International Journal of Multidisciplinary Research and Development
Online ISSN: 2349-4182 Print ISSN: 2349-5979
www.allsubjectjournal.com
Volume 3; Issue 1; January 2016; Page No. 420-422

Significance of reference on the web and semantic web

Kulbir Kumar

MCA, Maharshi Dayanand University, Rohtak, Haryana, India

Abstract
The Web is a universal information space for naming and accessing information via URIs. However, the classical philosophical problems of meaning and reference that have been the source of debate within the philosophy of language return when the Web is given as the foundation for a knowledge representation with the Semantic Web.
Debates on the Semantic Web about the meaning and referential status of a URI are explored as analogues to debates about the meaning and reference of names in the philosophy of language. Three main positions are inspected: the logical position, as exemplified by the descriptivist theory of reference, the direct reference position. These positions show that debates within the philosophy of language are alive and well on the Web, and so in the philosophy of computer science.

Keywords: world wide web, URI, semantic web

Introduction

The World Wide Web, often shortened to the 'Web,' consists of a space of names called Uniform Resource Identifiers (URIs), a unique identifier. Examples of URIs include http://www.google.com and http://www.wikipedia.org. URIs like http://www.wikipedia.com/wiki/Philosophy is a separate URI from http://www.wikipedia.org, although they both share a domain name, the root wikipedia.org.

URIs for usually used for accessing hypertext web-pages. In this regard, the Web can be considered a virtual space for protocols that actually ship the bits that compose the hypertext from the server to the client with the web browser. The functioning of the Web can be illustrated by the following example. An agent, such as a web browser, wishes to access some information about a resource such as the Eiffel Tower in Paris. (Grice *et al*. 2012) [6].

A resource is defined in the broadest of terms to be "anything that might be identified by a URI". A person can access the representation of a resource, such as its web-page, using its URI, like http://www.tour-eiffel.fr/. While this is all uncontroversial, the novelty of the Web lays in extending URIs beyond web-pages, as a telephone number can be given a URI such as tel:+1-816-555-1212. These URIs for things outside web-pages remained mostly a theoretical possibility until the advent of the Semantic Web, an ambitious project to the use of the Web itself as the infrastructure for a global knowledge representation system. The essential bet of the Semantic Web, a next generation of the Web "in which information is given well-defined meaning, better enabling computers and people to work in cooperation," is that decentralized agents will come to an agreement on using the same URI to name a thing, including things that aren't accessible on the Web, like people, places, and abstract concepts.

Suddenly, the question of what a 'resource' is, and how different agents can tell which resources a URI identify, transformed from a purely theoretical question into one with practical consequences for the further development of the Web. (Kripke *et al*. 2012) [8].

Having information in "machine readable forms" requires a knowledge representation language that has some sort of relatively content-neutral syntax.

Review of related literature

The parallel to knowledge representation in artificial intelligence is striking, as it also sought to find a syntax for encoding human intelligence. The second point, of "allowing links," means that the basic model of the Semantic Web will be a reflection of the Web itself: the Semantic Web is constituted by connecting resources by links. (Beckett *et al*. 2011) [1].

The Semantic Web is then easily construed as a descendant of semantic networks from symbolic artificial intelligence, where nodes are resources and arcs are links. Semantic networks refer declaratively to things in the world, but uses 'natural-language-like' labels on its nodes and edges. Yet semantic networks fell out of favor because of their use of ambiguous natural language terms to identify their nodes and arcs, which became a problem when semantic networks were transported between domains and different users, a problem that would be fatal in the decentralized and multi-lingual environment of the Web. (Berners *et al*. 2014) [2].

When researchers attempted to communicate or combine their knowledge representation schemes, no-one really knew what the natural language description meant except the author. As powerfully explained by Woods, the IS-A link in semantic networks was interpreted in at least three different ways, which could represent both sub-classing, instantiation, close similarity, and more. This led to an assault on semantic networks by champions of first-order logic like Hayes, who believed that by providing a formal semantics that defined 'meaning', first-order logic at least allowed knowledge representations to be transportable across domains, and that many alternative knowledge representations could be re-expressed in first order-logic. (Carnap *et al*. 1928) [3].

Under the aegis of the Web standards body the World Wide Web Consortium (W3C), the Resource Description Framework (RDF) was created as the first knowledge representation

language for the Semantic Web. It was clearly influenced by work in AI on semantic networks, and this should come as no surprise. (Dummett *et al*. 2013) [4].

Importantly, every component of the knowledge representation language is considered a resource, and thus can be given a URI, replacing what is believed to be ambiguous words in semantic networks with URIs. Since statements in a knowledge representation language are usually about the world outside the Web, this means that the Semantic Web crucially depends on the rather strange fact that URIs can refer to things outside the Web. (Frege *et al*. 1892) [5].

Research Work

On the Semantic Web, a whole new cluster of questions, dubbed the Identity Crisis, emerges. Can a URI for the Eiffel Tower be used to refer to the Eiffel Tower in Paris itself? If one just re-uses a URI for a web-page of the Eiffel Tower, then one risks the URI being ambiguous between the Eiffel Tower itself and a particular representation of the Eiffel Tower.

If one gives the Eiffel Tower qua Eiffel Tower its own URI, should that URI allow access to any information, such as a hypertext web-page? This cluster of questions has been dubbed the Identity Crisis of the Semantic Web. As regards any theory of meaning for URIs, in the realm of official Web standards, the jury is still out. In the specification of RDF, Hayes notes that "exactly what is considered to be the 'meaning' of an assertion in RDF or RDF(S) in some broad sense may depend on many factors, including social conventions, comments in natural language" so unfortunately "much of this meaning will be inaccessible to machine processing" such that a "a full analysis of meaning" is "a large research topic".

Unsurprisingly, the reason there is no standardized way to determine the meaning of a URI is because, instead of a single clear answer, there is a conceptual quagmire dominated by two positions.

Adherents of this position hold that the referent of a URI is ambiguous, as many different things can satisfy whatever model is given by the interpretation of some sets of sentences using the URI. This position is generally held by logicians, who claim that the Semantic Web is entirely distinct from the hypertext Web.

Direct reference on the web

The causal theory of reference is naturally close to the direct reference position. The causal theory of reference is uncontroversial, for in database schemas, what a term refers to is a matter best left to the expert designer of the database.

There is also an element of Grice in the direct theory of reference, for the intended interpretation and perhaps even purpose of the owner is the one that really matters to study, not any publicly accessible particular Web representation. However, ultimately study has far more in common with the causal theory of reference, since although the URI owner's intention determines the referent, after the minting of the new URI for the resource, the intended interpretation is somehow never supposed to vary. (Luntley *et al*. 2011) [9].

To apply the causal theory of reference as to URIs, baptism is given by the registration of the domain names, which gives a domain name and legally binding set of IP addresses, such as example.org, a legally binding owner. Of course, the natural question then would be if this study practice can then be extended to entire URIs such as http://www.example.org/Eiffel? For most domain names a specific policy given by the owner could set the allowed referents for the creation of URIs that involve the domain name in question, perhaps as embodied in some software system.

One could imagine several variations on this theme, from the URIs being controlled indirectly by systems-programmers or even outsourced to the general public in the form of a user generated URI registry with a single top-level domain. Regardless of the details, the referent of a URI is established by fiat by the owner(s), and then optionally can be communicated to others in a causal chain in the form of publishing web-page accessible from the URI or by creating Semantic Web statements about the URI.

Sense and reference Redux

The Semantic Web has still not experienced the tremendous growth of the hypertext Web, and the primary reason appears to be this impasse at the Identity Crisis. For the first few years of its existence (2011-2012), in general the arguments of Hayes prevailed, and the URIs used in RDF graphs did not access any web pages. However, in this phase of its existence, the Semantic Web did not progress beyond yet another little used knowledge representation language. (Needham *et al*. 2012) [10].

In the last few years, the Semantic Web has experienced phenomenal growth under the term 'Linked Data,' as position has had more acceptance and users have started deploying RDF using actual URIs.

This growth of estimated billions triples, including large scale projects by biomedical community and in government data in using the Semantic Web. Yet that is far from true; what is apparent from any analysis of the Semantic Web is that there appear to be too many URIs for some things, while no URIs for other things.

As differing users export their data to the Web in a decentralized manner, new URIs are always minted, and so running the risk of fracturing the Semantic Web into isolated 'semantic' islands instead of becoming a unified web, as the same URIs are not re-used.

The critical missing element of the Semantic Web is some mechanism that allows users to come to agreement on URIs and then share and re-use them, a problem ignored both by the logical and direct reference positions.

Search Engines

Search engines work via analysis of existing web pages, breaking them down into terms, and then mapping those terms and their frequencies in a given webpage into a large index. So, each URI can be thought of as collection of terms in this search engine index. As the collection of term frequencies gathered into this index grows, ranging over larger and larger sources of data like the Web, it approximates a sample of human language use, as has been shown by studies in computational linguistics. (Oren *et al*. 2011) [11].

Users of a search engine then enter certain terms, the search query, which are then mapped via certain algorithms against the index. This results in an unordered list of possibly relevant URIs, which for an index that covers the entire Web range from thousands to millions of URIs. In turn these URIs are then ranked and ordered using an algorithm such as Google's famous Page Ranking algorithm, possibly with user feedback.

To explicate how user feedback works, search engines usually keep track on what URIs are actually clicked on by users. This

stream of clicks by multiple users can then be stored in a "query log," and then this query log can then be used to improve the discovery and ranking of URIs by search engines.

By inspecting which terms lead to which URIs for multiple users, a set of terms that best describes a URI for users can be discovered. In this way, typing in terms into a search engine can be thought of as a sort of language-game, with success in this game being judged in terms of whether or not a given user can, using a particular Web search engine, discover a relevant URI. The terms themselves may be ambiguous, but it does not matter if a relevant URI is discovered. While the user may not be aware of it, as it appears that searching the Web using a search engine is a private experience, it is in fact mediated by a vast amount of web-pages stored in the search engine's index and the behavior of previous search users. In this way, the objective sense of a URI can be considered the search terms that can be used by multiple users to find a particular URI.

What a URI means is precisely the set of search terms that leads multiple users to discover the URI in the context of satisfying a particular information need. It should also be noted that this position does not overtly contradict the logical position, as the logical position radically undermines the referent(s) on the Web. Instead, one can imagine that the public language position allows a URI to be grounded in user-behavior using search terms, but that this meaning can be supplemented by logical inference.

Conclusion

The Semantic Web has yet to be widely deployed, and it could be precisely due to the persistent of it's the problems of the philosophy of language regarding meaning and reference that it attempts to build upon. We have argued that the debates over the meaning and reference of URIs can be seen as a return of the debate between the causal and descriptivist theories of reference in the philosophy language, with this time the subject being URIs rather than natural language names.

In this way, it has been shown that in the course of the practice of computer science, even in such a new under theorized and undisciplined frontier like the Web, robustly philosophical problems arise. By stumbling on the difficult philosophical problem of reference and meaning, it appears that the success of the Semantic Web, one of the most ambitious projects of knowledge representation so far, has been stymied.

Searching the large hypertext Web leads precisely to a 'statistical semantic web,' where the meaning of URIs is given by the activity of users. In this way, the bet of using URIs as a universal naming scheme for things can just as easily be tied to statistical methods from information retrieval as it can to logic-based knowledge representations.

In this way, the philosophy of computer science can even provoke further practical engineering work on the Web, and it is precisely the success of computational engineering that wins debates on the Web, rather than pure argumentation. Currently information retrieval engines like Google far outweigh the Semantic Web in terms of usage, which give us a clue as regards which philosophical position may be correct. Is there any way the massive amount of search terms and URIs harvested by search terms be used to boot-strap the Semantic Web?

One bet would be that when users want to actually find URIs data on the Semantic Web or even find URIs to re-use for labeling their data on the Semantic Web, they will have to employ search-engines over the Semantic Web with the same natural language driven searches currently employed by Web search engines over the hypertext Web. However, instead of hypertext web-pages being indexed, the index will range over decentralized knowledge representation stored in RDF.

References

1. Beckett. Turtle - Terse RDF Triple Language. Member submission, W3C, 2011.
2. Berners. World Wide Web Future Directions. Plenary Talk. http://www.w3.org/Talks/WWW94Tim/ (Last accessed on Oct. 5th 2014).
3. Carnap. The Logical Structure of the World. University of California Press, Berkeley, California, USA. Republished in 1928.
4. Dummett. What is a Theory of Meaning? In The Seas of Language, Oxford University Press, Oxford, United Kingdom. Originally published in Truth and Meaning: Essays in Semantics in, 2013, 1-33.
5. Frege. Uber Sinn und Bedeutung. Zeitshrift fur Philosophie and philosophie Kritic Reprinted in The Philosophical Writings of Gottlieb 1892; 100:25-50.
6. Grice. Meaning. The Philosophical Review 2012; 66:377–388.
7. Keller. Using the web to obtain frequencies for unseen bigrams. Computational Linguistics 2013; 29(3):459-484.
8. Kripke. Naming and Necessity. Harvard University Press, Cambridge, Massachusetts, USA, 2012
9. Luntley. Contemporary Philosophy of Thought. Blackwell, London, United Kingdom, 2011.
10. Needham. A method for using computers in information classification. In Proceedings of the IFIP Congress, Vienna, Austria, 2012, 284-287.
11. Oren. Sindice.com: A document-oriented lookup index for open linked data. International Journal of Metadata, Semantics, and Ontologies 2011; 3(1):37-52.