[5] G. Klyne and J. J. Carroll. Resource description framework (RDF): Concepts and abstract syntax. W3C Rec., 10 Feb. 2004. http://www.w3.org/TR/rdf-concepts.

[6] G. Klyne and J. J. Carroll. Resource description framework (rdf): Concepts and abstract syntax. W3C Recommendation, 10 Feb. 2004. http://www.w3.org/TR/2003/WD-rdf-concepts-20030123.

[7] F. Manola and E. Miller. RDF primer. W3C Recommendation, 10 Feb. 2004. http://www.w3.org/TR/rdf-primer.

[8] P. F. Patel-Schneider, P. Hayes, and I. Horrocks. OWL web ontology language semantics and abstract syntax. W3C Recommendation, 10 Feb. 2004. http://www.w3.org/TR/owl-semantics.

[9] F. van Harmelen and D. Fensel. Practical knowledge representation for the web. In D. Fensel, editor, *Proceedings of the IJCAI'99 Workshop on Intelligent Information Integration*, 1999.