FILOZOFIA I NAUKA
Studia filozoficzne i interdyscyplinarne
Tom 2, 2014

Maciej Sopek[1]
Instytut Filozofii
Uniwersytet Warszawski

THE SEMANTIC WEB—INTRODUCING MEANING TO THE INTERNET

STRESZCZENIE

Celem niniejszej pracy jest analiza wybranych aspektów filozoficznych sieci World Wide Web, ze szczególnym naciskiem położonym na technologie Sieci Semantycznej. Jest to umotywowane przekonaniem, iż Sieć dostarcza zagadnień interesujących z punktu widzenia filozofii, a jednocześnie pozostaje tematem, który rzadko jest analizowany z perspektywy filozoficznej. Praca zawiera krótki zarys rozwoju sieci World Wide Web i podkreśla konceptualną zmianę, jaką technologie semantyczne wprowadzają do dziedziny; zawiera także przykłady ich praktycznego zastosowania. Do pojęć, które pojawiają się w pracy należą znaczenie, referencja i kwestia ontologicznego statusu przedmiotów, z których składa się Sieć.

Słowa kluczowe: internet, sieć, semantyka, desygnator, znaczenie, odniesienie

1. INTRODUCTION

The aim of the paper is to explore a number of selected themes pertaining to the modern World Wide Web, with a special emphasis on the concept of Semantic Web. In its present incarnation, the Semantic Web is considered the second major change in the history of the network, after the arrival of its present incarnation known as Web 2.0, associated with the emergence of social media. However, I will argue that in contrast to Web 2.0, the Semantic Web marks the first technical and—what is important for the problem considered here—the first philosophical shift in the way we approach the Web. I attempt to prove that the arrival of the Semantic Web is the first conceptual change in the structure of the Web since its inception.

In the paper I emphasize both theoretical and technical issues pertaining to the storing of natural language data in the digital form. Then I sketch the conceptual nature of the Web, the Internet, and lay down the basic idea of

[1] Adres Autora: riven213@gmail.com, IF UW, ul. Krakowskie Przedmieście 3, 00-047 Warszawa.

the Semantic Web; I try to explain why the latter may be considered the first true revolution in our approach to the Web. Most importantly, however, I indicate parallels between one of the central notions of the Web and Saul Kripke's concept of rigid designators. On this basis, I formulate a metaphysical question on the nature of the Web. Finally, I explain what the terms "meaning" and "reference" mean in the context of the Semantic Web, and argue that it is this very technology that enables us to talk about them. In order to bring out the conceptual difference between the Semantic Web and the traditional approach, I will invoke John Searle's Chinese Room argument.

2. THE MAGNITUDE OF THE WEB AND THE PROBLEM IT POSES

The Web is a vast network of data; its amount is impossible to be stored or organized by any other means available to humans. It grows and transforms every second, spreading across millions of computers around the world in a nearly entirely decentralized fashion. It contains an immense amount of meaningful information that has to be structured in order to be accessible and comprehensible. Traditional methods of computer data organization are, in principle, based on documents, or equivalently, files. This means that the stored content, although meaningful for human users, remains semantically neutral for computers. This implies that any structuring of a meaningful, natural language content has to be done by men. The volume of this data is vast, so the structuring is a difficult, cumbersome and often non-effective process. Besides, automatic structuring can only reach a certain level, for instance, when one searches through documents by means of the Internet search engines.

2.1. Google Searle

Although in recent months several new features have been implemented in major search services (the most significant is Google; see below), the fundamental principle of operation of these services remains the same: the search engine software scans the content of the Web for crude strings of characters. This process provides results and, on the surface level, is very quick and precise, when it is taken into account how vast the number of items to be analyzed and returned is. Search engines use extremely sophisticated and mathematically ingenious algorithms that provide precise and relevant results.

There is, however, one important drawback—the Web, with all its abundance of information, lacks semantics. The Google search bot will look for sequences of letters and words, but, in essence, this is it. All that is obtained is a list of files, semantically neutral for the software. No serious organiza-

tion or interpretation of the textual content by the computers is possible. Moreover, documents have unique and extremely diverse formal structures (different page structures, different programming languages, different software and hardware platforms etc.), and the natural language content understandable for men stored in computers cannot be effectively categorized or extracted by means of automation. The most that can be done, as mentioned before, is a detection and comparison of strings of characters.

A useful philosophical parallel to this state of affairs is Searle's famous "Chinese Room" thought experiment. Searle conceived his seminal idea as an argument against the possibility of creating a strong artificial intelligence (AI), that is, such an AI that not only reflects human mental phenomena, but actually experiences them, just as sentient beings do.

Essentially, Searle's argument is against functionalism in the philosophy of mind. Functionalism regards mental states only in respect to their functions, that is, in respect to the way they operate and react to the input and output data. Since only the function is taken into consideration, functionalism holds that mental states can be realized by different systems, not only the human neural system. Thus, theoretically, mental states could also be realized by the computer software, insofar as the functional role of the software is similar to that of human mental states.

Searle's thought experiment can be sketched as follows. Let us imagine that a computer program has been created that behaves as if it actually would understood Chinese. It can receive input provided in Chinese and provide answers in a coherent and grammatically sound manner. It is logically (and technically) possible that such a piece of the software could be created, and perhaps even pass the Turing test. Therefore, in the functionalist perspective, it could be said that as the machine behaves as if it understood Chinese, it in fact it does understand it. What Searle questions here is the very notion of understanding: Does the machine actually understand the meanings of the words it uses? Or maybe it is a mere simulation, despite the fact that the output provided could not be distinguished from utterances produced by a human being? In order to prove that that the latter is true, Searle imagines a person that does not know Chinese at all and is isolated from the world in a room where he only receives strings of Chinese characters. The person conducts operations on these characters according to a set of instructions. What he essentially does is identical with the workings of the software; the only difference lies in the fact that it is done manually rather than automatically. The output provided by the person in the Chinese Room will also be perfectly sound and will appear as if a Chinese-speaking person produced it. However, it is clear that this man has no actual understanding of Chinese, he merely follows rudimentary instructions.

The argument is directed against strong AI, but in particular it attacks the thesis that the operations on syntax alone imply a veridical understanding.

"The Chinese Room argument [...] (does not) purport to show that machines cannot think"—Searle says that brains are machines, and brains think. "It is directed at the view that formal computations on symbols can produce thought."[2] The Semantic Web research can be considered as pertaining to the broader area of artificial intelligence study. However, it has never been claimed that introducing semantics to the Web would bring it closer to Strong AI. In any sense, it is still, in essence, procedure-following, and no claims of conscious understanding should be made.

The Chinese Room argument, however, may be used to emphasize the difference between the traditional, syntax-based approach to content on the Web, and the way it is dealt with within the Semantic Web. What search engines can accomplish today is what Searle's operator does, if he looks across volumes of incomprehensible texts—he just compares strings with other strings, without reasoning.

2.2. The numbers

Never in history has there been so much information gathered in a codified, systematic manner. Moreover, the Web expands and changes every second. There is a lot of meaningful information there, and the need for a method of automated analysis becomes apparent. A method is required that would increase efficiency of knowledge access by orders of magnitude, or rather, offer an entirely new approach to the problem—the Semantic Web makes such a promise.

3. THE WEB BEFORE SEMANTICS

I now present a brief sketch of the development of the World Wide Web prior to the arrival of the concept of the Semantic Web, in order to provide a context for understanding of the importance of semantic technologies and the philosophical novelty that they bring to the Web.

3.1. Internet versus the Web

The World Wide Web (or simply: the Web) should not be confused with the Internet. The terms are often used interchangeably, but their actual meanings are quite different.

The Internet is a worldwide network of computers interconnected by means of smaller, local networks. It is a technical foundation for other services, including e-mail, file sharing platforms and the World Wide Web it-

[2] D. Cole, *The Chinese Room Argument*, "The Stanford Encyclopedia of Philosophy", Edward N. Zalta (ed.), Winter 2009 Edition, http://plato.stanford.edu/archives/win2009/entries/chinese-room/.

self—that is, the system of linked documents containing text, image or multimedia contents.

3.2. Origins and Principles

The theoretical principles and technological foundations of the Web were developed at the end of the 1980s by Sir Tim Berners-Lee, a British computer scientist working at the time at CERN. Together with a team of fellow researchers, he established a way to merge existing Internet technologies, and introduced several techniques for what would later come to form the foundation of the Web as we know it today. These techniques include:

a) HTML — the HyperText[3] Markup Language

A computer language used for describing content of web pages. A given page is described by means of syntax which is then interpreted by a web browser. What we see on a website is a result of the browser's interpretation of the underlying code that defines appearance, structure and content of the site.

b) HTTP — the HyperText Transfer Protocol

A protocol of communication on the basis of which computers access each other within the World Wide Web. The essential method of navigation between pages is based on the so-called hyperlinks that bind different sites by means of addresses (e.g. http://www.google.com). An essential feature that differentiates HTTP from other existing protocols of communication is that hyperlinks are unidirectional, meaning that a given link may address another resource without any action required on the part of the target resource. This greatly facilitated the development of the Web since links could be attributed freely to any given site; the drawback of this solution is that links may expire or lead to incorrect addresses, thus resulting in the infamous "404" errors where the user reaches a dead end. Unlike other then-existing solutions of this kind, the HTTP protocol has always been free to use by anyone without any licensing fees, which is an aspect that has been instrumental to the widespread development of the Web.

c) URI — the Uniform Resource Identifier

From the philosophical point of view the URI is probably the most interesting of the three backbones of the Web. Tim Berners-Lee's, Roy Fielding's and Larry Masinter's document, a part of the Internet standards specifica-

[3] "Linking of related information by electronic connections in order to allow a user easy access between them. (..) Hypertext links create a branching or network structure that permits direct, unmediated jumps to related information." Merriam-Webster Dictionary, available at www.merriam-webster.com/, obtained on 2012-09-23.

tion, states that: “A Uniform Resource Identifier (URI) is a compact sequence of characters that identifies an abstract or physical resource.”[4]

The URIs are the most basic building blocks of the Web—they identify any online item (referred to most often as a resource) with a unique name. A URI can therefore refer to an e-mail account of a person, an image, a file, a whole website etc. The essential features of URIs are enclosed in the three constituting terms:

Uniform — in that "different types of resource identifiers can be used in the same context”;

Resource — “used in a general sense for whatever might be identified by a URI”,

Identifier — “embodies the information required to distinguish what is being identified from all other things within its scope of identification.”[5]

This brings to mind a familiar philosophical concept of proper name, in particular, of the Kripkean rigid designator. This notion is one of the most important in contemporary philosophy of language, with implications extending over many other fields of philosophical analysis. Coined by the Saul Kripke in his lectures, later published in *Naming and Necessity*, the term stands for a name that refers to a given entity in all possible worlds in which the entity exists. The relation of the rigid designator and the designated object must hold in all these possible worlds—if it holds at all. This is a standpoint rival to the descriptivist theory of reference (one of the main proponents of which was Bertrand Russell) that claims that names are either synonymous to descriptions, or are based on clusters of descriptions exclusive to a given entity.

Kripke objects to this approach since, as he claims, descriptions refer to contingent properties; if names are to be connected with objects by means of descriptions, any alteration in properties puts into question the reference of a given name and object. He uses the example of Aristotle: in the descriptivist perspective, if Aristotle died before conceiving any of his works (using possible worlds semantics: in a possible world in which he existed but his life ended prematurely), he would not be Aristotle—which is a suspicious thought.

By applying rigid designators we omit this difficulty. Since, in Kripke’s view, all proper names are rigid designators, Aristotle will remain Aristotle regardless of the possible changes in his traits or biography. Similarly, the term “the 37th President of the United State” may denote different people in different possible worlds—in some of them, where the 1968 election results turned out differently, this could be Hubert Humphrey. However, the rigid designator “Richard Nixon” always refers to Richard Nixon, regardless of his

[4] T. Berners-Lee, R. Fielding, U.C. Irvine, L. Masinter, *Uniform Resource Identifier (URI): Generic Syntax*, Network Working Group , January 2005.

[5] Ibidem.

political career or other possible factors that might have affected the infinity of possible worlds in which Nixon existed.

My claim is that URIs play a role in the Web similar to that of rigid designators. They do not carry descriptions of objects, and thus do not fit within the descriptivist theory of names. Just as a rigid designator always designates the same entity, regardless of its contingent features in a given world, URIs designate particular objects (or resources) irrespective of their current content or state. If my claim is sensible, this will spark especially interesting implications when considered in the light of the Semantic Web, in which not only items of the Internet, but any thing can be designated by URIs.

4. The Structure of the Web

The first incarnation of the Web, which is considered to have lasted until around 2004 (inaccurately so as I argue later on), was thus basically a web of interconnected documents, operating on the principles established by Tim Berners-Lee and his colleagues in the late 1980s. The late 1990s saw the explosion of Internet-based business, with giants such as Amazon, reaching new heights of financial success and inventing entirely new forms of business, ultimately proving that the Internet has become an integral part of the modern world. The enthusiasm was so great that it sparked the phenomenon known as the "dot-com bubble", which describes the immense expansion of online ventures. The so-called "dot-coms" were often vastly overrated by investors and in multiple cases ended with spectacular failure. Nevertheless, the online era has begun for good.

4.1. Web 2.0, Social Media

Web 2.0 is the term popularly used to describe the current era of the Internet. It basically refers to the shift from traditional, author-to-user model, towards user-generated content. This chiefly includes social media portals such as MySpace, Facebook or Twitter, but also encompasses other forms of online activity—such as wiki sites (the most famous is Wikipedia), blogs, video sharing portals (like YouTube) etc. However, this did not introduce or require any conceptual alterations. Web 2.0 is a purely social phenomenon, one of the most important developments of the modern age, and as such is a subject of tremendous importance to social philosophy, yet it does not bring about any substantial novelties pertaining to the way the Web functions as a structure.

4.2. The Semantic Web

The term was coined in the seminal paper entitled *The Semantic Web* by Tim Berners-Lee, James Hendler and Ora Lassila, published in 2001 in

"Scientific American". The basic technical innovation behind the Semantic Web is conceptually simple: it introduces computer—readable metadata to web pages that describes their natural language content. These tags enable automated mechanisms to organize entities appearing within the content, and make possible the conduct of inferential operations on them. This is achieved by means of the Resource Description Framework, (RDF).

4.3. Triples

"RDF provides a consistent way to represent data so that information from multiple sources can be brought together and treated as if they came from a single source."[6]

RDF is a model of the metadata description used to describe statements about resources (that is, entities). Its most essential expression is known as a triple of subject-predicate-object, each part constituted by an individual URI. The subject denotes a given resource, while the predicate describes its properties and defines the relationship of the subject and the object. For example, the sentence: "Tom is the son of John." expressed as an RDF triple consists of the subject Tom, the predicate is his son of and object—John. Triples themselves can become subjects and objects by means of an operation called reification. For example, the above sentence can be reified into a single object and a triple in a following way: " 'Tom is the son of John' is a true sentence'."

Thus, complex structures may be created out of simple entities denoted by URIs. This is the essential manner in which objects are incorporated into the semantic framework.

4.4. Ontologies

The term "ontology" is similar in its meaning to its philosophical counterpart. It is a scheme that serves as a template of concepts and principles that govern relationships between objects in the Semantic Web environment. It defines what kind of entities can be present in a given framework, and how these can be processed. The most popular language used to establish ontologies is the Web Ontology Language (OWL). The language is associated with and expressed in RDF, and is recommended by the World Wide Web Consortium W3C, the founder of which is Tim Berners-Lee.

Ontology can pertain to any set of entities. It can describe terms used in an encyclopedia, a collection of books in a bookstore, a biological research project, or a commercial offer of products. Theoretically, any portion of hu-

[6] D. Allemang, J. Hendler, *Semantic Web for the Working Ontologist. Effective Modeling in RDFS and OWL*, 1 edition, Morgan Kaufmann, Burlington 2008, p. 89.

man knowledge could be described this way. Even though a broad scale application of the Semantic Web technologies is yet to be seen, individual entities such as businesses, medical organizations or scientific research groups have already begun incorporating semantic ontologies into their work. One particularly interesting example of the application of a semantic ontology can be found in the field of bioinformatics. Results of this application were published in the article *Using Semantic Web Rules to Reason on an Ontology of Pseudogenes*.

> Recent years have seen the development of a wide range of biomedical ontologies. Notable among these is Sequence Ontology (SO) which offers a rich hierarchy of terms and relationships that can be used to annotate genomic data. Well-designed formal ontologies allow data to be reasoned upon in a consistent and logically sound way and can lead to the discovery of new relationships.[7]

The application of the semantic ontology not only allowed to infer new relationships in the already gathered data, but also enabled to publish the resulting knowledge base for public access in an interactive form.[8]

5. THE METAPHYSICS OF THE WEB

5.1. Things online

As mentioned earlier, URIs resemble Kripkean rigid designators in that they refer to particular resources regardless of their contingent properties or environment. They already pinpoint locations online, but the ambition is to be able to designate any possible object, be it a Web address, a person, a chemical compound, a book, etc.

This aspect of URIs, as described in the cited sentences: "The triples of RDF form webs of information about related things. Because RDF uses URIs to encode this information in a document, the URIs ensure that concepts are not just words in a document but are tied to a unique definition that everyone can find on the Web."[9] may serve as a fertile ground for philosophical investigation into the nature of things present online. I will venture to say that the Web, like software, is in some sense of dualistic character, one aspect of it being firmly grounded in the purely physical sphere (servers, networking infrastructure etc.), while the other—in the sphere of content in which the users operate—is a domain of virtuality that cannot be efficiently

---

[7] M.E. Holford, E. Khurana, K-H. Cheung, M. Gerstein, *Using Semantic Web Rules to Reason on an Ontology of Pseudogenes*, "Oxford Journal of Biometrics" 26(12), 2010, p. i71–i78.

[8] Available at: http://bioinformatics.oxfordjournals.org/content/26/12/i71.full

[9] T. Berners-Lee, J. Hendler, O. Lassila, *The Semantic Web*, "Scientific American", 17th May, 2001.

reduced to the underlying hardware. I am not attempting to suggest any kind of Cartesian dualism in the ontological sense, but for practical purposes, the Web is irreducible to physicality.

The crucial difference between the Web and software is that unlike traditional computer programs that let the user work with a limited set of functionalities, the Web is an open space of interaction for millions of conscious agents. A new content is generated on a constant basis, therefore it is in some sense a kind of "reality", or, rather, a theater of interaction for human users. The stage may be virtual, but the actions are very real. Hence the question is: if we decide to treat URIs as rigid designators, can we talk of a process of actualization of pure ideas? Using URIs, practically anything can be considered an object—be it an online resource or a mere thought. Within the space of the Web, these abstract notions obtain a kind of existence, and in the light of the Semantic Web, they become objects endowed with very concrete attributes which enable considering them. Alexandre Monnin, a researcher whose current endeavor is to establish an independent branch of philosophy of the Web, proposes a term "artifactualization":

> [...] a process where concepts become "embodied" in materiality—with lasting consequences. While such a process clearly predates the Web, we can from our present moment see within a single human lifetime the increasing speed at which it is taking place, and through which technical categories (often rooted inphilosophical ones) are becoming increasingly dominant over their previously unquestioned "natural" and "logical" counterparts.[10]

Moreover, not only is the content created consciously and intentionally—the very foundations that govern this stage of action can be modified—but also the Semantic Web is an example of such a principle-altering phenomenon. Note that even considering the vast conceptual novelty it brings about, it still operates within the already existing Web. It is not separate or alien to it, and it does not require any new technology to use by its participants. The Web preserves its numerical identity despite fundamental changes in its structure. Thus perhaps its essence is not to be sought in the way it organizes data, but in the fact that it is an interactive space of incredible flexibility.

5.2. Meaning and Reference—the Paradigm Shift

Hopefully, thanks to the above description, the difference between the original Web of documents and the Semantic Web becomes clear: the new Web makes use of meaning. The original paradigm—on the basis of which the overwhelming majority of the Web still functions today—is based on

[10] A. Monnin, H. Haplin, *Toward a Philosophy of the Web. Foundations and Open Problems*, 2012 "Metaphilosophy" 43(4), 2012, 361–379, p. 366.

documents. Conceptually speaking, it is a rather simple net of interconnected files; the connections and relations (or structures of connections) between them are always defined manually by humans. This means that they are chaotic and contingent, allowing no systematic—let alone automatic—analysis that would generate new knowledge.

The meaning-based paradigm that the Semantic Web hopes to promote is on the opposite side of the spectrum. The connections between items are not declared, but rather inferred from their properties on the basis of the assumed logical principles defined in ontologies. There can be talk of actual reference—between URIs and their referents. No such thing takes place within the traditional Web of documents that, from a mechanistic point of view, is a giant set of semantically neutral pieces whose main computable value can be reduced to similarities in sequences of characters.

Meaning in the Semantic Web may be conceived as resulting from these inferences. A single, separate resource does not carry any value for the computer, however, when it is bound by logical relations to other resources—that are, in turn, bound even further—there appears space for action, that is, for obtaining new knowledge. Obviously this knowledge is not *new* in a purely philosophical sense; after all, what takes place is basic logical operation. However, given the enormous amounts of data that semantic methods are to deal with, this still provides results that humans would be incapable of producing—one example of this benefit is the bioinformatics research project mentioned in the earlier section.

6. CONCLUSION: THE SEMANTIC WEB – THE REAL WEB 2.0?

The Semantic Web used to be classified as Web 3.0; I suppose that this is a category mistake. As it has been mentioned here, the common opinion holds that we are currently in the age of the so-called Web 2.0, defined by the dominating role of social media. However, the social revolution did not invoke any conceptual changes in the structure of the Web. Even the technical development in terms of website creation is more of a refinement of existing ideas rather than anything entirely new. Therefore I believe that the term "Web 2.0" should be reserved for the arrival of the Semantic Web, when—or if—it becomes the dominant paradigm for establishing the content in the Internet. In a sociological sense, the Web has been in fact transformed around 2004. In terms of "metaphysics of the Web," Web 1.0 never really ended and we are still stuck with the ideas established over 20 years ago.

The Semantic Web points to a road ahead and it yet remains to be seen if the ambitious idea of the Web of meaning takes hold beyond isolated, specialized implementations. If it does, it may constitute a tremendous leap in the way we access and make use of knowledge in general, something that

would provide benefits that are difficult to grasp today. However, even if it fails as a practical revolution, it reveals a fascinating new area for philosophical research.

ABSTRACT

The aim of the present article is to explore selected philosophical aspects of the World Wide Web, with particular emphasis on the Semantic Web technologies. This is motivated by the conviction that the Web as such provides grounds for philosophical investigation, and at the same time remains a subject that is rarely approached from philosophical perspective. The article contains a brief sketch of the development of the World Wide Web and emphasizes the conceptual novelty that the Semantic technologies introduce to the area; as well as examples of their practical application. Some of the notions approached in the article include meaning, reference and the question of the ontological status of objects being part of the Web.

Keywords: internet, web, semantics, designators, meaning, reference