

Article

# The Evolution of the Concept of Semantic Web in the Context of Wikipedia: An Exploratory Approach to Study the Collective Conceptualization in a Digital Collaborative Environment

Luís Miguel Machado <sup>1,\*</sup>, Maria Manuel Borges <sup>1</sup> and Renato Rocha Souza <sup>2</sup>

<sup>1</sup> Department of Philosophy, Communication and Information, University of Coimbra, 3004-530 Coimbra, Portugal; mmb@fl.uc.pt

<sup>2</sup> School of Applied Mathematics “EMAp”, Getúlio Vargas Foundation, Rio de Janeiro RJ 22250-900, Brazil; renato.souza@fgv.br

\* Correspondence: luismachado@fe.uc.pt; Tel.: +351-964-152-983

Received: 28 July 2018; Accepted: 29 October 2018; Published: 5 November 2018



**Abstract:** Wikipedia, as a “social machine”, is a privileged place to observe the collective construction of concepts without central control. Based on Dahlberg’s theory of concept, and anchored in the pragmatism of Hjørland—in which the concepts are socially negotiated meanings—the evolution of the concept of semantic web (SW) was analyzed in the English version of Wikipedia. An exploratory, descriptive, and qualitative study was designed and we identified 26 different definitions (between 12 July 2001 and 31 December 2017), of which eight are of particular relevance for their duration, with the latter being the two recorded at the end of the analyzed period. According to them, SW: “is an extension of the web” and “is a Web of Data”; the latter, used as a complementary definition, links to Berners-Lee’s publications. In Wikipedia, the evolution of the SW concept appears to be based on the search for the use of non-technical vocabulary and the control of authority carried out by the debate. As a space for collective bargaining of meanings, the Wikipedia study may bring relevant contributions to a community’s understanding of a particular concept and how it evolves over time.

**Keywords:** semantic web; Wikipedia; conceptual evolution; negotiated meanings

## 1. Introduction

Wikipedia can be described as one of the “abstract social machines” advocated by Berners-Lee and Fischetti [1], in processes enabled by the World Wide Web (WWW) where people do the creative work and the machines does the administration. The concept includes the software and systems framework that supports it, as well as the rules, policies, and organizational structure governing the participation of the actors in the same “machine” [2]. In the case of Wikipedia, the massive number of collaborators (more than 32 million registered users<sup>1</sup>) contributes to the hypothesis that it is the most comprehensive project in the scope of Digital Humanities [3]. Its dynamics make it used as a field for investigation of the interaction between humans and computational artefacts under several foci, such as sociological [4,5], informational [6,7], or educational [8,9]. A systematization of the research areas of studies related to Wikipedia can be found in Tramullas’ work [10].

<sup>1</sup> In addition to registered users (34,115,228 as of 26 July 2018) numerous other unregistered users participate in the Wikipedia, (<https://en.wikipedia.org/wiki/Wikipedia:About>).

Despite the sheer number of works on Wikipedia, it is possible to obtain a comprehensive view of their focus from extensive literature reviews [11–13], or from platforms like WikiLit or WikiPapers<sup>2</sup> that collect these works. Among the issues debated around Wikipedia is its relationship with the academy. A relation in apparent change, from distrust and denial of its use, to an attitude of cautious acceptance or of its assumed use, namely as a pedagogical tool [14,15]. For this change of attitude, we can find studies that point to the credibility of the information contained in Wikipedia [16,17], as well as the publication of experiments on its use in academia [18,19].

Considering that Wikipedia presents itself as a free encyclopedia where any Internet user can edit, it can be considered as a place where the collective construction of knowledge occurs. In this context, the present work intends to place the focus of the analysis on the evolution of a concept constructed in a collaborative way, represented in the respective entry in Wikipedia. Collective knowledge in this paper is understood in the sense given by Scardamalia and Bereiter, that is, the public knowledge available to be managed and used by others [20]. In the same way that the collective construction of knowledge was limited to its observable exteriorization, the collective construction of a concept will be restricted here to its verbal definition represented in the form of written statements in a collaborative way. Although it can be understood as a reductionist view, we believe that these verbal externalisations are the ways in which a body of people can work through building a concept. We consider that this point of view fits in with Hjørland's view that "concepts are dynamically constructed and collectively negotiated meanings" [21].

The semantic web concept was chosen for analysis because we could observe an instability and lack of consensual definitions over time, even in the community directly related to its provenance, the Computer Science field. One previous piece of research, focused on the statements of the World Wide Web Consortium (W3C) and its director, Berners-Lee's works, showed that "the concept of Semantic Web is ambiguous and misinterpreting, given its biasing connection with the term 'semantics' and the association to other terms such as Web of Data, Linked Data or even Web of Linked Data" [22]. The study describes the terminological and conceptual metamorphosis of the definition of semantic web, expressed in the documents analyzed. This condition of the semantic web concept confers on the collective construction space, materialized in the respective Wikipedia entry, a context conducive to the debate and negotiation of different personal perspectives. The existence of the previous study allows a comparison between the perspective described there and that of the editors of the Wikipedia article under analysis. We consider the approach of this study to be a relevant contribution both to the investigation of the relationship between Wikipedia and the academy, as well as to the field of Knowledge Organization, through empirical subsidies for the theoretical study of concept theories.

In this way, we intend to analyze the evolution of the semantic web concept in the English version of Wikipedia, treating this as a context of collective knowledge construction. For this purpose, the objective is to: (i) collect the different definitions presented in the "introductory section" of the Semantic Web article, from December 2001 (date of creation of the article) to December 2017; (ii) to analyze the definitions collected in relation to the concept in question; (iii) to diachronically compare the concepts among each other and between them and the analysis of the same concept based on the publications of Berners-Lee and W3C.

In the next section of this paper, we will present a brief theoretical framework for the current research. Section 3 presents the methodology used in the present study, followed by, in Sections 4 and 5 respectively, the presentation of the results and discussion, and in the sixth section, the major conclusions are summarized.

---

<sup>2</sup> Accessible, respectively, in <http://wikilit.referata.com> and in <http://wikipapers.referata.com>.

## 2. Background

Regardless of all the controversy surrounding Wikipedia, in particular as regards the quality of information [23,24], there are calls for the attention of the academic community in the sense of its importance, or even the inevitability of its use as a means of scientific dissemination aimed at a wider audience [13,19,25]. As Nielsen points out: “Universities expect researchers to make their work more widely known, and extending Wikipedia is one way to spread both researchers’ work as well as ordinary information seekers” [13].

The use of Wikipedia as a platform for scientific publication is restricted by its rule of non-admission of original research (NOR)<sup>3</sup>, based on the premise of the need for published sources that attest to the reliability of the information introduced. However, Wikipedia’s dynamics are pointed out as “the potential model for more rapid and reliable dissemination of scholarly knowledge” [26], as exemplified by wiki-based scholarly publishing Species-ID<sup>4</sup>. In addition to the NOR rule, another issue that may negatively influence scientific writing on Wikipedia, by experts, is the absence of “academic reward” [13]. One way to deal with these two issues can be found in the RNA Biology journal approach that requires authors of articles on new RNA families to submit them accompanied by the draft of a corresponding entry on Wikipedia, which then cites the original article [27].

Although there are some contact points between the academic environment and Wikipedia, its dynamics of production is quite different. Although several comparative studies involving Wikipedia have focused on the quality of their information content (cf. Table 1), few have addressed the evolution of the concepts presented. The focus on concepts arises in a different context, in studies that use Wikipedia as a source of textual data, with the purpose of extracting and using conceptual relations for processes of facilitation of information retrieval, natural language processing, and ontology construction [28].

---

<sup>3</sup> See: [https://en.wikipedia.org/wiki/Wikipedia:No\\_original\\_research](https://en.wikipedia.org/wiki/Wikipedia:No_original_research).

<sup>4</sup> Accessible in: <http://species-id.net>.

**Table 1.** Selection of Wikipedia quality studies.

<b>Work</b>	<b>Comparisons <sup>1</sup></b>	<b>Evaluation</b>	<b>Results/Conclusions</b>
[16] 2005	Enc. Britannica (42 entries)	Blinded experts	“Only eight serious errors, such as misinterpretations of important concepts, were detected in the pairs of articles reviewed, four from each encyclopaedia.”
[29] 2006	Enc. Encarta (52 entries); American National Biography (25 entries)	The author	“Wikipedia, then, beats Encarta but not American National Biography Online in coverage and roughly matches Encarta in accuracy.”
[30] 2006	Topics: pop culture, current affairs, science	Specialized library reviewers	“Despite its flaws, however, Wikipedia should not be dismissed. Although the writing is not exceptional, good content abounds.”
[31] 2007	Surgical procedures (35 entries)	Experts	“Wikipedia is an accurate though often incomplete medical reference, with a remarkably high level of internal validity.”
[32] 2007	“Top” Western philosophers (7 entries)	The author	“we were unable to uncover any outright errors. The sins of Wikipedia are more of omission than commission.”
[33] 2007	Enc. Brockhaus (50 entries)	Research institute	“On a scale of 1 to 6, with 1 being the best score, articles from Wikipedia’s German version received an average rating of 1.7, while Brockhaus’ average rating was 2.7.”
[34] 2007	Medscape Drug Reference (80 questions)	The authors	“Wikipedia has a more narrow scope, is less complete, and has more errors of omission than the comparator database.”
[35] 2007	Thomson Scientific Journal Citation Reports	The author	“The present number of structured outbound citations from Wikipedia is quite small compared to the total number of scientific citations found in current scientific literature.”
[36] 2008	Access Medicine; eMedicine; Up-ToDate	Blinded experts	“Wikipedia was found currently unsuitable for medical students in isolation from other medical information resources.”
[37] 2008	The Stanford Enc. of Philosophy; Internet Enc. of Philosophy	The author	“although the types of academics listed in Wikipedia are generally similar to those in the other encyclopaedias, their relative youth and their very numbers may still serve to give the user a very different impression on philosophy as a field.”
[38] 2008	Library of Congress categories; Enc. of Linguistics; New Princeton Enc. of Poetry and Poetics; Enc. of Physics	The authors	“Even in the least covered areas, because of its sheer size, Wikipedia does well . . . . It cannot be a coincidence that two areas that are particularly lacking on Wikipedia—law and medicine—are also the purview of licensed experts.”

Table 1. Cont.

Work	Comparisons <sup>1</sup>	Evaluation	Results/Conclusions
[39] 2011	Physician Data Query (US National Cancer Institute)	Medically trained personnel	“Although the wiki resource had similar accuracy and depth as the professionally edited database, it was significantly less readable.”
[40] 2011	US gubernatorial candidates and elections	The author	“Wikipedia is almost always accurate when a relevant article exists, but errors of omission are extremely frequent.”
[41] 2011	Enc. Britannica (“thousands” of biographies)	The authors	“While Wikipedia’s massive reach in coverage means one is more likely to find a biography of a woman there than in Britannica, evidence of gender bias surfaces from a deeper analysis of those articles each reference work misses.”
[42] 2014	Peer-review medical literature	Pairs of medicine residents or interns	“Most Wikipedia articles representing the 10 most costly medical conditions in the United States contain many errors when checked against standard peer-reviewed sources.”
[43] 2014	Text books (information for 100 drugs)	The authors	“Our study suggests that Wikipedia is an accurate and comprehensive source of drug-related information for undergraduate medical education.”
[44] 2014	Biographies of scientists (400 entries)	The authors	“We also did not find any evidence that the scientists with better WP representation are necessarily more prominent in their fields . . . . In each of the examined fields, Wikipedia failed in covering notable scholars properly.”

<sup>1</sup> Enc.—short for Encyclopedia. Selection from the extensive literature review, updated in 2017 [13].

Regarding the study of concepts, there is no consensus for what their nature is (mental representations or abstract entities?), or their constitution (bundles of features or they embody mental theories?). Different approaches, derived from Philosophy, Cognitive Science or Linguistics, 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 [45]. For these authors, concepts are mental representations and a theory with the necessary explanatory potential is only possible if one “admits different types of conceptual structure while tying them together by maintaining that concepts have atomic cores” [46].

From a perspective of Organization of Knowledge, we start from the pluralist epistemological position presented for the association of two distinct perspectives of two prominent authors of the area. 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 [47]. 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 [48].

### 3. Materials and Methods

In order to fulfil the defined objectives, an exploratory/descriptive qualitative study was designed, following an observational/comparative methodology [49]. For the operationalization of the empirical component of the study we chose the English version of Wikipedia, since this is the language used in the W3C and Berners-Lee reference documents about the semantic web. Thus, the “history”<sup>5</sup> of the Semantic Web entry was mapped to identify the semantic changes made to the support statement of the respective definition, presented in the “introduction” of the different versions of this Wikipedia article. During the analysis, it was used whenever deemed necessary the “discussion” page<sup>6</sup> in order to obtain contextual information to help clarify the definitions presented.

As an analytical technique, the categorization was applied “by collection”, that is, the categories resulted from the analogous and progressive classificatory process performed [50]. Subsequently, a procedure based on “time-series analysis” [51] was used for the content units considered in each category, for diachronic comparison. The conceptual analysis focused on the identification of generic terms and their specifying characteristics [48], in order to compare the definitions collected. In the determination of generic terms, we sought for the non-use of compound terms, for the sake of simplicity.

In situations where the definitions use evaluative terms or contextual interpretation (on the “discussion” page and descriptions appended to the respective changes), we used the contributions provided by the analysis of the discursive strategies, in particular the predicative, of intensification and of attenuation, as long as they provide indicators on the valuing of characteristics and the attitudes and positions of stakeholders [52].

### 4. Results

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 (the corresponding statements

---

<sup>5</sup> Record of all changes made to the page with the possibility of comparing the different versions, ([https://en.wikipedia.org/w/index.php?title=Semantic\\_Web&action=history](https://en.wikipedia.org/w/index.php?title=Semantic_Web&action=history)).

<sup>6</sup> Forum associated with each page where editors can justify their changes or question those of others ([https://en.wikipedia.org/wiki/Talk:Semantic\\_Web](https://en.wikipedia.org/wiki/Talk:Semantic_Web)).

are found in Table A1 Appendix A). In Table 2 we present the definitions grouped within each category, according to the respective generic term.

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 synchronogenetic nature of the element “of data” [53] 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 of the 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 2), is presented in Figure 1.

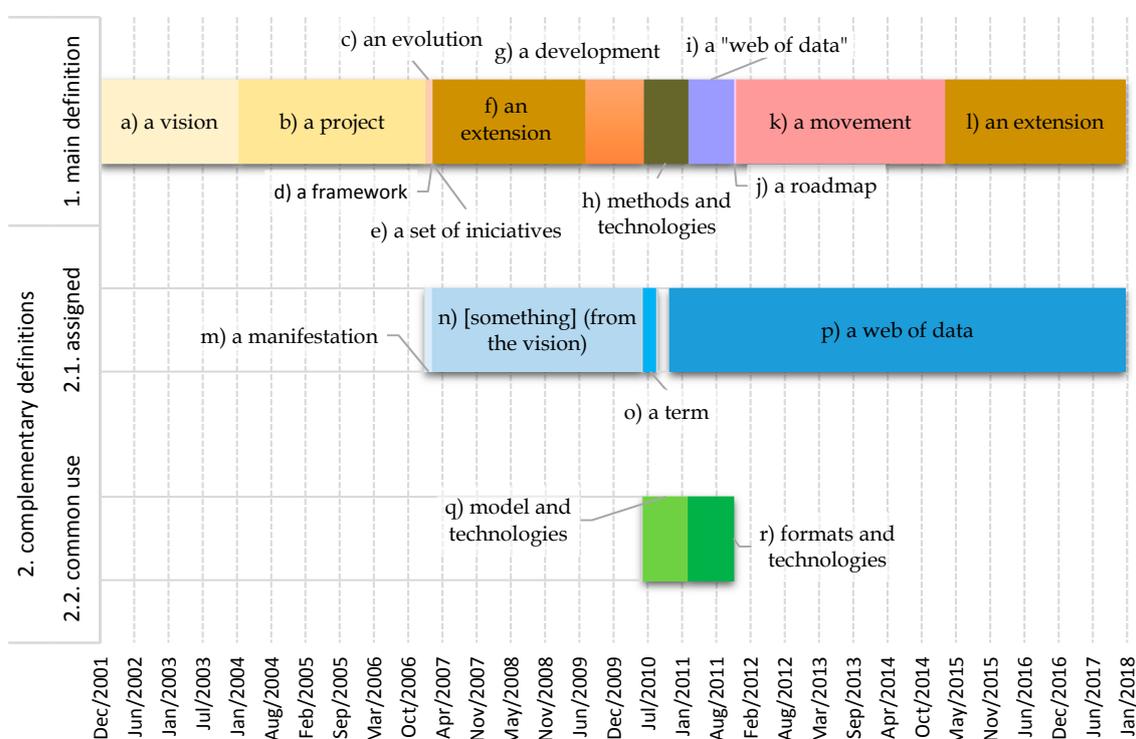


Figure 1. Temporal distribution of the definitions (group by the respective generic terms).

The diachronic visualization presents an enlightening overview of the evolution of the semantic web concept in Wikipedia’s context. Given the extended time span (December 2001 to December 2017) 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 3).

**Table 2.** Generic terms and respective content units retrieved from the identified definitions.

rf.GT <sup>1</sup>	Generic Terms (GT)	Content Units
<b>category 1. Main definitions</b>		
(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.
<b>category 2. Complementary definitions—sub-category 2.1. Assigned</b>		
(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.
<b>category 2. Complementary definitions—sub-category 2.2. Common use</b>		
(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.

<sup>1</sup> rf.GT—references associated to the generic terms (GT). The numerical references of the content units relate to the 26 definitions identified (see Table A1 in Appendix A).

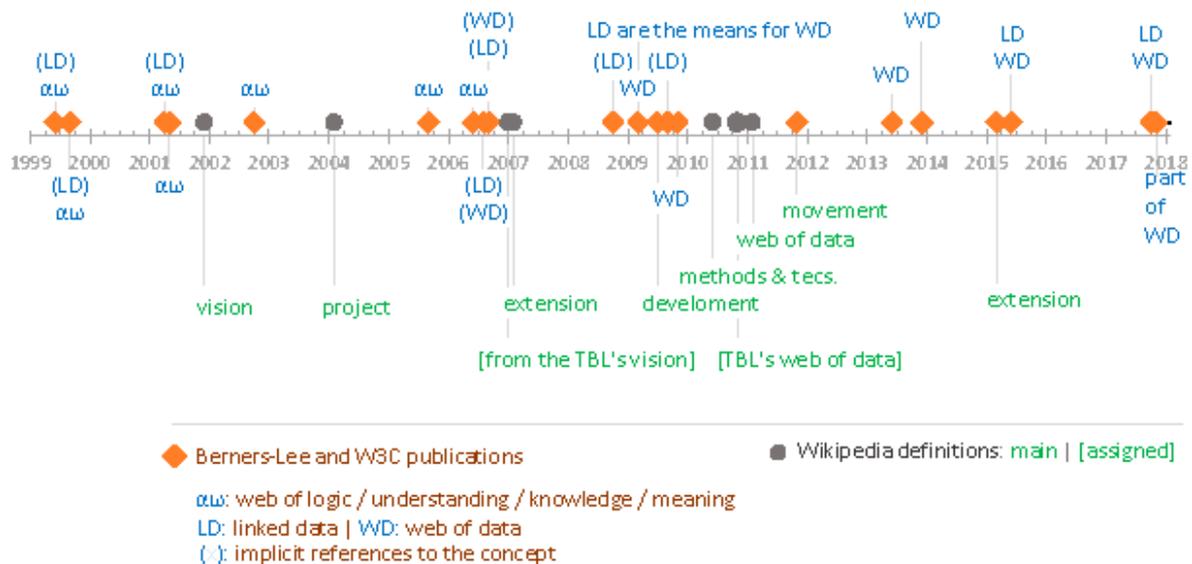
**Table 3.** Specific characteristics of generic terms.

rf.GT	Specifiers pre-GT	Generic Terms (GT)	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 3, 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 3. 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 in Wikipedia, with those resulting from the analysis of the same concept based on the publications of Berners-Lee and W3C, was also carried out in a diachronic perspective. For the sake of clarity and representativeness, 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 and W3C publications previously analyzed. For this matter, we repeat in Table 4 the content units of the cited study [22].

Referring to the Table 4, we can note that the term “extension” is used to 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.

**Table 4.** 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”	<ul style="list-style-type: none"> <li>a. The Web of understanding (7 June 1999); A universal web of knowledge (26 April 2001).</li> <li>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).</li> <li>c. A web of data with meaning (22 September 1999).</li> </ul>
2. Linked Data	<ul style="list-style-type: none"> <li>a. The Web of linked data (11 August 2006; 2015).</li> <li>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).</li> <li>c. Linked Data provides the means (March 2009).</li> </ul>
3. Web of Data	<ul style="list-style-type: none"> <li>a. A Web of Data (March 2009; 12 November 2009; 27 June 2013; 11 December 2013; 2015; 11 October 2017).</li> <li>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).</li> <li>c. Part of the Web of Data (2016).</li> </ul>

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 4, for 2006 and 2015 (sub-group 2.a.).

## 5. Discussion

The concept of the semantic web, presented in the respective entry of 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)<sup>7</sup> and in the discussion page. As an example, for the first case, “skewed the defn [sic] to an outsider’s (web user’s) point of view” (Vanished user kjsdion3i4jf, 23 February 2008); “Query users by better explaining ‘to web of data that can be processed by machines’” (Quercus solaris, 13 October 2017) and, for the second, “WP content is intended for a ‘general audience’, the wording should reflect that” (dr.ef.tymac, 20 February 2007)<sup>8</sup>.

Although the evolution of this concept presents points of contact and similarity between the two scopes (Wikipedia and the publications of Berners-Lee and 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). The present study 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<sup>9</sup>. In some changes made to the article, this attempt to promote clarification is explicitly stated, as in 21 November 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”<sup>10</sup>. 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” [54], 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 (see Table A1 in Appendix A). The only exception, reported on 20 February 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

<sup>7</sup> In: [https://en.wikipedia.org/w/index.php?title=Semantic\\_Web&action=history](https://en.wikipedia.org/w/index.php?title=Semantic_Web&action=history).

<sup>8</sup> See: [https://en.wikipedia.org/wiki/Talk:Semantic\\_Web/Archive\\_1#20070220meta\\_:\\_modifications\\_to\\_lead\\_section](https://en.wikipedia.org/wiki/Talk:Semantic_Web/Archive_1#20070220meta_:_modifications_to_lead_section).

<sup>9</sup> A buzzword is a word or expression that has become fashionable in a particular field and is being used a lot by the media, (<https://www.collinsdictionary.com/dictionary/english/buzzword>).

<sup>10</sup> Retrieved from *Semantic Web: Revision history*, (dynamic URL).

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)<sup>11</sup>.

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 [55], 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” [15].

## 6. Conclusions

Regarding the relationship between Wikipedia and academia, the study points to the target audience as a relevant difference factor. This feature of moderating the language used to reach a wider audience implies that Wikipedia, even without the NOR rule, should be taken as a complementary not an alternative medium for scientific dissemination.

Given the characteristics of Wikipedia, described and discussed throughout this paper, we can consider it as a place for collective bargaining of meanings, and it is therefore important to take it as an object of study for a community’s understanding of a concept in particular. This position is aligned with Hjørland’s quote: “Concepts have been understood as socially negotiated meanings that should be identified by studying discourses rather than by studying individual users or a priori principles” [21]. In this context, this research presents an approach, for the diachronic study of these discourses using the information source and features provided by Wikipedia.

Despite Wikipedia’s relevance to this study of the collective construction of meanings, other similar studies would be necessary to understand the importance of this phenomenon in a more comprehensive process of “dictionaryization” in which the content of a concept is fixed by its definitions [38]. It is possible, however, to draw a parallel between the conceptual evolutionary dynamics inherent in the workings of Wikipedia and Derqui’s assertion, that says that: “a social system is organized around definitions and redefinitions” [56].

**Author Contributions:** Conceptualization, L.M.M.; writing: original draft, L.M.M., M.M.B. and R.R.S.; writing: review and editing, L.M.M., M.M.B. and R.R.S.

**Funding:** This research received no external funding.

**Acknowledgments:** Thanks to all referees for valuable feedback on previous versions of this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

Definitions extracted from the “Semantic Web: Revision history” inserted in the context units (column *Context units*) with bold emphasis of the content units; the respective authors (column *Users*); the date in which the definition was entered (column *Start*); the date it was withdrawn (column *End*); the reference relative to 129 statements collected (column *ref.Tt.*) and the reference assigned to units of content (column *ref.Df.*), composed of the chronological number, followed by the generic term identifier and the category (main (1.) or complementary definitions (2.1 or 2.2)) where this has been classified (see Table 2. Generic terms and respective content units retrieved from the identified definitions.).

<sup>11</sup> See: [https://en.wikipedia.org/wiki/Talk:Semantic\\_Web](https://en.wikipedia.org/wiki/Talk:Semantic_Web).

**Table A1.** Definitions extracted from the “Semantic Web: Revision history”.

ref.Df	ref.Tt.	Start	End	Context units	Users
#01.a cat.1	t001	07-12-2001	13-01-2004	<i>is Tim Berners-Lee’s vision of the future of the WWW.</i>	The Anome
#02.a cat.1	t002	13-01-2002	10-02-2004	<i>is a vision of the future of the WWW proposed by Tim Berners-Lee,</i>	65.2.226.xxx
#03.b cat.1	t008	10-02-2004	23-07-2004	<i>is a current project under the direction of Tim Berners-Lee of the W3C to extend the ability of the WWW.</i>	ShaunMacPherson
#04.b cat.1	t012	23-07-2004	26-07-2004	<i>is a project underway that intends to create a universal medium for the exchange of information by giving meaning, in a manner understandable by machines, to the content of documents on the WWW.</i>	Mjb
#05.b cat.1	t013	26-07-2004	11-01-2007	<i>is a project that intends to create a universal medium for the information exchange by giving meaning, in a manner understandable by machines, to the content of documents on the WWW.</i>	Lou Quillio
#06.c cat.1	t020	11-01-2007	26-01-2007	<i>is an evolution of the current Web that seeks to provide granular access to the underlying data that fuels the WWW.</i>	KingsleyIdehen
#07.m cat.2.1	t020	11-01-2007	20-02-2007	<i>It’s a manifestation of the W3C chairman Tim Berners-Lee’s vision of the Web as a universal medium for Data, Information, and Knowledge exchange.</i>	KingsleyIdehen
#08.c cat.1	t026	26-01-2007	19-02-2007	<i>is an evolution of the WWW in which information is machine processable (rather than being only human oriented),</i>	71.68.198.237
#09.d cat.1	t028	19-02-2007	20-02-2007	<i>is a loosely defined and evolving framework of WWW based technologies that seek to augment human readable content with information that is machine processable,</i>	Numskll
#10.d cat.1	t029	20-02-2007	20-02-2007	<i>is a loosely defined and evolving framework intended to augment web content with machine processable metadata,</i>	Dreftymac
#11.n cat.2.1	t029	20-02-2007	12-06-2010	<i>It derives from W3C director Tim Berners-Lee’s vision of the WWW as a universal medium for data, information, and knowledge exchange.</i>	Dreftymac
#12.e cat.1	t031	20-02-2007	21-02-2007	<i>is a set of loosely-defined and evolving initiatives to extend web content into a framework that can be processed and interpreted by automata,</i>	Dreftymac
#13.f cat.1	t034	21-02-2007	13-07-2009	<i>is an evolving extension of the WWW in which Web content can not only be expressed in natural language, but also in a form that can be understood, interpreted and used by software agents,</i>	Cygri

Table A1. Cont.

ref.Df	ref.Tt.	Start	End	Context units	Users
#14.g cat.1	t057	13-07-2009	12-06-2010	<i>is an evolving development of the WWW in which web content can not only be expressed in natural language, but also in a form that can be understood, interpreted and used by software agents,</i>	Andy Dingley
#15.h cat.1	t067	12-06-2010	01-09-2010	<i>It describes methods and technologies to allow machines to understand the meaning—or “semantics”—of information on the WWW.</i>	Averell23
#16.o cat.2.1	t067	12-06-2010	01-09-2010	<i>is a term coined by W3C director Sir Tim Berners-Lee.</i>	Averell23
#17.q cat.2.2	t067	12-06-2010	23-02-2011	<i>it is mainly used to describe the model and technologies proposed by the W3C.</i>	Averell23
#18.g cat.1	t073	01-09-2010	23-02-2011	<i>is a group of methods and technologies to allow machines to understand the meaning—or “semantics”—of information on the WWW.</i>	Wikidemon
#19.p cat.2.1	t074	13-11-2010	21-11-2011	<i>Tim Berners-Lee defined the Semantic Web as “a web of data that can be processed directly and indirectly by machines”.</i>	99.41.179.96
#20.i cat.1	t085	23-02-2011	03-09-2011	<i>is a “web of data” that enables machines to understand the semantics, or meaning, of information on the WWW.</i>	Michael A. White
#21.r cat.2.2	t085	23-02-2011	12-11-2011	<i>is often used more specifically to refer to the formats and technologies that enable it.</i>	Michael A. White
#22.i cat.1	t093	03-09-2011	12-11-2011	<i>is a “man-made woven web of data” that facilitates machines to understand the semantics, or meaning, of information on the WWW.</i>	Wireless friend
#23.j cat.1	t096	12-11-2011	21-11-2011	<i>is the roadmap of a “man-made woven web of data” that facilitates machines to understand the semantics, or meaning, of information on the WWW.</i>	Karima Rafes
#24. cat.1k	t097	21-11-2011	06-06-2013	<i>is a collaborative movement led by the W3C that promotes common formats for data on the WWW.</i>	24.69.174.26
#25.o cat.2.1	t113	06-06-2013	-	<i>The term was coined by Tim Berners-Lee for a web of data that can be processed by machines.</i>	Nigelj
#26.l cat.1	t117	09-03-2015	-	<i>is an extension of the Web through standards by the W3C.</i>	Denny

## References

- Berners-Lee, T.; Fischetti, M. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web*; Harper Collins: New York, NY, USA, 1999; ISBN 9780062515872.
- Hendler, J.; Shadbolt, N.; Hall, W.; Berners-Lee, T.; Weitzner, D. Web Science: An interdisciplinary approach to understanding the web. *Commun. ACM* **2008**, *51*, 60–69. [[CrossRef](#)]
- Flores, P. 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 16 November 2016).
- Iñiguez, G.; Török, J.; Yasseri, T.; Kaski, K.; Kertész, J. Modeling social dynamics in a collaborative environment. *EPJ Data Sci.* **2014**, *3*, 7. [[CrossRef](#)]
- Westerman, W. Epistemology, the Sociology of Knowledge, and the Wikipedia Userbox Controversy. In *Folklore and the Internet: Vernacular Expression in a Digital World*; Blank, T.J., Ed.; University Press of Colorado, Urban Institute: Louisville, CO, USA, 2009; pp. 123–158. ISBN 978-0-87421-750-6.
- Kleeb, R.; Gloor, P.A.; Nemoto, K.; Henninger, M. Wikimaps: Dynamic maps of knowledge. *Int. J. Organ. Des. Eng.* **2012**, *2*, 204–224. [[CrossRef](#)]
- Biuk-Aghai, R.P. Visualizing Co-Authorship Networks in Online Wikipedia. In *International Symposium on Communications and Information Technologies*; IEEE: Piscataway Township, NJ, USA, 2006; pp. 737–742.
- Cress, U.; Kimmerle, J. A theoretical framework of collaborative knowledge building with wikis: A systemic and cognitive perspective. In Proceedings of the 8th International Conference on Computer Supported Collaborative Learning (CSCL'07), New Brunswick, NJ, USA, 16–21 July 2007; pp. 156–164.
- Kump, B.; Moskaliuk, J.; Dennerlein, S.; Ley, T. Tracing knowledge co-evolution in a realistic course setting: A wiki-based field experiment. *Comput. Educ.* **2013**, *69*, 60–70. [[CrossRef](#)]
- Tramullas, J. Wikipedia como objeto de investigación. *Anu. ThinkEPI* **2015**, *9*, 223–226. [[CrossRef](#)]
- Jullien, N. What We Know about Wikipedia: A Review of the Literature Analyzing the Project(s). *SSRN Electron. J.* **2012**, *86*. [[CrossRef](#)]
- Okoli, C.; Mehdi, M.; Mesgari, M.; Nielsen, F.Å.; Lanamäki, A. The People's Encyclopedia Under the Gaze of the Sages: A Systematic Review of Scholarly Research on Wikipedia. *SSRN Electron. J.* **2012**. [[CrossRef](#)]
- Nielsen, F.Å. *Wikipedia Research and Tools: Review and Comments*; Working Paper; The Technical University of Denmark: Lyngby, Denmark, 2017.
- Aibar, E. Wikipedia, Academia, and Science. In Proceedings of the 9th International Conference on Web and Social Media (ICWSM-15), Oxford, UK, 26–29 May 2015; pp. 2–5.
- 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; ISBN 978-0-472-07198-2.
- Giles, J. Internet encyclopaedias go head to head. *Nature* **2005**, *438*, 900–901. [[CrossRef](#)] [[PubMed](#)]
- Messner, M.; South, J. Legitimizing Wikipedia. *J. Pract.* **2011**, *5*, 145–160. [[CrossRef](#)]
- Writing History in the Digital Age*; Nawrotzki, K. (Ed.) University of Michigan Press: Ann Arbor, MI, USA, 2013; ISBN 978-0-472-07206-4.
- Jemielniak, D.; Aibar, E. Bridging the gap between wikipedia and academia. *J. Assoc. Inf. Sci. Technol.* **2016**, *67*, 1773–1776. [[CrossRef](#)]
- Scardamalia, M.; Bereiter, C. Knowledge building. In *Encyclopedia of Education*, 2nd ed.; Macmillan Reference: New York, NY, USA, 2003; pp. 1370–1373.
- Hjørland, B. Concept theory. *J. Am. Soc. Inf. Sci. Technol.* **2009**, *60*, 1519–1536. [[CrossRef](#)]
- Machado, L.M.O.; Rocha, R.S.; da Graça Simões, M. 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. *J. Assoc. Inf. Sci. Technol.* **2018**, in press.
- Lucassen, T.; Dijkstra, R.; Schraagen, J.M. Readability of Wikipedia. *First Monday* **2012**, *17*. [[CrossRef](#)]
- Roberts, P.; Peters, M.A. From Castalia to Wikipedia: Openness and Closure in Knowledge Communities. *E-Learn. Digit. Media* **2011**, *8*, 36–46. [[CrossRef](#)]
- Rush, E.K.; Tracy, S.J. Wikipedia as Public Scholarship: Communicating Our Impact Online. *J. Appl. Commun. Res.* **2010**, *38*, 309–315. [[CrossRef](#)]
- Black, E.W. Wikipedia and academic peer review. *Online Inf. Rev.* **2008**, *32*, 73–88. [[CrossRef](#)]

27. Mietchen, D.; Hagedorn, G.; Förstner, K.U.; Kubke, M.F.; Koltzenburg, C.; Hahnel, M.; Penev, L. Wikis in scholarly publishing. *Inf. Serv. Use* **2011**, *31*, 53–59. [CrossRef]
28. Medelyan, O.; Milne, D.; Legg, C.; Witten, I.H. Mining meaning from Wikipedia. *Int. J. Hum. Comput. Stud.* **2009**, *67*, 716–754. [CrossRef]
29. Rosenzweig, R. Can History Be Open Source? Wikipedia and the Future of the Past. *J. Am. Hist.* **2006**, *93*, 117–146. [CrossRef]
30. Miller, B.X.; Helicher, K.; Berry, T. I want my Wikipedia! *Libr. J.* **2006**, *6*, 122–124.
31. Devgan, L.; Powe, N.; Blakey, B.; Makary, M. Wiki-Surgery? Internal validity of Wikipedia as a medical and surgical reference. *J. Am. Coll. Surg.* **2007**, *205*, S76–S77. [CrossRef]
32. Bragues, G. Wiki-Philosophizing in a Marketplace of Ideas: Evaluating Wikipedia’s Entries on Seven Great Minds. *SSRN Electron. J.* **2007**. [CrossRef]
33. Jones, K.C. German Wikipedia Outranks Traditional Encyclopedia’s Online Version—InformationWeek. Available online: <https://www.informationweek.com/german-wikipedia-outranks-traditional-encyclopedias-online-version/d/d-id/1062250> (accessed on 26 October 2018).
34. Clauson, K.A.; Polen, H.H.; Boulos, M.N.K.; Dzenowagis, J.H. Scope, Completeness, and Accuracy of Drug Information in Wikipedia. *Ann. Pharmacother.* **2008**, *42*, 1814–1821. [CrossRef] [PubMed]
35. Nielsen, F.A. Scientific citations in Wikipedia. *First Monday* **2007**, *12*. [CrossRef]
36. Pender, M.P.; Lasserre, K.; Del Mar, C.B.; Kruesi, L.; Anuradha, S. Putting Wikipedia to the test: A case study. In Proceedings of the Special Libraries Association Annual Conference, Seattle, WA, USA, 15–18 June 2008.
37. Elvebakk, B. Philosophy Democratized? A comparison between Wikipedia and two other Web-based philosophy resources. *First Monday* **2008**, *13*. [CrossRef]
38. Halavais, A.; Lackaff, D. An Analysis of Topical Coverage of Wikipedia. *J. Comput. Commun.* **2008**, *13*, 429–440. [CrossRef]
39. Rajagopalan, M.S.; Khanna, V.K.; Leiter, Y.; Stott, M.; Showalter, T.N.; Dicker, A.P.; Lawrence, Y.R. Patient-Oriented Cancer Information on the Internet: A Comparison of Wikipedia and a Professionally Maintained Database. *J. Oncol. Pract.* **2011**, *7*, 319–323. [CrossRef] [PubMed]
40. Brown, A.R. Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage. *PS Political Sci. Politics* **2011**, *44*, 339–343. [CrossRef]
41. Reagle, J.; Rhue, L. Gender Bias in Wikipedia and Britannica. *Int. J. Commun.* **2011**, *5*, 21.
42. Hasty, R.; Garvalosa, R.; Barbato, V.; Valdes, P.; Powers, D.; Hernandez, E.; John, J.; Suci, G.; Qureshi, F.; Popa-Radu, M.; et al. Wikipedia vs Peer-Reviewed Medical Literature for Information About the 10 Most Costly Medical Conditions. *J. Am. Osteopath. Assoc.* **2014**, *114*, 368–373. [CrossRef] [PubMed]
43. Kräenbring, J.; Monzon Penza, T.; Gutmann, J.; Muehlich, S.; Zolk, O.; Wojnowski, L.; Maas, R.; Engelhardt, S.; Sarikas, A. Accuracy and Completeness of Drug Information in Wikipedia: A Comparison with Standard Textbooks of Pharmacology. *PLoS ONE* **2014**, *9*, e106930. [CrossRef] [PubMed]
44. Samoilenko, A.; Yasserli, T. The distorted mirror of Wikipedia: A quantitative analysis of Wikipedia coverage of academics. *EPJ Data Sci.* **2014**, *3*, 1. [CrossRef]
45. *Concepts: Core Readings*; Margolis, E.; Laurence, S. (Eds.) MIT Press: Cambridge, MA, USA, 1999; ISBN 9780262631938.
46. Laurence, S.; Margolis, E. Concepts. In *The Blackwell Guide to the Philosophy of Mind*; Warfield, T.A., Stich, S.P., Eds.; Blackwell’s: Malden, MA, USA, 2003; pp. 190–213.
47. Arboit, A.E. O processo de (re) construção da teoria do conceito no domínio de Organização do Conhecimento: Uma visão dialógica. *Scire* **2012**, *2*, 129–134.
48. Dahlberg, I. A referent-oriented, analytical concept theory for INTERCONCEPT. *Int. Classif.* **1978**, *5*, 143–151. [CrossRef]
49. Gil, A.C. *Métodos e Técnicas de Pesquisa Social*, 6th ed.; Atlas S.A.: São Paulo, Brazil, 2008; ISBN 9788522451425.
50. Bardin, L. *Análise de Conteúdo*; Almedina: São Paulo, Brazil, 2011; ISBN 978-85-62938-04-7.
51. Yin, R.K. *Case Study Research: Design and Methods*, 5th ed.; SAGE Publications: London, UK, 2014; ISBN 9781452242569.
52. *Manual de Análise do Discurso em Ciências Sociais*; Iñiguez, L. (Ed.) Vozes: Petrópolis, Brazil, 2004.
53. Stock, W.G. Concepts and semantic relations in information science. *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 1951–1969. [CrossRef]

54. Viégas, F.B.; Wattenberg, M.; Dave, K. Studying cooperation and conflict between authors with history flow visualizations. In Proceedings of the 2004 Conference on Human Factors in Computing Systems (CHI'04), Vienna, Austria, 24–29 April 2004; ACM Press: Vienna, Austria, 2004; Volume 6, pp. 575–582.
55. Vygotsky, L.S. *A Construção do Pensamento e da Linguagem*; Martins Fontes: São Paulo, Brazil, 2001.
56. Derqui, P.M. *Da Informação à Categorização: A Formação Sistêmica dos Conceitos*; Universidade de São Paulo: São Paulo, Brazil, 2014.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).