

Identity, Reference, and Meaning on the Web

Harry Halpin
Institute for Communicating and Collaborative Systems
University of Edinburgh
2 Buccleuch Place
Edinburgh, United Kingdom
H.Halpin@ed.ac.uk

ABSTRACT

Problems of reference, identity, and meaning are becoming increasingly endemic on the Web. We focus first on the convergence between Web architecture and classical problems in philosophy, leading to the advent of “philosophical engineering.” We survey how the Semantic Web initiative in particular provoked an “identity crisis” for the Web due to its use of URIs for both “things” and web pages and the W3C’s proposed solution. The problem of reference is inspected in relation to both the *direct object theory of reference* of Russell and the *causal theory of reference* of Kripke, and the proposed standards of new URN spaces and Published Subjects. Then we progress onto the problem of meaning in light of the Fregean slogan of *the priority of truth over reference* and the notion of logical interpretation. The popular notion of “social meaning” and the practice of tagging as a possible solution is analyzed in light of the ideas of Lewis on convention. Finally, we conclude that a full notion of meaning, identity, and reference may be possible, but that it is an open problem whether or not practical implementations and standards can be created.

1. PHILOSOPHICAL ENGINEERING

While the Web epitomizes the beginning of a new digital era, it has also caused an untimely return of philosophical issues in identify, reference, and meaning. These questions are thought of as a “black hole” that has long puzzled philosophers and logicians. Up until now, there has been little incentive outside academic philosophy to solve these issues in any practical manner. Could there be any connection between the fast-paced world of the Web and philosophers who dwell upon unsolvable questions? In a surprising move, the next stage in the development of the Web seems to be signalling a return to the very same questions of identity, reference, and meaning that have troubled philosophers for so long.

While the hypertext Web has skirted around these questions, attempts at increasing the scope of the Web can not: “The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation” [41]. *Meaning* is a thorny word: do we define meaning as “machine-readable” or “has a relation to a formal model?” Or do we define meaning as “easily understood by humans,” or “somehow connected to the world in a ro-

Copyright is held by the author/owner(s).

WWW2006, May 22–26, 2006, Edinburgh, UK.

bust manner?” Further progress in creating both satisfying and pragmatic solutions to these problems in the context of the Web is possible since currently many of these questions are left underspecified by current Web standards. While many in philosophy seem to be willing to hedge their bets in various ideological camps, on the Web there is a powerful urge to co-operate. There is a distinct difference between the classical posing of these questions in philosophy and these questions in the context of the Web, since the Web is a human artifact. The inventor of the Web, Tim Berners-Lee, summarized this position: “We are not analyzing a world, we are building it. We are not experimental philosophers, we are philosophical engineers” [2].

2. THE IDENTITY CRISIS OF URIS

The first step in the creation of the Semantic Web was to extend the use of a URI (Uniform Resource Identifier) to identify not just web pages, but anything. This was historically always part of Berners-Lee’s vision, but only recently came to light with Semantic Web standardization efforts and has caused disagreement from some of the other original Web architects like Larry Masinter, co-author of the URI standard [4]. In contrast to past practice that generally used URIs for web pages, URIs could be given to things traditionally thought of as “not on the Web” such as concepts and people. The guiding example is that instead of just visiting Tim Berners-Lee’s web page to retrieve a representation of Tim Berners-Lee via *http*, you could use the Semantic Web to make statements about Tim himself, such as where he works or the color of his hair. Early proposals made a chasm across URIs, dividing them into URLs and URNs. URIs for web pages (documents) are URLs (Uniform Resource Locators) that could use a scheme such as *http* to perform a “variety of operations” on a resource[5]. In contrast, URNs (Uniform Resource Names) purposely avoided such access mechanisms in order to create “persistent, location-independent, resource identifiers” [29]. URNs were not widely adopted, perhaps due to their centralized nature that required explicitly registering them with IANA. In response, URLs were just called “URIs” and used not only for web pages, but for *things not on the web*. Separate URN standards such as Masinter’s *tdb* URN space have been declared, but have not been widely adopted [27]. Instead, people use *http* in general to identify both web pages and things.

There is one sensible solution to get a separate URI for the thing if one has a URI that currently serves a representation of a thing, but one wishes to make statements about the

thing itself. First, one can use URI redirection for a URI about a thing and then resolve the redirection to a more informative web page [28]. The quickest way to do this is to append a “hash” (fragment identifier) onto the end of a URI, and so the redirection happens automatically. This is arguably an abuse of fragment identifiers which were originally meant for client-side processing. Yet according to the W3C, using a fragment identifier technically also identifies a separate and distinct “secondary resource” [23]. Regardless, this ability to talk about anything with URIs leads to a few practical questions: Can I make a statement on the Semantic Web about Tim Berners-Lee by making a statement about his home-page? If he is using a separate URI for himself, should he deliver a representation of himself? However, in all these cases there is the lurking threat of ambiguity: There is no principled way to distinguish a URI for a web page versus a URI for a thing “not on the Web.” This was dubbed the Identity Crisis, and has spawned endless discussions ever since [9].

For web pages (or “documents”) it’s pretty easy to tell what a URI identifies: The URI identifies the stream of bits that one gets when one accesses the URI with whatever operations are allowed by the scheme of the URI. Therefore, unlike names in natural language, URIs often imply the potential possession of whatever representations the URI gives one access to, and in a Wittgenstein-like move David Booth declares that there is only a *myth of identity* [7]. What a URI identifies or means is a question of use. “The problem of URI identity is the problem of locating appropriate descriptive information about the associated resource – descriptive information that enables you to make use of that URI in a particular application” [7]. In general, this should be the minimal amount of information one can get away with to make sure that the URI is used properly in a particular application. However, if the meaning of a URI is its use, then this use can easily change between applications, and nothing about the meaning (use) of a URI should be assumed to be invariant across applications. While this is a utilitarian and attractive reading, it prevents the one thing the Web is supposed to allow: a universal information space. Since the elementary building blocks of this space, URIs, are meaningless without the concrete context of an application, and each applications may have orthogonal contexts, there is no way an application can share its use of URIs in general with other applications.

3. URIS IDENTIFY ONE THING

Tim Berners-Lee has stated that URIs “identify one thing” [3]. This thing is a *resource*. The most current IETF RFC for URIs states that it does not “limit the scope of what might be a resource” but that a resource “is used in a general sense for whatever might be identified by a URI” such as “human beings, corporations, and bound books in a library” and even “abstract concepts” [4]. An earlier RFC tried to ground out the concept of a resource as “the conceptual mapping to an entity or set of entities, not necessarily the entity which corresponds to that mapping at any particular instance in time” in order to deal with changes in particular representations over time, as exemplified by the web sites of newspapers like <http://www.guardian.co.uk> [4]. If a URI identifies a conceptual mapping, in whose head in that conceptual mapping? The user of the URI, the owner of the URI, or some common understanding? While Tim

Berners-Lee argues that the URI owner dictates what thing the URI identifies, Larry Masinter believes the user should be the final authority, and calls this “the power of readers over writers.” Yet this psychological middle-man is dropped in the latest RFC that states that URIs provide “a simple and extensible means for identifying a resource,” and a resource is “whatever might be identified by a URI” [4]. In this manner, a resource and its URI become disturbing close to a tautology. Given a URI, what does it identify? A resource. What’s a resource? It’s what the URI identifies. According to Berners-Lee, in a given RDF statement, a URI should identify one resource. Furthermore, this URI identifies one thing in a “global context” [2]. This position taken to an extreme leads to problems: given two textually distinct URIs, is it possible they could identify the same thing? How can we judge if they identify the same thing?

The classic definition of *identity* is whether or not two objects are in fact, on some given level, the same. The classic formulation is Leibniz’s Law, which states if two objects have all their properties in common, then they are identical and so only one object [25]. With web pages, one can compare the representations byte-by-byte even if the URIs are different, and so we can say two mirrors of a web pages have the same content (in terms of text, images, or multimedia), but perhaps differ in terms of their content. Yet discovering whether the content of text or pictures are “about the same thing” is something computers are notoriously incapable of discovering. If the URI of a “thing” is just a redirection (perhaps an appended hash) or a URI that has no representation, this problem seems insurmountable without the introduction of explicit equivalence, as in *owl:sameAs*. Furthermore, one common criticism is that Leibniz’s Law can only be defined in a second order language between objects in the first order language [34]. A more radical criticism is that identity is only relative to some criterion of properties and so absolute identity does not exist [16].

Regardless, what criteria should the Web use (if any) for identity? How could these criteria be made explicit? A number of proposals exist that attempt to find particular schemes for distinguishing whether URIs are about “things” rather than web pages. The first proposal was by Pepper, coming out of the Topic Map community but applicable to a wider audience [33]. He makes a case that things (subject indicator) can be given URIs by putting some informative document there and demarcating it as a subject indicator by publishing it in a decentralized mechanism [33]. Ginsberg thinks of this as a sort of lexicon for the Web, complete with word sense disambiguation [17]. Over time, a common-ground of usable published subject indicators would emerge, imagined by Black as a “Semantic Wikipedia,” a project already being undertaken at Wikipedia [6]. A further concrete RDDL-like XHTML format called Expanded Web Proper Names was proposed as a representation for these URIs for things by having a human list other web-pages that were about the same thing [20]. In principle it was thought that machines would be able to automate when two URIs were about the same thing by simply comparing their links, and a shortcut using only a search engine and search terms was proposed to avoid explicitly listing all the web pages.

4. THE DIRECT THEORY OF REFERENCE

One can inspect the premise that one URI identifies one thing. When can anything directly refer to another thing? A theory where a symbol directly refers to an object is called a *direct theory of reference*. In its earliest incarnation, Russell states that “patches of sense-data” known through “direct acquaintance” allow one to ground the atoms of logical statements or create descriptions that can form the basis for names [35, 36]. This grounding can be done through the “acquaintance” theory of reference, and is usually signalled by pointing or the use of demonstratives in natural language, such as “That is red.”¹

How can we directly refer to things we are not acquainted with, such as imaginary or historical objects? Is not direct acquaintance incommunicable except through logical structure? Later work in natural language semantics thought that while most words did not directly refer to their object, proper names indeed did. This was further explored by Kripke in his *causal theory of naming*, in which a historical causal chain between a current speaker and past speaker allows the meaning of a name to be transmitted through time, with a name being given its original referent through a process Kripke calls “baptism,” which is just direct acquaintance with the referent and the action of naming the referent. In summary, *reference has primacy over meaning*, so the meaning of a statement can be given by the referents of the terms. For example, in RDF a triple would be a statement while its URIs would each be a term. From this viewpoint, the identity of a URI is established by fiat by the owner, and then communicated to others in a causal chain in the form of Semantic Web metadata and maybe publishing documents at the URI (or a redirection thereof). Strong proponents of what is generally considered “Good-Old Fashioned Artificial Intelligence” furthermore posit a sort of unified ontology of things, where that the world is a priori neatly divided up into things, properties, abstract classes, and relationships. Under this view, each of these can be given a distinct URI, although this viewpoint is controversial at best and thoroughly discredited at worse [39].

5. INTERPRETATION AND MEANING

In opposition to the direct theory of reference, the neo-Fregean tradition has as its slogan *the primacy of meaning over reference*, where the meaning of a statement and its possible referents are determined by the *sense* of a statement and formalized as the truth conditions of its interpretation. This sense is objective and sharable. The earliest proponent of this approach is Frege, who stated that two statements could be the same only if they shared the same sense [14]. This allows us to learn things about a referent by understanding its use in different senses. For example, the statements “Tim Berners-Lee is the Director of the W3C” and “Tim Berners-Lee was born in 1955” both have different senses but share one of the same referents. The referents of terms themselves makes no sense by themselves. Instead, what should taken into account is the patterns of their usage. This can be formalized by the use of statements that express sense, where each statement is composed of terms whose referent is determined by the meaning of the sense. Meaning

¹Further work has posited as basic particular individuals in space-time, and states that these are the basic ontology of the world from which we build abstractions [40].

is taken to be given by interpretations that define the truth conditions of the statement. This “logical grammar” is the bearer of meaning, and whatever satisfies the sense of the statement could be a referent. From some perspectives (such as resolving deixis), the meaning is a function from the context to interpretations [43], that in turn have truth values. To specify the matter in more detail, terms are in a language, and so any given statement in that language is given a well-defined syntax and a mapping to a domain through an interpretation. A statement without an interpretation is meaningless. A particular interpretation fixes the meaning of the statements in a (first-order) language. Therefore, statements can be considered identical if they have the same interpretation.

What is the domain of our interpretation? In general, one could imagine the interpretation stretches out from the syntax and into the world, and so selects objects in the world that fulfill the sense of the statement. This approach has been the one put forward in the analysis of natural language by situation theory [1]. However, one cannot reliably deal with such an informal notion of semantics in any consistent manner, especially in a manner useful to mathematics, logic, or by machines. Therefore, formal approaches to interpretation were developed that modelled the domain as a mathematical structure, and the study of these models is called model theory. In this reading, the interpretation of a statement in a language is a function that maps the terms and expressions in a syntax to a domain. This includes a mapping from the language’s relation symbols to specific relations on the domain. The use of model theory is considered *formal semantics* or *model-theoretic semantics*, which contrasts itself with the “informal semantics” or “real-world semantics” that in some manner connect the domain to the world. While the exact models vary (from rings in mathematics to typed lambda calculi for programming languages), once a model with well-known properties is chosen it becomes possible to prove useful properties about a language. Functions can then be defined as mappings from the domain to the truth values. However, how can a potentially infinite amount of syntactic statements be given an interpretation without iterating through an infinite number of individual syntactic statements? Tarski pioneered a solution by using inductive definitions to recursively specify the truth values of first-order predicate calculus [43]. In fact, a similar approach has been taken with the formal semantics of programming languages like XQuery.

This approach has been proven remarkably versatile. A formal semantics was given for a fragment of English by Montague [12], and in general the use of interpretation is considered standard practice in logic. After a first attempt at specifying RDF without formal semantics failed [24], RDF was given a formal semantics [21]. Further developments on the Semantic Web like OWL have kept in line with this tradition [32]. In the debate over the meaning of RDF, Pat Hayes held that the interpretation of URIs in a RDF statement were the only thing that could determine what the URIs identified (or denoted). Interpretations in general do not “constrain the meaning to only one thing,” although their constraints do put limits on the meaning of an expression in a language, as further explored by Hayes [22]. A possibly useful distinction for the discussion could be the distinction between *intension*, or all possible things a term

could describe, and *extension*, or all actual things that a term could describe. Berners-Lee brought up the point that he wanted the Semantic Web to function like mathematics, supposedly without any interpretation. In this case, there is a difference between interpretation in an informal sense and interpretation in a formal sense. In an informal sense mathematics is obviously about numbers, so the syntax of the number “2” is about the numeral two. This can be considered the difference between numbers (syntax) and numerals (semantics). There seems to be no natural or nomic connection between the syntax of “2” and the numeral two. The interpretation of numbers to numerals is arbitrary. “11” is three in binary and eleven in decimal. In fact, there was a historic debate over the origins of a model theory for arithmetic, the most trivial form of math. Peano arithmetic could be considered a model theory for arithmetic (and one that cannot be proven consistent) [42].

Is real-world ambiguity left by remaining at the level of formal semantics and their domains? Is this a good thing? Must we always rely on not only a formal interpretation? Do we need some full-fledged semantic agent such as a human to provide an interpretation of the formal semantics to the informal world? Or is that question even intellectually coherent? One objection could be that there is simply no alternative. If we wish for the Semantic Web to be machine understandable, we must proscribe only a formal semantics, and leave out any further discussion. This was the position adopted within the W3C as regards the “Social Meaning” of RDF debate. Yet a counter-objection could be that it is unclear if the notion of “truth” employed by formal semantics should be the defining primitive for meaning and identity on the Semantic Web. As regards statements in languages such as RDF that do not explicitly define false, there is a parallel to the basic idea of truth in the assertion of content itself and having behavior be rationally structured by the asserted content (such as triples) [11]. Finally, one response could be the Fregean notion of having meaning constrained not by reference but by the logical structure of the statement in particular and the language in general is not enough for the Web. Or is it? As noted by Wendy Hall, the Semantic Web is the “return of the link” and the power of a link “is that anything can link to anything” [41]. In this regard, the Semantic Web defines a richer structure of links with formal semantics that allow, through the process of using URIs as terms, a logical theory of meaning of identity and meaning to be implemented on the Web. In essence, it is the structure of the Web itself that determines meaning, not the URI. This is curiously reminiscent of Google’s earlier “page-ranking” algorithm. While the role of interpretation needs to be clarified and explicitly dealt with on the Semantic Web, it is clear that at least proposals that clarify the nature of interpretation on the Web are needed, as explained by Parsia and Patel-Schneider [31].

6. THE SOLUTION OF THE TAG

The Technical Architecture Group of the W3C, whose mission to clarify and formulate the guiding principles of web architecture, took on the Identity Crisis, calling it formally the *httpRange-14* problem. They phrased the Identity Crisis as the question “What is the range of the HTTP dereference function?” In their solution, they devised the class of “information resources” (since the original word “document” or the informal “web page” was too limiting). *Infor-*

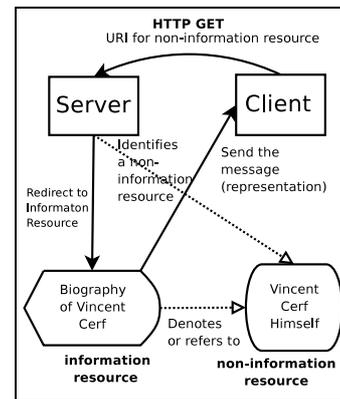


Figure 1: Picture of TAG resolution

mation resources are defined as anything whose “essential characteristics can be conveyed in a message” where these characteristics are “encoded, with varying degrees of fidelity, into a sequence of bits.” This leaves a wider class of things, such as “cars and dogs” as *non-information resources* “because their essence is not information” [23]. On this reading, URIs for information resources are like addresses, which imply that the address tells you exactly where to find a thing, while the URI of a non-information resource functions more like a name, and the possession of name by itself tells one little if nothing about the thing with the name.

The official resolution to *httpRange-14* is given as the following [13]:

1. If an *http* resource responds to a GET request with a 2xx response, then the resource identified by that URI is an information resource;
2. If an *http* resource responds to a GET request with a 303 (See Other) response, then the resource identified by that URI could be any resource;
3. If an *http* resource responds to a GET request with a 4xx (error) response, then the nature of the resource is unknown.

The diagram given as Figure 1 can clarify the TAG’s decision. First, a user sends a *http GET* statement to a URI for a non-information resource, like Vincent Cerf. Since the non-information resource is not an information resource, but it identifies Vincent Cerf himself, one does not want to host a representation there directly. Instead, one gets a 303 response code and redirect to another URI, this one an information resource also *about* Vincent Cerf that is then transmitted as a message (whose payload is a representation) to the client. This information resource could be quite a limited description. This information resource could be out of date, for it could tell you that Vincent Cerf was at one time a student at UCLA, it could not tell you that at the time this paper was written he worked at Google. The defining power of *some, but not all* resources then seem to be their representational power (text about Vincent Cerf) as opposed to things that are not representational on the Web (like a bunch of random bytes sent down the wire).

In this scheme, non-information resources should have distinct URIs but redirect to information resources. If it does

not redirect, it should give a distinct error code from the case where the URI definitely identified an information resource. In this scheme the distinction between a resource and a non-information resource is underdetermined. The user can not distinguish between the case of an information resource being redirected to another information resource and a non-information resource being redirected to an information resource. A further clarification of the pragmatic basis for this rationale is given using N3 by Dan Connolly [10]. It does seem as if something even more complex is going on behind the terminology of resources, and the paper by Gangemi and Presutti gives a complex Semantic Web ontology showing the various types of resources involved [15], which allows people to make precise statements about what type of resource they are using.

7. SOCIAL MEANING AND TAGGING

Matters of identity, reference, and meaning have already become incredibly multi-faceted. While these distinctions can provide deep insight, they can also lead to rather obscure technical solutions to practical problems, such as the key question “What does this web page mean?” Is there any way out besides the rather technical decisions of the TAG or the distinctions given by logic and philosophy? Noted philosopher of artificial intelligence Aaron Sloman stated that the problems of identity, reference, and meaning on the Web are fundamentally not different from those outside the Web, and so equally inscrutable. His prognosis of the situation is gloomy: “In such a system it is impossible to guarantee anything about either existence or uniqueness of reference of items referred to except in small, isolated subnets where all users share a deep common culture, including linguistic and terminological conventions” [38].

Sloman’s thesis has been proven wrong. On the Web there seems to be a surprising numbers of connections between these small and isolated “subnets,” and these groups suddenly find themselves to be large, fluid, and potentially in contact with many other groups. Since there is such a tremendous incentive for groups to exchange data over the Web, groups quickly establish conventions on how to exchange their data. In Web circles, this goes under the “social meaning” slogan. For example, an XML Schema and an associated human-readable document that explains the schema provide a “social contract” among all the parties that use that schema. In RDF this was a large debate about the relationship between the social meaning and the model-theoretic meaning on the Semantic Web, as given by a controversial and deleted section in the RDF specification on social meaning [30]. In the contest between social meaning and logic, logic won. While the W3C continued to not address the slippery concept of social meaning, social software took off.

The philosophical analysis of social meaning is generally given in terms of convention. In its classic game-theoretic formulation by David Lewis, a language is defined formally but established and used according to convention [26]. A convention is used by a community to solve some co-ordination problem, such as determining how to list dates so as to schedule meetings or record history, even if such a choice is arbitrary (the American convention of listing the month before the day as opposed to the European method of listing the day before the month). The identity and meaning of data on the Web could be viewed as a co-ordination problem.

A further general idea that falls out of a game-theoretic analysis is that people will in general use the minimum amount of convention to solve their co-ordination problem. This rule-of-thumb might explain the slowness of the Web community to embrace model-theoretic semantics.

Instead of using RDF to exchange data about meaning, “tagging” rose in popularity quickly with sites such as the bookmark-sharing *del.icio.us* and photo-sharing *flickr*. In essence, a group of “tags” (natural language words) are used to label data. Then users can search others people data to find more data with similar tags. An overlap of tags is given as a sign of similarity or identity, and two pieces of data that share the majority of the same tags are thought to be about the same thing. While this does avoid the cost of having a user type a few natural language words every time they wish to categorize data, millions of users have begun tagging on a regular basis. While it is unclear if such a technique can be subsumed by logic-based Semantic Web [18] or are a low-cost alternative to the Semantic Web [37], it seems that “tagging” is here to stay due to its large deployed user-base. Developers found it trivial to implement tagging-based solutions to data-sharing and identification problems, while many were probably unfortunately confused by the complex nature of Semantic Web standards, even though such standards like RDF are conceptually simple.

Although tagging is a surprisingly clever solution to the problems of identity and meaning, it has a number of problems. One contingent problem is that these tagging methodologies are generally employed by corporations attempting to make money out of the the data of users via some form of data-mining or targeted advertising, and so do not currently allow users to import and export their data in a service-neutral format. This causes a form of “data lock-in,” a Web 2.0 analogy of the fragmentation of the “browser wars.” A more fundamental problem is that while statistical natural language processing could provide some help, it seems like such a loose and unconstrained mechanism as tagging, while useful for simple things, makes it difficult for humans to express complex relationships. Unlike the Semantic Web, tagging is primarily for humans to interpret and nearly impossible for machines to interpret, and so the two initiatives may be orthogonal.

8. PRACTICAL SOLUTIONS

Is philosophical engineering possible on the Web? Can there be practical and theoretically sound solutions the problems of identity, reference, and meaning on the Web? There is clearly, as has been demonstrated, some degree of similarity between debates on the Web and previous debates in philosophy. The approach based based on a direct theory of reference as advocated by Tim Berners-Lee has some merit. It would be useful if we could know across contexts what URIs meant. Given their use as terms for both accessing representations and as the building blocks of the Semantic, it would make a certain amount of sense to try to provide some explanatory text in the representations for representations of non-information resources (and their redirects) [33].² This would solve the problem of not being able to tell an “ordinary” redirection from a redirection from a non-information resource, since a special sort of format was used by information resources at the end of the redirection. How-

²As done in <http://www.ihmc.us/users/phayes/PatHayes>

ever, any given application may be free to ignore these representations or not even access them [7]. While this practice would not break anything, it surely would not harm anything, and would allow people to use the “follow your nose” approach with regards to all resources, not just “information resources.” Furthermore, the URI in question could use content negotiation to return RDF, such as Minimum Self-Contained Graph, a formally-defined Concise-Bounded Description, or some other sort of graph [19]. While URIs can never identify one thing strictly on the Semantic Web without an interpretation, we can imagine a mutually beneficial relationship between the logic of the Semantic Web and the representations delivered by a URI that could help sort out the intended “real-world” domain of an interpretation.

One potent objection to this that the entire Identity Crisis will dissolve in time. While in 1996 this may have seemed unlikely, with the development of RFID tags for everyday objects (which could even be given URIs), the gulf between a thing (non-information resource) and its representation (information resource) on the Web may not be a gulf at all. Whether one makes an incorrect triple about you on the Semantic Web or manufactures a false biography on Wikipedia, representations on the Web are now part and parcel of reality. Even by just minting a URI for a thing, you have in a very real sense made that thing “on the Web.” More and more important facets of a thing, like personal identity, can be communicated by bits on the Web [8]. One day the question may be why can’t everything be primarily communicated over the Web?

Lastly, there is the problem of “social meaning.” The Web only works if people agree to abide by conventions to generally be truthful, to maintain their domain names, maintain the representations at their URIs, to not publish false statements, and so on. While in the face of the onslaught of spam this may seem Panglossian, it seems in general on the hypertext Web people are well-behaved, and exceptions like link farms are the minority. There are many conventions that people follow on the hypertext Web in order to make their information more valuable, leading to the exponential “network effect” of the Web [23]. So on the Semantic Web one can reasonably expect the same sort of behavior. Indeed, in order to form social meaning one needs to maximize the amount of information on how one wants a particular ontology or URI to be used, which is partially dependent on what the creator of the ontology believes it identifies. Instead of Masinter’s “power of readers over writers,” any reasonable reader will make an alliance with the writer in order to most effectively use the resource. However, people will in general only establish the minimum amount of convention needed to communicate for whatever purpose is at hand, and this makes the incentive for publishing information about things such as ontologies lower than tags. One bet might be that as the need to share data increases in an open and machine-readable manner, some sort of Semantic Web may naturally follow from social software. One could also assume that as soon as tags no longer fulfill the minimum requirements, the Semantic Web will rise to take their mantle.

Should we quietly be advocating a sort of neo-Davidsonian “semantic holism” for the Web [11]? The problem with holism is that it’s hard for machines, and even people, to deal with the whole complexity of any system, much less the Web. So any framework for meaning and identity must be easy-to-use, intuitive, and not require a background in

logic or the philosophy of language. As more people want to communicate – photos, bookmarks, anything – the benefits of deployment may eventually overcome the costs if the application developers and standard developers are skilled enough. Philosophical engineering is just taking previous work in logic and philosophy seriously enough to learn from it yet realizing the fundamental artificial nature of things like the Web that allow us to create and experiment with new ways of thinking, even if those fly in the face of philosophy. Practical solutions will doubtless require careful understanding of viewpoints from different backgrounds and mutual respect between disciplines. Yet as long as the all-too-human limitations of most users of the Web are kept in mind, such philosophical engineering could prove to be a crucial advantage in making the “Web of Meaning” a digital reality even if the solutions themselves remain open problems.

9. REFERENCES

- [1] J. Barwise and J. Perry. *Situations and Attitudes*. MIT Press, Boston, Massachusetts, 1983.
- [2] T. Berners-Lee. Message on www-tag@w3.org list, 2003. <http://lists.w3.org/Archives/Public/www-tag/2003Jul/0158.html>.
- [3] T. Berners-Lee. Message on www-tag@w3.org list, 2003. <http://lists.w3.org/Archives/Public/www-tag/2003Jul/0022.html>
- [4] T. Berners-Lee, R. Fielding, and L. Masinter. IETF RFC 3986 Uniform Resource Identifier (URI): Generic Syntax, January 2005. <http://www.ietf.org/rfc/rfc3986.txt>.
- [5] T. Berners-Lee, R. Fielding, and M. McCahill. IETF RFC 1738 Uniform Resource Locators (URL), 1994. <http://www.ietf.org/rfc/rfc1738.txt>.
- [6] J. Black. Creating a common ground for URI meaning using socially constructed web sites. In *Proceedings of Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/jblack.pdf>.
- [7] D. Booth. URIs and the myth of resource identity. In *Proceedings of Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/dbooth.pdf>.
- [8] H. Choi, S. Kruk, S. Grzonkowski, K. Stankiewicz, B. Davis, and J. Breslin. Trust models for community-aware identity management. In *Proceedings of Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/skruk2006.pdf>.
- [9] K. Clark. Identity crisis, 2002. <http://www.xml.com/pub/a/2002/09/11/deviant.html>.
- [10] D. Connolly. A pragmatic theory of reference for the Web. In *Proceedings of Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/dconnolly2006.pdf>.
- [11] D. Davidson. *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford, United Kingdom, 1984.
- [12] D. Dowty, R. Wall, and S. Peters. *Introduction to Montague Semantics*. Kluwer Academic, 1981.
- [13] R. Fielding. Message on www-tag@w3.org list, 2005. <http://lists.w3.org/Archives/Public/www-tag/2005Jun/0039.html>.

- [14] G. Frege. *Über sinn und bedeutung*. *Zeitschrift für Philosophie und philosophische Kritik*, (100):25–50, 1892.
- [15] A. Gangemi and V. Presutti. The bourne identity of a web resource. In *Proceedings of Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/skruk2006.pdf>.
- [16] P. Geach. Ontological relativity and relative identity. In *Logic and Ontology*. New York University Press, New York City, USA, 1973.
- [17] A. Ginsberg. The big schema of things. In *Proceedings of Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/aginsberg2006.pdf>.
- [18] T. Gruber. Folksonomy of ontology: A mash-up of apples and oranges. In *Proceedings of First on-Line conference on Metadata and Semantics Research (MTSR)*, 2005. <http://tomgruber.org/writing/mtrs05-ontology-of-folksonomy.htm>.
- [19] G. Tummarello, C. Morbidoni, P. Puliti, and F. Piazza. Signing individual fragments of an RDF graph. In *Proceedings of World Wide Web Conference Poster Track (WWW)*, 2005.
- [20] H. Halpin and H. Thompson. Web Proper Names: Naming referents on the web. In *Proceedings of The Semantic Computing Initiative Workshop at the World Wide Web Conference*, 2005.
- [21] P. Hayes. RDF semantics, W3C recommendation, 2004. <http://www.w3.org/TR/2004/REC-rdf-nt-20040210/>.
- [22] P. Hayes. In defense of ambiguity. In *Proceedings of Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/phayes.pdf>.
- [23] I. Jacobs and N. Walsh. Architecture of the World Wide Web. W3C Recommendation, 2004. <http://www.w3.org/TR/webarch/>.
- [24] O. Lassila and R. Swick. RDF Model and Syntax. Deprecated W3C Recommendation, 1999. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>.
- [25] G. Leibniz. *Discourse on Metaphysics and the Monadology*. Open Court, 1908.
- [26] D. Lewis. *Convention: A Philosophical Study*. Harvard University Press, 1969.
- [27] L. Masinter. Draft DURI and TDB URI namespaces based on dated URIs, 2004. <http://larry.masinter.net/duri.html>.
- [28] A. Miles. Working around the identity crisis, 2005. <http://esw.w3.org/topic/SkosDev/IdentityCrisis>.
- [29] R. Moats. IETF RFC 2141 URN syntax, 1997. <http://www.ietf.org/rfc/rfc2141.txt>.
- [30] B. Parsia. Message on www-rdf-comments@w3.org list, 2003. <http://lists.w3.org/Archives/Public/www-rdf-comments/2003JanMar/0366.html>.
- [31] B. Parsia and P. F. Patel-Schneider. Meaning and the semantic web. In *Proceedings of Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/bparsia2006.pdf>.
- [32] P. Patel-Schneider, P. Hayes, and I. Horrocks. OWL Web Ontology Language: Semantics and abstract syntax W3C Recommendation, 2004. <http://www.w3.org/TR/owl-semantics/>.
- [33] S. Pepper. The case for published subject. In *Proceedings of Identity, Reference, and the Web Workshop at the WWW Conference*, 2006. <http://www.ibiblio.org/hhalpin/irw2006/spepper.pdf>.
- [34] W. Quine. *From a Logical Point of View*. Routledge, New York City, USA, 1963.
- [35] B. Russell. On denoting. *Mind*, (14):479–493, 1905.
- [36] B. Russell. Knowledge by acquaintance, knowledge by description. *Proceedings of the Aristotelian Society*, (11):197–218, 1911.
- [37] C. Shirky. Ontology is overrated, 2005. <http://www.shirky.com/writings/ontologyoverrated.html>.
- [38] A. Sloman. Private communication, 2005.
- [39] B. C. Smith. The Owl and the Electric Encyclopedia. *Artificial Intelligence*, 47:251–288, 1991.
- [40] P. Strawson. *Individuals: An essay in descriptive metaphysics*. Routledge, 1959.
- [41] J. H. Tim Berners-Lee and O. Lassila. The Semantic Web. *Scientific American*, 2001.
- [42] J. van Heijenoort. *From Frege to Gödel: A sourcebook in mathematical logic*. Harvard University Press, 1967.
- [43] A. Tarski. The Semantical Concept of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research*, 4: 341–75, 1944.