Social Meaning on the Web: From Wittgenstein To Search Engines

Harry Halpin, H.Halpin@ed.ac.uk
School of Informatics, University of Edinburgh

One could hypothesize that the essential bet of the Web is that in a decentralized information space multiple agents can share the meaning of a URI. On the Semantic Web, does a URI get its meaning from its owner, or from the formal interpretation of statements that use it? Disagreement on this point crystallized as a debate between Berners-Lee and Pat Hayes, in which Berners-Lee put forward the hypothesis that a URI “identifies one thing” while Hayes responded that URIs are always ambiguous, as the interpretation given by the formal semantics of the Semantic Web assigns the same URI to different individuals in different models. This debate can be considered a return to a long-standing debate in the philosophy of language on the meaning of names, where URIs are thought of as just names on the Web. The position of Hayes can be considered analogous to Russell’s descriptive theory of names, where a name is considered functionally a set of descriptions that can be applied to different individuals. One refutation of the descriptive theory of names was given in the causal theory of names by Kripke, where the meaning of a proper name is given by an act of “baptism” and then causally transmitted through time, so that a name identifies a unique individual over all possible worlds. This position seems similar to that of Berners-Lee, with the idea of the purchase of domain name and “minting” of a URI being analogous to the act of “baptism.”

Yet there is a straightforward argument against the causal theory of names. In order for the identity of name to be transferred, there must be a successful communicative act using language, and a measure of ambiguity is always present in language. In Web terms, in order for the owner of URI to allow his or her intended meaning to be shared, the owner should host both human and machine-readable documents that allow its meaning to be shared across the widest-possible number of users. However, even with these public representations being accessible from a URI, there will always be ambiguity, as these representations are parasitic upon natural language and formal languages. In contrast to the ordinary hypertext Web, where a URI operationally identifies whatever hypertext the URI allows a browser to access, on the Semantic Web URIs are also given to non-accessible things. While techniques that can differentiate between these types of URIs are possible, they also fall prey to ambiguity. One solution, championed by initiatives like OKKAM, is to create a centralized “giant dictionary of meaning” that maps URIs to their referents. This type of simple solution falls victim to Quine’s famous argument of radical translation.

However, there has long been a third position, the public language theory of meaning, where names are fundamentally given their meaning by social and linguistic agreement, where some level of ambiguity is embraced. This position was first articulated by “late” Wittgenstein in a repudiation of his earlier strongly logicist viewpoint. If one reads Dummett carefully, this view is ultimately com-
compatible with Frege, wherein the meaning of any expression, including URIs, are grounded out not just in their formal truth values, but in their “sense.” The notion of “sense” can be reconstructed to be construed in terms of the socially-grounded norms that are necessary in order to grasp the use of a name across a language. Thus, the infamous slogan “meaning is use” arises. For a name in a public language to be successful, it must be adopted by a community, who in turn comes unto some sort of minimal, and so often ambiguous and unconscious, agreement on the use of a name. This social agreement ultimately grounds out the meaning of a URI in addition to any formal semantic theory of reference, for the sense of a name determines its referents. By restricting the Semantic Web to formal theories of reference and neglecting the more social notion of sense, the Semantic Web has been left with an impoverished theory of meaning. The central point for the evolution of the Web is to not view the Semantic Web only as a formal logic, but as a new form of language in of itself, where logic serves as a useful conjunct to social meaning.

The most revolutionary concept of Wittgenstein was the notion of a “forms of life” such that to “imagine a language is to imagine a form of life.” In other words, what is the primary activity that constitutes life on the Web? The answer is obvious: the use of search engines. In order for the Semantic Web to succeed, the meaning of a URI should have its formal meaning supplemented by its social meaning. We propose a methodology for having normal natural language searches in hypertext search engines to be supplemented by information from various URIs gathered from Linked Data, grounding these Semantic Web URIs in the aggregate search behavior of users. However, the problem is that the natural language terms are very ambiguous, and simply too much data is retrieved from the Semantic Web. By observing the behavior of the user in selecting certain hypertext web-pages, the precise information the user is interested in can often be detected via techniques from natural language processing. Then the hypertext web-pages can be used to approximate the additional social meaning of the query terms, and this can be used in combination with machine-learning techniques to both disambiguate Semantic Web URIs while maintaining the socially necessary ambiguity, putting these Semantic Web URIs in a “virtuous cycle” with the hypertext Web. An implementation based on real users, a search query-log, from a major search engine, and Linked Data will be discussed.

If the Web is to be considered a first-rate subject of science, then issues of social meaning can not be thrown by the wayside as somehow being unscientific. As the social sciences have shown in the large, and our experiment with search-engines and the Semantic Web has shown in the small, these issues of social meaning can be studied in depth, and the collective behavior of users on the Web provides historically unparallelled opportunities to study large-scale social phenomenon. At the same point, the Web is not a wholly new subject, and arguments about names and meaning that have emerged in philosophy can now longer be avoided, and far from being purely academic concerns. These concerns over meaning are now of critical importance for practical engineering. As put by Berners-Lee, “We are not analyzing a world, we are building it. We are not experimental philosophers, we are philosophical engineers.”