Collective Construction of a Concept

Subjects: Information Science & Library Science

Contributor: Luís Machado

Collective construction of a concept can be understood as the negotiation of a meaning represented in the form of written definitions in a collaborative way. Although it can be understood as a reductionist view, these verbal externalizations will be necessary for a group of people to collectively construct a concept.

concepts collective construction Wikipedia meanings definitions social context semantic web

1. Introduction

There is no consensus in the study of concepts, what is their nature (mental representations or abstract entities?), or their constitution (bundles of features, or they embody mental theories?). Different approaches have resulted in distinct theories of which stand out: the Classical Theory, the Prototype Theory, the Neoclassical Theory, the Theory-Theory, and Conceptual Atomism. All, according to Margolis and Laurence, present difficulties in explaining certain aspects involving concepts, among which, issues related to analyticity, compositionality or ignorance and error[1]. For these authors, concepts are mental representations and a theory with the necessary explanatory potential is only possible by "admits different types of conceptual structure while tying them together by maintaining that concepts have atomic cores".[2]

For the study of the collective construction of a concept as defined here, there are benefits in adopting a pluralist epistemological position associating the pragmatic positioning of Hjørland, based on the Theory of Theory, with Dahlberg's "theory of analytical concept of reference" within a neoclassical epistemic position. Dahlberg does not consider the influence of the social context in the formation of concepts, like Hjørland does, but takes it into account when it comes to their organization and representation.[3] In this perspective, Dahlberg's theory of concept approaches the position of Hjørland with respect to the representation of concepts, so that the theory provides a reference for the characterization, categorization and decomposition of concepts.[4] As for the social context, since, according to Hjørland, concepts "should be identified by studying discourses rather than by studying individual users or a priori principles",[5] collective open-spaces such as Wikipedia have the appropriate characteristics to do this type study.

Wikipedia can be described as one of the "abstract social machines" advocated by Berners-Lee and Fischetti, in processes enabled by the World Wide Web where people do the creative work and the machines do the administrative counterparts.[6] The massive number of Wikipedia's collaborators (more than 32 million registered

users) contributes to this being the most comprehensive project in the scope of Digital Humanities.[7] Considering that Wikipedia presents itself as a free encyclopedia where any Internet user can edit, it can be considered as a space where the collective bargaining of meanings occurs and it is therefore a privileged place for the diachronic study for a community's understanding of a concept in particular.

2. Case Study: the Concept "Semantic Web"

The history page of the Semantic Web entry in the English version of Wikipedia was mapped to identify the semantic changes made to the support statement of the respective definition, presented in the introduction section of the different versions of this article (from December 2001 to December 2017). During the analysis, it was used whenever deemed necessary the discussion page in order to obtain contextual information to help clarify the definitions presented.

There were 129 changes in the introductory part of Wikipedia´s entry titled Semantic Web, in which 26 definitions with some degree of semantic difference were identified. In Table 1 we present the definitions grouped within each category, according to the respective generic term.

Table 1. Generic terms and respective content units retrieved from the identified definitions.

rf.GT	Generic Terms	Content Units
category 1. Main definitions		
(a)	vision	(#01) is Tim Berners-Lee’s vision of the future of the WWW; (#02) is a vision of the future of the WWW.
(b)	project	(#03) is a current project; (#04) is a project underway; (#05) is a project.
(c)	evolution	(#06) is an evolution of the current Web; (#08) is an evolution of the WWW.
(d)	framework	(#09) is a loosely defined and evolving framework of WWW based technologies; (#10) is a loosely defined and evolving framework.
(e)	initiatives	(#12) is a set of loosely-defined and evolving initiatives.
(f)	extension	(#13) is an evolving extension of the WWW.

(g)	development	(#14) is an evolving development of the WWW.
(h)	methods and technologies	(#15) it describes methods and technologies; (#18) is a group of methods and technologies.
(i)	web of data	(#20) is a "web of data"; (#22) is a "man-made woven web of data".
(j)	roadmap	(#23) is the roadmap of a "man-made woven web of data".
(k)	movement	(#24) is a collaborative movement.
(l)	extension	(#26) is an extension of the Web.
category 2. Complementary definitions—sub-category 2.1. Assigned		
(m)	manifestation	(#07) is a manifestation of Tim Berners-Lee's vision of the Web.
(n)	(something)	(#11) it derives from Tim Berners-Lee's vision of the WWW.
(o)	term	(#16) is a term coined by Tim Berners-Lee.
(p)	web of data	(#19) Tim Berners-Lee defined the Semantic Web as "a web of data"; (#25) the term was coined by Tim Berners-Lee for a web of data.
category 2. Complementary definitions—sub-category 2.2. Common use		
(q)	model and technologies	(#17) it is mainly used to describe the W3C's model and technologies.

(r)	formats and technologies	(#21) is often used to refer to the formats and technologies that enable it.

In a generic terms list there is an exception for the use of the compound term "web of data", which was considered necessary because of the syncronogenematic nature of the element "of data"[8] and its necessity for the meaning intended with the term in question.

The option for two categories, "main definition" and "complementary definition", was necessary since in some versions of the Semantic Web entry two or three definitions coexisted. In these cases, the analysis of their statements revealed two patterns: in one, an assignment of the definition to Berners-Lee (subcategory 2.1.), and, on the other, a relation to the common usage of the term (subcategory 2.2.). Units #01 and #02 (rf.GT (a)) were considered within the category 1 as the main definition, despite their close relationship with Berners-Lee, given that in these initial versions of the article they are the only definitions. The temporal distribution of the groupings, by generic term (see Table 1), is presented in Figure 1.

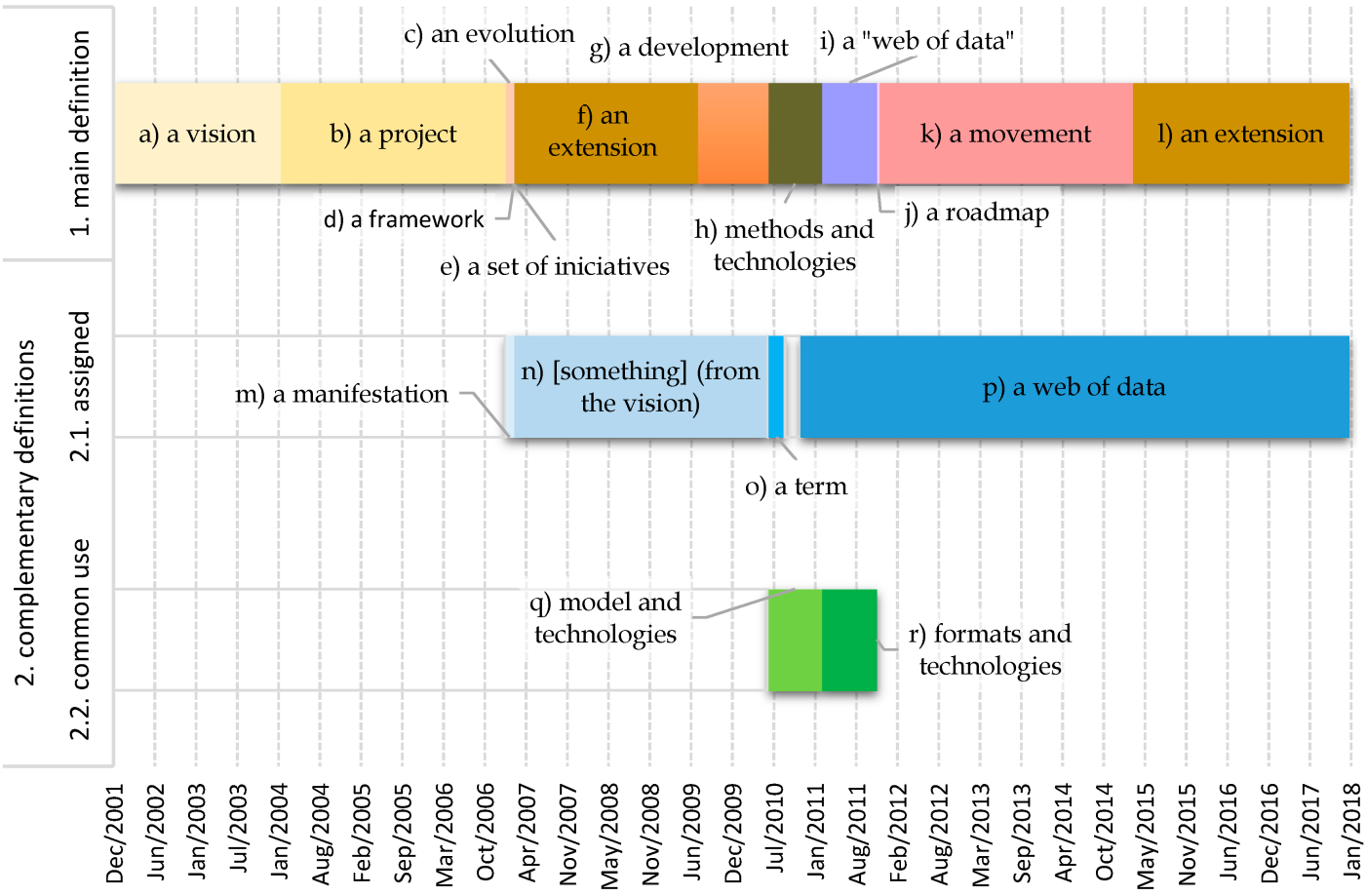

Figure 1. Temporal distribution of the definitions (group by the respective generic terms).

The diachronic visualization presents an enlightening overview of the evolution of the semantic web concept in Wikipedia´s context. Given the extended time span (16 years) it is natural that definitions with little longevity are

less noticeable, as is the case with those referred to with (d), (e) and (j), whose duration is less than 10 days.

The analysis of the definitions revealed conceptual variations due to the introduction or alteration of the specific characteristics attributed to the generic term (see Table 2).

Table 2. Specific characteristics of generic terms.

rf.GT	Specifiers pre-GT	Generic Terms	Specifiers post-GT
(a)		(#01; #02) vision	(#01) of Berners-Lee of the future of the WWW; (#02) of the future of the WWW
(m); (n)		(#07) manifestation; (#11) it (derives from)	(#07) of Berners-Lee's vision of the future of the WWW; (#11) Berners-Lee's vision of the WWW
(b)	(#03) a current	(#03; #04; #05) project	(#04) underway
(c)		(#06; #08) evolution	(#06) of the current WWW; (#08) of the WWW
(d)	(#09; #10) a loosely defined and evolving	(#09; #10) framework	(#09) of WWW based technologies
(e)	(#12) a loosely defined and evolving set of	(#12) initiatives	
(f); (g); (l)	(#13; #14) an evolving	(#13; #26) extension; (#14) development	(#13; #14; #26) of the WWW
(h)	(#18) a group of	(#15; #18) methods and technologies	

(q); (r)		(#17) model and technologies; (#21) formats and technologies	(#17) proposed by W3C; (#21) that enable it [the SW]
(i)	(#22) a man-made woven	(#20; #22) web of data	
(j)		(#23) roadmap	(#23) of a man-made woven web of data
(k)	(#24) a collaborative	(#24) movement	

In some cases, the conceptual drift only occurs in the qualifiers, as is the case in group (b) of Table 2, where a single generic term, "project", includes three variations: first the project is objectivized with the qualifier "current" (#03), then with the term "underway" (#04), and finally it loses its adjectivation (#05).

In an inverse situation are the supplements that serve as a link between the different generic terms, as occurs in groups (c) to (g) of Table 2. The variation between the five terms becomes gradual when framed by the specifiers that are maintained or little altered, such as pertaining to WWW membership in these groups. Another example is visible in the change from the term "evolution" (#08) to "framework" (#09 and #10), where the former becomes part of the specifying characteristics of the second, an "evolving framework". This specifier, "evolving", accompanies the following three terms: "set of initiatives" (#12), "extension" (#13), and "development" (#12).

The comparison between the definitions of the semantic web concept, identified on Wikipedia, with those resulting from the analysis of the same concept based on the publications of Berners-Lee and 3WC,[9] was also carried out from a diachronic perspective. For the sake of clarity and representatives, we have opted to restrict the analysis to variations with a duration of more than 90 days, and not to include the two complementary definitions of common use (subcategory 2.2), since they would only add "noise" to this comparison. Applying these criteria result in eight main definitions and two complementary definitions (Figure 2).

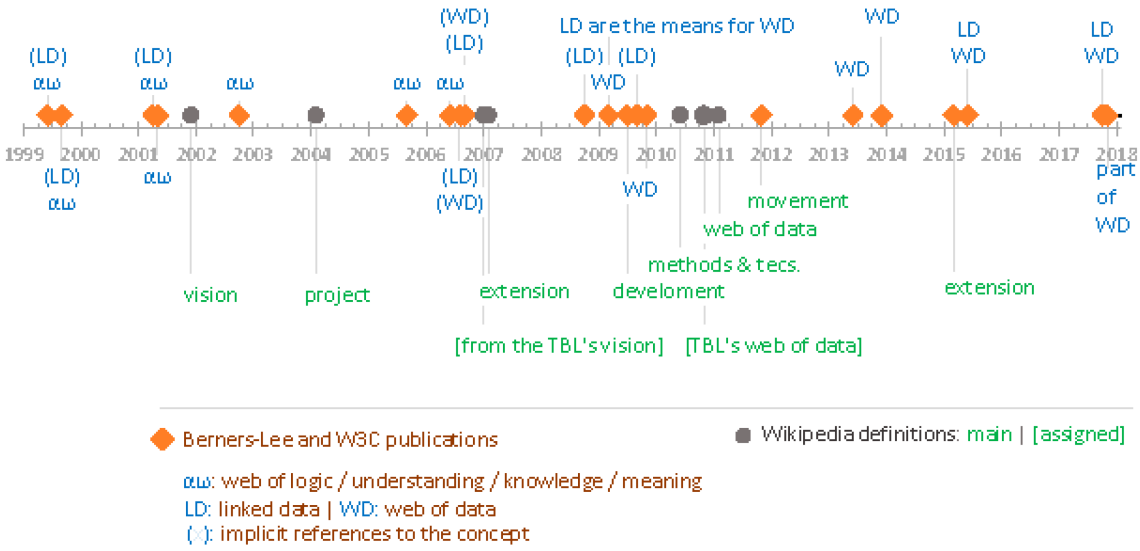

Figure 2. Comparative temporal distribution between the definitions of semantic web from the two sources (Wikipedia and publications of Berners-Lee/ World Wide Web Consortium (W3C)).

From the observation of the temporal distribution, presented in Figure 2, two situations stand out, the first being related to the variations of the main definition with the generic term "vision" and "project", to coincide with the period in which publications with definitions that have terms like "logic", "understanding", "knowledge", or "meaning" (αω). The second situation concerns to the term "web of data", both in the main definition (in 2011) and in the complementary (in 2010), after this term is used explicitly (in 2009) in the analyzed Berners-Lee / W3C publications. Another potential relation is to verify if we consider the descriptions present in the Berners-Lee / W3C publications previously analyzed. For this matter, we repeat in Table 3 the content units of the cited study.[9]

Table 3. Groups and respective content units considered in the analyses of the publications of Berners-Lee and W3C

Groups	The Semantic Web is…
1. Descriptions that include "semantics"	a. The Web of understanding (7 June 1999); A universal web of knowledge (26 April 2001).
	b. An extension of the current Web in which information is given well-defined meaning (May 2001; October 2002); A web of logic (13 September 2005); A Web of actionable information derived from data through a semantic theory for interpreting the symbols (June 2006).
	c. A web of data with meaning (22 September 1999).

2. Linked Data	a. The Web of linked data (11 August 2006; 2015).
	b. A new data model to support the linking of data from many different models (7 June 1999); The web of connections between different forms of data (22 September 1999); A world of trusted information shared along collaborating groups of users (26 April 2001); An open web of inter-referring resources (11 August 2006); A type of extension of the Web to extend the Web to cover linked data (September 2006); A network of data on the Web (October 2008); The world of linked data (22 October 2009).
	c. Linked Data provides the means (March 2009).
3. Web of Data	a. A Web of Data (March 2009; 12 November 2009; 27 June 2013; 11 December 2013; 2015; 11 October 2017).
	b. One extension of the Web moving from text documents to data resources (11 August 2006); Is intended to function in the context of the relational model of data (September 2006).
	c. Part of the Web of Data (2016).

Referring to the Table 3, we can note that the term "extension" is used define the semantic web in two moments. Initially, it appears in two documents (of 2001 and 2002, subgroup 1.b.) very close to the beginning of the article in Wikipedia (December, 2002) and then (August and September, 2006; subgroups 3.b. and 2.b., respectively). The same term was used in the Wikipedia definitions in February, 2007 ("an evolving extension"), very close, though, to the second occurrence in the publications.

Unlike the definition of the semantic web as the "Web of Data", verified in the two sources, we did not find in the definitions of Wikipedia mentions that could be understood as the "Web of Linked Data", as it appears explicitly in two publications in Table 3, for 2006 and 2015 (sub-group 2.a.).

3. Discussion

The concept of the semantic web, presented on Wikipedia, shows an evolution that seems to oscillate between the search for a more concrete definition and the use of terms accessible to the common layman. Explanations

regarding the need to adapt the vocabulary to the non-specialist user by the editors can be found in both the descriptions of the changes (available in the article history) and in the discussion page.

Although the evolution of this concept presents points of contact and similarity between the two scopes (Wikipedia and the publications of Berners-Lee / W3C), the differences detected go beyond that imposed by the type of support, *continuum* in the first (once it is continuously open since all contributions can be reversed at any time) and, in the second, composed by *discrete units* (which are closed to changes at the time of publication). This leads to the conclusion that the search for adaptation to non-specialist readers by Wikipedia editors marks a significant difference between the two scopes. The adaptation referred to above may also give rise to the need for additional definitions, since it is thus possible to present in an integrated form more than one point of view concerning the same concept.

The search for a clearer and more specific definition is, we believe, responsible for the elimination of dubious expressions or buzzwords. In some changes made to the article, this attempt to promote clarification is explicitly stated, as in 11-21-2011, where the segment "that facilitates machines to understand the semantics, or meaning, of information on the World Wide Web" was taken from the definition and classified as "obscure". Also, in the change from the generic term "project" to "framework", as well as in the change from the latter to "extension", we can identify this double intention of clarification and adapting to non-specialist readers. This belief is reinforced by the debate around this last change (from "framework" to "extension"), shown in the respective discussion page, where it is possible to find, in the editors' debate, the search for the balance between the personal understandings of the given concepts and the adequacy to the general readers. The discussion we are referring to is not a unique example of negotiation processes for the terms to be used in the definitions, detected on the discussion page. On the other hand, there were no occurrences in the history of the Semantic Web entry, of the repeated and systematic alternation between versions, known as "edit wars",[10] as we can see in several entries of Wikipedia.

In fact, regarding the authorship of the changes to the definition presented in the Semantic Web article, they are characterized by debate and diversity. In the 26 definitions registered, there are 16 different users registered and four unregistered. In addition, users with more than one definition make their contribution in the same edition and with definitions that fall into different categories; one main and one attributed and/or common use. The only exception, reported on 20-02-2007, occurred in the context of what could have originated an "edit war" between two editors (Dreftymac and Cygri). However, the debate was transferred to the appropriate channel, the discussion page, where the predominant position of the two editors was the negotiation of a consensus between the two different visions. A negotiation, where the perception of the multiple meanings that the semantic web concept can take for different people is present: "we deal with a much-hyped term that is used to mean quite different things by different people" (Cygri, 21 February 2007).[11]

Despite this, the last definition ("is an extension of the WWW") has remained stable for almost three years, in parallel with the definition attributed to Berners-Lee: "The term was coined by Tim Berners-Lee for web of data that can be processed by machines". The scope of this term, "extension", may contribute to the stability of the definition, but does not contribute to a specification of the concept that it intends to define. From this point of view,

the semantic web concept can be seen as being in a "pseudo-concept" phase which, according to Vygotsky,[12] is characterized by an intermediate stage between the general or complex notions and the fully developed concept.

Another issue that may create some kind of restraint in changing the definition is the link (academic and professional) of the author of the last definition to the semantic web. However, we are not giving to this influence too much weight because, in Wesch's words: "Authorized information is not beyond discussion on Wikipedia, information is authorized through discussion".[13]

References

1. Eric Margolis; Stephen Laurence. Concepts: Core readings; MIT Press: Cambridge, 1999; pp. 4-81.
2. Stephen Laurence; Eric Margolis; Concepts. In The Blackwell Guide to the Philosophy of Mind; Warfield, T. A., Stich, S. P., Eds.; Blackwell's: Malden, 2003; pp. 190–213.
3. Aline Arboit; O processo de (re) construção da teoria do conceito no domínio de Organização do Conhecimento: uma visão dialógica. *Scire* **2012**, *2(18)*, 129-134, http://hdl.handle.net/11449/73995.
4. Ingetraut Dahlberg; A referent-oriented, analytical concept theory for INTERCONCEPT. *Knowledge Organization* **1978**, *5*, 142-151, 10.5771/0943-7444-1978-3-142.
5. Birger Hjørland; Concept theory. *Journal of the American Society for Information Science and Technology* **2009**, *60*, 1519-1536, 10.1002/asi.21082.
6. Tim Berners-Lee; Mark Fischetti. Weaving the Web: : The original design and ultimate destiny of the World Wide Web; Harper Collins: New York, 1999; pp. 157-175.
7. Pepe Flores; Is Wikipedia the largest-ever digital humanities project? Exploring an emerging relationship Available online: https://blog.wikimedia.org/2016/08/17/wikipedia-largest-digital-humanities-project/ (accessed on Oct 25, 2018).
8. Wolfgang Stock; Concepts and semantic relations in information science. *Journal of the American Society for Information Science and Technology* **2010**, *61*, 1951-1969, 10.1002/asi.21382.
9. Luís Machado; Renato Souza; Maria Simões; Semantic Web or Web of Data? A Diachronic Study (1999 to 2017) of the Publications of Tim Berners-Lee and the World Wide Web Consortium. *Journal of the Association for Information Science and Technology* **2019**, *70*, 701-714, 10.1002/asi.24111.
10. Fernanda Viégas; Martin Wattenberg; Kushal Dave; Studying cooperation and conflict between authors with history flow visualizations. *Proceedings of the 2004 Conference on Human Factors in Computing Systems* **2004**, *6*, 575-582, 10.1145/985692.985765.

11. Talk: Semantic Web/Archive 1 . Wikipedia. Retrieved 2020-11-11
12. Lev Vygotsky. A Construção do Pensamento e da Linguagem; : , 2001; Martins Fontes: São Paulo, Brazil, 2001; pp. 229.
13. Michael Wesch. From Knowledgeable to Knowledge-able. In Hacking the Academy: New Approaches to Scholarship and Teaching from Digital Humanities; Cohen, D.J., Scheinfeldt, T., Eds.; University of Michigan Press: Ann Arbor, MI, USA, 2013; pp. 69-77.