



UNIVERSIDADE
ESTADUAL DE LONDRINA

GIAN CARLO DECARLI

**O CONHECIMENTO ENCICLOPÉDICO COLABORATIVO E
SUAS INFLUÊNCIAS NO DESENVOLVIMENTO DA *WEB*
PRAGMÁTICA**

Londrina
2017

GIAN CARLO DECARLI

**O CONHECIMENTO ENCICLOPÉDICO COLABORATIVO E
SUAS INFLUÊNCIAS NO DESENVOLVIMENTO DA *WEB*
PRAGMÁTICA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação (PPGCI) da Universidade Estadual de Londrina (UEL), como requisito parcial à obtenção do título de Mestre em Ciência da Informação.

Orientadora: Profa. Dra. Silvana Drumond Monteiro.

Londrina
2017

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Decarli, Gian Carlo.

O conhecimento enciclopédico colaborativo e suas influências no desenvolvimento da *Web Pragmática* / Gian Carlo Decarli. - Londrina, 2017.

154f. il.

Orientador: Silvana Drumond Monteiro.

Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual de Londrina, Centro de Educação Comunicação e Artes, Programa de Pós-Graduação em Ciência da Informação, 2017.

Inclui bibliografia.

1. Conhecimento enciclopédico - Tese. 2. Wikipédia - Tese. 3. Web Pragmática - Tese. 4. Mecanismos de busca - Tese. I. Monteiro, Silvana Drumond. II. Universidade Estadual de Londrina. Centro de Educação Comunicação e Artes. Programa de Pós-Graduação em Ciência da Informação. III. Título.

GIAN CARLO DECARLI

**O CONHECIMENTO ENCICLOPÉDICO COLABORATIVO E SUAS
INFLUÊNCIAS NO DESENVOLVIMENTO DA *WEB* PRAGMÁTICA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Mestre em Ciência da Informação.

BANCA EXAMINADORA

Profa. Dra. Silvana Drumond Monteiro
Universidade Estadual de Londrina - UEL

Profa. Dra. Renata Gonçalves Curty
Universidade Estadual de Londrina - UEL

Profa. Dra. Silvana Aparecida Borsetti Gregório
Vidotti
Universidade Estadual Paulista – UNESP

Londrina, 05 de maio de 2017.

Dedico este feito àquela que guardo no peito, meu amor, Dayane.

AGRADECIMENTOS

Aos meus familiares, em especial ao meu pai, que mesmo sem ter acesso a uma educação formal em sua vida, soube indicar com sabedoria este caminho ao longo de minha vida. Mais que isso, mostrou apoio incondicional ao meu ingresso e permanência no mestrado, ainda que não soubesse bem ao certo o que eu estava fazendo.

Ao meu amor, esposa, amiga, porto seguro e eterna namorada, Dayane, por me incentivar e apoiar minhas decisões ao longo desses oito anos, acreditando que sempre posso ir além.

Meus mais sinceros agradecimentos à minha orientadora, Dra. Silvana Drumond Monteiro, por compartilhar seus conhecimentos e experiências acadêmicas ao longo deste período como discente do PPGCI-UEL. A propósito, por ter sido uma pessoa fundamental nesta guinada acadêmica de minha vida, pois muitas vezes um orientador não tem ideia do poder de mudança que exerce sobre seu orientando.

Às ricas contribuições dos professores do PPGCI-UEL que enriqueceram minha trajetória acadêmica. Às professoras componentes da banca examinadora, Dra. Silvana Aparecida Borsetti Gregório Vidotti, por aceitar compor a banca de avaliação desta dissertação bem como suas valiosas considerações, além da receptividade em seu grupo de pesquisa e em eventos promovidos pelo PPGCI-UNESP (Marília); à professora Dra. Renata Gonçalves Curty, por também aceitar compor esta banca de avaliação e por suas expressivas contribuições para o desenvolvimento desta dissertação.

Aos amigos, Carlos Eduardo de Lima, por promover ricas e descontraídas discussões acadêmicas (ou não) interdisciplinares; João Arlindo dos Santos Neto, pela amizade desde a graduação e a experiência acadêmica compartilhada; ao calouro e amigo Gustavo Lunardelli Trevisan, por compartilhar dos mesmos anseios que circundam a vida de um discente de pós-graduação, além de incentivar a prática esportiva.

Ao Mozart, vulgo “Mozi”, por sua lealdade e companheirismo em dias solitários de escrita e por também não ter comido minha dissertação. Ainda, aos seus “pais”, Airton Leitzke e Milene Leitzke, por sua guarda compartilhada e amizade sincera.

Aos colegas do PPGCI-UEL e do Grupo de Pesquisa Informação e Conhecimento no Ciberespaço pelas discussões e trabalhos desenvolvidos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Fundação Araucária pelas bolsas de estudo; e à Universidade Estadual de Londrina (UEL) por toda minha formação de nível superior.

"Spam será uma coisa do passado em dois anos."
William Henry "Bill" Gates III, 2004.

DECARLI, Gian Carlo. **O conhecimento enciclopédico colaborativo e suas influências no desenvolvimento da Web Pragmática**. 2017. 154f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual de Londrina, Londrina, 2017.

RESUMO

Mediante a comunicação oral e escrita, o conhecimento enciclopédico ou de mundo transpassou por diferentes períodos da humanidade. Na história recente, as enciclopédias têm desempenhado plenamente a tarefa de registrar conhecimentos episódicos e declarativos de culturas, faunas, floras, indivíduos entre outros tantos fenômenos ao redor do mundo, convertendo-os em conhecimento enciclopédico. Constituída no início dos anos 2000, a Wikipédia é uma das plataformas que melhor representa o ideal de colaboração da *Web Social*. Além disso, introduz o conceito de enciclopédia livre, na qual todos podem colaborar para sua expansão por meio da participação na produção, no monitoramento e na disseminação de seus verbetes. A Wikipédia agregou ao conhecimento enciclopédico gerado o *status* de colaborativo, tornando-se uma grande base internacional do conhecimento. Na camada de origem da Wikipédia, a *Web Social*, houve a ruptura de um modelo linear de produção e consumo de conteúdo instaurado pela camada anterior, denominada de *Web 1.0*. Assim, quem consumia informação também a produzia e vice-versa. Entretanto, a liberdade de criação proporcionada pela camada Social trouxe problemas aos processos de recuperação de informações na *Web*. A *Web Semântica*, sua sucessora, emergiu com o propósito de estruturar esses dados para que pudessem fazer sentido tanto para humanos quanto para máquinas. Deste modo, se formou na camada Semântica as *Knowledge Bases*, responsáveis por ligar dados correlacionados de diferentes páginas da *Web* com o objetivo de gerar sentido/significado para determinados temas. A camada seguinte, a *Web Pragmática*, agrega todo o legado estrutural das camadas antecessoras para explorar o significado a partir da percepção do usuário, ou seja, considera que o sentido se dá a partir do sujeito em múltiplos contextos. Dessa forma, os Mecanismos de Busca atuais parecem cada vez mais engajados sob essa concepção de uma *Web Pragmática*. Nessas circunstâncias, tecnologias como o *Knowledge Graph* auxiliam os Mecanismos de Busca a melhor estruturar e exibir informações fatuais sobre determinados temas em suas *Search Engine Results Pages*. Ao observar essa conjuntura tecnológica, esta pesquisa propôs analisar se o conhecimento enciclopédico colaborativo produzido pela Wikipédia estava de alguma maneira influenciando no desenvolvimento da *Web Pragmática*. Para tanto, o percurso metodológico se dividiu em três etapas; a primeira descreveu o curso do conhecimento enciclopédico colaborativo na Wikipédia, a segunda analisou como as *Knowledge Bases* estruturavam esse conhecimento e a terceira investigou se os mecanismos de busca Bing, Google e Yahoo! usufruíam do conhecimento produzido e estruturado colaborativamente para aprimorar seus resultados e sustentar o conceito de *Web Pragmática*. Os resultados apontaram que o conhecimento enciclopédico colaborativo advindo da Wikipédia e semanticamente estruturado por *Knowledge Bases* é utilizado pelos principais mecanismos de busca no estabelecimento de uma *Web Pragmática*, com influências positivas a esse construto.

Palavras-chave: Conhecimento Enciclopédico. Wikipédia. *Web Pragmática*. Ciberespaço. Mecanismos de busca.

DECARLI, Gian Carlo. **The collaborative encyclopedic knowledge and their influences on the development of Pragmatic Web**. 2017. 154p. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual de Londrina, Londrina, 2017.

ABSTRACT

Through written and oral communication, encyclopedic or world knowledge has passed through different mankind periods. In recent history, encyclopedias have fully played the task of recording episodic and declarative knowledge of cultures, faunas, floras, individuals among many other phenomena around the world, converting them into encyclopedic knowledge. Established in the early 2000s, Wikipedia is one of the platforms that best represents the ideal of Social Web collaboration. In addition, it introduces the concept of free encyclopedia, in which everyone can collaborate for its expansion through production, monitoring and dissemination of its entries. Wikipedia has added the status of collaborative to the generated encyclopedic knowledge, becoming a great international knowledge base. In the Wikipedia source layer, named Social Web, there was a rupture of a linear model of content production and consumption, introduced by the previous layer, called Web 1.0. Thus, who consumed information also produced it and vice versa. However, the freedom creation provided by the Social layer brought problems to the information retrieval processes on the Web. The next layer, the Semantic Web, emerged with the purpose of structuring the data so that it could make sense for both men and machines. So, the Knowledge Bases were formed in the Semantic layer, responsible for linking correlated data from different Web pages in order to generate meaning for certain themes. The next layer, the Pragmatic Web aggregates the entire structural legacy of the predecessor layers to explore the meaning from the user's perception, regarding that the meaning is given from the subject in multiple contexts. In this way, the current Search Engines seem increasingly engaged under the Pragmatic Web idea. Technologies such as the Knowledge Graph help Search Engines to better structure and display factual information about certain topics in their Search Engine Results Pages. When observing this technological conjuncture, this research proposed to analyze whether the collaborative encyclopedic knowledge produced by Wikipedia was in some way influencing the Pragmatic Web development. For that, the methodological course was divided into three stages: The first described the course of collaborative encyclopaedic knowledge on Wikipedia; the second analyzed how Knowledge Bases structured knowledge; and the third investigated whether the search engines Bing, Google and Yahoo! availed the knowledge produced and structured collaboratively to improve their results and sustain the concept of Web Pragmatics. The results pointed out that the Wikipedia's collaborative encyclopedic knowledge, semantically structured by Knowledge Bases, is used by the main search engines in the Pragmatic Web establishment, with positive influences to this construct.

Key words: Encyclopedic Knowledge. Wikipedia. Pragmatic Web. Cyberspace. Search engines.

LISTA DE FIGURAS

Figura 1	– As dimensões da linguagem	22
Figura 2	– The semantic Web “layer cake”	45
Figura 3	– As dimensões Sintática, Semântica e Pragmática e as camadas da Web	55
Figura 4	– Síntese do percurso metodológico.....	71
Figura 5	– Auto-suggest para o termo “Tim Berners-Lee” na Wikipédia	73
Figura 6	– Conexões semânticas a partir do termo “Tim Berners-Lee” na Wikipédia.....	74
Figura 7	– Página “view source” do artigo de “Tim Berners-Lee” semi-protegida	79
Figura 8	– Infobox do verbete “Tim Berners-Lee” na Wikipédia	81
Figura 9	– Temas correlacionados a “Tim Berners-Lee” na seção “see also”	82
Figura 10	– Auto-suggest para o termo “England” na Wikipédia	83
Figura 11	– Resultados para “containing... England” na Wikipédia	84
Figura 12	– Infobox do verbete “England” na Wikipédia	87
Figura 13	– Resultados para “containing... Semantic Web” na Wikipédia	88
Figura 14	– Exemplo de aplicação de triplas RDF para o termo “Tim Berners-Lee” na DBpedia	97
Figura 15	– Relacionamentos semânticos entre as entidades de “Tim Berners-Lee” e “England” na DBpedia	99
Figura 16	– Relacionamentos semânticos entre as entidades de “Tim Berners-Lee”, “England” e “Semantic Web” na DBpedia	102
Figura 17	– Página do “item” de “Tim Berners-Lee” no Wikidata	105
Figura 18	– Auto-suggest do Bing para o termo “Tim Berners-Lee”	112
Figura 19	– Mosaico de SERPs do Bing para o termo “Tim Berners-Lee”	113
Figura 20	– Knowledge Graph do Bing para “Tim Berners-Lee”	114
Figura 21	– Auto-suggest do Bing para o termo “England”	116
Figura 22	– Mosaico de SERPs do Bing para o termo “England”	117
Figura 23	– Knowledge Graph do Bing para “England”	118
Figura 24	– Auto-suggest do Bing para o termo “Semantic Web”	119
Figura 25	– Mosaico de SERPs do Bing para o termo “Semantic Web”	120

Figura 26	– Knowledge Graph do Bing para “Semantic Web”	121
Figura 27	– Auto-suggest do Google para o termo “Tim Berners-Lee”	122
Figura 28	– Mosaico de SERPs do Google para o termo “Tim Berners-Lee”	123
Figura 29	– Knowledge Graph do Google para “Tim Berners-Lee”	124
Figura 30	– Auto-suggest do Google para o termo “England”	125
Figura 31	– Mosaico de SERPs do Google para o termo “England”	126
Figura 32	– Knowledge Graph do Google para “England”	127
Figura 33	– Auto-suggest do Google para o termo “Semantic Web”	128
Figura 34	– Mosaico de SERPs do Google para o termo “Semantic Web”	129
Figura 35	– Auto-suggest do Yahoo! para o termo “Tim Berners-Lee”	130
Figura 36	– Mosaico de serps do Yahoo! para o termo “Tim Berners-Lee”	131
Figura 37	– Knowledge Graph do Yahoo! para “Tim Berners-Lee”	132
Figura 38	– Auto-suggest do Yahoo! para o termo “England”	133
Figura 39	– Mosaico de SERPs do Yahoo! para o termo “England”	134
Figura 40	– Knowledge Graph do Yahoo! para “England”	135
Figura 41	– Auto-suggest do Yahoo! para o termo “Semantic Web”	136
Figura 42	– Mosaico de SERPs do Yahoo! para o termo “Semantic Web”	137

LISTA DE DIAGRAMAS

Diagrama 1 – Trajetória dos dados de “Tim Berners-Lee” da Wikipédia para DBpedia	96
Diagrama 2 – Trajetória dos dados de “England” da Wikipédia para DBpedia	98
Diagrama 3 – Trajetória dos dados de “Semantic Web” da Wikipédia para DBpedia	101
Diagrama 4 – Statements de “Tim Berners-Lee” no Wikidata	106
Diagrama 5 – Statements de “England” no Wikidata.....	108
Diagrama 6 – Statements de “Semantic Web” no Wikidata.....	109

LISTA DE QUADROS

Quadro 1	–	Requisitos De Um Projeto Enciclopedista.....	24
Quadro 2	–	Dados De Uma Nf-E Estruturados Em Xml.....	35
Quadro 3	–	Web 1.0 → Web 2.0.....	36
Quadro 4	–	Formatos De Serialização Rdf	44
Quadro 5	–	Síntese De Fundamentos Teóricos.....	65
Quadro 6	–	Classificação Do Verbetes “Tim Berners-Lee” Na Lista “Vital Articles”	76
Quadro 7	–	Esquema De Avaliação Da Qualidade De Um Artigo Wikipédia	77
Quadro 8	–	Seções De Conteúdo Do Verbetes “England” Na Wikipédia	85
Quadro 9	–	Classificação Do Verbetes “England” Na Lista “Vital Articles”	86
Quadro 10	–	Seções De Conteúdo Do Verbetes “Semantic Web” Na Wikipédia	89
Quadro 11	–	Dbpedia Extractors.....	94

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
APA	American Psychology Association
ARPANET	<i>Advanced Research Programs Agency Network</i>
ASCII	<i>American Standard Code for Information Interchange</i>
BD	Base de Dados
BLD	<i>Basic Logic Dialect</i>
CC	Ciência da Computação
CERN	<i>Conseil Européen pour la Recherche Nucléaire</i>
CI	Ciência da Informação
HTML	<i>HyperText Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IA	Inteligência Artificial
IRI	<i>Internationalized Resource Identifiers</i>
JSON	<i>JavaScript Object Notation</i>
KB	<i>Knowledge Base</i>
KBs	<i>Knowledge Bases</i>
KG	<i>Knowledge Graph</i>
KGs	<i>Knowledge Graphs</i>
LD	<i>Linked Data</i>
LOD	<i>Linked Open Data</i>
MB	Mecanismo de Busca
MBs	Mecanismos de Busca
OWL	<i>Web Ontology Language</i>
PRD	<i>Production Rule Dialect</i>
RDF	<i>Resource Description Framework</i>
RDFS	<i>Resource Description Framework Schema</i>
RI	Recuperação da Informação
RIF	<i>Rule Interchange Format</i>
RSS	<i>Really Simple Syndication</i>
SERP	<i>Search Engine Results Page</i>
SERPs	<i>Search Engine Results Pages</i>
SGML	<i>Standard Generalized Markup Language</i>
SPARQL	<i>Simple Protocol and RDF Query Language</i>
SRI	Sistema de Recuperação de Informação
TCP/IP	<i>Transmission Control Protocol/Internet Protocol</i>
TED	<i>Technology, Entertainment and Design</i>
TIC	Tecnologias de Informação e Comunicação
UCS	<i>Universal Coded Character Set</i>
URI	<i>Uniform Resource Identifier</i>
WWW	<i>World Wide Web</i> ou <i>Web</i>
XML	<i>eXtensible Markup Language</i>

SUMÁRIO

1	INTRODUÇÃO	15
1.1	OBJETIVO GERAL	18
1.1.1	Objetivos Específicos	18
2	CONHECIMENTO ENCICLOPÉDICO	19
3	CIBERESPAÇO	27
3.1	WEB SINTÁTICA	28
3.1.1	Wikipédia	37
3.2	WEB SEMÂNTICA.....	40
3.2.1	Knowledge Bases	48
3.2.1.1	DBpedia	50
3.2.1.2	Wikidata	51
3.3	WEB PRAGMÁTICA	52
3.4	MECANISMOS DE BUSCA	56
3.4.1	Knowledge Graphs	61
4	PERCURSO METODOLÓGICO	66
5	APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	72
5.1	WIKIPÉDIA	72
5.2	DBPEDIA E WIKIDATA	92
5.2.1	DBpedia	103
5.2.1	Wikidata	103
5.3	BING, GOOGLE E YAHOO	110
5.3.1	Bing	111
5.3.2	Google	122
5.3.3	Yahoo!	130
6	CONSIDERAÇÕES FINAIS	139
	REFERÊNCIAS	142

1 INTRODUÇÃO

Nas últimas décadas, as Tecnologias de Informação e Comunicação (TIC) passaram por constantes transformações. Entre elas, destaque para os computadores, que antes ocupavam a robustez física de andares de prédios acadêmicos e corporativos, mas que posteriormente, em versões “*micro*”, tornaram-se presentes também no cotidiano das pessoas. Por serem passíveis de aplicação aos mais variados contextos de uso, os computadores são importantes ferramentas de trabalho, estudo, entretenimento, entre outras atividades impulsionadas, em especial, pelo aprimoramento da *World Wide Web* (WWW ou *Web*).

Atualmente, os computadores possuem as mais variadas configurações físicas e lógicas para múltiplas finalidades. Dessa forma, algumas de suas ramificações como *smartphones*, *smartwatches*, *smart TVs*, *tablets*, *ultrabooks*, entre outras TIC, possibilitam que seus usuários acessem de qualquer lugar serviços e informações *Web* por navegadores ou aplicativos específicos. Esses aspectos de mobilidade, acessibilidade, pluralidade de sistemas e conectividade se tornaram familiares apenas recentemente, a partir da evolução dos dispositivos computacionais, mas essencialmente pelas mudanças ocorridas no contexto *Web*.

Desde seu surgimento, no início dos anos 1990, a *Web* passou por inúmeras readequações estruturais para que pudesse acompanhar a evolução de tecnologias análogas como os computadores. Seu propósito inicial de interoperabilidade transpassou por diferentes períodos, comumente interpretados como dimensões da *Web*; são elas: *Web* Sintática, *Web* Semântica e *Web* Pragmática. Contudo, não é uma regra compreendê-las dessa maneira, pois essas nomenclaturas correspondem a associações realizadas para cada dimensão da *Web* com base em aspectos linguísticos.

A *Web* Sintática abrange as duas primeiras camadas da *WWW*; são elas: *Web* 1.0 e *Web* 2.0 ou Social. A terminologia “Sintática” advém da forma como a informação é indexada e recuperada nessas camadas, por considerar apenas aspectos referentes à formação de palavras e frases nesses processos, restringindo aos humanos a tarefa de dar sentido às informações. Na *Web* 1.0, o fluxo de produção e consumo de conteúdo era linear, ou seja, do produtor para o receptor; na *Web* 2.0 houve a ruptura desse modelo, pois o consumidor também se tornou o produtor e vice-versa.

Originada em 2001, a Wikipédia é uma das plataformas que melhor representa o propósito da *Web Social*. Em seu *website* pessoas do mundo inteiro que dispõem de acesso à *internet* podem colaborar na criação e/ou expansão de conteúdos de seus verbetes. Isso possibilita que diferentes sujeitos colaborem para a construção de um mesmo texto, seja escrevendo ou monitorando seu conteúdo. Assim, o conhecimento de mundo de cada indivíduo corrobora para tornar a Wikipédia uma grande base internacional de conhecimento enciclopédico colaborativo na *Web*.

Na *Web Semântica* ou 3.0, o enfoque está em corrigir as lacunas deixadas pela camada anterior, isto é, busca-se ir além do modo formal que se volta para a maneira como são apresentadas palavras, frases e expressões, e vislumbra a importância do significado por elas produzido para que seja passível de interpretação por humanos e máquinas por meio de dados estruturados e ligados na rede. Nesse sentido, as chamadas *Knowledge Bases (KBs)* são exemplos de iniciativas correspondentes à dimensão da *Web Semântica*, pois estruturam em seu âmago o conhecimento disperso na rede, mediante conexões semânticas entre diferentes informações.

A *Web Pragmática*, por sua vez, busca agregar todas as melhorias realizadas e o legado estrutural de suas antecessoras com o propósito de explorar o significado segundo a perspectiva do usuário, ou seja, considera que a produção do sentido se dá a partir do sujeito em múltiplos contextos. Dessa forma, os Mecanismos de Busca (MBs) atuais estão cada vez mais engajados nessa concepção de uma *Web Pragmática*. Portanto, além de resultados, os MB buscam aperfeiçoar a interação humano-computador com a implementação de políticas e tecnologias complementares que os tornam mais rápidos, coerentes, contextualizados e com resultados que façam sentido aos seus usuários.

Nessas circunstâncias, tecnologias como o *Knowledge Graph (KG)* auxiliam os MBs a melhor estruturar e exibir informações fatuais sobre determinados temas em suas *Search Engine Results Pages (SERPs)*. Sua composição é apenas o produto de um longo processo de estruturação de informações na *Web*, iniciado de maneira colaborativa na Wikipédia como um verbete que posteriormente passa pelas *KBs* como DBpedia e Wikidata, nas quais é transformado semanticamente, mediante ligações com outros dados da rede, em uma entidade que, por fim, é aproveitada pelas *SERPs* dos MBs, tal como Bing, Google e Yahoo!, entre outros.

A observação dessa conjuntura tecnológica presente no Ciberespaço fez emergir a inquietação para elaborar esta pesquisa, mais precisamente, em buscar responder o seguinte questionamento: quais as influências do conhecimento enciclopédico colaborativo para o desenvolvimento da *Web Pragmática*? A elaboração deste estudo se firmou sob os fundamentos da Ciência da Informação (CI). Nesse enquadramento, muitos pontos aqui levantados são publicados há anos por pesquisadores da CI. Porém, o aspecto que tornou esta investigação inédita para área é o modo como esses assuntos são tratados no decorrer da dissertação, visto que essas tecnologias foram abordadas a partir de uma visão holística de seu funcionamento no contexto *Web*.

O ineditismo desta pesquisa harmoniza com a perspectiva de Santos e Vidotti (2009, não paginado), a qual propõem que a CI disponibilize “[...] mais espaços de investigação que permitam compreender as Tecnologias de Informação e Comunicação [...]”, com intuito de apresentar “Um olhar diferente! Não porque se olhava errado, mas por que se pode olhar diferente.” Assim, um único objeto de estudo pode se desmembrar em infinitas possibilidades de pesquisa ao ser analisado sob diferentes ângulos. Nesse sentido, o “olhar diferente” proposto nesta investigação observou como distintas tecnologias da *Web* confluem sinergicamente para a instauração de uma nova camada da *WWW*.

Esta pesquisa também tomou como base os pilares responsáveis pela existência e manutenção da CI que, segundo Saracevic (1996), são: a interdisciplinaridade, o elo com a tecnologia da informação e a dimensão social e humana. Assim, por intermédio da CI tornou-se possível realizar breves incursões em disciplinas de diferentes áreas do conhecimento para a assimilação de temáticas e de determinados fenômenos aqui abordados, fato esse que evidenciou o caráter interdisciplinar desta investigação.

A ligação com a tecnologia da informação está presente ao longo de toda a dissertação, seja por efeito dos levantamentos teóricos sobre a temática ou até devido ao próprio contexto do objeto de análise. Esta pesquisa interpretou os acontecimentos colaborativos internacionais que envolvem a construção dos verbetes da Wikipédia e a conversão dessas informações em entidades DBpedia e Wikidata como aspectos de sua dimensão social e humana. Dessa maneira, o emprego da tríade de sustentação da CI ao contexto desta pesquisa foi substancial para seu desenvolvimento.

Nesse sentido, o percurso metodológico adotado subdividiu-se em três etapas de levantamento e análise de dados, nas quais foram monitorados termos distintos, mas convergentes semanticamente entre si, são eles: “Tim Berners-Lee”, “*England*” e “*Semantic Web*”. Na primeira fase, os termos acima tiveram seu percurso monitorado na Wikipédia, sob a condição de verbetes; na segunda etapa foram submetidos à análise na DBpedia e Wikidata enquanto entidades. No terceiro estágio, os termos foram inseridos nas caixas de pesquisa dos seguintes MBs: Bing, Google e Yahoo!. Assim, o ciclo de análise e levantamento de dados desta dissertação foi finalizado, ao vislumbrar as formas de utilização do conhecimento enciclopédico colaborativo pelos MBs elencados.

Com base nos procedimentos de análise realizados nesta investigação, foi constatado que o conhecimento enciclopédico colaborativo advindo da Wikipédia e semanticamente estruturado por *Knowledge Bases*, que conseqüentemente é empregado na apresentação dos resultados de mecanismos de busca, mostrou influenciar positivamente o estabelecimento de uma *Web Pragmática*.

1.1 OBJETIVO GERAL

Analisar o conhecimento enciclopédico colaborativo e suas influências no desenvolvimento da *Web Pragmática*.

1.1.1 Objetivos Específicos

- a) descrever o curso do conhecimento enciclopédico colaborativo na *Web*;
- b) analisar como as *Knowledge Bases* estruturam o conhecimento enciclopédico colaborativo;
- c) investigar se os principais mecanismos de busca da atualidade utilizam o conhecimento enciclopédico colaborativo estruturado pelas *Knowledge Bases*.

2 CONHECIMENTO ENCICLOPÉDICO

No indivíduo, o significado é produzido a partir de interações com o mundo que o cerca. Esse construto não se dá apenas na virtualidade do sujeito, mas com e pela linguagem, reflexo das diferentes experiências humanas. Marcondes (2010, p. 8) esclarece que a linguagem é “[...] uma característica humana que é pré-requisito para evolução da cultura [...]” e tem se tornado gradativamente “[...] mais sofisticada no sentido de ser capaz de expressar o pensamento simbólico, ou seja, capaz de contar, descrever, reapresentar, socializar situações na ausência delas [...]”. Nesse sentido, a linguagem transpassa barreiras temporais, geográficas e culturais, mediante o compartilhamento de experiências comunicacionais entre os humanos.

O conhecimento de mundo que habita o indivíduo, em consonância com suas capacidades cognitivas, transforma o significado em algo enciclopédico, ao considerar a familiaridade com a coisa e não apenas a sua definição (GEERAERTS, 2006). Ao definir, o indivíduo preocupa-se com o significado em si e não com processos de referência a um estado de mundo (ECO, 2014). Assim, definir se refere a dizer o que é uma coisa e assumir de fato o que ela deve ser, mas sem que seja necessário provar, portanto, o conhecimento de mundo é substancial para a assimilação de um conceito.

Segundo Croft e Cruse (2004), o significado de uma palavra se dá a partir de uma breve inferência descritiva do conceito, complementado pelo conhecimento de mundo do sujeito. Para Koch e Elias (2008, p. 42), o conhecimento de mundo ou enciclopédico se refere aos

[...] conhecimentos gerais sobre o mundo – uma espécie de *thesaurus* mental – bem como a conhecimentos alusivos a vivências pessoais e eventos espaço-temporalmente situados, permitindo a produção de sentidos.

Nessa perspectiva, o conhecimento enciclopédico está intimamente relacionado a fatores empíricos e ao saber contextual para além de descrições genéricas. De acordo com Geeraerts (2006), o conhecimento enciclopédico não se restringe apenas à definição geral do que é, por exemplo, um pássaro, pois considera também a língua, a identidade cultural e social de cada grupo. Dessa

forma, a junção desses fatores influencia no conhecimento que as pessoas associam à palavra pássaro, que pode diferir conforme a cultura à qual pertencem.

Para Koch (1977 *apud* KOCH; ELIAS, 2008, p. 63-64) o conhecimento enciclopédico se divide em dois tipos, o primeiro, denominado declarativo é o “[...] conhecimento que recebemos pronto, que é introjetado em nossa memória ‘por ouvir falar’ [...]”, o segundo, chamado de episódico, está relacionado ao “[...] conhecimento adquirido através da convivência social [...] sobre as diversas situações e eventos da vida cotidiana.”

Pertinente, essa visão dicotômica do conhecimento enciclopédico (o que nos contam e o que vivenciamos) se mistura em meio à linguagem formal e informal na busca pela compreensão. Na linguagem formal, aqui entendida como a escrita, as enciclopédias têm desempenhado durante anos o papel de transformar o conhecimento episódico em declarativo, por intermédio da representação e, assim, disseminá-los ao redor do globo. Contudo, antes de adentrar a discussão sobre o conhecimento enciclopédico registrado, no caso as enciclopédias, torna-se importante apresentar, ainda que brevemente, um discernimento sobre enciclopédias e dicionários, se eles trabalham de modo separado ou são, aos olhos da teoria, o mesmo objeto.

Os verbetes dos dicionários descrevem o conhecimento lexical. Porém, para alguns nomes próprios, acaba-se recorrendo ao conhecimento enciclopédico para que o leitor, além do significado, tenha também a identificação do referente. Nessa perspectiva, o dicionário assume uma posição intermediária entre o conhecimento lexical e o enciclopédico (KIEFER, 1988). Com o propósito de esclarecer sobre o que é o conhecimento lexical, Meijs e Vossen (1991, p. 113, tradução nossa) o definem como “[...] o tipo de conhecimento que pode ser expresso em palavras [...]”

Cavazza e Zweigenbaum (1995) argumentam que nos dicionários o conhecimento linguístico em associação com outras palavras (sinônimos, antônimos, hipônimos etc.) auxilia na definição de um termo. Assim, o conhecimento linguístico abrange

[...] o conhecimento gramatical e lexical. Baseados nesse tipo de conhecimento, podemos compreender: a organização do material linguístico na superfície textual; o uso dos meios coesivos para efetuar a remissão ou sequenciação textual; a seleção lexical adequada ao tema ou aos modelos cognitivos ativados. (KOCH; ELIAS, 2008, p. 40).

Segundo Nunberg (1978 *apud* MURPHY, 2003), os dicionários contemplam apenas o significado, e não os sentidos nos quais as palavras podem ser empregadas. Nessa esteira, Murphy (2003, p. 17, tradução nossa) ressalta que

[...] o léxico deve incluir apenas as informações de definição (núcleo semântico), deixando a informação enciclopédica (conceitual) na esfera do que sabemos sobre as coisas, e não o que nós sabemos sobre as palavras para as coisas. Assim, por exemplo, que os cães são usados como puxadores de trenó pode ser algo que eu sei sobre cães, mas não faz parte do significado de cão.

Devido a isso, o dicionário apresenta informações relevantes para que um termo e seu conceito possam ser diferenciados dos demais. Interessa ao dicionário expressar o que é um cão, mas não especificar atributos a ele vinculados como o latido, pois essa é uma designação da enciclopédia (ECO, 2014).

Desse modo, faculta-se ao dicionário a tarefa de apresentar teoricamente para cada verbete listado informações que sejam pertinentes, pontuais, relevantes, resumidas e gerais, porém precisas, que auxiliarão o leitor na construção do significado, sem que seja necessário contextualizar esse conhecimento. Nessa direção, um dicionário não define um termo, mas apenas indica a maneira logicamente correta para empregá-lo (ECO, 2014).

Ao dicionário, incube-se a tarefa de armazenar informações sintáticas formais, morfológicas e especificidades semânticas, mas ele também pode ser considerado parte da enciclopédia por incluir algumas informações enciclopédicas. O dicionário não deve ser visto apenas como um inventário de palavras, mas interpretado como um índice cultural da linguagem. Ao contrário do dicionário, a enciclopédia parece não invadir a lexicografia (ALLAN, 2006).

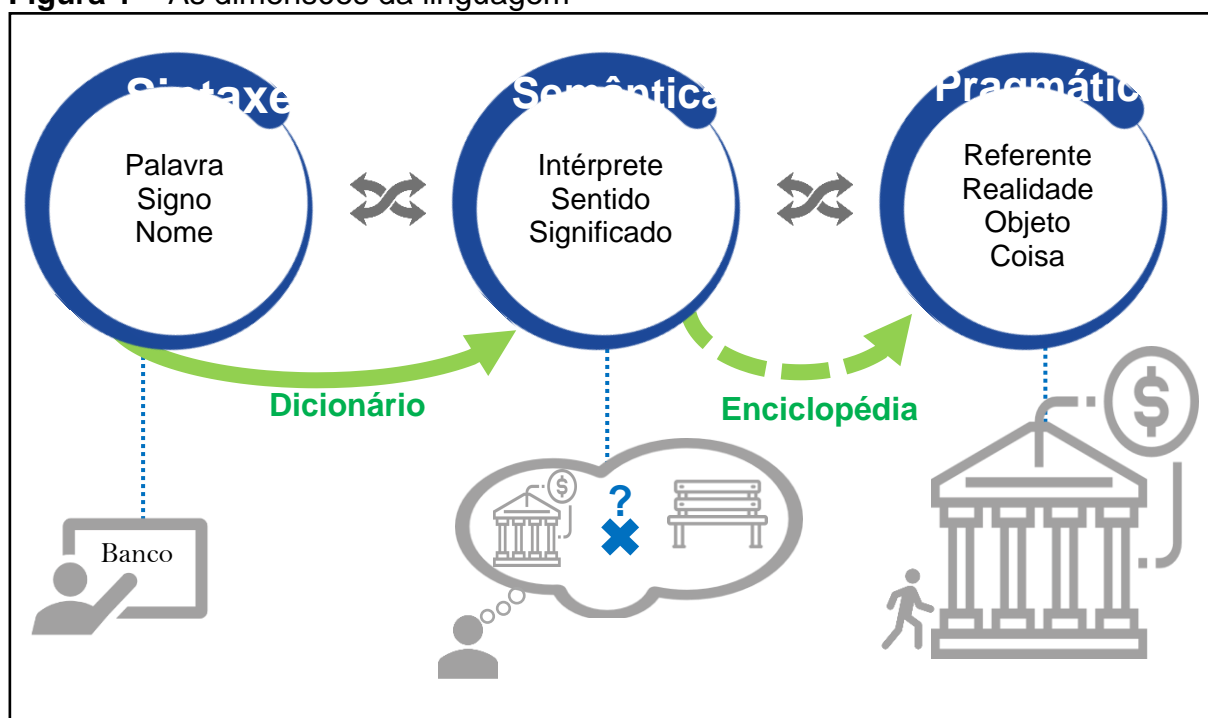
Pombo (2002) adverte que, apesar de semelhantes em alguns aspectos, dicionários e enciclopédias não podem ser confundidos. Enquanto o dicionário busca codificar de forma consistente a linguagem, a enciclopédia almeja o significado e a referência ao representar coisas e eventos do mundo. Nessa perspectiva, a Figura 1 sinaliza as semelhanças e divergências entre dicionários, enciclopédias, Sintaxe, Semântica e Pragmática.

O caminho percorrido pelo conhecimento linguístico existente no dicionário parte da palavra para o significado, isto é, ao indicar a palavra “Banco” o dicionário deve disponibilizar subsídios informacionais suficientes para que o

“Intérprete” seja capaz de compreender o significado desse termo, sem que seja necessário um contato anterior com o mesmo. A perspectiva Semântica adotada na Figura 1 corresponde a uma de suas concepções, portanto não deve ser compreendida como a única. Dessa maneira, sua descrição pode indicar que “Banco” é uma instituição financeira ou pode ser também um objeto desenvolvido para que as pessoas possam se sentar.

Em um dicionário o fluxo de compreensão perpassa pela Sintaxe (palavra) e a Semântica (significado), enquanto na enciclopédia predomina a relação entre a palavra e Pragmática (referente). Isso não os torna excludentes, pois na enciclopédia também há o conhecimento semântico, trata-se apenas do enfoque dado para cada caso.

Figura 1 – As dimensões da linguagem



Fonte: o próprio autor com base em Haiman (1980).

A enciclopédia busca um nível de detalhamento maior do contexto que envolve seus verbetes. Dessa maneira, enciclopédia e Pragmática (coisa) trabalham sob o mesmo ciclo de assimilação, ou seja, o sentido se manifesta a partir da proximidade com o objeto e necessita de maiores detalhes do contexto. Para Haiman (1980), a Pragmática tem capacidade de expansão para além do usuário, contempla também o contexto, o mundo em geral.

O dicionário foi considerado para este estudo um dispositivo tradutor e organizador da linguagem em sintaxe, que busca listar palavras e apresentar descrições concisas e objetivas, além de indicar o emprego correto destas em relação ao objeto por elas representado. Com base nisso, são apresentados adiante algumas considerações a respeito de movimentos enciclopedistas de maior notoriedade nos últimos tempos e suas contribuições para o desenvolvimento de projetos subsequentes.

O termo “enciclopédia” é de origem grega, mas se popularizou durante o século XVI, quando era comumente utilizado para se referir à totalidade de conhecimentos do mundo. No entanto, durante o Império Romano há uma obra que merece destaque: intitulada de “*Elder’s Historia Naturalis*”, é referente aos 37 livros de Plínio que descreviam aproximadamente 20 mil fatos com a colaboração de 500 autores. A importância dessa obra está na contribuição a projetos enciclopedistas subsequentes, pois inseriu o formato de organização do geral para o específico, baseando-se em princípios aristotélicos (ECO, 2014).

Precursor, Francis Bacon trouxe com o “*Novum Organum* (1620)” uma ordem razoavelmente lógica, expondo os assuntos do geral para o específico. Depois de Bacon, a enciclopédia se expandiu para a descoberta para além do que já é conhecido. Nessa perspectiva, o pensador Emanuele Tesauro’s considera um índice como um grande labirinto, no qual é possível realizar conexões com diferentes objetos. Nesse labirinto não há divisões lógicas, o que possibilita a descoberta de novos elementos e formulação de novos conhecimentos (ECO, 2014). Para a época, a ideia de labirinto parecia até confusa e utópica, mas, avançando para a atualidade, os hipertextos e *hiperlinks* parecem realizar sem muitos empecilhos a complexidade do labirinto do conhecimento.

Outro importante legado enciclopedista advém das formulações de Leibniz, que tinha consciência de que a enciclopédia não era apenas uma exposição exaustiva de todo o conhecimento, mas que deveria ter também capacidade para ordenar informações dispersas e caóticas, eliminar redundâncias e repetições para que pudesse registrar os pensamentos e experiências alcançados ao longo dos séculos (POMBO, 2002). Esses ideais de Leibniz podem ser interpretados como políticas para o desenvolvimento de um conteúdo mais coerente, que evite confusões no momento da recuperação dessas informações.

As pesquisas de Leibniz também eram focadas na diferenciação entre dicionários e enciclopédias. Condescendente com a proposta de Francis Bacon de que uma enciclopédia é um trabalho em constante construção, também considerava a inclusão do conhecimento que não está escrito, aquele que está disperso em meio as pessoas. Para Leibniz, a enciclopédia ideal deveria conter referências cruzadas para que os elementos pudessem ter diferentes relações e serem interpretados por diferentes pontos de vista. Essas características podem ser interpretadas como uma enciclopédia polidimensional com inúmeras provisões para várias conexões transversais (ECO, 2014).

Leibniz militou em prol da ciência como unidade, de aspiração filosófica, mas que somente se concretizaria por intermédio da enciclopédia. Em seu discurso, pregou pela colaboração para a construção de uma comunidade de conhecimento, na qual a enciclopédia seria um conglomerado de conhecimentos a serviço do progresso da ciência. Paralelo a isso, também deu especial atenção à construção de uma linguagem científica a partir de observações sobre a enciclopédia. As explanações de Leibniz a respeito do ideal de enciclopédia possivelmente venham ao encontro das propostas atuais de enciclopédias eletrônicas, do hipertexto e da *internet* (POMBO, 2002).

A enciclopédia é o modelo efetivo da colaboração entre inteligências, pois não há enciclopédia sem uma cooperação organizada (POMBO, 2006). Nessa perspectiva, no Quadro 1, Pombo (2002) aponta nove requisitos que devem compor um projeto enciclopédico para que possa ser considerado bem estruturado.

Quadro 1 – Requisitos de um projeto enciclopedista

1	A enciclopédia destina-se a ser uma exposição completa de todo o conhecimento conquistado pela humanidade e disponível a um certo momento histórico.
2	Enciclopédia não é dicionário.
3	Apesar de não ser dicionário, os dois são parecidos. Possuem textos descontínuos feitos de segmentos ou entradas independentes, em ordem alfabética ou em quadros conceituais ou temáticos.
4	A objetividade material da enciclopédia tem uma condição ilimitada , pois apesar de finita, seus assuntos, principalmente com a inserção de hipertexto, a torna ilimitada.

Quadro 1 – Requisitos de um projeto enciclopedista

(conclusão)

5	A natureza descontínua da enciclopédia é reforçada pelo fato de que ela frequentemente inclui materiais não-lineares como desenhos, diagramas, gravuras, ilustrações, mapas, estatísticas, planos, esquemas, fotografias e tabelas de todos os tipos.
6	Enciclopédia exerce forte papel cultural e educativo .
7	Enciclopédia é trabalho coletivo .
8	Enciclopédia nunca é uma quantidade de elementos descontínuos vindos a partir de diferentes fontes, não é uma miscelânea, mas uma apresentação ordenada .
9	Enciclopédia tem sempre um papel prospectivo .

Fonte: o próprio autor com base em Pombo (2002).

Segundo Pombo (2002), as ideias de Leibniz sobre a enciclopédia contemplam todos os requisitos que um projeto enciclopédico deveria conter. Esses ideais são absorvidos, mais tarde, por D' Alembert na elaboração da *Encyclopédie*. Audacioso, esse projeto enciclopedista foi desenvolvido durante o período Iluminista com o propósito de apresentar a totalidade do conhecimento humano sobre uma perspectiva crítico-científica (ECO, 2014). Os iluministas que participaram de sua construção buscavam desmistificar o senso relacionado com a autoridade, com os preconceitos e com as superstições e sugerir uma nova memória mais racional e humana, formada pelos conhecimentos em ciências, artes e ofícios (SERRA, 1999). Assim, a enciclopédia portava a importante característica do Iluminismo, a de disseminar a perspectiva crítica do conhecimento (MORAES, 2013).

Historicamente, as novidades que acompanharam o fim do século XX foram intimidadoras ao movimento enciclopedista. Entretanto, enquanto se pensava em encerrar um ciclo, a enciclopédia ressurgiu para renovar e reorganizar suas estruturas e interesses. Na enciclopédia eletrônica (*on-line* ou CDRom), por exemplo, a leitura foi substituída pela navegação, a manipulação entre volumes pelos “cliques” e a linearidade pela possibilidade de percorrer diversos caminhos livremente. Além disso, ela parece disponibilizar todo o conhecimento para todas as perguntas possíveis, com respostas instantâneas (POMBO, 2006).

Provavelmente, os pensadores envolvidos com o movimento enciclopedista ao longo dos séculos teriam boa parte de suas aspirações atendidas com o desenvolvimento das enciclopédias eletrônicas. Assim, segundo Pombo (2006), mais do que manter a informação, a *Web* indexa e organiza o conhecimento

mundial. Em sua hipertextualidade, disponibiliza mecanismos que possibilitam filtrar e selecionar a informação disponível, que, de certo modo, se relaciona com a enciclopédia de Diderot e D' Alembert na ordenação e encadeamento de conhecimentos. A própria *Web* pode ser considerada uma enciclopédia, por ser

[...] um sistema hipertextual integrado de milhares de sub-sistemas interconectados, que não contém apenas sumários e palavras-chave, mas artigos de toda a espécie, sem limite de extensão, cartas, gráficos, mapas, discos, bandas sonoras, microfilmes, tabelas e massas formidáveis de conhecimentos científicos de todo o tipo, médicos, demográficos, económicos, técnicos, políticos, informações relativas a negócios, direito ou viagens, artigos constantemente actualizáveis por equipas de especialistas a nível planetário, estudos especializados e de carácter geral, representativos, em cada momento, do estado total do conhecimento. (POMBO, 2006, p. 23).

Ao longo dos séculos, a enciclopédia passou por diferentes propostas de organização, classificação, estruturação e apresentação de seu conteúdo. No entanto, ideais como o de Plínio, Francis Bacon, Leibniz, Diderot e D' Alembert, entre outros pensadores, prevalecem até as versões mais modernas dessas obras, como a colaboração, a busca pela totalidade do conhecimento de mundo, o ponto de vista crítico, a imparcialidade e a liberdade do leitor em explorar o conteúdo. Atualmente, essas aspirações parecem sustentar o *modus operandi* de enciclopédias colaborativas *on-line*, em especial, a Wikipédia, por sua abrangência, o que será abordada na seção 3.1.1 deste trabalho.

3 CIBERESPAÇO

Em um local impreciso de outra realidade, existe um universo não palpável propiciado pela *internet*, sobre o qual não se pode dizer se deixa de existir ao desligar de um computador, mas que também é um espaço de significações, interações, comunicações e de vida em sociedade, chamado de Ciberespaço (MONTEIRO, 2007). O Ciberespaço não deve ser visto apenas como uma infraestrutura material para a comunicação digital, mas sim como um espaço informacional usufruído pelas pessoas que nele navegam (LÉVY, 1999).

Segundo Santaella (2004, p. 40), o ciberespaço pode ser entendido como “[...] um mundo virtual global coerente, independentemente de como se acede a ele e como se navega nele [...] é o espaço que se abre quando o usuário conecta-se com a rede.” Nesse sentido, parece haver inúmeros acessos distribuídos pelo mundo real (não virtual) com os quais se torna possível adentrar esse espaço, seja por meio do uso de um computador, *tablet*, *smartphone* ou *smartwatch* conectado à rede. Acrescentando à noção de Ciberespaço, para Ho Kim (2004, p. 216)

[...] quanto mais humanizamos e tornamos ‘amigável’ a nossa relação com o ciberespaço, por meio de simulações que imitam a nossa realidade não-virtual, mais nos tornamos cibernéticos. A contrapartida da naturalização do ciberespaço é que nos tornamos, também, extensão dele: à medida que a virtualidade se transforma em campo de ação prática, cada vez mais a realização total do ser humano prescinde de sua inserção como coisa virtual do ciberespaço.

Desse modo, é inegável a ubiquidade do Ciberespaço no cotidiano das pessoas, haja vista que, ao passo que se tenta humanizar o ciberespaço, ele cordialmente transforma humanos em parte de sua estrutura para que, com fluidez, possam transitar discretamente entre duas realidades. Por outro lado, fora do contexto de relacionamento ciborgue com o Ciberespaço, Monteiro e Fidencio (2013, p. 38) vislumbram-no como “[...] um espaço semântico/semiótico, onde o signo se dá em várias semióticas, desterritorializado, nômade, em escrita espacializada e com a memória em constante modificação.” Essa perspectiva denota toda a fluidez do ciberespaço, um espaço informacional extremamente dinâmico.

Segundo Lazzarin, Netto e Sousa (2015, p. 24), com o ciberespaço abre-se “[...] espaço para novas relações e valores entre as pessoas,

disponibilizando um fluxo de informações em diversos níveis, assim como potencializando o acesso a outros mundos [...]” Atualmente, o ciberespaço se tornou mais onipresente no cotidiano dos indivíduos, podendo ser acessado a qualquer momento por um dispositivo de bolso ou um relógio inteligente. A *Web*, tecnologia desenvolvida em seu âmago, é um veículo informacional mundial onde se configuram novas relações de produção e consumo de conhecimento entre humanos e entre máquinas e humanos.

3.1 WEB SINTÁTICA

Na década de 1960, algumas pesquisas militares resultaram na criação da *Advanced Research Programs Agency Network* (ARPANET), que aperfeiçoou a comunicação científica estadunidense, tendo como ponto mais forte a troca de *e-mails*. Além disso, incorporou-se o TCP/IP (*Transmission Control Protocol/Internet Protocol*), responsável pela conexão de redes locais com redes amplas de comunicação, que também era utilizado para especificar ações como as de transferência de arquivos ou envio de *e-mails* (ALESSO; SMITH, 2006).

Em 1980, os computadores tornaram-se familiares para pessoas e empresas, não contemplando apenas o contexto científico. Assim, em 1990 a ARPANET foi desativada, deixando para trás uma rede de conexões repleta de usuários, denominada atualmente de *Internet* (ALESSO; SMITH, 2006).

Porém, em 1989, Tim Berners-Lee, então cientista do CERN (*Conseil Européen pour la Recherche Nucléaire*), apresentou um projeto visando aprimorar e simplificar a troca de informações entre os pesquisadores do instituto. Assim, surgiu o “Bones”, nome dado ao protótipo do que viria a ser mais tarde conhecido como a *World Wide Web* (WWW). Seu objetivo era propiciar uma navegação hipertextual no intuito de minimizar a perda de dados em processos de gestão da informação no instituto (BERNERS-LEE, 1989; 1996).

A implementação da *Web* no Ciberespaço estabeleceu uma série de novos padrões de comunicação entre pessoas e máquinas no ambiente virtual. Para Berners-Lee (1996) esses padrões eram necessários para obter o máximo de flexibilidade e interoperabilidade entre diferentes sistemas. Assim, criou-se o *Uniform Resource Identifier* (URI), uma sequência compacta de caracteres que identifica um determinado recurso na rede; o *Hypertext Transfer Protocol* (HTTP), que percorre a

Web como um protocolo de requisição de informações hipertextuais, por meio do qual o usuário requisita uma determinada URI e; o *Hypertext Markup Language* (HTML), formato que padroniza o modo como os hipertextos devem ser disseminados na *Web* (BERNERS-LEE, 1996; BERNERS-LEE; FIELDING; MASINTER, 2005).

Khare e Rifkin (1997) complementam essa caracterização ao afirmarem que a linguagem HTML desenvolvida por Berners-Lee se apropriou de conceitos de marcação estrutural da *Standard Generalized Markup Language* (SGML), uma norma vigente do ano de 1986 que era utilizada para a validação de documentos e a gestão de grandes repositórios de informações.

A *Web* cresceu de maneira exponencial rapidamente. Assim, questões jurídicas que envolviam a posse dessa tecnologia necessitaram ser repensadas. Para atingir seu verdadeiro potencial sem que pessoas precisassem pedir permissão ou pagar para usá-la, foi necessária a quebra de sua patente em 1993, o que liberou seu uso em escala global (WEB FOUNDATION, 2017).

Com o passar do tempo, a solução apresentada para a resolução de problemas internos do CERN foi popularizada com e pela *Internet*. Aghaei, Nematbakhsh e Farsani (2012) esclarecem que a WWW e a *Web* são a mesma coisa, mas não devemos confundí-las com a *Internet* que, por sua vez, é um sistema tecno-social de interação humana baseada em redes tecnológicas. Nesse sentido, Monteiro e Fidencio (2013, p. 38) atentam que “[...] a *Web* não é tudo, mas é o principal constructo (dobra) do ciberespaço e vem crescendo a passos largos.”

Desde sua adoção internacional, provida pela evolução das TIC, a *Web* recebeu diferentes adjetivações de acordo com o modo como as informações são produzidas, organizadas, disseminadas e recuperadas no Ciberespaço. Nesta pesquisa, essas fases da *Web* são interpretadas na perspectiva de Vechiato e Vidotti (2014) que, com base em Monteiro (2012), as caracterizam a partir do conceito de dobras semióticas como pertencentes as dimensões da linguagem, compreendendo-as como *Web* Sintática (*Web* 1.0 e 2.0 ou Social), *Web* Semântica (*Web* 3.0) e *Web* Pragmática (*Web* 4.0).

Nesse sentido, uma breve incursão na área da Linguística se fez necessária para vislumbrar os aspectos que envolvem a formação dessas camadas. A Linguística não se atenta a uma língua ou ao conjunto delas, nem muito menos a padrões normativos para a preservação de uma língua; trata-se, sim, de uma ciência

empírica que estuda a “[...] língua como um fenômeno natural.” (VIOTTI, 2008, p. 8). A Linguística é “[...] o estudo dos códigos usados pelas pessoas para se comunicarem, e da capacidade inata que nos permite levar a efeito essa atividade [...]” (PERINI, 2010, p. 5). Portanto, ao contemplar aspectos orais e escritos, a Linguística abrange a linguagem sob diferentes perspectivas, produzindo estudos complexos e profundos sobre a língua.

Bem como acontece em outras ciências, a Linguística também possui subáreas responsáveis pela produção de estudos específicos sobre muitas temáticas, o que não será aprofundado aqui, pois, conforme indicado anteriormente, a aproximação com a Linguística nesta pesquisa decorre da necessidade de evidenciar as principais diferenças entre as camadas da *Web*. Dessa forma, a Sintaxe pertence a uma tripartição composta também por Semântica e Pragmática, que são apresentadas por Morris (1933 *apud* ILARI, 2000, p. 112) com o objetivo de indicar “[...] três pontos de vista distintos que podem orientar o estudo de uma linguagem.” Assim,

Ao olhar para uma linguagem qualquer, podemos querer inventariar as expressões simples que lhe são próprias e estabelecer como se combinam para formar expressões complexas: estaremos analisando aspectos de sua **sintaxe**. Além da combinatória possível das expressões, isto é, das relações que os sinais mantêm entre si, podemos querer considerar as relações que eles mantêm com os objetos e com as situações do mundo: estaremos estudando esses mesmos sinais **semanticamente**. Podemos, por fim, querer incluir em nosso estudo mais um fator de complexidade: os interlocutores e a interação entre os interlocutores: estaremos adotando uma perspectiva **pragmática**. (ILARI, 2000, p. 112-113, grifo nosso).

Isso posto, a Sintática, aqui também interpretada como Sintaxe para esclarecimentos teóricos, corresponde ao estudo das “[...] propriedades de combinação de certas expressões linguísticas [...]” (VIOTTI, 2008, p. 59). De acordo com Blühdorn (1997), substituir o termo Sintática por Gramática é mais apropriado, pois abrange dois tipos de relacionamento entre signos, o sintagmático (trata da sintaxe e da formação de textos) e paradigmático (referente a morfologia e lexicografia, ou seja, flexão e formação das palavras). Ao vinculá-lo ao contexto evolutivo da *Web*, mediante a aplicação das dimensões da linguagem às camadas da WWW, esta investigação a compreende apenas como Sintática.

Para Blühdorn (1997, p. 153) a sintaxe é “[...] a arte de conectar palavras a fim de formar orações [...]” Dessa maneira, à sintaxe está designado o estudo sobre o encaixe e a ordenação das palavras no processo de formação de uma frase para que seja passível de interpretação. Nesse contexto, atribuir a nomenclatura “Sintática” para identificar as duas primeiras camadas da *Web* corresponde ao modo como a informação era indexada e recuperada nessas camadas, por considerar apenas aspectos referentes à formação lógica de palavras e frases, sem necessariamente atribuir-lhes significado, reservando aos humanos essa tarefa.

Fuchs et al. (2010) interpretam essas camadas como qualidades presentes na WWW que são usadas para descrever e caracterizar dinâmicas sociais e processos informacionais que fazem parte da *internet*. Ainda de acordo com os autores, elas não devem ser entendidas em uma escala evolutiva, mas sim como qualidades de um sistema tecno-social que visa melhorar a cognição, comunicação e cooperação humana.

Na *Web 1.0* os *websites* em HTML eram constantemente atualizados, porém estáticos, monodirecionais e destinados à leitura. Sem interação, a presença de usuários e visitantes não impactavam em seu funcionamento. Desse modo, *websites* corporativos apenas apresentavam o catálogo de produtos e o contato para possíveis interessados, sem ter um público-alvo em vista, objetivando apenas marcar presença *online* (AGHAEI; NEMATBAKHS; FARSANI, 2012). Assim, na *Web 1.0*, poucas pessoas produziam conteúdo, pois era necessário ser especialista. Existiam muitos usuários passivos, os quais apenas consumiam a pouca informação produzida pelos técnicos (CORMODE; KRISHNAMURTHY, 2008; SILVA, 2010).

Fuchs et al. (2010) complementam que a cognição, termo representativo da *Web 1.0*, refere-se ao entendimento sistêmico subjetivo de uma pessoa ligada à outra pessoa por meio de um sistema de mediação. Dessa maneira, a *Web 1.0* (cognição) é assim denominada pelo fato do usuário consultar determinada informação produzida por outro indivíduo para a complementação de suas concepções, utilizando um servidor (*web*) como elemento mediador do processo de troca de informações. Na perspectiva de Monteiro (2015), a *Web 1.0* é *Web Sintática* por relacionar-se com as dimensões da linguagem, pois a linguagem é legível por máquinas, mas apenas compreensível por humanos.

A *Web* não possui uma escala de tempo como aquela utilizada para determinar os períodos históricos do planeta Terra (cretáceo, jurássico, etc.), pois é dinâmica demais para se ter um consenso sobre marcos de transição de suas qualidades, e nem mesmo é possível responder precisamente se quando uma nova *Web* se inicia a antiga deixa de existir ou se o legado da anterior é reaproveitado ou ainda se funcionam de forma simultânea.

No entanto, para Aghaei, Nematbakhsh e Farsani (2012), o fim da *Web* 1.0 está vinculado à declaração do vice-presidente da O'Reilly Media, Dale Dougherty, em um evento internacional realizado no ano de 2004. Nessa data, ele utiliza o termo *Web* 2.0 para se referir a uma *Web* que revolucionaria os negócios no mundo da informática. A introdução dessa nova fase da *Web* não alterou apenas o modo de funcionamento dos negócios *online*, mas também transformou a forma como os indivíduos consumiam e produziam conteúdo por ela veiculado.

Na literatura, inúmeros são os termos que nominam de forma menos técnica a *Web* 2.0, como, por exemplo, rede de comunicação, *Web* sabedoria, *Web* centrada em pessoas, *Web* participativa, *software* social, *Web* de leitura/escrita, *Web* social, entre outras mais (AGHAEI, NEMATBAKHSH, FARSANI, 2012; FUCHS et al., 2010; MURUGESAN, 2007). Apesar de diferentes termos serem utilizados para designar essa fase da *Web*, a temática que as circunda é a mesma, a do caráter social e cooperativo a ela vinculada.

Nesse sentido, Berners-Lee e Fischetti (2000) esclarecem que a *Web*, de forma geral, foi projetada para causar um efeito social e não para apenas solucionar uma questão técnica. A natureza social da *Web* supracitada e visionada pelos autores somente se objetivou a partir da *Web* Social, quando se tornou a rede de comunicação.

Murugesan (2007, p. 34, tradução nossa) identifica a *Web* 2.0 como “[...] uma coleção de tecnologias, estratégias de negócios e tendências sociais.” Além disso, o autor também a caracteriza como dinâmica e interativa por permitir que os usuários de conteúdo também contribuam para ele e consigam se manter atualizados mesmo sem acessar diretamente as páginas da *Web*. Outro aspecto levantado pelo autor supracitado é que essa *Web* permite um desenvolvimento facilitado de aplicações, pois elas conseguem se basear em dados, informações e serviços já disponíveis na *internet*.

Aghaei, Nematbakhsh e Farsani (2012) consideram que, diferente da 1.0, a *Web 2.0* tem de ser vista como uma plataforma na qual os usuários têm mais interação com menos controle. Boulos e Wheelert (2007, p. 10, tradução nossa) complementam que na *Web 2.0* “[...] o conteúdo pode ser mais facilmente gerado e publicado pelos usuários, e a inteligência coletiva dos usuários incentiva o uso mais democrático.” Nesse contexto, a implementação da *Web 2.0* para Murugesan (2007) representa mais do que uma nova roupagem para a antiga versão. Para o autor, os itens indicados abaixo podem evidenciar essa diferenciação:

- a) design flexível da *Web*, reuso criativo de informações e constantes atualizações;
- b) interface rica e ágil para o usuário;
- c) facilidade de criar e editar conteúdo colaborativo;
- d) cria novas e também reutiliza e combina diferentes aplicações *Web*;
- e) forma redes sociais de pessoas com interesses em comum;
- f) suporta a colaboração e ajuda a reunir inteligência coletiva.

A compilação dos fatores supracitados favorece a construção de um ambiente propício para comunicação na *Web*. Dessa maneira, Fuchs et al. (2010) consideram que o caráter social da *Web 2.0* é determinado conforme a finalidade de uso de uma aplicação, por exemplo, o processador de texto *Microsoft Word* tem a função colaborativa de controle de alterações, porém não é substancialmente social, diferente de um texto redigido *on-line* no qual há a participação e a interação de um grupo. Assim, pode-se inferir que a essência da *Web Social* está na interação entre indivíduos e grupos, mediada por tecnologias, serviços e ferramentas desenvolvidas para esse fim.

Dessa forma, o enfoque desta pesquisa está nas tecnologias que foram desenvolvidas na *Web Social*, potencializadas pela *Web Semântica* e que, conseqüentemente, se interconectaram na *Web Pragmática*, envolvidas com a organização, a disseminação e a recuperação de informações no Ciberespaço. Nesse sentido, o desenvolvimento e a implementação dessas tecnologias na *Web 2.0* contribuíam expressivamente para o início da estruturação da informação veiculada na rede. Assim, o *eXtensible Markup Language* (XML) é provavelmente a

tecnologia base para o desenvolvimento da maioria das ferramentas e funcionalidades na *Web*.

A aplicação do XML permite estruturar dados de um *website* para que programas possam interpretá-los, porém são só informações sem significado (ALESSO; SMITH, 2006). A interoperabilidade presente no XML é resultado da utilização de *tags* do SGML que não são pré-definidas. Esse aspecto promove a flexibilidade e a extensibilidade necessárias para que ele possa apoiar diferentes serviços da *Web* (ALESSO; SMITH, 2006).

De acordo com Bosak e Bray (1999), o XML é uma metalinguagem de marcação, desenvolvida com base na SGML, tecnologia que também foi utilizada por Tim Berners-Lee para o desenvolvimento do HTML. Para os autores, o XML é uma versão simplificada do SGML, que usa linguagem de marcação para descrever informações sobre dados, denominados de metadados. Nessa concepção, segundo Yu (2007, p. 10, tradução nossa), metadados podem ser definidos como “[...] dados sobre dados [...]” e descrevem de forma sistemática (estruturada) recursos informacionais, fato que viabiliza a interpretação dessas informações por máquina.

Essas informações adicionais dos metadados são identificadas mediante a aplicação de linguagens de marcação que acompanham o dado principal. Assim, o XML suporta uma ampla variedade de aplicações, pois seu objetivo principal é partilhar dados entre diferentes sistemas, permitindo que *softwares* compreendam e troquem informações formatadas nessa linguagem (KNIGHTS, 2007).

O XML foi desenvolvido para ser autodescritivo, armazenar e transportar dados sem *tags* pré-definidas, como no caso do HTML, pois as *tags* podem ser customizadas pelo desenvolvedor (W3SCHOOLS, 2016). Dessa forma, no Quadro 2 há um exemplo da aplicabilidade da tecnologia XML no cotidiano das pessoas, por meio de uma Nota Fiscal Eletrônica (NF-e) utilizada em território nacional para processos de compra e venda de produtos e serviços.

Quadro 2 – Dados de uma NF-e estruturados em XML

Estrutura XML	Tags	Dados
<emit>	<emit>	
<CNPJ>57.456.602/0001-75</CNPJ>	<CNPJ>	57.456.602/0001-75
<xNome>Empresa XYZ</xNome>	<xNome>	Empresa XYZ
<xMun>Londrina</xMun>	<xMun>	Londrina
<UF>PR</UF>	<UF>	PR
<CEP>86050280</CEP>	<CEP>	86050280
<xPais>Brasil</xPais>	<xPais>	Brasil
<fone>4399368160</fone>	<fone>	4399368160
</emit>	</emit>	

Fonte: o próprio autor com base em Fazenda (2017).

No Quadro 2, ao relembrar Knights (2007), Yu (2007) e Bosak e Bray (1999), é possível observar que, de fato, as *tags* são as “informações sobre os dados”, ou seja, são os metadados que identificam o tipo de informação que está descrito entre elas. Dessa maneira, o uso do XML na transmissão de dados estruturados para aplicações *Web* fez com que a *WWW* deixasse de ser apenas um espaço universal de informações para se tornar uma rede do conhecimento (KHARE; RIFKIN, 1997). Nessa perspectiva, o XML auxilia no processo de formação do sentido para agentes computacionais, propiciando mais autonomia operacional para as máquinas.

Além do XML, a *Web Social* implantou uma série de outros elementos que alteraram o *modus operandi* dessa dimensão. Nesse sentido, o Quadro 3 aponta de forma progressiva comparativa as principais modificações ocorridas entre as camadas que compõem a *Web Sintática*. Entretanto, vale salientar, que essas mudanças ocorreram não apenas em prol do estabelecimento de uma nova camada, mas também por consequência da própria evolução das TIC.

Quadro 3 – Web 1.0 → Web 2.0

Web 1.0	Web 2.0
Leitura	Leitura/Escrita
Empresas	Comunidades
Cliente-Servidor	P2P (<i>Peer-to-Peer</i>)
HTML, Portais	XML, RSS
Taxonomia	<i>Tags</i>
Possuir	Compartilhar
IPOs	Vendas do comércio
Netscape	Google
Formulários <i>Web</i>	Aplicações <i>Web</i>
<i>Screen scraping</i>	APIs
Conexão discada	Banda larga
Custos com <i>hardware</i>	Custos com banda larga
Palestras	Conversação
Propaganda	<i>Word of mouth</i>
Serviços vendidos pela <i>Web</i>	Serviços <i>Web</i>
Portais de informações	Plataformas

Fonte: adaptado de Aghaei, Nematbakhsh e Farsani (2012).

Ao confluir os elementos de cada coluna do Quadro 3, tornou-se possível traçar um panorama envolvendo as principais características atreladas a cada camada da *Web* Sintática. Dessa maneira, a *Web* 1.0 é composta por passividade de uso, restrita quase que exclusivamente a corporações, linear, hierárquica, proprietária, com conteúdo estático, catálogos *on-line*, baixa velocidade de conexão e elevados custos de *hardware*. Por outro lado, na *Web* Social, há a participação ativa dos usuários, formação de comunidades *on-line*, compartilhamento de informações, tecnologias adequadas para a autonomia de agentes computacionais, indexações sociais e aumento de custos relacionados à conectividade.

Apesar de não explicitadas no Quadro 3, duas importantes tecnologias contribuíram significativamente para a expansão do contexto social que envolve a *Web* 2.0.; são elas: *weblogs (blogs)* e *wikis*. Assim, com a proposta de editar conteúdo sem necessariamente possuir um conhecimento técnico, essas tecnologias tiveram grande disseminação no ano de 2003, o que possibilitou que sujeitos e grupos reivindicassem seu espaço na *Web* e impulsionou ainda mais o aspecto social de sua camada 2.0 (MIKA, 2007).

Nesse sentido, os *blogs* ligavam pessoas com interesses em comum, resultando assim em comunidades *on-line*. Além disso, o usuário poderia se inscrever em diferentes *blogs* para receber atualizações de conteúdo por *e-mail*,

sem que houvesse a necessidade de ter que acessar a página do *weblog*. As wikis, por sua vez, são poderosos sistemas para a criação e a edição de conteúdo, nos quais as pessoas podem criar, editar e revisar artigos *on-line* colaborativamente (MURUGESAN, 2007).

A versatilidade de uso de uma *wiki* é o seu ponto mais forte. A Wikipédia, por exemplo, apesar de considerada por Chatfield (2009) como apenas a ponta do *iceberg*, é a que melhor representa a potencialidade dessa tecnologia. Para Leuf e Cunningham (2001) outro atrativo das *wikis*, quando comparadas às outras ferramentas de colaboração, é o fato de serem informais e fáceis de usar por não necessitarem de *logins* e senhas.

A criação dessas tecnologias remonta uma trajetória semelhante à de Tim Berners-Lee com o desenvolvimento da *Web*. Nos anos de 1994 e 1995, Ward Cunningham desenvolveu e publicou o primeiro motor *wiki*. Na época, seu objetivo era solucionar problemas de compartilhamento de ideias em sua empresa. Para isso, criou um conjunto simples de regras para deixar o mais fácil possível a criação e compartilhamento de conteúdo em *websites* e o nomeou de WikiWikiWeb (WOODS, THOENY, 2007).

A invenção da WikiWikiWeb, segundo Lih (2009), foi considerada insana por conflitar diretamente com um senso comum da época, que acreditava que, para ser seguro, tinha de ser bloqueado e que, para se ter qualidade, era necessário restringir quem participaria da edição do conteúdo na *Web*. Entretanto, o que parecia ser uma vulnerabilidade em seu funcionamento tornou-se o atrativo principal, isto é, um sistema de edição internacional, simples e aberto a qualquer usuário interessado. Assim, entre outros tantos projetos decorrentes dessa iniciativa, a Wikipédia, ao que tudo indica, parece ser o de maior representatividade em nível mundial, carregando em seu desenvolvimento colaborativo a tecnologia *wiki* e os princípios a ela vinculados.

3.1.1 Wikipédia

Em 2001, Larry Sanger e Jimmy Wales aplicaram a tecnologia WikiWikiWeb para solucionar problemas referentes à liberdade de edição do projeto Nupedia, a enciclopédia para a *Internet* (LIH, 2009). A Nupedia era um *website* de conteúdo livre e seus artigos eram construídos colaborativamente por especialistas

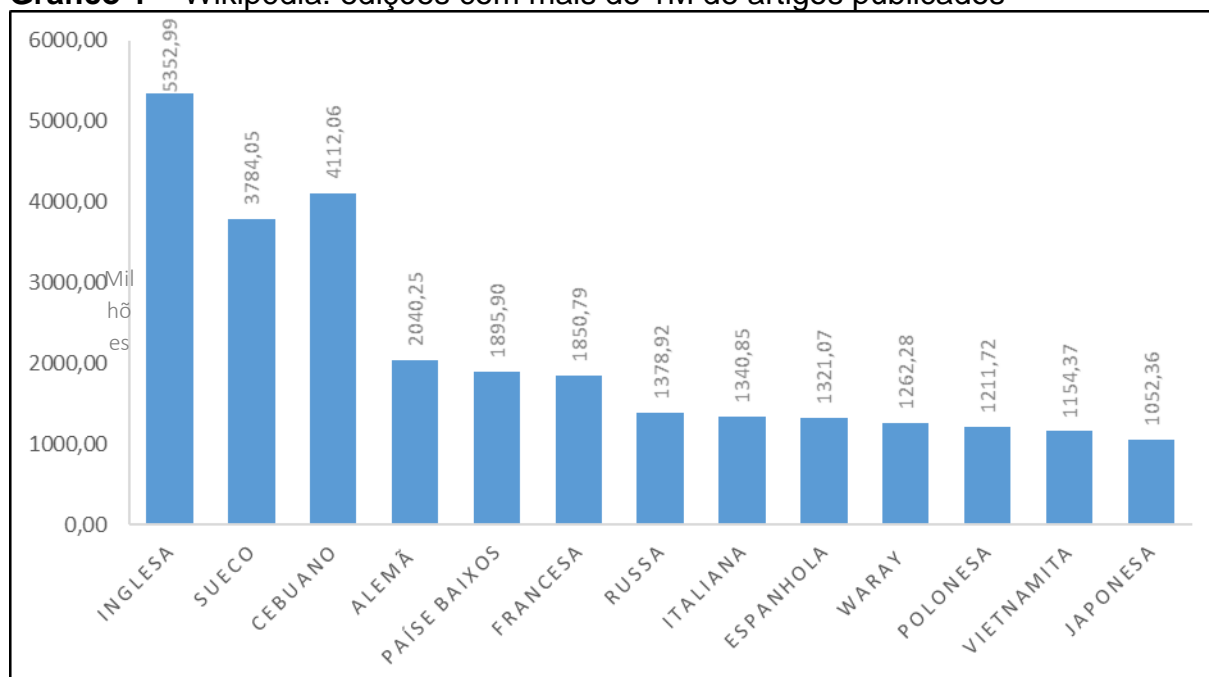
de diversos campos do conhecimento. Porém, o extenso processo de revisão pelos pares ocasionou entraves em seu processo produtivo (AYERS, MATTHEWS, YATES, 2008).

Desse modo, ao perceber que a revisão por pares estava causando muitos problemas no processo de produção, Larry Sanger, na época o editor-chefe, decidiu incorporar a tecnologia *wiki* para a criação de conteúdo enciclopédico (JEMIELNIAK, 2014). No entanto, a crença de que uma boa enciclopédia não poderia ser criada usando uma tecnologia *wiki* fez com que seus fundadores, receosos com a aceitação do público, criassem um domínio específico desvinculado da Nupedia, o endereço da URL era: *wikipedia.com* (ANDERSON, 2011).

Segundo Jemielniak (2014), no início a Wikipédia foi introduzida apenas como um alimentador para o banco de dados da Nupedia. No entanto, ela caiu no gosto de novos e antigos editores da Nupedia, sendo que seu principal atrativo é a possibilidade de editar conteúdo sem ter que utilizar credenciais de acesso para se autenticar no *website* (ANDERSON, 2011).

Desse modo, em seu primeiro ano de funcionamento, a Wikipédia atingiu a marca de vinte mil artigos publicados, sendo esse valor cinco vezes maior ao final do segundo ano. Os números alcançados pela Wikipédia decretaram, em 2003, o fim da Nupedia com apenas 24 artigos publicados e 74 em processo de produção (JEMIELNIAK, 2014).

Formalmente, a Wikipédia foi iniciada em 15 de janeiro de 2001. Desde então, a quantidade de leitores, colaboradores e artigos aumentou progressivamente. Dessa maneira, em 2007, a plataforma atingiu a marca de 2 milhões de artigos publicados na língua inglesa e se tornou a maior enciclopédia já construída (MILLER, 2010). Na atualidade, a quantidade de verbetes/artigos produzidos aumentou significativamente. No Gráfico 1, constam os números alcançados para cada edição da Wikipédia com mais de um milhão de publicações.

Gráfico 1 – Wikipédia: edições com mais de 1M de artigos publicados

Fonte: o próprio autor com base em Wikipédia (2017a)¹.

Ao comparar os dados sobre a produção de artigos em língua inglesa do Gráfico 1 com os apresentados por Miller (2010) sobre o ano de 2007, constata-se o aumento de aproximadamente 3,3 milhões de verbetes publicados no decorrer de dez anos. Essas informações indicam a popularidade atingida pela Wikipédia nos últimos anos, não só na língua inglesa, mas também em idiomas de menor representatividade. Metas nunca foram estipuladas para que esses números fossem alcançados; toda a publicação que atenda aos padrões solicitados é armazenada na plataforma (AYERS; MATTHEWS; YATES, 2008).

Nesse contexto, artigos são entendidos como páginas da Wikipédia que apresentam informações enciclopédicas, formados por aproximadamente 500 caracteres e ao menos uma ligação interna via *hyperlink* para outra página da Wikipédia (AYERS; MATTHEWS; YATES, 2008). Além disso, pode ser considerado como um artigo o documento hipertextual que defina ou descreva uma entidade ou evento e possua *hyperlinks* para páginas internas e externas da Wikipédia para fornecer informações adicionais sobre determinado tema (MIHALCEA; CSONAI, 2007).

Além de indivíduos dispostos a colaborar, a fórmula para o sucesso da Wikipédia também é o fato de ser composta por um *software*, responsável por

¹ Levantamento dos dados em 09 de março de 2017.

sustentar todas as configurações disponíveis na plataforma, denominado de MediaWiki. Inicialmente, com base na tecnologia WikiWikiWeb de Cunningham, desenvolveu-se por meio da colaboração o *software* UseModWiki com a finalidade de monitorar alterações básicas em artigos. Ao longo do tempo, o sistema foi otimizado e uma importante atualização resultou no MediaWiki. Entre suas principais funcionalidades, destaca-se a possibilidade de controle de versionamento de artigos para recuperação em casos de vandalismos e atualizações equivocadas de conteúdo (ANDERSON, 2011).

A distribuição do MediaWiki é gratuita, qualquer pessoa pode baixar, instalar, executar e compartilhar sem custos. O *software* também possui inúmeras opções de configuração, é de fácil uso, além de ser *open source*, ou seja, usuários com habilidades em programação podem alterar o código-fonte do programa, buscando sanar possíveis necessidades e realizar melhorias. Atrativos como esses fazem do MediaWiki um dos *softwares wiki* mais utilizados em todo o mundo (BARRETT, 2009).

Desse modo, o uso do MediaWiki como programa base da Wikipédia contribuiu fortemente para seu crescimento, cuja única regra para contribuir era ser ousado. Assim, Sanger encorajou as pessoas para que não se preocupassem em cometer erros em suas contribuições, pois os erros poderiam ser corrigidos e desfeitos por intermédio do MediaWiki (ANDERSON, 2011). Apesar da liberdade dada para incentivar a colaboração, algumas regras tiveram que ser impostas, mas nada que viesse a coibir a produção. Segundo a autora supracitada, tratam-se de três regras: ponto de vista neutro, verificabilidade (sempre citar as fontes) e sem pesquisas inéditas (não apresentar teorias próprias).

De acordo com Barrett (2009), apesar do MediaWiki ser uma ferramenta ideal para o desenvolvimento de forças coletivas de trabalho, sua aplicação é mais propícia para alguns casos do que para outros. Para o autor a ferramenta se aplicaria para contextos de: compartilhamento do conhecimento informal; retorno rápido; comunidades de pessoas com interesses semelhantes; comunidades globais; facilidade de administração e confiabilidade. Por outro lado, o autor não a recomenda para os casos de: aplicações que necessitam de controle de acesso restrito; gestão de conteúdo geral; usuários com conhecimento técnico limitado e parágrafos de textos simples.

Iniciativas como a Wikipédia, *weblogs*, ferramentas de redes sociais, entre outras plataformas colaborativas, aumentaram significativamente a quantidade de informações existentes na *Web*. Além disso, houve também mais interação e liberdade para a produção de conteúdo na rede por grupos e indivíduos. No entanto, a liberdade excessiva de criação não seguiu padrões de indexação para uma boa recuperação futura. Desse modo, emergiu a *Web Semântica*, a terceira camada da *WWW*, cujo propósito é tentar estabelecer padrões para que as informações produzidas e veiculadas pela *Web Social* possam ser indexadas, recuperadas e interpretadas por humanos e máquinas.

3.2 WEB SEMÂNTICA

Em sua trajetória, a *Web* se tornou uma tecnologia referência no compartilhamento ilimitado de informações entre as pessoas. No entanto, houve uma sobrecarga de informações que tornou difícil indicar quais informações eram relevantes. Assim, a *Web Semântica* objetiva ser a base para o desenvolvimento de aplicações inteligentes que possibilitem o uso eficiente da informação (SCHOOP; MOOR; DIETZ, 2006).

Fuchs et al. (2010, p. 53, tradução nossa) consideram a *Web Semântica* ou *Web 3.0* “[...] um sistema de colaboração *on-line* que permite a formação de comunidades virtuais, conhecimento e trabalho cooperativo [...]” Na percepção de Aghaei, Nematbakhsh e Farsani (2012), a *Web 3.0* pode conectar, integrar e analisar dados de diferentes fontes, a fim de obter um novo fluxo informacional. Ainda de acordo com os autores, esta *Web* pode melhorar a gestão de dados, dar apoio à acessibilidade na *internet* móvel, estimular a criatividade e inovação, aumentar a satisfação de clientes e organizar a colaboração na *Web Social*.

Berners-Lee, Hendler e Lassila (2001) referem-se à *Web 3.0* como a “*Web Semântica*”. De acordo com os autores, ela é uma extensão da anterior, em que a informação tem seu significado bem definido para que pessoas e computadores possam trabalhar em cooperação. A relação homem-máquina, sugerida pelos autores, vai além da mecanização e automatização de funções e tarefas, e consiste na capacidade do computador em compreender os devires da comunicação humana.

Na *Web Semântica* há a possibilidade de aplicar avançadas tecnologias de extração do conteúdo de páginas da *Web* que irão auxiliar o computador a processar, compreender, argumentar e responder às consultas humanas (MIKA, 2007). O objetivo dessa *Web* é se tornar legível para as máquinas, diferente do que acontecia na *Web 2.0*, cujo foco era na comunicação entre os usuários, pois se preocupa com o conjunto de dados interligados (AGHAEI; NEMATBAKHSI; FARSANI, 2012).

Tecnologias como o XML e o *Resource Description Framework* (RDF) são essenciais para conectar dados na *Web Semântica*. O XML estrutura a informação de uma página *Web* por meio de *tags* ocultas, que podem ser interpretadas por *softwares*, porém são dados sem significado. Em contrapartida, o RDF utiliza dados estruturados pelas *tags* do XML para atribuir propriedades e valores a um sujeito, por meio de URI interligadas (BERNERS-LEE; HENDLER; LASSILA, 2001). Assim, segundo Breitman (2010, p. 20), o RDF pode ser definido como

[...] uma linguagem declarativa que fornece uma maneira padronizada de utilizar o XML para representar metadados no formato de sentenças sobre propriedades e relacionamentos entre itens da *Web*. Esses itens, chamados de recursos, podem ser virtualmente qualquer objeto (texto, figura, vídeo e outros), desde que possuam um endereço *Web*.

Desse modo, a utilização do RDF permite descrever coisas do mundo, tal como pessoas, locais e conceitos complexos, como também de que maneira estes se relacionam (HEATH; BIZER, 2011). Diferente de Breitman (2010), Legg (2008) argumenta que o RDF não pode ser considerado uma linguagem, mas um modelo de dados para descrição semântica de recursos *Web* que possam ser interpretáveis por máquinas. Apesar de divergências entre os autores em determinar se o RDF é linguagem ou modelo, nesta investigação ele é compreendido das duas maneiras, pois tem regras próprias de sintaxe para a descrição dos recursos, configurando-o como uma linguagem. Além disso, é um modelo para a descrição de recursos na *Web* que pode ser adotado pelos desenvolvedores.

Segundo Schoop, Moor e Dietz, (2006, p. 75, tradução nossa), o RDF é aplicado para que os “[...] dados e as regras para raciocinar sobre dados e informações sejam sistematicamente descritos [...]” e, posteriormente, possam ser

facilmente compartilhados e utilizados por pessoas e agentes de *software*. Para Bax (2012, p. 12) o RDF é uma linguagem, dentre outros padrões propostos pela *World Wide Web Consortium (W3C²)*, que possibilita a representação do conhecimento na *Web Semântica* e, conseqüentemente, a interpretação por máquinas, por meio de uma *Semântica Formal*.

Dessa maneira, ao empreender novamente uma breve incursão na Linguística, a Semântica busca “[...] descrever o ‘significado’ das palavras e das sentenças [...]”. Em contraste com a Sintaxe, abordada na camada Sintática da *Web*, a Semântica não se preocupa em analisar como uma frase está estruturada, mas sim se volta para o significado que pode ser interpretado e disseminado por seu intermédio. Nesse sentido, o significado é tratado como uma unidade de diferença, assim, uma palavra é definida por não ter o mesmo significado que outra (OLIVEIRA, 2006, p. 17).

Os estudos contemporâneos no Brasil que envolvem a Semântica podem ser desmembrados em até três subdisciplinas: a Semântica Formal, a Semântica da Enunciação e a Semântica Cognitiva. Na Semântica Formal é necessário a união de sentido e referência para a criação do significado, haja vista que o lógico precisa de evidências do mundo palpável para gerar o sentido. Por outro lado, a Semântica da Enunciação considera que as pessoas são imersas na linguagem e que qualquer referência de mundo é uma ilusão criada pela própria linguagem, isto é, o mundo é criado conforme a linguagem e não como é de fato. A última, a Semântica Cognitiva, forma o significado a partir do contato com a coisa, assim é necessário tocar e manipular um objeto para formular um conceito sobre ele (OLIVEIRA, 2006).

Dessa forma, de volta ao contexto da *WWW*, tornou-se possível assimilar o funcionamento da *Web 3.0* com a Semântica Formal indicada por Bax (2012) anteriormente. Nesse sentido, da mesma maneira que a Semântica Formal necessita de um vínculo com o referente para a produção do sentido, a *Web Semântica*, por meio de suas tecnologias, em especial o *RDF*, conecta *hyperlinks* correlatos para que em conjunto auxiliem no processo de significação de dados para humanos e máquinas na *Web*.

² <https://www.w3.org/>

A descrição de um recurso é feita por meio das triplas *RDF*, as quais têm a função de espelhar a estrutura básica de uma frase e são formadas por sujeito, predicado e objeto (HEATH; BIZER, 2011). Ao utilizar as triplas *RDF*, uma URL deixa de ser apenas um *hyperlink* e passa a transmitir informações fatuais por meio de dados interligados, característica esta que otimiza aspectos de encontrabilidade e usabilidade de recursos *Web* (LANDIS, 2014). Dessa maneira, segundo Yu (2007, p. 40, tradução nossa), o *RDF* “[...] é para *Web Semântica* o que o HTML tem sido para *Web* [...]”, ou seja, um elemento substancial à concretização do ideal semântico que circunda o desenvolvimento da *Web 3.0*.

Além disso, ao adotar o *RDF* para a publicação de dados, torna-se necessária a escolha antecipada de um “*RDF syntax*”, também conhecido como formato de serialização; são eles: *RDF/XML*, *RDFa*, *Turtle*, *N-Triples* e *RDF/JSON*, conforme especificado no Quadro 4 (HEATH; BIZER, 2011).

Quadro 4 – Formatos de serialização RDF

Formatos	Descrição
RDF/XML	É o principal formato utilizado para a publicação de dados <i>Web</i> . Sua sintaxe é de difícil compreensão para humanos, não sendo indicada para desenvolvimentos que envolvam muita participação humana.
RDFa	Neste formato as triplas <i>RDF</i> são incorporadas diretamente ao código HTML da página como parte integrante do código.
Turtle	Formato simples de serialização, ideal para ler e escrever triplas manualmente, pois tem um nível de detalhamento maior.
N-Triples	Trata-se de um subconjunto do <i>Turtle</i> , conhecido por ter um alto nível de redundância nas declarações. É indicado para o carregamento de arquivos grandes, como <i>backups</i> e espelhamentos.
RDF/JSON	Destinado a atender uma demanda por serialização do <i>JSON (JavaScript Object Notation)</i> , é equivalente ao funcionamento do <i>XML</i> (um formato de transferência e intercâmbio de dados). Sua adoção evita importação e adaptação de formatos diferentes de interoperabilidade pelos desenvolvedores <i>Web</i> .

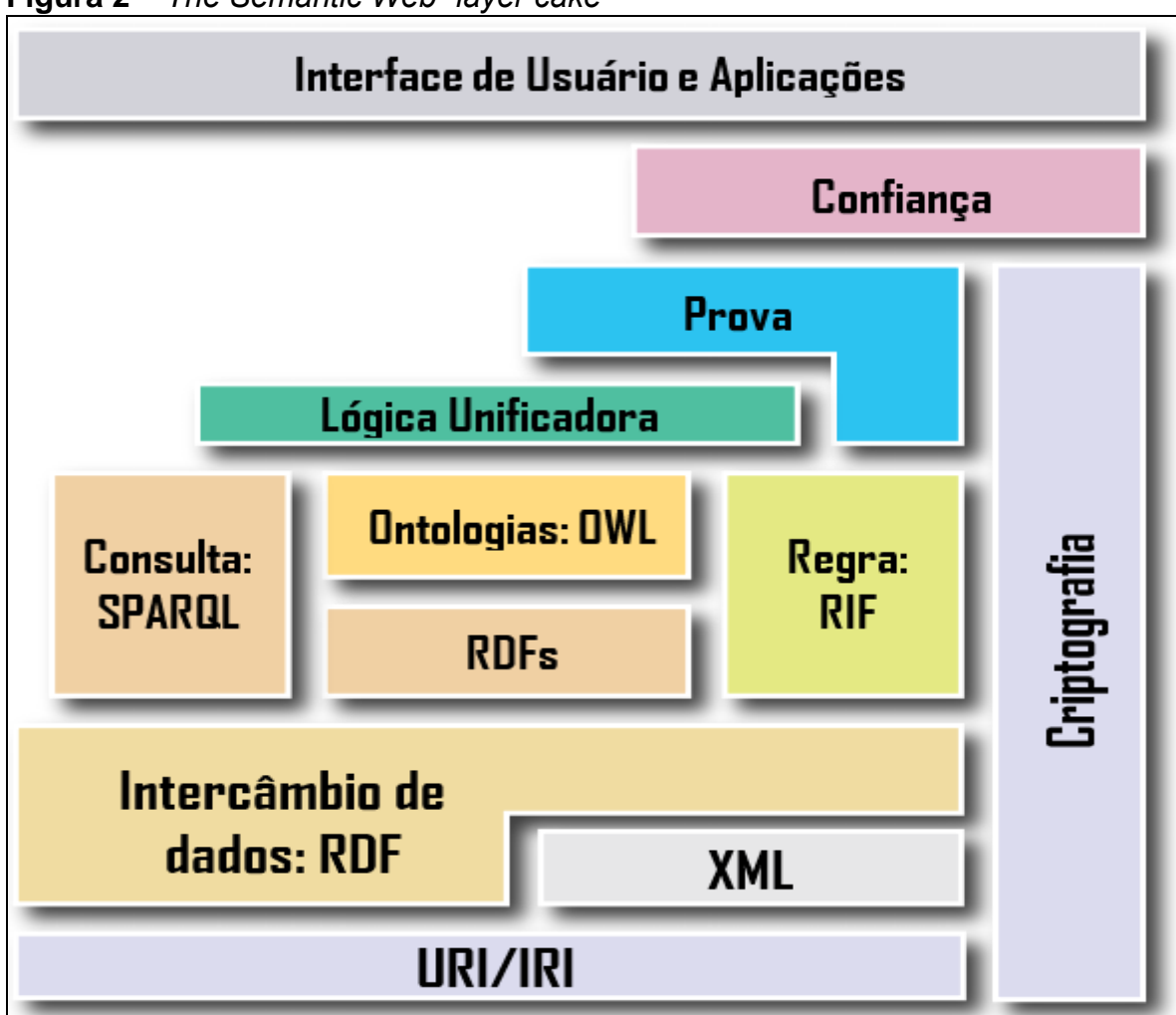
Fonte: o próprio autor com base em Heath e Bizer (2011).

Dessa forma, as serializações *RDF* podem ser interpretadas como formatos específicos para diferentes situações que envolvam a descrição de recursos. Assim, seu modo de funcionamento é semelhante aos padrões adotados para a formatação de textos científicos, pois, ao editá-los, por exemplo, sob as normas da Associação Brasileira de Normas Técnicas (ABNT) ou American

Psychological Association (APA), entre outras, a essência do texto não é alterada, mas o conteúdo é disposto e referenciado de formas diferentes.

Apesar de sua substancialidade para o desenvolvimento e manutenção da *Web Semântica*, o RDF é apenas um dos vários componentes do conglomerado de tecnologias responsáveis por estruturar as informações da rede nessa camada. Dessa maneira, a Figura 2 apresenta o clássico “*layer cake*” utilizado para exemplificar, por meio de camadas semelhantes às de um bolo (confeitaria), as tecnologias e conceitos que sustentam o propósito semântico da *Web 3.0*.

Figura 2 – *The Semantic Web “layer cake”*



Fonte: adaptado de W3C (2013a).

Nesse contexto, essas tecnologias da *Web Semântica* permitem a criação e armazenamento de dados, a construção de vocabulários e de regras para o tratamento dessas informações (W3C, 2017a, não paginado). Entretanto, em conformidade com o propósito desta investigação, nem todas as camadas da Figura

2 serão especificadas ou conceitualizadas a seguir, pois se buscou compreender apenas as camadas em que os dados são estruturados e consultados. Assim, como algumas das camadas foram abordadas em parágrafos anteriores, a exemplo, o XML e o RDF, o enfoque recaiu sobre as seguintes tecnologias: URI/IRI, RIF, RDFS, *Web Ontology Language (OWL)*, e *Simple Protocol and RDF Query Language (SPARQL)*.

De acordo com Yu (2007), o *Internationalized Resource Identifiers (IRI)* é uma versão internacionalizada do URI, um novo padrão, que permite a inserção de uma quantidade maior de caracteres do que a existente no URI. Isso ocorre pelo fato do URI utilizar a codificação *American Standard Code for Information Interchange (ASCII)*, limitada aos caracteres presentes na língua inglesa. Por outro lado, o IRI emprega como padrão o *Universal Coded Character Set (UCS)* que contempla caracteres advindos de diferentes idiomas (IRI, 2017).

Por existir um número elevado de regras declarativas e de produção na *Web Semântica*, o *Rule Interchange Format (RIF)* foi criado para a padronização dessas divergências. Trata-se de uma linguagem XML que disponibiliza um formato para o intercâmbio dessas regras (W3C, 2013b). Além disso, o RIF é uma linguagem complementar o OWL, sendo composta pelas seguintes sublinguagens: RIF- *Basic Logic Dialect (BLD)*, RIF- *Production Rules Dialect (PRD)* e RIF-Core (DOMINGUE; FENSEL; HENDLER, 2011). Assim, o RIF-Core é o “dialeto” principal, considerado simples e de ampla aceitação, e destina-se a *Web* como um todo. No RIF-BLD o Core é complementado com a implementação de funções lógicas e nomeação de argumentos. O último, o RIF-PRD, apresenta a mesma estrutura dos anteriores, mas trabalha com regras de encadeamento que permitem expandir ou retrair informações (W3C, 2013b).

Segundo Yu (2007, p. 74, tradução nossa), o *Resource Description Framework Schema (RDFS)* é “[...] uma linguagem que pode ser usada para criar um vocabulário para a descrição de classes, subclasses e propriedades de recursos RDF [...]”, pode também ser interpretado como uma extensão do RDF que permite que os dados da *Web* sejam facilmente processados por máquinas. Para Domingue, Fensel e Hendler (2011), o vocabulário do RDFS pode ser expandido com o emprego do OWL que, por meio de descrições lógicas, estabelece um sistema de inferências inversas, transitivas, simétricas e funcionais, o que possibilita a ele relacionar classes, hierarquia de classes e propriedades com maior flexibilidade.

Dessa forma, de acordo com Santarém Segundo, Souza e Coneglian (2015, p. 6, grifo nosso), as *OWL* são compostas por classes, propriedades e indivíduos, sendo

[...] as **classes** [...] representações concretas de um conceito ou de uma entidade. Os **indivíduos** são os objetos de um domínio. E as **propriedades** são relações entre as classes e atributos destas classes. Possibilitando também inserir restrições e realizar inferências sobre as relações que existem.

Segundo Schoop, Moor e Dietz (2006, p. 75, tradução nossa), as ontologias são os principais componentes para o bom funcionamento da *Web Semântica*, pois elas “[...] representam conceitos e as relações entre os conceitos [...]” que podem ser estabelecidas de diferentes maneiras.

Todas as tecnologias apresentadas até o momento estão relacionadas a questões de organização e representação de dados para *Web*. Dessa maneira, a última delas de interesse desta investigação, indicada na Figura 2, é a do *SPARQL*. Confere-se ao *SPARQL*, enquanto linguagem de consulta, recuperar informações de diversas bases de dados da *Web*, desde que estejam no formato RDF, pois é designado especificamente para consultar dados nesse formato (W3C, 2013c). Assim, as camadas abordadas do “*layer cake*” estão designadas a processos de descrição, representação, tradução, vinculação e consulta de dados na *Web Semântica*.

Nesse sentido, o ideal da *Web Semântica* foi recentemente enriquecido com a aplicação do conceito de *Linked Data (LD)* ao seu contexto, uma vez que visa tratar de maneira simples e organizada grandes quantidades de dados interligados. Para que as tecnologias/ferramentas da *Web Semântica* tenham acesso e consigam gerenciar grandes quantidades de dados e seus relacionamentos, é necessário que haja padronização dessas informações. Desse modo, um formato em comum, a exemplo o RDF, permite a execução do *LD* em diferentes instâncias e bases de dados da *Web* (W3C, 2017b).

Desse modo, o *LD* refere-se à adoção de melhores práticas para a publicação e conexão de dados na *Web*, fato que expande a potencialidade da rede, permitindo-o conectar domínios tematicamente diferentes na busca pelo significado (BIZER; HEATH; BERNERS-LEE, 2009). Nesse contexto, o DBpedia é considerado o exemplo mais proeminente da aplicação dos princípios de *LD*, por extrair o conhecimento representado no formato RDF da Wikipédia e publicá-lo de forma

interligada na *Web* (KONTOKOSTAS et al., 2012). Dessa forma, Monteiro (2015, p. 5) aponta que a Semântica existente nesse contexto tecnológico se refere

[...] a construção de uma infraestrutura adequada para os agentes inteligentes percorrerem a *web* para extrair o conhecimento sobre algo ou alguma coisa (entidades) e executarem ações complexas. Trata-se de representar significados, que associam o conhecimento e coloca-os para trabalhar de forma a tornar a experiência com ciberespaço mais relevante, útil e agradável. Nesse sentido, observa-se a criação de *knowledge database*, ou base de conhecimento livre, como a Wikidata, Freebase e o DBpedia.

Há na *Web Semântica* um constante movimento de estruturação e padronização de dados produzidos e veiculados na rede. Nesse contexto, tecnologias de tradução de linguagens e formatos de descrição, declaração e consulta buscam estabelecer um modelo singular, compreensível por humanos e máquinas. Entretanto, sua tangibilidade depende da interconexão desses variados elementos em um único ambiente.

Dessa forma, em meio às tecnologias e linguagens apresentadas nesta seção, a *Web Semântica* cumpre seu principal objetivo que, segundo Schoop, Moor e Dietz (2006, p. 75, tradução nossa), é de ser a “[...] a base para aplicações inteligentes que possibilitem um uso mais eficiente da informação, não apenas um conjunto de documentos vinculados, mas uma coleção de repositórios do conhecimento.” Assim, na próxima seção, serão abordadas, ainda sob a perspectiva da *Web Semântica*, as *Knowledge Bases (KBs)*, que são os espaços equivalentes a grandes fábricas, que coletam os dados, tal como se fosse sua matéria-prima, para transformá-los mediante a aplicação de tecnologias semânticas em representações do conhecimento que podem ser facilmente disseminadas pela *Web*.

3.2.1 *Knowledge Bases*

A quantidade extraordinária de artigos produzidos e armazenados nas diferentes edições da Wikipédia torna-a uma grande base internacional do conhecimento enciclopédico colaborativo na *Web*. Entretanto, limita-se a buscas apenas por palavras-chave, além de apresentar inconsistências decorrentes da duplicidade de conteúdo entre suas diferentes edições (MORSEY et al. 2009). Nesse contexto, apesar de sua grandiosidade, a Wikipédia tem necessidade por

tecnologias que tornem seu conteúdo cada vez mais acessível e disponível ao restante da *Web*. Assim, tecnologias como as *Knowledge Bases (KBs)* auxiliam no processo de representação e disseminação do conhecimento interconectado na rede.

Segundo Sterckx et al. (2016, p. 1, tradução nossa), as *KBs* são “[...] bancos de dados contendo milhões de fatos sobre pessoas, organizações, eventos, produtos, etc.” Algumas delas se baseiam na Wikipédia (Yago, DBpedia e Freebase) e outras, na própria linguagem natural (ConceptNet³ e SenticNet⁴) como fonte do conhecimento que estruturam e representam. Assim, as *KBs* contribuem de forma expressiva para tornar a *Web* mais inteligente, sendo a Wikipédia a principal fonte de conhecimento (BIZER et al., 2009). Dessa forma, as *KBs* conectam em seu âmago

[...] milhões de entidades, incluindo pessoas, universidades, cidades, organizações, ou obras de arte. Estas entidades estão estruturadas em uma taxonomia de classes, onde as classes mais gerais (como pessoa) podem se dissipar em classes mais específicas (como cantora). [...] também contêm centenas de milhões de fatos sobre estas entidades, tais como onde a pessoa nasceu, onde, que o cantor cantou qual canção, ou qual cidade está localizada em qual país. Ao contrário da Wikipédia, as *Knowledge Bases* armazenam esse conhecimento de forma estruturada em triplas sujeito-predicado-objeto, que permite consultá-las como em um banco de dados. (MAHDISOLTANI; BIEGA; SUCHANEK, 2015, p.1, tradução nossa).

As *KBs* extraem as informações estruturadas para que possam ser criadas representações a partir delas, mediante a aplicação de triplas RDF (sujeito, predicado e objeto), apresentadas no formato de declarações. Desse modo, o cruzamento dessas informações permite o relacionamento entre entidades, que resulta em respostas mais concisas para quando se necessita saber qual a cidade de nascimento de uma celebridade ou em quais filmes já atuou (BIEGA; KUZEY; SUCHANEK, 2013). As tecnologias que viabilizam esses processos de organização e representação do conhecimento em *KBs* se referem as mesmas apresentadas na Figura 2 como base da *Web Semântica*.

A *Web Semântica* conecta diferentes *KBs* pela rede, sejam elas comerciais ou acadêmicas. Nessa direção, grandes empresas do ramo de tecnologia

³ <http://conceptnet.io/>

⁴ <http://sentic.net/>

desenvolveram suas próprias *KBs* para aperfeiçoar os resultados de seus mecanismos, como o Google (*Knowledge Vault*), a Microsoft (*Satori*) e a Amazon (*Evi*) (SUCHANEK, 2016). Apesar da notoriedade das *KBs* supracitadas, em sua maioria, elas são projetos destinados à melhoria dos serviços de recuperação de informações prestados por suas corporações. O acesso a essas bases poderia ser limitado ou totalmente restrito por questões de sigilo de mercado. Então, buscou-se nesta investigação por *KBs* de acesso público, edição colaborativa e que usassem como fonte principal de dados a Wikipédia. Assim, DBpedia e Wikidata atenderam a estes requisitos.

3.2.1.1 DBpedia

Constituída em 2007 na Alemanha, a DBpedia concretizou-se a partir da cooperação entre as Universidades de Leipzig e Mannheim e a *OpenLink Software*. Seu objetivo principal é extrair dados das diferentes páginas da Wikipédia para combiná-los e interligá-los para proporcionar novas opções de navegação e uso dessas informações. Segundo Bizer et al. (2009), os textos dos artigos são construídos de forma livre e a linguagem de marcação *wiki* utilizada na Wikipédia possibilita que a DBpedia extraia essas informações e as converta em uma grande base de conhecimento. Esses dados são convertidos em entidades, compostas por propriedades e valores.

As informações extraídas das diferentes edições da Wikipédia tornam a DBpedia uma grande base de conhecimento multilíngue. São extraídas informações em mais de 120 idiomas, totalizando 38,3 milhões de coisas descritas em sua *KB*, sendo 4,5 milhões correspondentes apenas aos dados extraídos da versão em língua inglesa da Wikipédia (DBPEDIA, 2016). Abrangente, essa base desperta o interesse da comunidade científica que, atualmente desenvolve inúmeros experimentos envolvendo-a. Além disso, a DBpedia pode ser considerada um grande servidor de dados vinculados da *Web*, fato que contribui para o sucesso do *Linked Open Data (LOD)* (LEHMANN et al., 2012; KONTOKOSTAS et al., 2012).

A interface da DBpedia é orientada para receber os dados da Wikipédia, boa parte de sua manutenção é realizada pelos mesmos editores, autores e programadores da Wikipédia. Seu propósito é combinar as informações das várias páginas da Wikipédia para formar um grande banco de dados para que

essas informações possam ser reutilizadas de formas inovadoras e interessantes (DBPEDIA, 2016). Mais informações sobre o funcionamento da DBpedia serão apresentadas na seção de análise desta investigação.

3.2.1.2 Wikidata

Iniciado em 2012 pela Wikimedia Foundation⁵, o Wikidata contava com um contingente pequeno e restrito de editores que podia criar e conectar itens com a Wikipédia. No ano seguinte, com mais de três milhões de itens criados, realizaram-se integrações com outros idiomas e, posteriormente, em 2014, o acesso foi liberado para mais de 40.000 (quarenta mil) wikipedistas, desses 3.500 (três mil e quinhentos) contribuindo mensalmente com uma média de cinco edições. Esses números fazem do Wikidata um dos projetos mais ativos da Wikimedia Foundation (VRANDECIC; KRÖTZSCH, 2014), que pode ser definido como:

[...] um banco de dados secundário, multilíngue, colaborativo e livre, que coleta dados estruturados para fornecer suporte à Wikipédia, ao Wikimedia Commons, aos demais projetos da Wikimedia [...] (WIKIDATA, 2017, não paginado).

O Wikidata é considerado um BD secundário por compilar as instruções e fontes de informações externas em sua *KB*, o que o torna detentor de um conhecimento muito diversificado, mas ao mesmo tempo verificável, justamente por reter suas fontes. O aspecto multilíngue advém de sua disponibilidade em diversos idiomas e de encorajar a comunidade a editar em qualquer idioma. A colaboração segue a mesma linha de produção e edição adotada na Wikipédia, e se considera livre por adotar a licença “Creative Commons Public Domain Dedication 1.0⁶” para a publicação mais flexível de seus dados (WIKIDATA, 2017).

No Wikidata as entidades são compreendidas como “*Itens*”, também compostos por propriedades e valores, sendo que a propriedade é uma categoria de dados e o valor é a própria descrição do “*Item*” (WIKIDATA, 2017). Atualmente, o Wikidata passa por um processo de migração de conteúdo, não do seu, mas

⁵ <https://wikimediafoundation.org/wiki/Home>

⁶ <https://creativecommons.org/publicdomain/zero/1.0/>

recepcionando os dados advindos do Freebase⁷. Isso se dá devido à representatividade do Wikidata na *Web*. A Google encerrou a Freebase e está auxiliando a comunidade Wikidata no processo de migração do conteúdo Freebase para o Wikidata (TANON et al., 2016). A assistência da Google é extremamente necessária, pois os dados estão armazenados, mapeados e licenciados de maneiras diferentes (proprietário e público), necessitando de muita colaboração para essa integração.

Esta seção encerra a conceituação da *Web Semântica* e seus desdobramentos tecnológicos. As *KBs* mostram-se ao mesmo tempo produtos e também grandes responsáveis pela aplicação, pelo uso e pela disseminação das regras e tecnologias instauradas na *Web 3.0*. Além disso, possibilitam, por meio de padronizações e interoperabilidade, o entendimento entre humanos e máquinas no Ciberespaço. Na próxima seção, serão abordados os princípios que envolvem o desenvolvimento de sua camada subsequente, a *Web Pragmática*.

3.3 WEB PRAGMÁTICA

Nas seções anteriores, foram apresentados os conceitos e tecnologias que circundam o desenvolvimento das dimensões Sintática (*Web 1.0* e Social) e Semântica (*Web 3.0*) da *WWW*. Dessa maneira, serão abordados nesta seção os pressupostos que envolvem a formação de uma terceira dimensão, denominada como *Web Pragmática*. Assim, segundo Monteiro (2015, p. 3), essa dimensão também pode ser qualificada como: “[...] *Web 4.0*, *Web Ubíqua*, *Web Social Semântica* e *Web Semiótica*.”

A origem dessa nomenclatura está ligada aos tradicionais estudos dos signos (sintaxe, semântica e pragmática) realizados pela Semiótica. Em vista disso, o sentido pragmático empregado a essa dimensão decorre do fato de ter que tratar do significado no contexto, pois há uma tendência da área de Tecnologia da Informação em desenvolver aplicativos cada vez mais sensíveis ao contexto (WEIGAND; PASCHKE, 2012). Assim, sob outra perspectiva, Singh (2002a) identifica essa dimensão como Semiótica, pois a considera um sistema simbólico,

⁷ <https://developers.google.com/freebase/>

em que os conteúdos das páginas em conjunto com as linguagens de marcação podem ser interpretados como símbolos.

Na mesma direção, Pietarinen (2003) sugere que os aspectos conceituais que envolvem a *Web Semântica* (lógica de dados) e a *Web Pragmática* (perspectiva humana) não devem ser segregados, mas trabalhados em conjunto para a formação de uma nova camada *Web*, a *Semiótica*. Nesta camada, segundo Monteiro (2015, p. 11), “[...] os significados e o conhecimento compartilhados são constructos de homens e máquinas.”, ou seja, trata-se da *Web Semântica* em seu plano ideal, no qual agentes computacionais e humanos compartilham da mesma linguagem. Nessa perspectiva, Singh (2002a, p. 4, tradução nossa, grifo nosso) justifica que a *Semiótica* pode ser aplicada ao contexto *Web*, pois

[...] nos ajuda a pensar sistematicamente sobre os símbolos que a constituem. Na *Web*, a **sintaxe** refere-se a *tags* (como *tags* HTML ou XML); **semântica** se refere ao que essas *tags* denotam (inventários de peças, por exemplo), e **pragmática** refere-se aos aspectos sensíveis ao contexto do significado (por exemplo, datas de estoque e tempos ou processos que afetam o tamanho do inventário).

Nesse sentido, Singh (2002b, p. 5, tradução nossa) argumenta que é necessária uma *Semântica* especial, ou seja, a *Pragmática* para “[...] incentivar o surgimento de comunidades de interesse e prática que desenvolvam seu próprio conhecimento consensual de base que irá padronizar suas representações.” Nesse contexto, o exemplo do próximo parágrafo, proposto por Schoop, Moor e Dietz (2006) torna mais simples a assimilação da forma como as comunidades podem influenciar no estabelecimento de um conhecimento consensual coletivo.

Em uma situação hipotética, tem-se uma conversa de negócios entre três estranhos de localidades distintas em prol de um projeto que será realizado em conjunto. Entretanto, o assunto em pauta possui divergência conceitual entre os três participantes, fato que dificulta a compreensão de detalhes do projeto. Assim, cada indivíduo compartilha seu conhecimento com os demais para alcançar um consenso coletivo sobre o conceito, porém não haverá uma descrição objetiva do conceito por envolver fatores tácitos, de difícil formalização (SCHOOP; MOOR; DIETZ, 2006). Nessa perspectiva, a *Web Pragmática* é complementar a *Web Semântica* e visa

[...] aumentar efetivamente a colaboração humana por meio de tecnologias apropriadas, tais como sistemas para negociação de ontologias, negócios baseados em interações de ontologias e construção de ontologias pragmáticas por comunidades [...] melhorando a qualidade e a legitimidade de discursos colaborativos e orientados a metas das comunidades. (SCHOOP; MOOR; DIETZ, 2006, p. 76, tradução nossa).

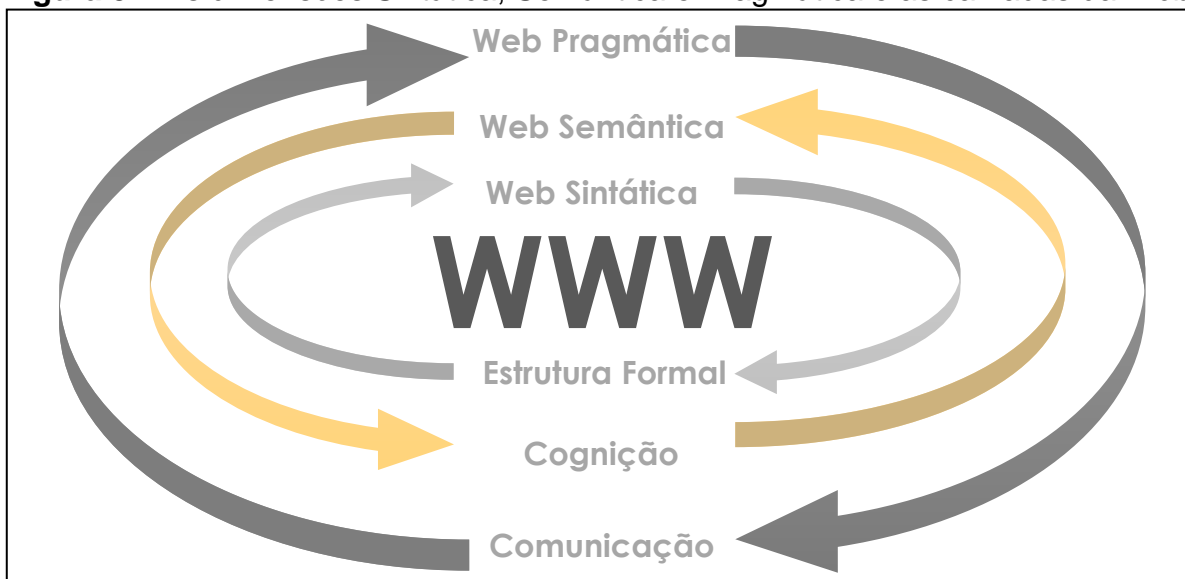
Dessa forma, a *Web Pragmática* parece ser a junção da *Web Social* com a *Web Semântica*, regida pela colaboração em torno do consenso. Assim, Vaidhyanathan (2011, p. 74 *apud* MONTEIRO, 2015, p. 10-11) relaciona esse sentido de a comunidade preocupar-se com a padronização da representação com o *modus operandi* do Google:

Teoricamente, o modo como o Google confia na avaliação coletiva dos milhões de usuários da rede parece concretizar uma das mais influentes teorias da epistemologia: o pragmatismo norte-americano. Como Charles Sanders Peirce e William James formularam na década de 1890. [...] a verdade é gerada por um processo de experimentação, descoberta, *feedback* e consenso.

Dessa maneira, a Pragmática advém do Pragmatismo⁸ e seu foco está na relação entre os signos e seus intérpretes (MORRIS, 1976). Nesse sentido, a Pragmática diferencia-se de Sintaxe e Semântica ao buscar “[...] o uso concreto da linguagem, com vistas em seus usuários [...] na prática linguística [...]” (PINTO, 2006, p. 47). Portanto, concerne à Pragmática o “[...] estudo das relações entre os signos e seus usuários.” (ORLANDI, 2009, p. 53). Assim, na Figura 3 as dimensões da linguagem (Sintática, Semântica e Pragmática) estão associadas às camadas da *Web* para que se tenha uma interpretação de seu modo de funcionamento.

⁸ Corrente de ideias que prega que a validade de uma doutrina é determinada pelo seu bom êxito prático [É esp. aplicado ao movimento filosófico norte-americano baseado em ideias de Charles Sanders Peirce (1839-1914) e William James (1842-1910).]” (PRAGMATISMO, 2009).

Figura 3 – As dimensões Sintática, Semântica e Pragmática e as camadas da *Web*



Fonte: o próprio autor com base em Blühdorn (1997).

Na Figura 3, a disposição cíclica das dimensões da linguagem aplicadas para a interpretação das camadas da *Web* visa esclarecer que, apesar de remeterem à um funcionamento independente para cada camada, elas possuem resquícios tecnológicos, operacionais e conceituais umas das outras e, assim, mesclam-se na composição e expansão da *Web*.

Desse modo, a *Web Pragmática* utiliza os agentes e serviços computacionais da *Web Semântica* para acessar dados e ontologias com as quais fará inferências para a tomada de decisões autônomas. Além disso, essa pragmática aplicada a *Web* tem o mesmo sentido compartilhado pela Linguística e pela Semiótica: o de buscar compreender como o contexto influencia na interpretação do significado. Dessa maneira, em uma comunidade de colaboração, os significados individuais compõem uma estrutura cognitiva que está em sintonia com outras por meio de um processo comunicativo. Isso permite que a máquina possa lidar com questões de ambiguidade, semântica, relevância e sobrecarga de informações (PASCHKE et al., 2007).

A Pragmática na *Web* reside no fato de pôr recursos semânticos em contextos pragmáticos, os quais devem compor-se da junção entre conceitos comuns (senso comum) e individuais para que faça sentido sua adoção na comunidade (MOOR, 2005). Assim, Santaella (2012, p. 41) argumenta que na *Web Pragmática*

[...] as páginas não serão acessadas somente por seu significado, mas também por sua utilização, como, por exemplo, pelas características dos grupos sociais e políticos de que o usuário participa, ou pelo perfil de pesquisa acadêmica a que pertence etc.

Sob uma perspectiva mais imersiva, mas não utópica, Choudhury (2014) a compreende como *Web Simbiótica*, ao passo que seria produto de interações altamente inteligentes entre humanos e máquinas, atingindo o ponto de a máquina conseguir interpretar o conteúdo da *Web* e julgar de maneira autônoma o que é mais relevante para o usuário. Nesse sentido, remete ao conceito de “Simbiose” cunhado na Biologia, no qual há “[...] uma relação mutuamente vantajosa entre dois ou mais organismos vivos de espécies diferentes.” (SIMBIOSE, 2017, não paginado). Assim, humanos e máquinas estabeleceriam uma relação bilateral, imersiva e com muita Inteligência Artificial (IA) no Ciberespaço.

No entanto, a perspectiva adotada nesta investigação é de uma *Web Pragmática* pautada no conhecimento consensual da comunidade sobre um mesmo objeto. Assim, contempla não apenas o significado, mas também a maneira como ele é empregado por diferentes grupos. Dessa maneira, faz a junção entre a lógica dos dados estruturados, a comunidade *Web*, o significado consensual e múltiplos contextos.

3.4 MECANISMOS DE BUSCA

Atualmente, parece não haver mais dúvidas que não possam ser sanadas a partir de uma breve consulta a Mecanismos de Busca (MB) da *Web*. Esses mecanismos tornam-se cada vez mais atrativos para a resolução de problemas corriqueiros e até para questões mais complexas. Talvez esse sucesso seja fruto da simplicidade com a qual se pode questionar uma máquina sobre diferentes temáticas e obter respostas, resultados e redirecionamentos imediatos e concisos.

Essa notoriedade dos MB foi alcançada apenas recentemente com a evolução dessas ferramentas. Nesse sentido, nos primórdios da *Web*, logo que foi lançada por Tim Berners-Lee com a proposta de ser um espaço comum para a troca e acesso a diferentes tipos de informações, os MB dedicavam-se a replicar páginas da *Web* em HTML simples para suas bases de dados. Assim, esse processo não

contemplava a *Web* de maneira geral, além de restringir e controlar os resultados veiculados a apenas documentos HTML, desconsiderando outros formatos de dados existentes (SHERMAN; PRICE, 2001).

Dessa maneira, a cada nova camada, a *Web* emprega diferentes formas de organização, representação e recuperação do conteúdo produzido e veiculado em sua extensão. Com tal característica, a etapa de recuperação das informações parece ser o ponto validador de seu progresso, pois, ao considerá-lo em um movimento cíclico que envolve as fases de organização, representação e recuperação de dados, esta última é que torna legítimo todo o processo anterior de padronização de informações. Antes de adentrar aos pressupostos relacionados aos MB para a *Web*, torna-se necessária uma breve incursão entre as tecnologias e conceitos que circundam o funcionamento de um Sistema de Recuperação da Informação (SRI), fundamento estrutural de um MB.

Em CI, a Recuperação da Informação (RI) já era abordada em estudos anteriores ao surgimento da *Web*. Os primeiros passos, na década de 1940, referenciavam-se aos SRI, cujo objetivo principal era apresentar informações relevantes aos usuários (SARACEVIC, 1975). Assim, segundo Fachin (2009), entre os anos de 1940 a 1970, o enfoque foi automatizar a informação e as unidades de informação; já de 1980 a 1990, com a ampliação das TIC, houve o surgimento de novas ferramentas para tratamento, disseminação e recuperação da informação, que resultaram em importantes avanços científicos para as áreas de Computação, Biblioteconomia e CI.

Pioneiras, as bibliotecas inovaram com a adoção dos SRI, que, de início, apenas automatizaram as tradicionais buscas por autor e título. No desenrolar dos anos, as palavras-chaves e os cabeçalhos de assunto foram inseridos, assim como houve a otimização da interface gráfica, inserção de formulários, recursos de hipertexto e arquitetura de sistemas abertos (BAEZA-YATES; RIBEIRO-NETO, 1999).

Apesar da constante transformação, os SRI ainda mantêm o mesmo desafio, o de tornar seus resultados relevantes aos usuários. Esse entrave, na realidade, acaba por estimular o desenvolvimento de SRI mais inteligentes. No entanto, o problema está no processo de tradução da noção humana de relevância para a formalizada da máquina, pois a relevância difere entre as pessoas, seja por crenças, subjetividade, pessoalidade e/ou imparcialidade.

Para Saracevic (2015) compreender a relevância de forma intuitiva não requer muita explanação. Entretanto, apesar dos MB considerarem questões culturais, sociais e contextuais na apresentação de seus resultados, ainda há a fragilidade de que a noção de relevância é humana e não técnica. Esse fato a torna de difícil formalização por possuir muitas variáveis. Em concordância, Silva, Santos e Ferneda (2013, p. 37) consideram que a relevância consiste em

[...] mostrar os resultados possivelmente mais relevantes em forma de ranque (*ranking*), do mais relevante ao menos relevante. Entretanto, o conceito de relevância é subjetivo e inexato, não podendo ser definido por fórmulas matemáticas e implementadas em sistemas computacionais.

Desse modo, a construção de SRI eficazes somente será possível quando se traduzir a noção de relevância intuitiva em formalismo escrito (LAVRENKO, 2009). Enquanto humanos usam o contexto para a resolução de problemas, os mecanismos usam algoritmos e processos tecnológicos aliados à probabilidade na apresentação de seus resultados (SARACEVIC, 2012).

Nessa perspectiva, a Ciência da Computação (CC) e a CI são a base para dos estudos de RI. Na CC a preocupação está em avaliar e evoluir tecnicamente a RI, enquanto em CI o enfoque permeia o usuário e a avaliação deste sobre o sistema. Os motores de busca para *Web* desafiam essas áreas a trabalhar a interdisciplinaridade, pois seu uso na rede desperta também o interesse de outras disciplinas, em virtude do impacto social desses mecanismos (LEWANDOWSKI, 2012).

Na CI, a relevância tornou-se central quando as preocupações práticas e teóricas se voltaram para a RI e não apenas na organização de informações. A Biblioteconomia envolve-se com a temática, com a organização das informações, enquanto a CI se preocupa com a recuperação, trabalhando com a relevância (SARACEVIC, 2012). Dessa forma, os índices dos motores de busca *Web* são muito semelhantes aos usados pelos bibliotecários do século passado. No entanto, há três características que os diferenciam. A primeira se refere ao barateamento do acesso às diversas fontes de informação. A segunda diz respeito aos avanços das TIC que promoveram o acesso à rede. Já a terceira versa sobre a liberdade de uso e publicação de conteúdo na *Web* (BAEZA-YATES; RIBEIRO-NETO, 1999).

Dessa maneira, com os avanços das TIC, mais pessoas começaram a usufruir desses SRI na *Web* e, conseqüentemente, os MB se tornaram cada vez mais necessários, pois, na medida em que o acesso aumentou, a quantidade de conteúdo veiculado também cresceu, ligados a uma grande dificuldade de recuperação dessas informações. Por conseqüência, na *Web*, os MB são responsáveis por concentrar boa parte do tráfego da *Internet*. Além disso, exercem determinadas tarefas que, a partir de uma analogia, podem ser comparadas a papéis exercidos por diferentes profissionais, por exemplo, o de um bibliotecário que dá “[...] ordem ao caótico acúmulo de informações *on-line*.”, ou até mesmo a de um crítico que determina quais serão os conteúdos apropriados para exibição e aqueles que nunca serão apresentados aos usuários do MB (GRIMMELMANN, 2007, p. 3, tradução nossa).

Nesse contexto, essas ferramentas para recuperação de informações da rede ou em plataformas específicas da *Web* são entendidas nesta pesquisa da seguinte maneira:

[...] mecanismos de busca, sistemas de busca, motores de busca, buscadores, *search engine*, são os termos utilizados para indicar um conjunto organizado de computadores, algoritmos, bases de dados e índices reunidos com a função de analisar e indexar as páginas da *Web* e apresentá-las de forma organizada aos usuários. (OLIVEIRA; ARAÚJO, 2012, p. 64-65).

Esses motores de busca permitem a organização do conhecimento no Ciberespaço por meio da indexação que realizam (MONTEIRO, 2006). Além disso, os MB correspondem à aplicação de práticas e técnicas de RI voltadas para grandes coleções de informações. Na *Web*, MB como Google e Yahoo! são destinados a rastrear e capturar milhões de *terabytes* de dados e convertê-los em respostas instantâneas para diversas consultas (CROFT; METZLER; STROHMAN, 2015).

Um MB é uma ferramenta de pesquisa que apresenta *websites* relevantes instantaneamente com base em requisições dos usuários. Dessa maneira, os resultados recuperados são apresentados em uma *Search Engine Results Page* (*SERP*) em ordem de relevância. Para indexar essas informações, robôs denominados “*spider*” ou “*crawler*” são enviados pelo MB para pesquisar e indexar esses dados (SHENOY; PRABHU, 2016).

Dessa maneira, um motor de busca faz parte de um sistema maior, que utiliza aplicativos (*crawler*, *spider* ou robô) para capturar palavras-chaves e frases que identifiquem o conteúdo de uma página *Web* para indexar e armazenar em sua base de dados. Além disso, conta com uma interface destinada ao usuário para consultar essas informações mediante a inserção de termos de pesquisa, que irão acionar o algoritmo de busca para exibir os resultados recuperados da base de dados (LEDFORD, 2009).

Segundo Shenoy e Prabhu (2016), MB *Web* mais modernos são compostos por alguns procedimentos específicos de funcionamento, são eles: *Web crawling*, *indexing* e *searching*. Nesse contexto, esses requisitos podem ser interpretados como componentes da “anatomia” de um MB, sendo o “*crawling*” responsável por ‘varrer’ o Ciberespaço para a coleta de páginas da *Web*, o “*indexing*” indexa ou gera índices a partir dessa coleta e o “*searching*” se refere à interface de busca destinada ao usuário (MONTEIRO, 2009). Sobre a etapa de “*crawling*”, os “*spiders*” são enviados pelos MB para visitar os *websites* para a coleta de dados de um arquivo denominado como “*robots.txt*”. Feito isso, a partir das informações que coletaram, os “*spiders*” as transformam em um índice de termos e *websites* que contém esses termos (SHENOY; PRABHU, 2016).

Os robôs indexadores dos MB usam as tecnologias estruturais da *Web Semântica* para circular entre os *websites* e coletar diversas informações que serão convertidas em resultados de busca. Nesse contexto, há toda uma complexidade de desenvolvimento (*crawling*, *indexing* e *searching*) dessas ferramentas para que possam funcionar harmonicamente na *Web*. No entanto, depois de estabelecer esse modo de operação primário de coleta, processamento e geração de índices para consultas, ou seja, por em atividade um MB na *Web*, ainda é necessário se ater a questões de performance e avaliar constantemente o funcionamento desses MB com base na experiência do usuário.

Os MB devem estar atentos ao cumprimento de alguns requisitos para uma boa RI; são eles: algoritmo de ranqueamento eficaz, avaliação e interação com o usuário. Além disso, há questões mais críticas, relacionadas à performance de MB; são elas: “*response time*” - tempo de resposta entre a requisição do usuário e a disponibilização da lista de resultados; “*query throughput*” - mensura a quantidade de requisições que podem ser realizadas simultaneamente; “*indexing speed*” - o tempo gasto para um documento se transformar em índice; “*coverage*” -

mensura a existência do dado indexado e há quanto tempo está na base; “*scalability*” - capacidade do sistema de se adequar a crescente de dados e usuários; “*customizable and adaptable*” - para se ter escalabilidade é necessário ser customizável e adaptável; “*spam*” - controlar páginas de falsos-positivos que influenciam negativamente o ranqueamento, seus desenvolvedores indexam informações de má fé para permanecer por mais tempo na lista de resultados (CROFT; METZLER; STROHMAN, 2015).

Como apresentado nesta seção, a RI é uma preocupação anterior a implementação das tecnologias *Web*, iniciada nas áreas de CI e CC para aperfeiçoar diferentes SRI. O interesse na RI perdura até os dias atuais em virtude da dificuldade de recuperação de informações relevantes produzidas e veiculadas na *Web*.

Assim, os MBs têm uma constante necessidade de aprimoramento das tecnologias que permeiam seu funcionamento para que possam manter a competitividade em meio a tantos outros MB, mas essencialmente para indicar resultados cada vez mais precisos e com sentido para o contexto do sujeito que os requisita. Nesse sentido, tecnologias como os *Knowledge Graphs (KBs)* auxiliam no processo de transformação da RI no contexto *Web*.

3.4.1 *Knowledge Graphs*

Os *KGs* são importantes recursos para aplicações *Web* e *mobile* que utilizam IA para processos de recuperação de informações. Podem ser interpretados como grafos multirrelacionais compostos por entidades e seus relacionamentos (WANG et al., 2014). Assim, pessoas e organizações são exemplos de entidades, e seus relacionamentos e informações fatuais são obtidos a partir de dados estruturados em *Knowledge Bases* como Freebase e Wikidata (ROSPOCHER, et al. 2016). Antes da implementação do *KG*, as pesquisas nos MB estavam pautadas apenas na descrição textual para a recuperação, algo que não evidenciava o que de fato as pessoas sabiam sobre um termo, isto é, seu significado (SINGHAL, 2012).

Segundo Arenas et al. (2016, p. 1, tradução nossa), os *KGs* correspondem a “[...] grandes coleções de entidades interconectadas e enriquecidas com anotações semânticas, tornando-os recursos essenciais para melhorias em pesquisas [...].” Dessa maneira, os *KGs* são empregados para melhorar os

resultados fornecidos pelos MB com base em informações estruturadas em *KB* como Freebase, DBpedia, Yago e Wikidata (ROSPOCHER et al., 2016). Segundo Arenas et al. (2016), os exemplos a seguir são proeminentes *KGs* de larga escala: Yago, Freebase, Google's Knowledge Graph, Facebook's Graph Search, Microsoft's Satori, e Yahoo's Knowledge Graph.

Esses repositórios concentram conhecimento enciclopédico e informações fatuais vinculados ao nome e sobrenome de pessoas famosas, data e local de nascimento, como também a atividade profissional desempenhada pelo indivíduo. Entretanto, notícias recentes nem sempre são exibidas pelos *KGs*, visto a escassez de informações estruturadas sobre esse tipo de evento (ROSPOCHER et al., 2016). Sullivan (2012) compreende os *KGs* como "*knowledge panels*" destinados a fatos e não a ações, isto é, se procurar por um determinado ponto turístico serão exibidas informações fatuais sobre a localidade, porém não haverá nenhum *hyperlink* de ação para que o usuário possa reservar uma passagem ou hotel.

Muitos *KGs* são disponibilizados por meio do *LOD*, fato que possibilita exportar conjuntos de dados em *RDF*, como também aprimorá-los com a aplicação de *OWL*. Além disso, o conjunto *RDF*, *OWL* e *SPARQL* torna os *KGs* acessíveis para aplicações na *Web* (ARENAS et al., 2016). Nesse sentido, ao adotar as tecnologias da *Web Semântica* para a estruturação e compartilhamento de seus dados, os *KGs* simplificam o acesso e consulta de seu conteúdo para uso de outras tecnologias. Para Monteiro (2015, p. 5), o *KG* "[...] é um agente inteligente, que busca metadados e algoritmos: busca o conhecimento estruturado e vinculado, [...] reconhece entidades [...] e consulta bases referenciais de conhecimento."

Dessa maneira, os MBs foram as tecnologias responsáveis por alavancar a popularidade dos *KGs*, visto que motores como Google e Bing passaram a empregá-los constantemente na apresentação de seus resultados (ROSPOCHER et al., 2016). Porém, Wang et al. (2014) apontam a existência de duas grandes dificuldades relacionadas a implantação de *KGs*: a primeira refere-se ao fato do *KG* ser um sistema lógico e simbólico, enquanto o restante das aplicações *Web* estão baseadas em computação numérica; a segunda corresponde a dificuldade em agregar o conhecimento mundial sobre um grafo. Entretanto, esses fatos não parecem ter sido empecilhos para o desenvolvimento pioneiro do Google's Knowledge Graph.

No ano de 2012, o *KG* do Google foi iniciado com o intuito de apresentar resultados mais precisos e úteis, além de também influenciar no “*auto-complete*” de pesquisas preditivas, aquelas em que o MB conclui a busca para o usuário. Além disso, para algumas temáticas reconhecidas pelo *KG*, outro elemento é evidenciado: o “*answer box*”, que apresenta caixas com respostas às requisições realizadas na própria SERP, sem ser necessário acessar o *website* (BUTZBACH, 2017). Dessa forma, segundo Singhal (2012, não paginado, tradução nossa), a presença do *KG* no MB Google

[...] permite pesquisar coisas, pessoas ou lugares que o Google conhece – marcos, celebridades, equipes esportivas, prédios, características geográficas, filmes, objetos celestiais, obras de arte e muito mais – e obter informações relevantes instantaneamente para sua pergunta.

Subsequentemente, no ano de 2013, o Bing (Microsoft) implementou em seu MB o Satori⁹, um *KG* de natureza equivalente aos outros, porém com um diferencial: segundo Qian (2013), além de pessoas famosas e celebridades, o Satori consegue apresentar também profissionais por meio de consulta a dados da Wikipédia e LinkedIn¹⁰. No ano de 2014, foi a vez da Yahoo! implantar a sua versão do *KG* que segue os mesmos padrões de desenvolvimento de *KGs* convencionais, com apenas uma peculiaridade em comparação aos outros, que é a utilização do mesmo motor do Bing com apresentação de resultados de forma diferente (SCHWARTZ, 2015).

Portanto, os *KGs* são a exibição gráfica/textual de informações semanticamente interligadas em *KBs*, mediante o uso e a aplicação de tecnologias e conceitos estabelecidos na *Web Semântica*, como *LD*, *RDF*, *OWL*, *SPARQL*, etc. Assim, têm por objetivo fazer com que os resultados apresentados nos MB tenham mais sentido para além do significado, isto é, buscam envolver o contexto e perfil de uso do sujeito informacional na apresentação dos resultados.

Dessa maneira, com base na revisão de literatura realizada ao longo de três principais seções de assuntos, esta investigação conseguiu reunir importantes fatos, teorias e conceitos de diferentes áreas do conhecimento que subsidiaram o desenvolvimento da análise, apoiada no percurso metodológico

⁹ Significa “entendimento” em japonês (QIAN, 2013).

¹⁰ <https://www.linkedin.com/>

proposto na próxima seção deste estudo. Assim, com vistas a simplificar o entendimento sobre a composição desta fundamentação teórica, o Quadro 5 apresenta a estrutura de seções temáticas em conjunto com os autores empregados para os esclarecimentos sobre cada assunto nesta pesquisa.

Quadro 5 – Síntese de fundamentos teóricos

Seção Principal	Subseções	Autores
Conhecimento Enciclopédico		Marcondes (2010); Geeraerts (2006); Eco (2014); Croft e Cruse (2004); Koch e Elias (2008); Koch (1977 <i>apud</i> KOCH; ELIAS, 2008); Kiefer (1988); Meijs e Vossen (1991); Cavazza e Zweigenbaum (1995); Koch e Elias (2008); Nunberg (1978 <i>apud</i> MURPHY, 2003); Murphy (2003); Allan (2006); Pombo (2002); Haiman (1980); Pombo (2006); Serra (1999); Moraes (2013).
Ciberespaço	Ciberespaço	Monteiro (2007); Lévy (1999); Santaella (2004); Ho Kim (2004); Monteiro e Fidencio (2013); Lazzarin, Netto e Sousa (2015).
	Web Sintática	Alesso e Smith (2006); Berners-Lee (1989; 1996); Berners-Lee, Fielding e Masinter (2005); Khare e Rifkin (1997); WEB FOUNDATION (2017); Aghaei, Nematbakhsh e Farsani (2012); Monteiro e Fidencio (2013); Vecchiato e Vidotti (2014); Monteiro (2012); Viotti (2008); Perini (2010); Morris (1933 <i>apud</i> ILARI, 2000); Ilari (2000); Blühdorn (1997); Fuchs et al. (2010); Cormode e Krishnamurthy (2008); Monteiro (2015); Murugesan (2007); Berners-Lee e Fischetti (2000); Boulos e Wheelert (2007); Bosak e Bray (1999); Yu (2007); Knights (2007); W3SCHOOLS (2016); Fazenda (2017); Mika (2007); Chatfield (2009); Leuf e Cunningham (2001); Woods e Thoeny (2007); Lih (2009).
	Wikipédia	Lih (2009); Ayers, Matthews e Yates (2008); Jemielniak (2014); Anderson (2011); Miller (2010); Mihalcea e Csonai (2007); Barrett (2009).
	Web Semântica	Schoop, Moor e Dietz (2006); Fuchs et al. (2010); Aghaei, Nematbakhsh e Farsani (2012); Berners-Lee, Hendler e Lassila (2001); Mika (2007); Breitman (2010); Heath e Bizer (2011); Legg (2008); Bax (2012); Oliveira (2006); Landis (2014); Yu (2007); W3C (2013a; 2013b; 2017a; 2017b); IRI (2017); Domingue, Fensel e Hendler (2011); Santarém Segundo, Souza e Coneglian (2015); Bizer, Heath e Berners-Lee (2009); Kontokostas et al. (2012); Monteiro (2015).
	Knowledge Bases	Morsey et al. (2009); Sterckx et al. (2016); Bizer et al. (2009); Mahdisoltani, Biega e Suchanek (2015); Biega, Kuzey e Suchanek (2013); Suchanek (2016).
	DBpedia	Bizer et al. (2009); DBPEDIA (2016); Lehmann et al. (2012); Kontokostas et al. (2012).
	Wikidata	Vrandečić e Krötzsch (2014); WIKIDATA (2017a); Tanon et al. (2016).
	Web Pragmática	Monteiro (2015); Weigand e Paschke (2012); Singh (2002a; 2002b); Pietarinen (2003); Schoop, Moor e Dietz (2006); Vaidyanathan (2011, <i>apud</i> MONTEIRO, 2015); Morris (1976); Pinto (2006); Orlandi (2009); Pragmatismo (2009); Blühdorn (1997); Paschke et al. (2007); Moor (2005); Santaella (2012); Choudhury (2014); Simbiose (2017).
	Mecanismos de Busca	Sherman e Price (2001); Saracevic (1975; 2012; 2015); Fachin (2009); Baeza-Yates e Ribeiro-Neto (1999); Silva, Santos e Ferneda (2013); Lavrenko (2009); Lewandowski (2012); Grimmelmann (2007); Oliveira e Araújo (2012); Monteiro (2006; 2009); Croft, Metzler e Strohman (2015); Shenoy e Prabhu (2016); Ledford (2009).
	Knowledge Graphs	Wang et al. (2014); Rospocher et al. (2016); Singhal (2012); Arenas et al. (2016); Sullivan (2012); Monteiro (2015); Butzbach (2017); Qian (2013); Schwartz (2015).

Fonte: o próprio autor.

4 PERCURSO METODOLÓGICO

De natureza básica, esta investigação centrou seus esforços para a expansão do conhecimento acerca do Ciberespaço em Ciência da Informação no Brasil. Em vista disso, não se propôs o desenvolvimento de *softwares* ou soluções aplicáveis como possíveis produtos dos fenômenos aqui estudados. Assim, corroborou-se a perspectiva de Connaway e Powell (2010, p. 6) de que uma das principais atribuições de uma pesquisa básica é a de “[...] criar novos conhecimentos.”

Sem a pretensão de quantificar o grandioso volume informacional existente na *Web*, este estudo adotou a abordagem qualitativa para compreender as influências que podem ser ocasionadas a partir da “amarração” entre as diferentes tecnologias aqui abordadas. Nesse sentido, Connaway e Powell (2010, p. 77) argumentam que a pesquisa qualitativa consiste em “[...] uma abordagem mais holística e natural para a resolução de um problema [...]” Holewa (2010, p. 680) complementa que a abordagem qualitativa se refere a uma “[...] metodologia não-numérica que tem como objetivo descrever e compreender [...]” Dessa maneira, a partir de uma visão holística sobre o conhecimento enciclopédico colaborativo e as tecnologias *Web* relacionadas, tornou-se possível descrever e analisar as influências desses fenômenos no desenvolvimento da *Web* Pragmática.

Para atender aos objetivos específicos desta investigação, optou-se pelo delineamento documental, que tem por base o uso de

[...] qualquer um dos suportes de informação decorrentes de momentos anteriores à pesquisa, quer em andamento, quer relatadas, ou então de informação resultantes do Fazer Humano ligado à outras áreas que não a ciência [...] cujos objetivos ou hipóteses podem ser verificados através da análise de documentos bibliográficos ou não-bibliográficos [...] (WITTER, 1990, p. 19-22).

Dessa maneira, a escolha pelo delineamento documental propiciou o estreitamento de laços entre *corpus* teórico (revisão de literatura) e metodológico (Wikipédia, Knowledge’s Base e Mecanismos de Busca) com o objetivo de responder a seguinte questão: quais as contribuições do conhecimento enciclopédico colaborativo no desenvolvimento da *Web* Pragmática? Assim sendo, sua aplicação se deu tanto para a construção teórica da revisão de literatura, quanto

para a própria análise em si. Esse delineamento se mostrou adequado em virtude da dinamicidade de produção e constante atualização de conteúdos na *Web*, aspectos estes que a ciência nem sempre pode acompanhar devido à necessidade de validação e reavaliação de textos científicos. Nessa perspectiva, por se tratar de um tema atual, tornou-se inevitável consultar fontes de informações que não tenham passado por uma revisão científica criteriosa dos pares no desenvolvimento do *corpus* teórico e metodológico desta investigação, a exemplo, o conteúdo de *websites* não acadêmicos.

Para as etapas de levantamento e análise de dados, esta pesquisa subdividiu-se em três fases associadas aos objetivos específicos propostos anteriormente. Dessa forma, em cumprimento do **OBJETIVO ESPECÍFICO A**, que visou descrever o curso do conhecimento enciclopédico colaborativo na *Web*, deu-se início ao levantamento de possíveis unidades de análise que poderiam contemplar o aspecto do conhecimento enciclopédico colaborativo na rede. Dessa maneira, a partir de dados obtidos do *website* da Wikipédia¹¹, se chegou ao número de 162 enciclopédias *on-line*.

Porém, diante do elevado número de enciclopédias recuperadas, tornou-se necessário o uso de restrições mediante a aplicação de filtros para garantir a exequibilidade desta pesquisa. Dessa forma, buscou-se por enciclopédias gerais, multilíngues, colaborativas, de acesso aberto e que fizessem conexões com as bases de dados semânticos DBpedia e Wikidata. Assim, 53 enciclopédias foram enquadradas como gerais; destas, apenas quatro consideradas multilíngues, colaborativas e de acesso aberto. Em vista disso, o fator determinante para a escolha do *corpus* metodológico que viria a compor esta investigação foi a possibilidade de conexão com bases de dados semânticos. Com esse requisito, apenas a Wikipédia foi selecionada para a análise.

Dessa maneira, credenciais de acesso foram criadas no intuito de aprofundar o conhecimento sobre o *modus operandi* (pesquisa, edição, organização, validação, publicação e recuperação de conteúdo) da Wikipédia, além do próprio conhecimento enciclopédico nela registrado. Além disso, focou-se também em como essas informações eram estruturadas, quais regras e linguagens eram utilizadas. Assim, por se tratar de uma enciclopédia, considerou-se importante que a

¹¹ Acesso a lista: https://en.wikipedia.org/wiki/List_of_online_encyclopedias

análise/observação fosse realizada mediante o monitoramento de alguns verbetes que a compõe.

A escolha dos termos monitorados ocorreu a partir de possíveis especulações associativas que os mecanismos de busca poderiam vir a realizar entre eles por meio de dados ligados. Dessa forma, os termos se dividiram em três qualidades distintas, mas convergentes entre si semanticamente, são elas: Tim Berners-Lee (pessoa), *England* (lugar) e *Semantic Web* (coisa). Para confirmar a possibilidade de rastreamento dos termos selecionados por entre as plataformas elencadas, utilizou-se o *website* BabelNet¹², autointitulado como “[...] um dicionário enciclopédico multilíngue [...]”. Assim, tornou-se possível pesquisar pelos termos listados e verificar se possuíam referências na Wikipédia, DBpedia e Wikidata para validar o ciclo que seria realizado na etapa de análise desta investigação.

Destinada ao cumprimento do **OBJETIVO ESPECÍFICO B**, que se propôs a analisar como as bases de conhecimento semântico estruturam o conhecimento enciclopédico extraído da Wikipédia, a segunda etapa do percurso metodológico buscou determinar quais bases de conhecimento seriam utilizadas para o processo de análise desta pesquisa. Dessa maneira, apoiou-se sobre a recente pesquisa de Färber *et al.* (2015), em que se realiza um estudo comparativo entre as principais bases de conhecimento da atualidade; são elas: DBpedia, Freebase, OpenCyc, Wikidata e Yago.

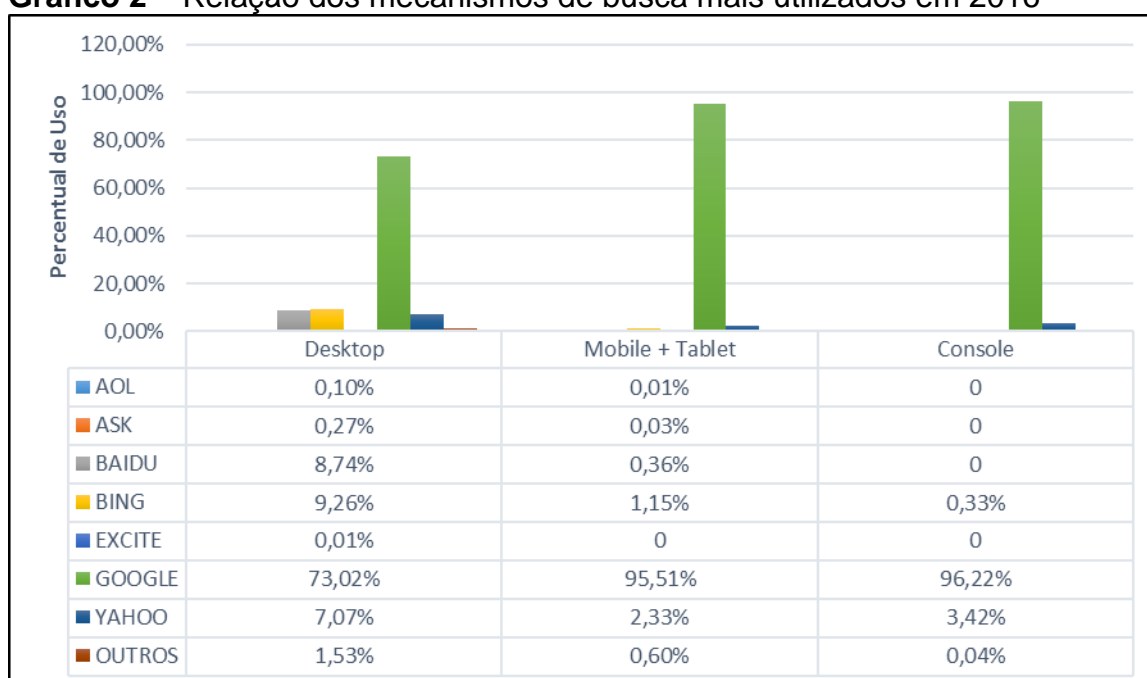
No entanto, esta pesquisa focou apenas em bases que estivessem completamente disponíveis *on-line*, além de serem colaborativas e de acesso aberto. Assim, apenas DBpedia e Wikidata se classificaram, pois a Freebase foi desativada pela Google e seus dados estão em processo de migração para Wikidata. Por outro lado, OpenCyc e Yago disponibilizam o acesso a suas bases, porém é necessário realizar *download* e instalação em uma máquina local, fator este que inviabilizaria a pesquisa, visto que não se contava com computadores com os requisitos necessários para tal. Posteriormente, retomou-se o processo de acompanhamento dos termos, desde verbetes da Wikipédia até sua conversão em entidade pelas bases de conhecimento semântico DBpedia e Wikidata.

A terceira e última etapa atendeu ao **OBJETIVO ESPECÍFICO C** de investigar se os principais mecanismos de busca da atualidade utilizam e de qual

¹² <http://babelnet.org/>

maneira o conhecimento enciclopédico colaborativo obtido da Wikipédia por intermédio das *Knowledge Bases*: DBpedia e Wikidata. Neste passo, a partir de dados obtidos do website “<https://netmarketshare.com/>”, tornou-se possível elencar quais mecanismos de busca seriam utilizados para o processo de análise desta investigação. Desse modo, foram reunidas informações sobre os mecanismos de busca atuais em diferentes contextos de uso como *desktop*, *mobile*, *tablet* e *consoles* (*videogames*, *smart TVs*, entre outros), conforme apresentado no Gráfico 2.

Gráfico 2 – Relação dos mecanismos de busca mais utilizados em 2016



Fonte: o próprio autor com base em NETMARKETSHARE (2016).

Dessa maneira, os mecanismos de busca que apresentaram percentuais nulos de utilização em ao menos um dos dispositivos levantados foram desclassificados para esta pesquisa; são eles: Aol, Ask, Baidu e Excite. Aplicou-se esse critério, pois se entende que, ao estudar a *Web Pragmática*, é necessário analisar mecanismos de busca presentes em diferentes contextos de uso. Portanto, foram selecionados os seguintes mecanismos: Bing¹³, Google¹⁴ e Yahoo!¹⁵. Tanto a Wikipédia quanto DBpedia, Wikidata e os mecanismos de busca foram analisados em domínios internacionais, pertencentes à língua inglesa, pois esta, conforme

¹³ <http://www.bing.com/>

¹⁴ <https://www.google.com/ncr>

¹⁵ <https://us.yahoo.com/>

apresentado no Gráfico 1, possuía um elevado número de verbetes e, conseqüentemente, um número maior de entidades nas bases de conhecimento semântico. Assim, buscou-se na Figura 4 simplificar o percurso metodológico aqui apresentado, tornando-o uma seqüência de etapas a serem cumpridas.

Figura 4 – Síntese do percurso metodológico



Fonte: o próprio autor.

5 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Nesta seção, são apresentados os resultados obtidos a partir da análise realizada com base no percurso metodológico proposto anteriormente. Ao acessar a edição em língua inglesa (Estados Unidos) do *website* da Wikipédia¹⁶, deu-se início ao processo de análise desta dissertação. Na plataforma, o objetivo foi descrever o curso do conhecimento enciclopédico colaborativo na *Web*, mediante o monitoramento dos seguintes termos: “Tim Berners-Lee”, “England” e “*Semantic Web*”.

Posteriormente, as *Knowledge Bases* DBpedia e Wikidata também foram acessadas no intento de compreender como o conhecimento enciclopédico colaborativo extraído da Wikipédia era estruturado nessas bases, ao converter verbetes em entidades. A última etapa do processo de análise investigou se os mecanismos de busca Bing, Google e Yahoo! utilizavam o conhecimento enciclopédico colaborativo extraído da Wikipédia e estruturado pelas *Knowledge Bases* DBpedia e Wikidata e, se sim, de que maneira isso era feito. Assim, as próximas subseções apresentam as especificidades de cada uma das etapas de análise desta investigação.

5.1 WIKIPÉDIA

Na Wikipédia, credenciais de acesso permitiram um desenvolvimento mais imersivo do procedimento de análise desta pesquisa, pois, ao se autenticar como usuário, o pesquisador também se tornou parte da comunidade *on-line*, permitindo-lhe observar de forma mais aprofundada seu *modus operandi*. Entretanto, vale salientar que a política pregada pela Wikipédia (2017a) é de uma “[...] enciclopédia livre que todos podem editar [...]”, logo, não era necessário criar uma conta para contribuir ou simplesmente navegar entre seus verbetes. Porém, ser um membro da comunidade possibilitou compreender de uma maneira mais participativa seu funcionamento, além de se ter acesso a ferramentas e funcionalidades exclusivas a usuários autenticados.

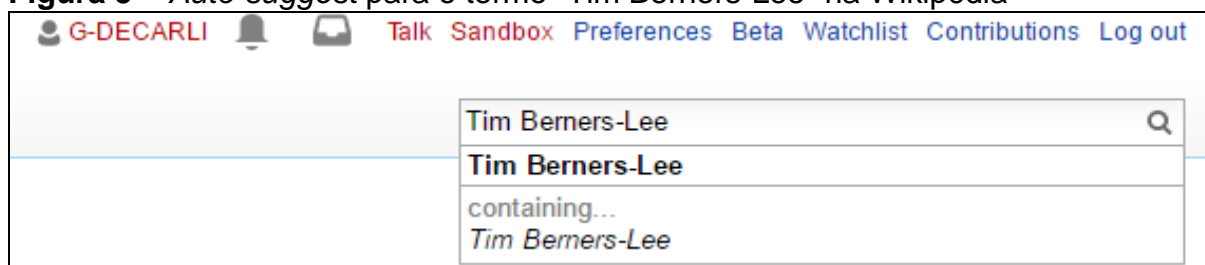
Dessa forma, o acesso à Wikipédia objetivou descrever o curso do conhecimento enciclopédico colaborativo na *Web*. Após a autenticação de usuário, o

¹⁶ https://en.wikipedia.org/wiki/Main_Page

processo de monitoramento dos seguintes termos foi iniciado: “Tim Berners-Lee” (pessoa), “*England*” (lugar) e “*Semantic Web*” (coisa/conceito). Para cada um desses termos há um artigo na Wikipédia que pode ser compreendido também como um verbete. Assim, os termos elencados para a análise foram recuperados mediante consultas realizadas no MB nativo da Wikipédia, visto a inexistência de outras ferramentas para a recuperação de informações na comunidade.

Sem a obrigatoriedade de uma ordem específica de avaliação dos termos, o procedimento de análise foi iniciado mediante a inserção do termo “Tim Berners-Lee” na caixa de busca, disponibilizada no canto superior direito de diferentes páginas do *website* da Wikipédia. Assim, após digitar o termo exato e não iniciar a pesquisa, o MB sugeriu duas opções complementares de requisição, por meio da função *auto-suggest*¹⁷, conforme apresentado na Figura 5.

Figura 5 – *Auto-suggest* para o termo “Tim Berners-Lee” na Wikipédia



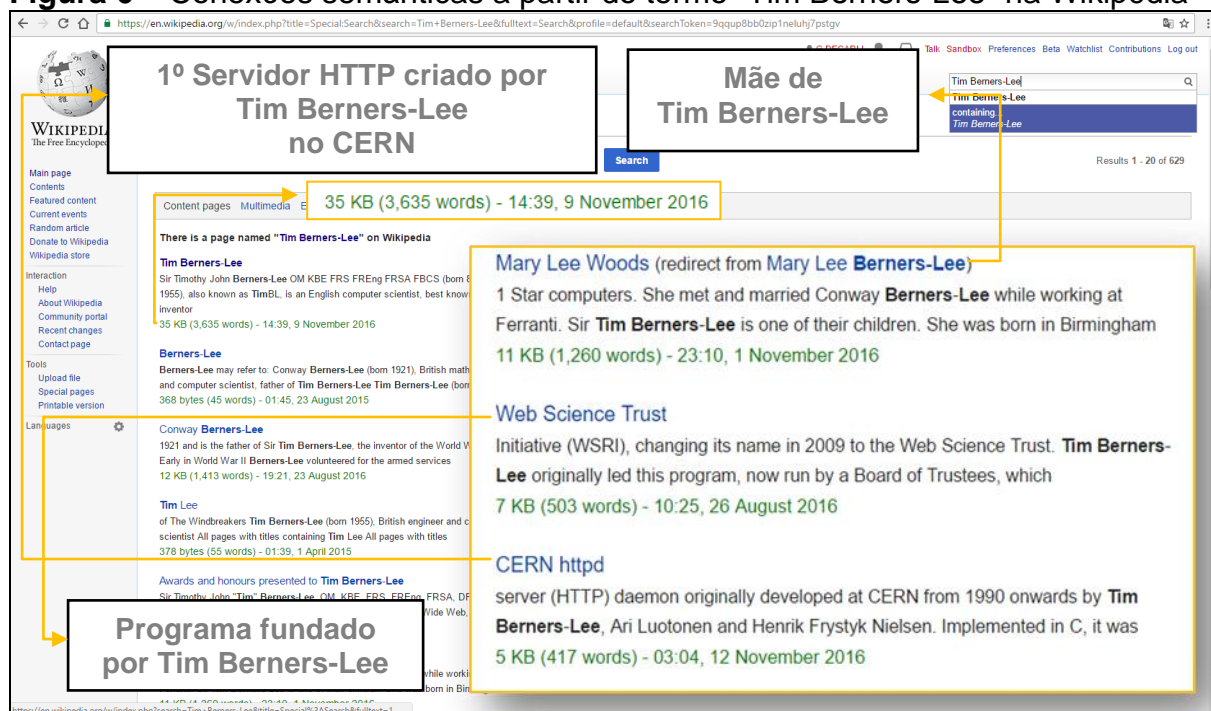
Fonte: adaptado de Wikipédia (2017b).

Ao escolher por “Tim Berners-Lee” em negrito, o usuário é redirecionado para a página do verbete sobre “Tim Berners-Lee”. No entanto, ao selecionar a opção “*containing... Tim Berners-Lee*”, a Wikipédia exibe uma “*Special page*”, na qual são expostas todas as páginas que se relacionam direta ou indiretamente com o termo pesquisado. Além disso, a “*Special page*” disponibiliza várias opções de filtragem para esses vínculos. Apesar de haver esta opção, o enfoque nesta investigação recaiu sobre a opção “*Everything*” para que se pudesse ter ideia do global. O termo “Tim Berners-Lee” retornou 2.068 correspondências, dentre as quais constam artigos de honrarias em seu nome, verbetes sobre seus pais, irmãos, esposas e filhos, entre outras informações a ele relacionadas, presentes em diferentes artigos da Wikipédia.

¹⁷ São sugestões ou previsões de pesquisa baseadas nas atividades dos usuários (SULLIVAN, 2011).

Ao criar ligações entre termos e conteúdos presentes na Wikipédia, a opção “*containing...*”, ainda que restrita apenas ao ambiente da comunidade, consegue estabelecer conexões semânticas eficazes com base no uso e relacionamentos de conteúdo criados pelos próprios usuários, conforme indicado na Figura 6.

Figura 6 – Conexões semânticas a partir do termo “Tim Berners-Lee” na Wikipédia



Fonte: adaptado de Wikipédia (2017b).

Em retomada ao procedimento de monitoramento dos verbetes, o artigo sobre “Tim Berners-Lee”¹⁸ foi acessado na íntegra. Ao visualizar a página, o usuário se depara com seis novas guias que compõem o artigo em exibição; são elas: “*Article*”, “*Talk*”, “*Read*”, “*Edit source*”, “*New section*” e “*View history*”. As guias “*Article*” e “*Read*” ficam simultaneamente habilitadas, pois as duas são responsáveis pela exibição do conteúdo do verbete. A guia “*Talk*” redireciona para uma página de discussão pertencente ao artigo, destinada apenas para melhorias específicas do verbete atual, não entrando no mérito de discussões gerais sobre o conteúdo. Além disso, também apresenta *hyperlinks* sobre boas práticas e políticas para o envio e recepção de conteúdos de outros membros, dispostos como sugestões para o bom convívio, conforme indicado na Figura 7.

¹⁸ https://en.wikipedia.org/wiki/Tim_Berners-Lee

Mais abaixo, a guia “*Talk*” também apresenta informações sobre a língua utilizada para a construção do artigo que, neste caso está em inglês britânico e deve seguir a política mantida no “*Manual of style*” que prima por manter um padrão linguístico consensual entre os membros da comunidade.

Figura 7 – Guia “*Talk*” do verbete de “Tim Berners-Lee” na wikipedia

The image shows a screenshot of the Wikipedia talk page for the article "Tim Berners-Lee". The page is titled "Talk:Tim Berners-Lee" and includes a sidebar with navigation links, a main content area with a table of contents, and a section for "WikiProjects".

Annotations on the page include:

- Hyperlinks de políticas e boas práticas para envio e recebimento de conteúdo.** (Blue arrow pointing to the top navigation bar)
- Informa que o artigo está escrito em inglês britânico e que não deve ser alterado para outra variação do inglês sem o consentimento da comunidade.** (Yellow arrow pointing to the "This article is written in British English" notice)
- Informa uma importante classificação alcançada pelo artigo na comunidade.** (Yellow arrow pointing to the "Vital article" notice)
- Políticas de edição para biografia de pessoas vivas e medidas para identificação e denúncia de conteúdo difamatório.** (Yellow arrow pointing to the "Biographies of living persons" notice)
- Wikiprojects aos quais esse artigo pertence.** (Red arrow pointing to the "WikiProjects" section)

Fonte: adaptado de Wikipédia (2017b).

Um verbete também pode alcançar um determinado *status* a partir de classificações realizadas pela própria comunidade. O artigo de “Tim Berners-Lee” consta na lista dos “*Vital articles/Expanded/People*¹⁹” que elenca os artigos mais importantes da Wikipédia. O verbete de “Tim Berners-Lee” está em meio a 1.999

¹⁹ https://en.wikipedia.org/wiki/Wikipedia:Vital_articles/Expanded/People

artigos importantes sobre pessoas, e sua posição do geral para o específico é apresentada no Quadro 6.

Quadro 6 – Classificação do verbete “Tim Berners-Lee” na lista “*Vital articles*”

1 People (1.999 articles)
1.14 Scientists, inventors and mathematicians (235 articles)
1.14.7 Computer scientists (11 articles)
1.14.7.2 Tim Berners-Lee (1 article)


Fonte: adaptado de Wikipédia (2017c).

O modo hierárquico como os artigos estão classificados no “*Vital articles*” assemelha-se aos tradicionais sistemas internacionais de classificação de obras em bibliotecas. Entretanto, essa representação não visa classificá-los tematicamente como no caso das bibliotecas, mas sim criar um *ranking* dos artigos mais importantes para a comunidade Wikipédia, com enfoque na qualidade e confiabilidade de seus conteúdos. Assim, conforme apresentado na Tabela 1, o “*Vital articles*” possui 9.887 artigos classificados em seu inventário e busca atingir o número de 10.000 artigos somando todas suas onze seções, sendo que, para algumas delas a meta já foi até ultrapassada.

Tabela 1 – Lista de classificação do *Vital Articles*

Número	Seção	Contagem de dezembro	Meta
1	<i>People</i>	1.999	2.000
2	<i>History</i>	680	675
3	<i>Geography</i>	1.161	1.200
4	<i>Arts</i>	671	670
5	<i>Philosophy and religion</i>	414	425
6	<i>Everyday life</i>	480	500
7	<i>Society and social sciences</i>	907	900
8	<i>Biology and health sciences</i>	1.453	1.480
9	<i>Physical sciences</i>	1.075	250
10	<i>Technology</i>	751	750
11	<i>Mathematics</i>	296	300
TOTAL		9.887	10.000






Fonte: o próprio autor com base em Wikipédia (2017c).

Dessa maneira, as seções principais também se subdividem em outras classes mais específicas, como no caso de “Tim Berners-Lee” indicado no Quadro 6. Além disso, o artigo de “Tim Berners-Lee” possui um selo “ C-Class” que, se refere a um indicativo de qualidade do conteúdo apresentado no verbete.

Essa classificação é estendida para todos os artigos da Wikipédia e não somente para aqueles classificados no “*Vital articles*”. Há nove classes oficiais, com diferentes critérios, perfis de leitores e sugestões de edição, os quais delimitam e indicam a qual classe cada um dos artigos deve ser indexado, conforme apresentado no Quadro 7.




Quadro 7 – Esquema de avaliação da qualidade de um artigo Wikipédia

(continua)

Selo/ícone	Classe	Critérios
	FA (<i>Featured Articles</i>)	Artigos “FA” possuem: - Padrões profissionais de escrita, apresentação de conteúdo e confiabilidade de fontes; - Boa escrita; - Abrangente; - Bom aprofundamento de pesquisa; - Neutralidade; - Estabilidade de edição; - Conteúdo multimídia; - Tamanho de texto adequado; - Citações consistentes.
	A-Class articles	Artigos “A-Class” possuem: - Boa organização e conteúdo completo; - Revisão de revisores imparciais; - Estrutura adequada; - Tamanho de texto adequado; - Fontes amplamente confiáveis; - Ilustrado; Cumprimento à direitos autorais; - Pendências de estilo que não o deixam ser o FA.
	GA (<i>Good Articles</i>)	Artigos “GA” possuem: - Avaliação oficial de editores; - Boa escrita; - Referências verificáveis, sem plágio; - Neutralidade; - Estabilidade de edição; - Ilustrado.
	B-Class articles	Artigos “B-Class” possuem: - Conteúdo completo, mas necessitam de padrões; - Fontes confiáveis e devidamente referenciadas; - Não contém omissões óbvias ou imprecisões no conteúdo; - Estrutura bem definida (seções, tópicos, tabelas, etc); - Poucos erros gramaticais e fluência textual; - Materiais de apoio; - Conteúdo exposto de forma compreensível.
	C-Class articles	Artigos “C-Class” possuem: - Conteúdo substancial, mas necessita de mais fontes de informação; - Estilo enciclopédico razoável; - Referências e fontes confiáveis; - Lacunas e elementos em falta; - Edições são realizadas no intuito de deixá-lo mais claro, equilibrado e fluído.

Quadro 7 – Esquema de avaliação da qualidade de um artigo Wikipédia

(conclusão)

Selo/ícone	Classe	Critérios
	Start-class articles	Artigos “Start” possuem: - Bom conteúdo, mas são fracos em desmembramentos do assunto; - Referências inadequadas; - Pode possuir conteúdo não enciclopédico.
	Stub-class articles	Artigos “Stub” possuem: - Descrições básicas; - Texto curto; - Conteúdo incipiente.
	FL (<i>Featured List</i>)	Artigos “FL” possuem: - Padrões profissionais de escrita; - Condução de texto (início, meio e fim); - Abrangente (fornece um panorama sobre o assunto abordado); - Estruturado (seções, tópicos, tabelas, quadros, etc.); - Estilo (atende aos padrões de cores de conteúdo); - Estabilidade (o assunto não gera conflitos de edição entre ideias divergente).
Sem ícone	List	Artigos “List” possuem: - Conteúdo em forma de lista com <i>hyperlinks</i> para outros artigos.

Fonte: o próprio autor com base em Wikipédia (2017c).

Nesse sentido, apesar do artigo sobre “Tim Berners-Lee” constar entre os mais importantes da Wikipédia, ainda é necessário atingir uma série de requisitos para que se torne um artigo exemplar para a comunidade e alcance o *status* “*Featured Articles*”. O restante das informações presentes na página “*Talk*”, indicadas na Figura 6, é constituído principalmente por notas e avisos sobre atividades que foram e que ainda devem ser cumpridas pela comunidade para a manutenção do artigo de “Tim Berners-Lee”.

Ao acessar a página da guia “*View source*”, um aviso incomum na maioria das outras páginas de artigos da Wikipédia alerta que o artigo de “Tim Berners-Lee” é semi-protegido e tem edição restrita aos membros da comunidade. Porém, nem todos os usuários autenticados podem editar o conteúdo semi-protegido, pois, para editá-lo, alguns pré-requisitos são necessários. São exigências como a de possuir uma conta há mais de quatro dias e de ter realizado mais de dez edições de conteúdo, entre outras peculiaridades, que fazem parte do pacote de requisitos para poder editar artigos semi-protegidos.

Figura 7 – Página “View source” do artigo de “Tim Berners-Lee” semi-protegida

The screenshot shows the 'View source' page for the article 'Tim Berners-Lee' on Wikipedia. The page is semi-protected, as indicated by a lock icon and the text: 'This page is currently semi-protected so that only established registered users can edit it.' Several callout boxes highlight key information:

- O cadeado indica que se trata de uma página com edição semi-protegida.** (The lock icon indicates that this is a page with semi-protected editing.)
- Motivos que levaram a aumentar a segurança de edição.** (Reasons that led to increasing the security of editing.)
- What can I do?** (A list of actions a user can take, such as logging in, creating an account, or submitting an edit request.)
- Requisitos necessários para editá-la.** (Requirements necessary to edit it.)

The main content area shows the source code for the article, including a template for a scientist and a list of work places.

Fonte: adaptado de Wikipédia (2017b).

Apesar de ir à direção contrária do que versa o próprio *slogan* da Wikipédia, de ser “a enciclopédia livre que todos podem editar”, esse adendo se fez necessário em virtude de diversos ataques de vandalismo realizados ao conteúdo do artigo de “Tim Berners-Lee”, fato que corrobora a visão de Anderson (2011) sobre o controle de versionamento disponibilizado pelo MediaWiki par este fim. A página especial “*Protection log*” vinculada ao verbete de “Tim Berners-Lee” registrou sete atividades de vandalismo atreladas ao conteúdo do verbete.

Nesse aspecto, a adoção de políticas de restrição de edição a determinados grupos de usuários mantém vivo o propósito de construção colaborativa do conhecimento da Wikipédia. Ainda assim, é possível ter acesso e copiar por inteiro o código/fonte utilizado na construção do artigo, conforme indicado na Figura 6, mas com a ressalva de que, para alterar o conteúdo original, é necessário cumprir todos os requisitos anteriores.

A guia “*View history*” é responsável por exibir todo o histórico de revisões realizadas ao longo dos anos no artigo. Dessa maneira, importantes informações sobre a construção do verbete constam nessa guia, como, por exemplo, o fato de o artigo de “Tim Berners-Lee” ter sido criado no ano de 2001 pelo usuário “Claudine”. Todas as alterações ficam devidamente registradas por ordem

cronológica, desde a inserção de longos conteúdos até a simples contribuição de um usuário em deixar a sigla “Tim Berners-Lee” em negrito.

No histórico de edições há informações de data, horário e usuário responsável pelas modificações, se foi contribuição textual ou abertura de tópico de discussão, além da quantidade em “bytes” de dados inseridos. Ferramentas externas também podem auxiliar no processo de compreensão dessas revisões, como a “*Revision history statistics*²⁰”, que indicou aproximadamente 400 revisões, realizadas por cerca de 1.900 usuários desde sua criação, no ano de 2001. As informações contidas no histórico de revisões tornam a Wikipédia um modelo efetivo de enciclopédia que preza pela colaboração entre inteligências mediante uma cooperação organizada (POMBO, 2006).

O conhecimento enciclopédico presente na Wikipédia é constituído essencialmente pela colaboração exercida entre seus membros, seja para a escrita de um conteúdo, formatação de seções, criação de *bots* (robôs) de monitoramento dos textos e até a criação de grupos com a indicação de responsáveis para o controle e validação das informações veiculadas nos artigos.


A realização desses processos de produção, organização, controle e disseminação de conhecimentos na Wikipédia envolve pessoas dispostas a colaborar, sem a necessidade de uma remuneração, diferente do que ocorria nas enciclopédias tradicionais que contavam com uma equipe de profissionais altamente qualificados para seu desenvolvimento. Dessa forma, títulos acadêmicos e patentes profissionais se tornam irrelevantes e são apenas adicionais para aqueles que buscam contribuir na Wikipédia, pois o substancial é cooperar.

De volta à guia “*Article*”, algumas considerações a respeito da organização e apresentação do verbete de “Tim Berners-Lee” puderam ser tecidas. Seu conteúdo está disposto em nove seções, denominadas de: “1 *Early life and education*”, “2 *Carrer*”, “3 *Current work*”, “4 *Awards and honours*”, “5 *Personal life*”, “6 *See also*”, “7 *References*”, “8 *Further reading*” e “9 *External links*”, apoiadas em 68 referências. Além disso, imagens sobre “Tim Berners-Lee” também compõem o conteúdo. Outro importante recurso presente entre as seções é “*Infobox*²¹”, indicado na Figura 8.

²⁰ https://tools.wmflabs.org/sigma/articleinfo.py?page=Talk:Tim_Berners-Lee&server=enwiki

²¹ Infobox é um modelo usado para coletar e apresentar um subconjunto de informações sobre o seu assunto, como um documento. (INFOBOX, 2017).

Figura 8 – Infobox do verbete “Tim Berners-Lee” na Wikipédia

Sir Tim Berners-Lee	
	
Berners-Lee in 2008	
Born	Timothy John Berners-Lee 8 June 1955 (age 61) ^[1] London, England, UK
Institutions	World Wide Web Consortium University of Southampton Plessey MIT
Alma mater	The Queen's College, Oxford (BA)
Notable awards	OM (2007) KBE (2004) FRS (2001) ^[2] FREng (2001) FRSA (2001) DFBCS (1995) See full list of honours
Spouse	Nancy Carlson (m. 1990; div. 2011) Rosemary Leith (m. 2014)
Children	2

Born: exibe nome completo, data e local de nascimento.

Institutions: exibe a lista das principais instituições que TBL possui vínculo.

Alma mater: exibe o local de formação do sujeito do verbete.

Notable awards: exibe uma relação de prêmios notórios relacionados ao indivíduo.

Spouse: exibe nome e duração de seus relacionamentos amoroso.

Children: exibe a quantidade de filhos.

Fonte: adaptado de Wikipédia (2017b).

Dessa maneira, o *infobox* vinculado à página de “Tim Berners-Lee” apresenta sucintas informações fatuais a respeito de acontecimentos ocorridos ao longo da vida de “Tim Berners-Lee”. Ao todo, o *infobox* do verbete de “Tim Berners-Lee” é composto por seis campos, identificados como: “*Born*”, “*Institutions*”, “*Alma mater*”, “*Notable awards*”, “*Spouse*” e “*Children*”.

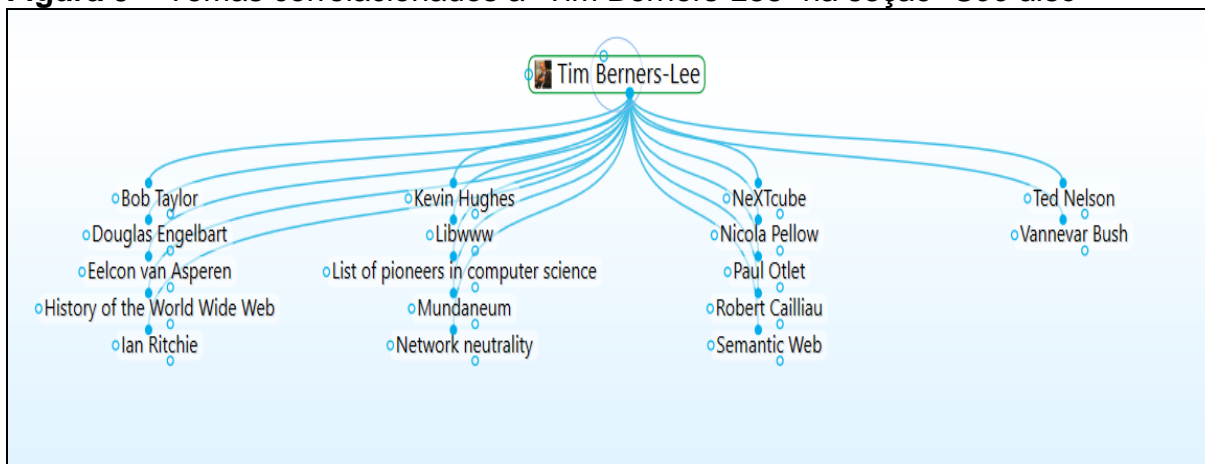
A maioria dos campos que compõem o *infobox* de “Tim Berners-Lee” possui *hyperlinks* que redirecionam o usuário para acessar o conteúdo de um verbete na íntegra. Apenas os campos “*Spouse*” e “*Children*” não possuem verbetes criados na Wikipédia que permitam um redirecionamento interno. Após verificar a guia “*Talk*”, constatou-se que há uma discussão em curso sobre a possibilidade de inserção de dados mais concisos, mais precisamente com a criação de verbetes para suas antiga e atual companheiras.

Com o auxílio do buscador auxiliar, originado da combinação das teclas “Ctrl + F” no navegador, tornou-se possível identificar no texto de “Tim Berners-Lee” a presença do termo “*England*” que será o próximo a ser monitorado.

Porém, curiosamente não há *hyperlink* para redirecionar o usuário até a página do verbete “*England*” da Wikipédia.

Mais abaixo, na seção “*See also*” do mesmo verbete, existe uma interessante lista de assuntos correlacionados a “Tim Berners-Lee”, criada pelos próprios usuários da Wikipédia, conforme exposto no Figura 9.

Figura 9 – Temas correlacionados a “Tim Berners-Lee” na seção “*See also*”



Fonte: adaptado do *software* “The Brain²²” com base em Wikipédia (2017b).

De maneira direta e/ou indireta pessoas, projetos, conceitos, entre outros assuntos abordados na Figura 9 estão conectados a vida profissional de “Tim Berners-Lee”. Entre eles, destaca-se para esta pesquisa a presença do termo “*Semantic Web*”, terceiro elemento a ser analisado nesta seção. Assim, esse contexto versa sobre a perspectiva de Pombo (2006) em relação a uma enciclopédia eletrônica, de que a leitura foi substituída pela navegação, a manipulação de volumes por “cliques” e a linearidade pela possibilidade de percorrer diversos caminhos.

Em continuidade ao procedimento de análise realizado no âmbito da Wikipédia, o termo referente a lugar “*England*” também foi monitorado. Dessa maneira, a procura pelo artigo sobre “*England*²³” se deu pela inserção exata da palavra no mecanismo de busca nativo. Porém, por se tratar de uma localidade e conter mais assuntos vinculados, o *auto-suggest* da caixa de pesquisa apresentou outras dez opções de pesquisas relacionadas ao termo “*England*”, além de exibir a opção “*Containing... England*”, conforme indicado na Figura 10.

²² <http://www.thebrain.com/>

²³ <https://en.wikipedia.org/wiki/England>

Figura 10 – *Auto-suggest* para o termo “*England*” na Wikipédia



Fonte: adaptado de Wikipédia (2017d).

As outras dez sugestões de pesquisa da caixa de busca incluem: “*England*”, “*England runestones*”, “*England national football team*”, “*England cricket team*”, “*England in the Late Middle Ages*”, “*England national rugby union team*”, “*England in the Middle Ages*”, “*England national under-21 football team results*”, “*England national football team manager*” e “*England women's national football team*”. O elevado número de sugestões apresentadas pelo mecanismo de busca da Wikipédia se deve ao fato de que o termo “*England*” consta como parte do verbete de cada uma das opções apresentadas pelo *auto-suggest*.

Por outro lado, ao selecionar a opção “*Containing... England*” com a aplicação do filtro “*Everything*” para vislumbrar todas suas possíveis ligações na Wikipédia, o resultado obtido é de 1.710.665 itens que envolvem a palavra “*England*” na comunidade, conforme indicado na Figura 11. Esse fato mostra um número absurdo se comparado aos 2.068 ligados a “*Tim Berners-Lee*”, porém justificável, pois o termo “*England*” pode ser amplamente utilizado para a confecção de diferentes tipos de artigos como lugares, pessoas, grupos, instituições, períodos históricos, entre outras temáticas, algo que não ocorre com o verbete de “*Tim Berners-Lee*” por se tratar de uma pessoa.

Figura 11 – Resultados para “containing... England” na Wikipédia

The image shows a screenshot of a Wikipedia search results page for the term "England". The search bar at the top contains "England" and the search button is highlighted. The results list includes several entries, each with a title, a brief description, and a size/date stamp. Yellow arrows point from the search results to labels on the right side of the image:

- 215 KB (21,195 words) - 20:05, 3 February 2017**: Points to the main article "England".
- Artigo sobre a Inglaterra**: Points to the first result, "England (disambiguation)".
- Artigo sobre a Inglaterra e País de Gales**: Points to the second result, "England and Wales".
- Artigo sobre a Nova Inglaterra**: Points to the third result, "New England".
- Artigo sobre a Igreja Anglicana**: Points to the fourth result, "Church of England".
- Artigos sobre a Geografia da Inglaterra**: Points to the fifth result, "Geography of England".

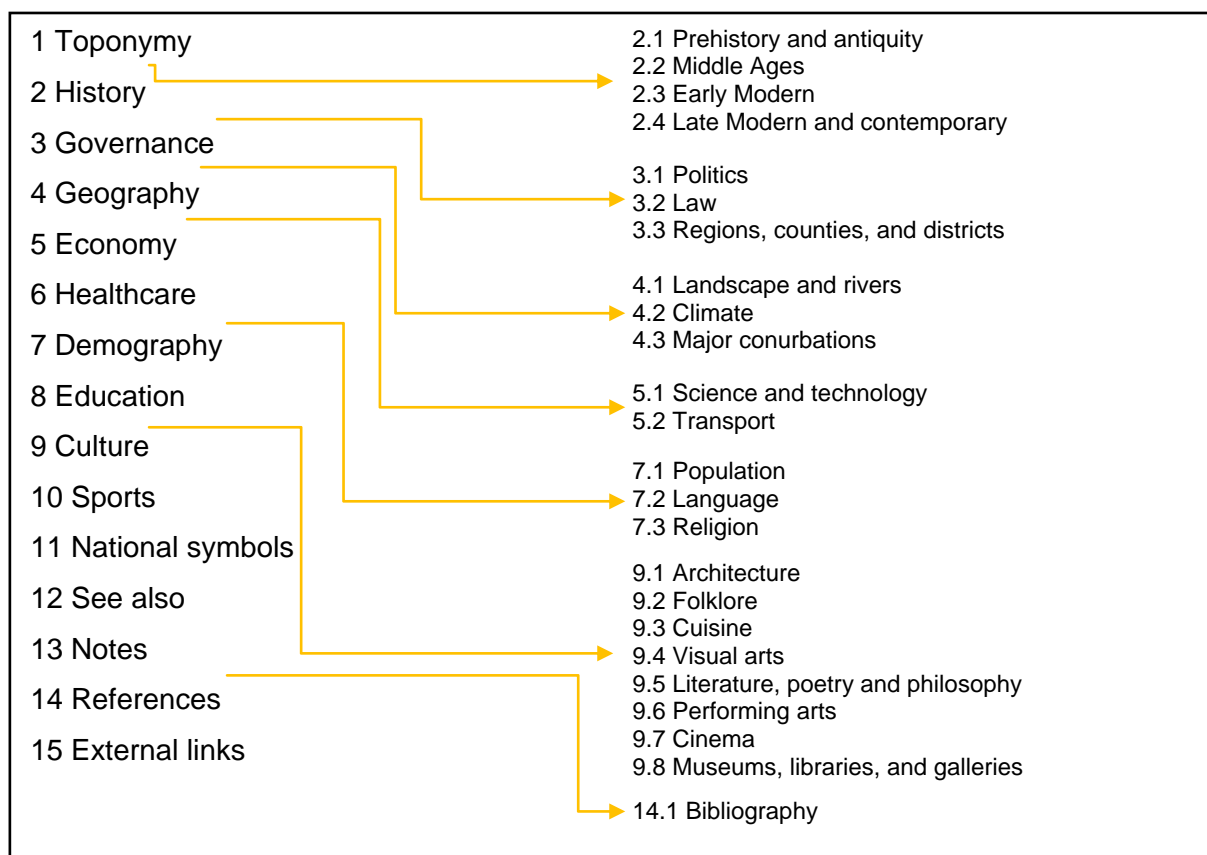
The left sidebar of the Wikipedia page is visible, showing navigation links like "Main page", "Contents", "Featured content", etc. The top right corner shows the user's name "G-DECARLI" and various utility links.

Fonte: adaptado de Wikipédia (2017d).

De estrutura semelhante ao artigo de “Tim Berners-Lee”, o conteúdo da página “Article/Read” do verbete “England” se mostrou volumoso perante todos os detalhes e acontecimentos que circundam sua história. A primeira frase do texto reflete a preocupação da comunidade em orientar seus leitores de que aquele conteúdo corresponde a um artigo sobre um país, mediante a expressão “*This article is about the country.*” Além disso, sugere o acesso a outra página denominada como “*England (disambiguation)*”, em que constam outros artigos que podem envolver o termo “England”. Este fato corresponde diretamente aos anseios de Leibniz de que uma enciclopédia deveria ser capaz de eliminar redundâncias e repetições para que possa registrar pensamentos e experiências alcançadas ao longo dos séculos (POMBO, 2002).

Devido à grande quantidade de informações existentes, seu conteúdo está seccionado em quinze subdivisões e algumas dessas ainda possuem outras subseções, como indicado no Quadro 8.

Quadro 8 – Seções de conteúdo do verbete “England” na Wikipédia



Fonte: adaptado de Wikipédia (2017d).

Essas seções são extremamente necessárias por terem a capacidade de indicar especificidades temáticas em meio a um panorama sobre a Inglaterra. Além disso, o conteúdo possui riqueza de imagens sobre diferentes assuntos como artes, religião, história, governos, música, monumentos, gastronomia, entre outros elementos nacionais. Diante de tantos temas que compreendem sua construção, o verbete possui exatas 351 referências, além de 83 bibliografias adicionais.

Na seção “5.1 *Science and technology*” do Quadro 8 se constatou uma singela conexão com o artigo que versa sobre “Tim Berners-Lee”. Essa ligação está disposta em formato de um *hyperlink* atrelado ao nome de “Tim Berners-Lee”, que é citado em meio a outras tantas figuras proeminentes do campo da Ciência e Matemática da Inglaterra. Apesar de ser apenas um *hyperlink* de redirecionamento para outra página da Wikipédia, pequenas ações como essa mostram a simplicidade com a qual se pode relacionar verbetes da Wikipédia, considerando o contexto.

Equivalente ao conteúdo da guia “*Talk*” de “Tim Berners-Lee”, o verbete “*England*” apresenta sugestões de boas práticas e políticas para envio e recebimento de edições do conteúdo. Outra similaridade com o verbete anterior é o fato de também se tratar de uma página “*Semi-protected*” devido ao alto índice de vandalismo textual. Além disso, o artigo “*England*” também pertence à lista “*Vital articles*”, exibida anteriormente na Tabela 1, alocado na classe “*3 Geography*”²⁴ em meio a outros 1.161 artigos sobre o tema, conforme apresentado no Quadro 9.

Quadro 9 – Classificação do verbete “*England*” na lista “*Vital articles*”

<p style="text-align: center;">1 Geography (1.161 articles)</p> <p style="text-align: center;">1.5 Regions and country subdivisions (103 articles)</p> <p style="text-align: center;">1.5.3 Europe and Russia (28 articles)</p> <p style="text-align: center;">1.5.3.9 UK (5 articles)</p> <p style="text-align: center;">1.5.3.9.1 England (1 article)</p>

Fonte: adaptado de Wikipédia (2017d).




A respeito da qualidade de seu conteúdo, o verbete “*England*” possui um *status* mais alto do que o de “Tim Berners-Lee”, sendo rotulado como “+ GA – *Good Articles*”, cujos critérios de avaliação podem ser observados no Quadro 7. Ainda sobre a página “*Talk*”, outro item chama que a atenção é a indicação do verbete como parte de um projeto denominado “*Portal: Contents/Outlines*”²⁵. Trata-se de um portal na Wikipédia destinado a reunir esboços/resumos sobre diferentes conhecimentos de mundo. Dessa maneira, o artigo sobre “*England*” compõe a categoria “*Geography and places*” ao lado de outros territórios globais.

Na página “*Article/Read*”, semelhante ao encontrado no texto do verbete de “Tim Berners-Lee”, o artigo sobre “*England*” também exibe em seu conteúdo o *infobox*. Trata-se de uma versão com mais informações, composta por uma sequência de 25 campos de descrição textuais e imagéticos, conforme indicado na Figura 12.

²⁴ https://en.wikipedia.org/wiki/Wikipedia:Vital_articles/Expanded/Geography

²⁵ <https://en.wikipedia.org/wiki/Portal:Contents/Outlines>

Figura 12 – Infobox do verbete “England” na Wikipédia

England	
	
Flag	Royal Banner
Anthem: Various Predominantly "God Save the Queen"	
	
Location of England (dark green) – in Europe (green & dark grey) – in the United Kingdom (green)	
Status	Country
Capital and largest city	London 51°30′N 0°7′W
National language	English
Regional languages	Cornish
Ethnic groups (2011)	85.4% White 7.8% Asian 3.5% Black 2.3% Mixed 0.4% Arab 0.6% Other ^[1]
Religion	Church of England
Demonym	English
Type	Constituent unit
Sovereign state	United Kingdom
Legal jurisdiction	England and Wales
Government	Part of a constitutional monarchy
• Monarch	Elizabeth II
Parliament of the United Kingdom	
• House of Commons	533 MPs (of 650)
Establishment	
• Anglo-Saxon settlement	5th–6th century
• Unification	10th century
• Union with Scotland	1 May 1707
Area	
• Land	130,279 km ² (50,301 sq mi) ^[2]
Population	
• 2015 estimate	54,786,300
• 2011 census	53,012,456
• Density	420.5/km ² (1,089.1/sq mi)
GDP (nominal)	2015 ^[3] estimate
• Total	\$2,340 billion
• Per capita	\$56,890
Currency	Pound sterling (GBP)
Time zone	GMT (UTC)
• Summer (DST)	BST (UTC+1)
Date format	dd/mm/yyyy (AD)
Drives on the	left
Calling code	+44
Patron saint	Saint George
ISO 3166 code	GB-ENG

Hino: existem vários, mas o que predomina é o "God Save the Queen"

Status: País

Capital: Londres

Língua nacional: Inglês

Línguas regionais: Cornish

Grupos étnicos (2011): 85,4% Brancos | 7,8% Asiáticos | 3,5% Preto | 2,3% Misto | 0,4% Árabe | 0,6% Outros.

Religião: Igreja Anglicana

Gentílico: Inglês

Tipo: Divisão administrativa

Estado Soberano: Reino Unido

Jurisdição: Inglaterra e Wales

Governo: parte de uma monarquia constitucional

Monarca: Elizabeth II

Parlamento do Reino Unido: House of Commons 533 deputados

Estabelecimento: Anglo-Saxônico – Sec. V – VI
Unificação – Sec. X
União com a Escócia - 1707

Área: terreno - 130.279 km² (50.301 milhas quadradas)

População: Estimativa para 2015 - 54.786.300
Censo de 2011 - 53.012.456
Densidade 420,5 / km² (1,089,1 / sq mi)

PIB: Nominal 2015 estimativa
Total: US \$ 2.340 trilhões
Per capita: \$ 56.890

Moeda: Libra esterlina (GBP)

Formato da data: dd/mm/aaaa

Fuso horário: GMT (UTC) - Verão (DST) - BST (UTC +1)

Tráfego: esquerda

Código de chamada: +44

Código ISO 3166: GB-ENG

Santo padroeiro: São Jorge

Fonte: adaptado de Wikipédia (2017d).

Os campos de descrição presentes no *infobox* do verbete “England” conectam-no por meio de *hyperlinks* a outros artigos da Wikipédia que versam sobre sua bandeira e brasão oficial, santo padroeiro, etnia e quantidade da população,

localização geográfica, religião principal, entre outros dados elucidativos de sua cultura.

O último verbete a ser analisado na Wikipédia correspondeu ao termo “*Semantic Web*”. Dessa maneira, como aconteceu com os termos anteriores, ao inseri-lo no mecanismo de busca nativo, a opção *auto-suggest* indicou nove sugestões de pesquisa, incluindo a opção “*Containing... Semantic Web*”, conforme indicado na Figura 13. Assim, as sugestões de pesquisa foram: “*Semantic Web*”, “*Semantic Web service*”, “*Semantic Web Rule Language*”, “*Semantic Web Stack*”, “*Semantic Web Stack*”, “*Semantic Web Advanced Development*”, “*Semantic web data space*”, “*Semantic web publishing*”, “*Semantic webbing*” e “*Containing... Semantic Web*”. Ao acessar essa última opção e aplicar o filtro “*Everything*”, a recuperação indicou 14.320 itens correspondentes ao termo “*Semantic Web*”.

Figura 13 – Resultados para “*containing... Semantic Web*” na Wikipédia

The screenshot shows the Wikipedia search interface. The search bar contains the text "Semantic Web" and a blue "Search" button. Below the search bar, there are tabs for "Content pages", "Multimedia", "Everything", and "Advanced". The search results section displays a list of articles. The first article, "Semantic Web", is highlighted with a yellow box. It shows a size of "38 KB (4,821 words) - 19:47, 6 February 2017". Below this, there are several other articles listed, including "Semantic Web Stack", "Semantic Web service", and "Extended Semantic Web Conference". On the right side of the page, a dropdown menu shows search suggestions, with "containing... Semantic Web" selected and highlighted in yellow. The Wikipedia logo and navigation menu are visible on the left side of the page.

Fonte: adaptado de Wikipédia (2017e).

Entre o conteúdo listado pelo “*Containing... Semantic Web*”, o primeiro resultado, composto por 4.821 palavras, se refere ao verbete “*Semantic Web*”. Ao acessá-lo, nota-se a existência de um *hyperlink* para a o verbete de “Tim Berners-Lee” logo em seus primeiros parágrafos, conexão feita justamente por “Tim Berners-Lee” ser reconhecido como o responsável por cunhar o termo “*Semantic Web*”.

Na guia “*Article/Read*”, a estrutura textual apresentada é semelhante à dos verbetes anteriores. Assim, seu conteúdo está disposto em onze grandes classes de assuntos, conforme indicado no Quadro 10.

Quadro 10 – Seções de conteúdo do verbete “*Semantic Web*” na Wikipédia

1	Example
2	Background
	2.1 Limitations of HTML
	2.2 Semantic Web solutions
	2.3 Web 3.0
3	Challenges
4	Standards
	4.1 Components
	4.2 Current state of standardization
5	Applications
6	Skeptical reactions
	6.1 Practical feasibility
	6.2 Censorship and privacy
	6.3 Doubling output formats
7	Research activities on corporate applications
8	See also
9	References
10	Further reading
11	External links

Fonte: adaptado de Wikipédia (2017e).

A estrutura para disposição dos assuntos do verbete é similar as dos termos anteriores, porém não há um elemento substancial para a identificação de dados objetivos, o *infobox*. Provavelmente, a ausência desse componente seja em decorrência de ser um conceito, ao qual não é possível relacionar dados fatuais, como no caso de *infoboxes* para pessoas e lugares.

Na guia “*Talk*” há uma caixa textual no topo da página, informando alguns reconhecimentos obtidos no desenvolvimento do artigo. Nela, consta que o artigo está vinculado ao “WikiProject *Internet*”²⁶, responsável por fazer cumprir as várias regras de qualidade dos verbetes. Outra informação se refere à classificação do artigo em relação à escala de qualidade apresentada no Quadro 7, categorizado como “🟡 *C-Class*”. Além disso, está presente em outra lista, a “*Importance Scale*”²⁷, designado com “*High-importance*”, nessa escala a comunidade classifica os artigos de acordo com sua importância internacional dentro de seu escopo, tornando crucial o constante desenvolvimento de seu conteúdo.

²⁶ WikiProject Internet é um ambiente de colaboração destinado a grupos de editores que buscam a melhorar a cobertura da Wikipédia na Internet. (WIKIPROJECT INTERNET, 2017).

²⁷ https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Internet/Assessment#Importance_scale

Outro ponto que o diferencia dos verbetes anteriores está nas guias “*Edit source*” e “*New section*”, em que reside a liberdade de edição, pois no artigo sobre “*Semantic Web*” não há a restrição “*semi-protection*” existente nos artigos de “Tim Berners-Lee” e “*England*”. Assim, não é necessário ser um usuário autenticado para a modificação do conteúdo, cabe apenas editar o artigo sob o aceite de termos e condições, incluindo textos que não violem direitos autorais e que tenham fontes verificáveis.

Na página “*View history*”, como esperado, são apresentadas todas as revisões realizadas no conteúdo do verbete desde sua criação em fevereiro de 2002 por meio de um “*Conversion script*”, isto é, o conteúdo foi transferido/traduzido automaticamente de uma outra edição da Wikipédia em outro idioma por um “*bot*”²⁸. Assim, ao acessar a ferramenta “*Article revision statistics*”²⁹, foi possível verificar que o verbete passou por 217 revisões, realizadas por aproximadamente 140 usuários.

Nesta seção, buscou-se descrever o contexto que envolve a criação, edição e manutenção do conhecimento enciclopédico presente na Wikipédia em torno da análise de três termos distintos: “Tim Berners-Lee”, “*England*” e “*Semantic Web*”. Ao acessar as páginas dos respectivos verbetes, uma breve descrição do funcionamento, estrutura e políticas que regem a produção desses artigos da Wikipédia foi realizada como base para esta análise.

Nesse sentido, as guias “*View history*” listam todo o processo de criação e revisão de cada artigo, informando usuários participantes, suas contribuições, datas das alterações, quantidade de dados inseridos, entre outras informações importantes para fins de registro. Mostraram ser, sob a perspectiva desta investigação, um grande indicativo sobre o curso do conhecimento enciclopédico colaborativo na Wikipédia. Este contexto compactua com a visão de Leibniz apresentada por Pombo (2002), na qual busca a unicidade da ciência por intermédio de uma enciclopédia, regida pela colaboração existente em uma comunidade do conhecimento.

Desse modo, os diversos *hyperlinks*, dispostos entre o conteúdo de cada artigo, permitem a ampliação do entendimento do sujeito que os consulta sobre a temática que o circunda. Assim, as conexões geradas pelos *hyperlinks* remetem à

²⁸ “**Bot**” é um apelido comum para reconhecer robôs de *software* da Wikipédia, eles trabalham de forma semi ou totalmente automatizada modificando conteúdos de artigos (BOT, 2017).

²⁹ https://tools.wmflabs.org/sigma/articleinfo.py?page=Talk:Semantic_Web&server=enwiki

concepção de enciclopédia polidimensional de Leibniz que, segundo Eco (2014), corresponde às referências cruzadas que realizam ligações transversais para que possam ser interpretadas sob diferentes perspectivas.

As diretrizes, políticas, regras e o bom senso que rodeiam a produção e manutenção do conteúdo de verbetes na Wikipédia se aproximam dos preceitos que envolveram o desenvolvimento da “*Encyclopédie*” por D’Alembert e Diderot durante o período Iluminista. Porém, vale salientar que são contextos diferentes de produção coletiva do conhecimento. No Iluminismo os grandes pensadores envolvidos clamavam pela produção envolvendo a razão, sem a interferência de dogmas religiosos, além da proximidade geográfica dos membros. Já na Wikipédia há a liberdade na produção dos conteúdos e seus editores são pessoas comuns de diferentes regiões do mundo em prol de uma construção coletiva do conhecimento enciclopédico voluntariamente.

Os requisitos sugeridos por Pombo (2002) para o estabelecimento de um projeto enciclopedista são quase que totalmente contemplados pela Wikipédia, visto que ela mantém um papel prospectivo por estar em constante produção/revisão, além de ser um trabalho coletivo que exerce um forte papel cultural e educativo na exposição de todo o conhecimento possível e, essencialmente por sua condição ilimitada, propiciada pelo emprego de *hyperlinks* nos artigos.

Apesar da liberdade de edição, a própria comunidade Wikipédia busca estabelecer padrões de escrita, desde recomendações básicas como neutralidade e inserção de fontes verificáveis e confiáveis, até políticas concisas de qualidade e importância, por exemplo, “*Vital articles*” e “*Importance Scale*”, respectivamente. Assim, como na “*Encyclopédie*”, a Wikipédia apresenta uma proposta de desmistificação do senso comum, de busca pela totalidade do conhecimento, algo mais racional e humano, além de implementar uma perspectiva crítica do conhecimento (ECO, 2014; MORAES, 2013; SERRA, 1999).

Por fim, com origem na *Web* Sintática, a Wikipédia tornou-se o melhor exemplo prático da segunda camada desta dimensão, a *Web* Social. Em parte, por conseguir implementar, com o auxílio de tecnologias *wiki*, um sistema de colaboração *on-line* para a construção e disseminação do conhecimento enciclopédico colaborativo mundial na *Web*. Nesse contexto, a próxima seção aborda o prosseguimento deste pensamento cooperativo para além das extensões

da Wikipédia, irá vislumbrar os mesmos termos aqui analisados sob a perspectiva de funcionamento das *Knowledge Bases*: DBpedia e Wikidata.

5.2 DBpedia e Wikidata

Em continuidade ao procedimento de análise desta investigação, esta seção se destinou a analisar como as duas *Knowledge Bases* selecionadas - DBpedia³⁰ e Wikidata³¹ - estruturam o conhecimento enciclopédico colaborativo advindo da Wikipédia. Dessa forma, os termos “Tim Berners-Lee”, “*England*” e “*Semantic Web*” foram novamente submetidos a um processo individual de monitoramento simultâneo nas *KB* DBpedia e Wikidata. Assim, tornou-se possível analisar o processo de conversão de verbetes Wikipédia em entidades nas *Knowledge Bases*.

5.2.1 DBpedia

A análise foi iniciada mediante acesso ao *website* da DBpedia, que disponibiliza informações sobre os diversos procedimentos que envolvem a transformação de um verbete da Wikipédia em uma entidade de sua *KB*. A DBpedia é uma comunidade *on-line* destinada à extração de informações estruturadas da Wikipédia para torná-las disponíveis ao restante da *Web*, possibilitando a realização de questionamentos de pesquisa mais complexos a MBs (DBPEDIA, 2016). Ainda, segundo o próprio *website*, a DBpedia é uma interface orientada aos dados da Wikipédia e sua infraestrutura depende da constante colaboração de wikipedistas e desenvolvedores na integração de seu banco de dados.

Nem todas as informações de um artigo Wikipédia são extraídas pela DBpedia, apenas as consideradas relevantes, ou seja, aquelas contidas em “*Infoboxes*” e “*Tables*” dos verbetes. Entretanto, há divergências nesse processo, que carece de uma dedicação maior da comunidade para que haja o bom funcionamento desse sistema. Este problema consiste no modo como as diferentes edições da Wikipédia se referem à descrição de um mesmo objeto, por exemplo, o

³⁰ <http://wiki.dbpedia.org/>

³¹ https://www.wikidata.org/wiki/Wikidata:Main_Page

campo designado para local de nascimento na Wikipédia pode ser encontrado de duas maneiras: “*birthplace*” ou “*placeofbirth*” (DBPEDIA, 2016).

A comunidade identifica esse tipo de divergência como “atributos sinônimos”, e sua existência causa muitos conflitos em processos de extração automática de informações da Wikipédia. Assim, como não é possível regularizar o legado de verbetes da Wikipédia, a solução encontrada pela DBpedia foi de criar a “DBpedia Mapping Language (*DML*)³²”. Trata-se de uma linguagem própria de padronização multilíngue para a conversão dessas divergências em ontologias DBpedia (DBPEDIA, 2016).

Antes de iniciar o procedimento de extração de dados da Wikipédia é preciso verificar a existência e a configuração da “*Ontology Schema Definition*”, elemento responsável por definir as classes dos “*namespace*”³³ DBpedia e todas suas propriedades. Na sequência, após aplicar o *DML*, o “*Schema Parser*” irá analisar e validar o esquema ontológico e convertê-lo em *OWL* (JENTZSCH et al., 2010).

Depois de padronizados, os dados considerados relevantes de verbetes da Wikipédia são extraídos por intermédio da “DBpedia Extraction Framework”³⁴. Nessa plataforma, o conteúdo semiestruturado (*infoboxes*, *hyperlinks* e *categorias*) é extraído por agentes computacionais denominados como “*Extractors*”, conforme indicado no Quadro 11 (DBPEDIA, 2016).

³² https://github.com/dbpedia/extraction-framework/raw/master/core/doc/mapping_language/DBpedia_Mapping_Language.pdf

³³ <https://dbpedia.org/sparql?nsdecl>

³⁴ <http://wiki.dbpedia.org/documentation>

Quadro 11 – DBpedia Extractors

Extractor	Função	Idioma
LabelExtractor	Extrai rótulos para artigos com base em seu título	Todos
MappingExtractor	Extrai dados estruturados com base em mapeamentos manuais de usuários da Wikipédia	Todos
InfoboxExtractor	Extrai todas as propriedades dos <i>infoboxes</i>	Todos
WikiPageExtractor	Extrai <i>hyperlinks</i> para artigos correspondentes da Wikipédia	Todos
PageLinksExtractor	Extrai <i>hyperlinks</i> internos de paginação de artigos Wikipédia para convertê-los em instancias DBpedia;	Todos
GeoExtractor	Extrai coordenadas geográficas	Todos
ArticleCategoriesExtractor	Extrai <i>hyperlinks</i> de conceitos usando o vocabulário SKOS	Inglês
CategoryLabelExtractor	Extrai rótulos para categorias DBpedia	Inglês
ImageExtractor	Extrai a primeira imagem de um artigo Wikipédia	Inglês
ExternalLinksExtractor	Extrai <i>hyperlinks</i> para páginas externas do verbete	Todos
HomepageExtractor	Extrai <i>hyperlinks</i> para a página inicial da DBpedia	Alemão, Francês e Inglês
DisambiguationExtractor	Extrai <i>hyperlinks</i> de desambiguação;	Todos
PersondataExtractor	Extrai informações sobre pessoas (data e local de nascimento, etc.) da Wikipédia em Inglês e Alemão, representando-as com o vocabulário FOAF ³⁵ ;	Alemão e Inglês
PndExtractor	Extrai informações do “Personennamendatei” (PND), um arquivo publicado pela Biblioteca Nacional Alemã, que possui a relação de pessoas com nome, nascimento, ocupação etc.	Alemão e Inglês
SkosCategoriesExtractor	Extrai informações sobre qual conceito uma categoria pertence e como elas se relacionam;	Inglês
RedirectExtractor	Extrai <i>hyperlinks</i> de redirecionamento da Wikipédia.	Todos

Fonte: adaptado de DBpedia (2016).

A extração de dados da Wikipédia envolve uma série de procedimentos que devem ser realizados pela DBpedia. Dessa maneira, após aplicar linguagens, modelos e protocolos de padronização e transmissão de dados,

³⁵ FOAF (*Friend of a Friend*) – é um vocabulário ontológico legível por máquina para a descrição de pessoas (FOAF, 2017).

as informações presentes nos verbetes da Wikipédia são extraídas para composição de entidades e propriedades das ontologias da DBpedia.

Diante disso, o *website* da DBpedia foi novamente acessado para verificar a disponibilidade de alguma ferramenta de recuperação de informações de sua *KB* para que os termos elencados pudessem ser analisados. Entre outras formas, duas chamaram a atenção por seus modos de operação; a primeira, mais técnica, permite a elaboração de requisições mais complexas com a utilização da linguagem de consulta *SPARQL* na interface Virtuoso *SPARQL Query Editor*³⁶. A segunda, disponibilizada na interface OpenLink Virtuoso³⁷, conta com facilidades de pesquisa, visto que seu *layout* está apresentado de maneira equivalente ao de MBs, fato que permitiu inserção, recuperação e análise mais simples dos termos.

O primeiro termo submetido à análise foi o de “Tim Berners-Lee”, que recuperou como entidade principal a classe “*scientist*”. Na propriedade “*rdf:type*”, presente na mesma página³⁸, também há vinculação com as classes “*person*” e “*agent*”. Desse modo, a entidade de “Tim Berners-Lee” é composta por 41 propriedades/declarações relacionadas a 245 valores de descrição.

Em decorrência de um número elevado de propriedades e valores na composição de uma entidade, avaliá-las por completo tornaria o procedimento de análise inviável. Assim, para o caso de “Tim Berners-Lee”, foram consideradas apenas as informações presentes na *infobox* de seu verbete da Wikipédia, conforme exposto anteriormente na Figura 8. Além disso, na DBpedia, a propriedade “*dbo:abstract*” também foi considerada para todos os termos, pois corresponde ao texto inicial de artigos Wikipédia.

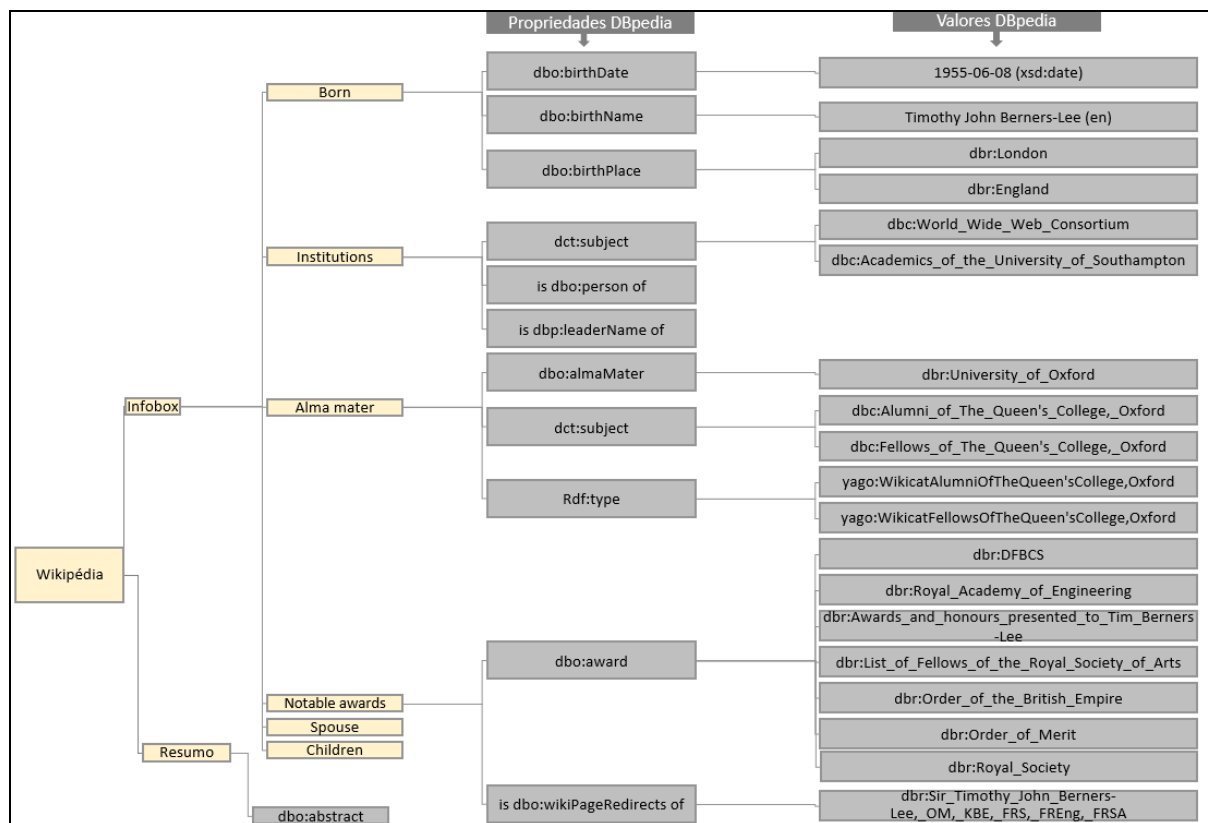
Dessa maneira, o Diagrama 1 apresenta a origem dos dados extraídos do verbete de “Tim Berners-Lee” na Wikipédia e sua conversão em propriedades e valores de uma entidade na DBpedia. A respeito dos dados do *infobox*, apenas os campos “Spouse” e “Children” não tiveram suas correspondências localizadas na composição da entidade de “Tim Berners-Lee” na DBpedia.

³⁶ <http://dbpedia.org/sparql>

³⁷ <http://dbpedia.org/fct/facet.vsp>

³⁸ http://dbpedia.org/page/Tim_Berners-Lee

Diagrama 1 – Trajetória dos dados de “Tim Berners-Lee” da Wikipédia para DBpedia



Fonte: o próprio autor com base em Wikipédia (2017b) e DBpedia (2017a).

As propriedades são categorias para a descrição de uma entidade. Dessa forma, propriedades como “`dbo:birthDate`”³⁹, “`dbo:birthName`” e “`dbo:birthPlace`” tiveram seus valores extraídos do campo “Born” advindo do *infobox* da Wikipédia. Nas duas primeiras propriedades, os dados foram convertidos em valores apenas de formato textual, sem conexão com outros elementos da DBpedia. Já no caso da propriedade “`dbo:birthPlace`”, os valores “`dbr:London`”⁴⁰ e “`dbr:England`” se tornaram novos recursos, isto é, além de descreverem a entidade de “Tim Berners-Lee”, elas também são entidades.

Além disso, foram identificadas propriedades em formato de declarações no Diagrama 1; são elas: “`is dbo:person of`”, “`is dbp:leaderName of`”⁴¹ e “`is dbo:wikiPageRedirects of`”, o que remete à aplicação das triplas RDF na *KB* da DBpedia. Segundo Berners-Lee, Hendler e Lassila (2001), o RDF usa o XML para

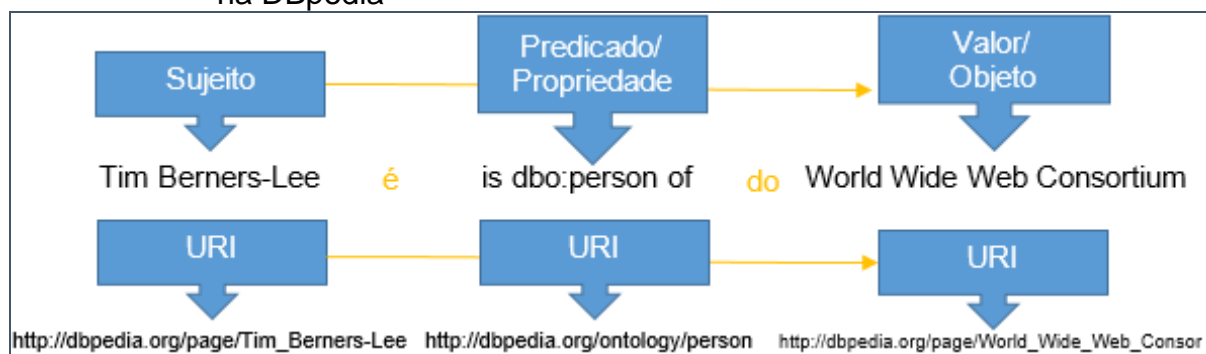
³⁹ DBpedia Ontology (dbo)

⁴⁰ DBpedia Resource (dbr)

⁴¹ DBpedia Property (dbp)

atribuir propriedades e valores a um sujeito mediante URI interligadas, conforme indicado na Figura 14.

Figura 14 – Exemplo de aplicação de triplas RDF para o termo “Tim Berners-Lee” na DBpedia



Fonte: o próprio autor com base em DBpedia (2017a).

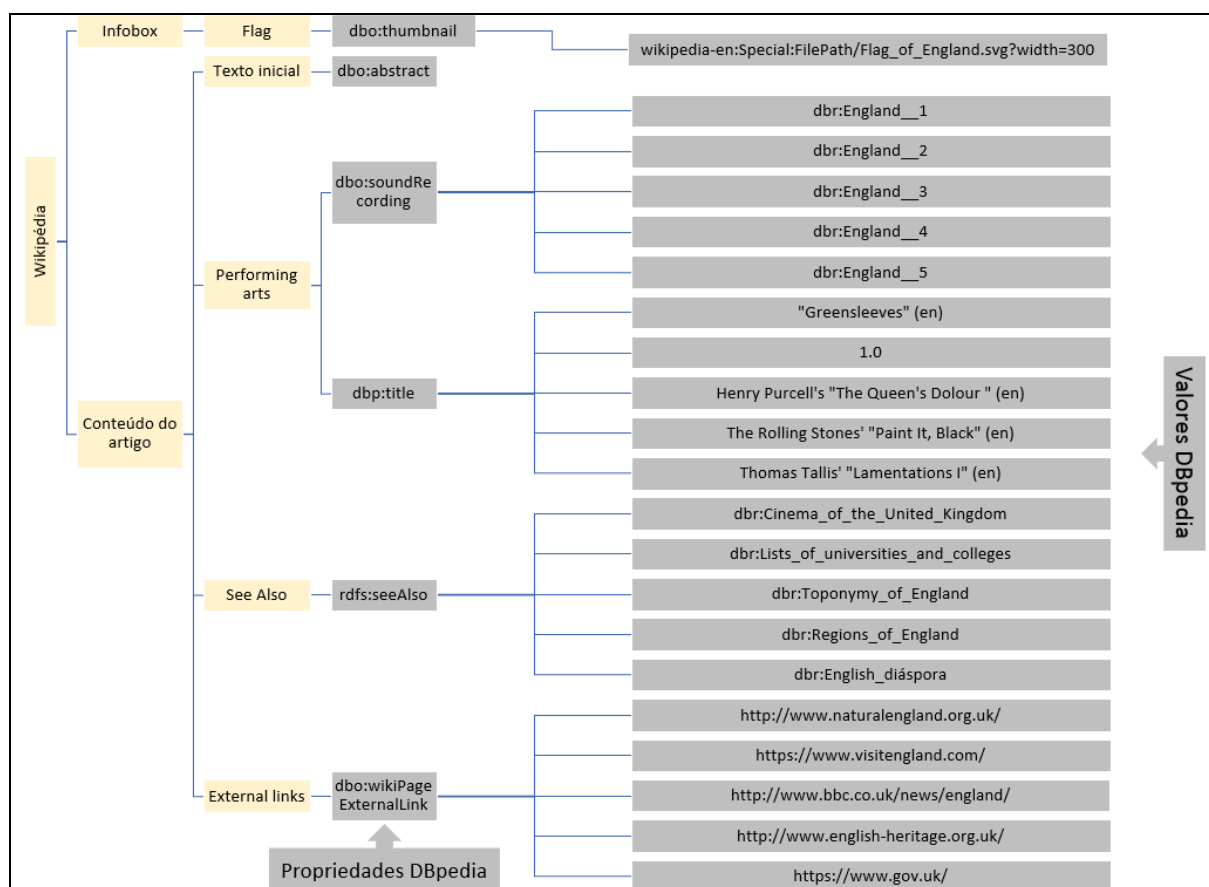
Nesse contexto, a interconexão dessas URIs representa a aplicação do conceito de *Linked Data* indicado por Kontokostas et al. (2012), no qual os dados são extraídos da Wikipédia, representados em triplas (sujeito, predicado e objeto) RDF e disseminados de maneira interligada na *Web*. Assim, as declarações promovem a construção do significado formal para agentes computacionais na *Web*, mediante os relacionamentos estabelecidos entre as entidades com o uso de URIs.

Na continuação desta análise, “England” foi o segundo termo investigado na *KB* da DBpedia. Em sua página⁴² DBpedia, verificou-se que o termo está classificado como uma entidade do tipo “Country108544813”, correspondente a classes ontológicas advindas do Yago⁴³, que são utilizadas pela DBpedia para informações de localidades relacionadas a países e distritos administrativos. Semelhante ao acontecido com o termo anterior, “England” apresentou uma volumosa quantidade de elementos na composição de sua entidade. Assim, diante das 185 propriedades/declarações e 9.786 valores, tornou-se necessário a seleção de algumas destas para ilustrar o Diagrama 2.

⁴² <http://dbpedia.org/page/England>

⁴³ <https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/>

Diagrama 2 – Trajetória dos dados de “England” da Wikipédia para DBpedia



Fonte: o próprio autor com base em Wikipédia (2017d) e DBpedia (2017b).

Ao analisar a entidade referente ao termo “England” na DBpedia, verificou-se que a maioria dos dados existentes no *infobox* apresentado na Figura 12 sobre o respectivo verbete na Wikipédia não foram extraídos e convertidos em propriedades e valores da KB. Em contrapartida, houve uma presença superior de informações advindas do corpo do artigo, por exemplo, “Performing arts”, “See also” e “External links”. Assim, grande parte desses dados está disposta no formato de *hyperlinks* e/ou em tabelas de assunto ao longo do texto.

Apesar de não constarem no Diagrama 2, também foram encontradas muitas propriedades no formato de declarações, por exemplo, a “`isPartOf of`” que relaciona “England” com o valor/entidade “`dbr:London`”. As ligações transversais realizadas pelos *hyperlinks* existentes em propriedades, valores e declarações de entidades DBpedia, estabelecem a configuração necessária para assimilação do sentido por máquinas. Dessa maneira, a Semântica Formal indicada por Oliveira (2006) e Bax (2012) é evidenciada, ao utilizar triplas RDF para a união de sentido e referência na *Web*.

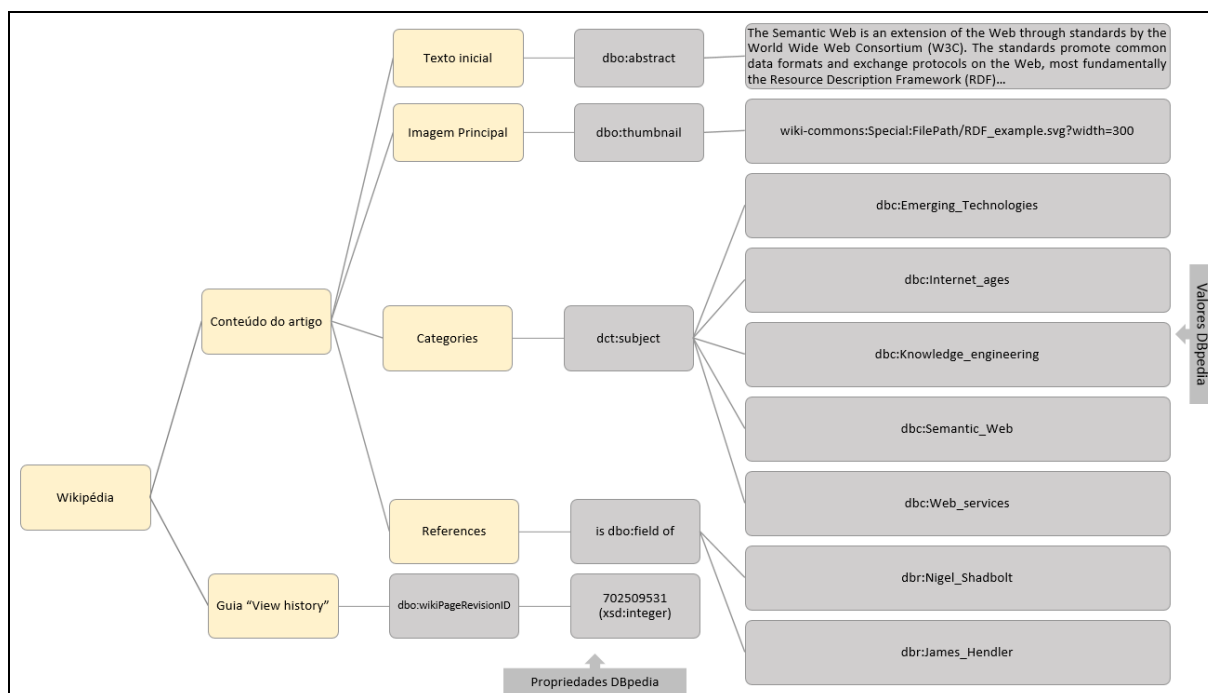
O termo “*Semantic Web*” foi a última entidade analisada na *KB* da DBpedia. Na página⁴⁶ recuperada sobre o respectivo termo, consta que se trata de uma entidade do tipo “Event100029378”, porém não há mais informações a respeito das regras que designam uma entidade a esse tipo de categorização. Em continuidade, a entidade “*Semantic Web*” tem em sua composição 29 propriedades/declarações com aproximadamente 170 valores de descrição.

Equivalente a maneira como os termos anteriores foram analisados, tornou-se necessário reduzir a quantidade de elementos que compõem a entidade para que fosse possível exemplificar seu funcionamento. Poucas propriedades mostraram possuir valores que pudessem ser diretamente relacionados ao conteúdo do verbete “*Semantic Web*” na Wikipédia. Diante disso, apenas as propriedades/declarações a seguir foram envolvidas na construção do Diagrama 3: “dbo:abstract”, “dbo:thumbnail”, “dbo:wikiPageRevisionID”, “dct:subject” e “is dbo:field of”.

A propriedade “dbo:abstract” mostrou-se presente em todas as entidades analisadas. Sua função é extrair o texto que comumente antecede a tabela “*Contents*” dos artigos das diferentes edições da Wikipédia. A “dbo:thumbnail” está designada a obter o *hyperlink* da primeira imagem que aparece no artigo, seja no *infobox* ou no próprio corpo do texto. Já a “dbo:wikiPageRevisionID” traz o código de revisão do verbete existente na guia “*View history*” para que seja possível saber de qual versão do artigo essas informações foram retiradas. Há também a “dct:subject” que extraiu os dados da última seção do artigo na Wikipédia e se refere a assuntos *hyperlinks* internos relacionados ao tema “*Semantic Web*”.

⁴⁶ http://dbpedia.org/page/Semantic_Web

Diagrama 3 – Trajetória dos dados de “*Semantic Web*” da Wikipédia para DBpedia



Fonte: o próprio autor com base em Wikipédia (2017e) e DBpedia (2017c).

A declaração “is dbo:field of” extrai algumas das informações contidas na classe “*References*” presente no conteúdo do artigo sobre “*Semantic Web*” na Wikipédia. Nessa categoria, são indexadas as referências utilizadas pelos wikipedistas para o desenvolvimento do verbete. Assim, valores como “James Hendler” e “Nigel Shadbolt” correspondem a autores que produzem estudos sobre a temática e foram empregados para a construção do verbete. Apesar do verbete “*Semantic Web*” não possuir um *infobox* para a extração de informações fatuais provavelmente por se tratar de um conceito, os dados obtidos são suficientes para relacioná-lo com outras entidades DBpedia na busca pela significação.

Por fim, empregou-se novamente a ferramenta RelFinder para vislumbrar os relacionamentos semânticos existentes entre os três termos investigados na DBpedia. As conexões obtidas como resultados a partir da consulta realizada na RelFinder são apresentadas na Figura 16. De maneira simples, muitas conexões entre os três termos foram evidenciadas; porém, devido a quantidade de itens recuperados, a nitidez da imagem foi prejudicada. Dessa forma, algumas marcações auxiliam no entendimento sobre os relacionamentos estabelecidos entre essas entidades.

“London”, que conseqüentemente se liga com a entidade “England”, utilizando a declaração “is dbo:isPartOf of”. O sentido inverso também as conecta; então, se o ponto de partida for de “England”, o mesmo caminho de significação poderá ser percorrido.

Outro ponto de conexão é realizado mutuamente entre “Tim Berners-Lee” e “Semantic Web” através da propriedade “dbo:wikiPageExternalLink”, que os une por intermédio do valor *http://computemagazine.com/man-who-invented-world-wide-web-gives-new-definition/*, ou seja, é a ponte entre o conceito e seu inventor. Entretanto, a conexão entre o conceito “Semantic Web” e a localidade “England” não se dá por intermédio de “Tim Berners-Lee”, mas por meio da declaração “is dbo:knowFor of”, na qual apresenta o autor “Nigel Shadbolt”, usado como referência no artigo da Wikipédia, como o valor de sua propriedade. Assim, o autor é a ligação entre “*Semantic Web*” e “England”, pois aborda o tema em seus estudos, e a propriedade “dbo:placeOfBirth” o conecta à “England” por ser seu local de nascimento.

Nesta seção foram abordados os três termos elencados para etapa de análise desta investigação no ambiente da DBpedia. Dessa maneira, apresentou-se o *modus operandi* dessa *KB* e algumas particularidades que envolvem o estabelecimento de cada um desses termos como entidades. A próxima seção irá apresentar o modo de funcionamento do Wikidata e, ao fim dela, será traçado um paralelo entre as duas *KBs*.

5.2.1 Wikidata

Diferente do modelo de *website* adotado pelo DBpedia, o Wikidata possui muitas semelhanças de interface e funcionamento com a Wikipédia, em virtude de compartilharem da tecnologia e ideologia “*wiki*” no desenvolvimento e manutenção dessas plataformas por sua mantenedora, a Wikimedia Foundation. Diante disso, acessar e compreender o modo de funcionamento do Wikidata mostrou-se uma tarefa mais simples por conta da familiarização com o padrão de interface. Até mesmo o usuário criado na etapa de análise da Wikipédia pode ser reaproveitado no Wikidata, pois o registro também é compartilhado entre as plataformas.

Existem apenas algumas mudanças de nomenclaturas em relação aos elementos que foram apresentados na DBpedia. Assim, enquanto uma “*Entity/Entidade*” poderia ser interpretada tanto como propriedade/declaração ou valor, no Wikidata uma “*Entity/Entidade*” pode ser compreendida como “*Item*” ou, dependendo do idioma, até como objeto, assunto ou elemento, que corresponde a um objeto, conceito ou evento do mundo real, seguido de propriedades/declarações e valores, como ocorria na DBpedia (WIKIDATA, 2017).

Além disso, os dados podem ser inseridos manualmente por seus editores ou automaticamente por “*bots*” programados para tal tarefa. Por meio de procedimentos semelhantes aos da DBpedia, o Wikidata extrai as informações da Wikipédia e também fornece dados para a complementação de seus verbetes, isto é, há uma relação inversa, pois ao mesmo tempo que extrai também fornece informações, conforme o interesse da comunidade (WIKIDATA, 2017).

Desse modo, o procedimento de análise se iniciou pela inserção do termo “Tim Berners-Lee” na caixa de pesquisa nativa do Wikidata. Com o modo de operação equivalente ao da Wikipédia, a caixa de busca indicou cinco pesquisas relacionadas por meio do *auto-suggest*, são elas: “Tim Berners-Lee (British computer scientist)”, “Tim Berners-Lee: A Magna Carta for the Web (TED2014)”, “Tim Berners-Lee: The next Web (TED2009)”, “Tim Berners-Lee: The year open data went worldwide (TED2010)” e “*containing...* Tim Berners-Lee”.

O primeiro resultado aponta para o “*Item*” sobre “Tim Berners-Lee”, os outros três seguintes também correspondem a “*Items*”, mas sobre palestras por ele ministradas à conferência *Technology, Entertainment and Design*⁴⁷ (TED), nos anos de 2009, 2010 e 2014. Para o último “*Item*” foram aplicados os mesmos parâmetros da análise realizada na Wikipédia, que recuperou 339 correspondências do termo na *KB* do Wikidata.

A página do “*Item*” também possui guias semelhantes às usadas na Wikipédia, sendo formada pela sequência: “*Item/Read*”, “*Discussion*”, “*View source*” e “*View history*”. Na primeira é possível visualizar e editar as declarações da entidade; em “*Discussion*” podem ser abertos tópicos sobre melhorias/correções de conteúdo; na terceira a fonte pode ser visualizada e editada; e, em “*View history*” é exibido o histórico de revisões e os usuários responsáveis, de maneira idêntica ao

⁴⁷ <https://www.ted.com/>

funcionamento da Wikipédia. Em “*View history*” consta que a entidade de “Tim Berners-Lee” foi criada no ano de 2012 por um usuário não autenticado, permanecendo registrado apenas o número de IP do computador em que a edição foi realizada. Desde então, foram realizadas 559 edições nessa entidade. Além disso, outra semelhança corresponde à proteção das páginas (*semi-protected*) para casos de vandalismo de conteúdo.

Diferente da página apresentada na DBpedia, a correspondente ao “*Item*” de “Tim Berners-Lee”⁴⁸ no Wikidata parece ter sua interface destinada de fato ao usuário final, com informações bem organizadas e claras. Os “*Items*” são identificados de duas formas: a primeira, chamada de “*Label*”, é destinada a visualização por humanos e remete ao próprio nome “Tim Berners-Lee”; a segunda é denominada de “*Unique identifier*” e corresponde a uma codificação do texto que identifica o “*Item*” para que seja facilmente processado por máquinas. No caso de “Tim Berners-Lee” é usado o “Q80”, conforme apresentado na Figura 17.

Figura 17 – Página do “*Item*” de “Tim Berners-Lee” no Wikidata

The image shows a screenshot of the Wikidata page for the entity 'Tim Berners-Lee' (Q80). The page is annotated with several grey boxes and yellow arrows pointing to specific UI elements:

- Label**: Points to the main title 'Tim Berners-Lee (Q80)'.
- Unique identifier**: Points to the identifier '(Q80)'.
- Aliases**: Points to the 'Also known as' section.
- Description**: Points to the main description 'British computer scientist'.
- Property**: Points to the 'Property' label in the left sidebar.
- Rank**: Points to the 'Rank' label in the left sidebar.
- Value**: Points to the 'value' column in the 'Statements' table.
- Qualifiers**: Points to the 'Qualifiers' label in the left sidebar.
- Opened references**: Points to the '+10 references' link in the 'Statements' table.
- Collapsed reference**: Points to the '+ 0 references' link in the 'Image' statement.

The 'Statements' table shows the following data:

Property	Value	Qualifiers
employer	CERN	
start time	June 1980	
end time	December 1980	
position held	independent contractor	

Fonte: adaptado de Wikidata (2017b).

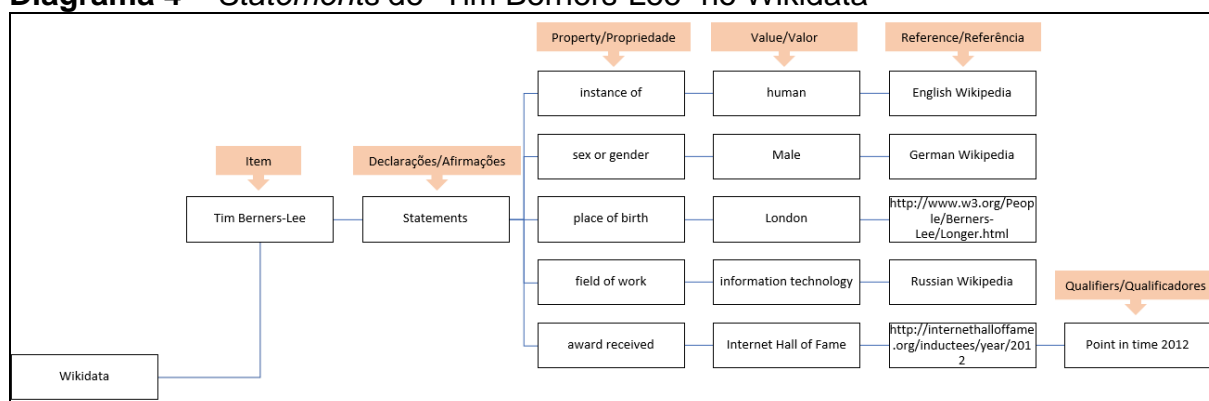
⁴⁸ <https://www.wikidata.org/wiki/Q80>

As informações existentes no “*Item*” de “Tim Berners-Lee” advêm de diferentes fontes, como *hyperlinks* para *websites* externos, além de Wikipédia, Wikisource, Wikibooks, Wikiquote, entre outros projetos da Wikimedia Foundation. A relação direta com a Wikipédia é evidente, porém não há como precisamente indicar seu fluxo de extração de dados da mesma forma que ocorreu com as entidades da DBpedia. Dessa maneira, a análise no Wikidata se atentou ao modo como as informações são inseridas, editadas e interligadas.

Todos os campos apresentados na Figura 17 são passíveis de edição e podem ser modificados por usuários autenticados ou não. A organização visual dos elementos que compõem o “*Item*” torna o processo de alimentação e edição rápida e simples, diferente da DBpedia, em que é necessário um conhecimento técnico, principalmente sobre a linguagem de consulta *SPARQL* para poder contribuir.

Dessa maneira, as informações apresentadas pelos campos “*Label*”, “*Unique identifier*”, “*Description*” (designando a breve descrição da entidade) e “*Aliases*” (correspondente a outras nomenclaturas que também podem identificar o “*Item*”) podem ser editadas através da tabela apresentada logo na sequência dos campos, que atualmente descreve a entidade de “Tim Berners-Lee” em aproximadamente 120 idiomas. Na segunda seção do “*Item*” há os agrupamentos de “*Statements*” ou declarações/afirmações, nos quais estão alocadas as “*Property/Propriedades*” e os “*Value/Valor*”, além de “*Qualifiers/Qualificadores*” para declarações adicionais sobre os próprios valores, conforme indicado no Diagrama 4.

Diagrama 4 – *Statements* de “Tim Berners-Lee” no Wikidata



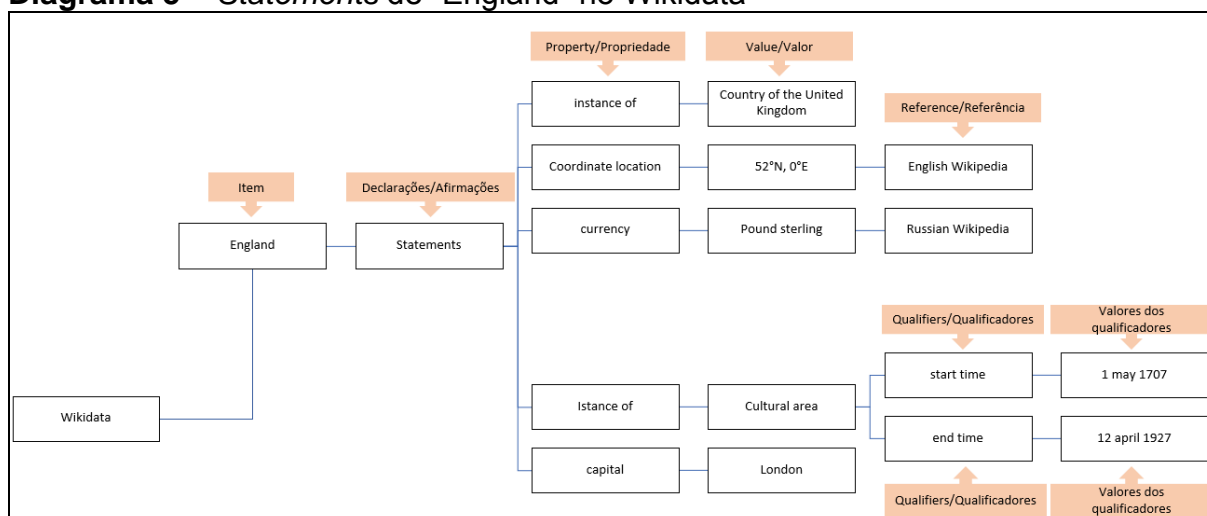
Fonte: o próprio autor com base em Wikidata (2017b).

No Diagrama 4, o campo “*Qualifiers*” informa o ano de obtenção da premiação “*Internet Hall of Fame*”, derivada da propriedade “*award received*”, correspondendo de fato a um dado adicional sobre o valor. Para as propriedades “*instance of*”, “*sex or gender*” e “*field of work*”, há a indicação de referência de que os dados foram extraídos respectivamente das versões inglesa, alemã e russa da Wikipédia, porém não há como especificar a localização de onde as informações foram extraídas no corpo do verbete. Além disso, nas propriedades “*place of birth*” e “*award received*” são indicados *hyperlinks* de *websites* externos como referência dos valores atribuídos.

A segunda entidade analisada correspondeu ao termo “England”. Após sua inserção na caixa de pesquisa do Wikidata, o *auto-suggest* foi ativado e indicou um número elevado de buscas complementares, incluindo também a opção “*containing... England*”, que apontou a existência de mais de 1M de ocorrências desta palavra na *KB* do Wikidata. Assim, a entidade “England” foi acessada mediante a seleção do primeiro resultado indicado pelo mecanismo de busca, que se referia ao emprego da palavra no sentido de “country in north-west Europe, part of the United Kingdom”.

Dessa maneira, ao acessar a guia “*View history*”, existente na página sobre “England”, verificou-se que sua criação foi no ano de 2012 pelo usuário “Reedy”, e seu conteúdo passou, desde então, por mais de 600 edições. De volta à página principal do “*Item*”, o termo “England” é descrito no campo “*Label*” e seu código “*Unique identifier*” é o “Q21”. Essas informações, juntamente com as dos campos “*Description*” e “*Aliases*”, estão descritas em aproximadamente 200 idiomas diferentes. Além disso, a entidade também é composta por 43 propriedades, 86 valores e alguns qualificadores. Assim, no Diagrama 5 são expostos alguns desses elementos para exemplificar a estrutura da entidade “England”.

Diagrama 5 – Statements de “England” no Wikidata



Fonte: o próprio autor com base em Wikidata (2017c).

Multivalorada, a propriedade “*instance of*” incorpora o valor “*Country of the United Kingdom*” e também o “*Cultural area*”; este segundo ainda se divide em dois qualificadores denominados de “*start time*” e “*end time*”, com os respectivos valores “1 may 1707” e “12 april 1927”. Nesses casos, a declaração da entidade pode ficar ainda maior caso também conste a referência utilizada, o que não ocorreu neste exemplo. As propriedades “*coordinate location*” e “*currency*” tiveram seus valores extraídos das versões em russo e em inglês da Wikipédia.

O último elemento analisado na *KB* do Wikidata foi o termo “*Semantic Web*”. Novamente, o mecanismo de busca nativo foi empregado para a recuperação das informações acerca dessa entidade. Como ocorreu no caso de “England”, muitas pesquisas complementares foram indicadas pelo *auto-suggest*, porém apenas a opção “*containing... Semantic Web*” foi acessada com o propósito de verificar a recorrência do termo entre os conteúdos do Wikidata. Assim, a recuperação acusou a existência de 850 correspondências em sua *KB*.

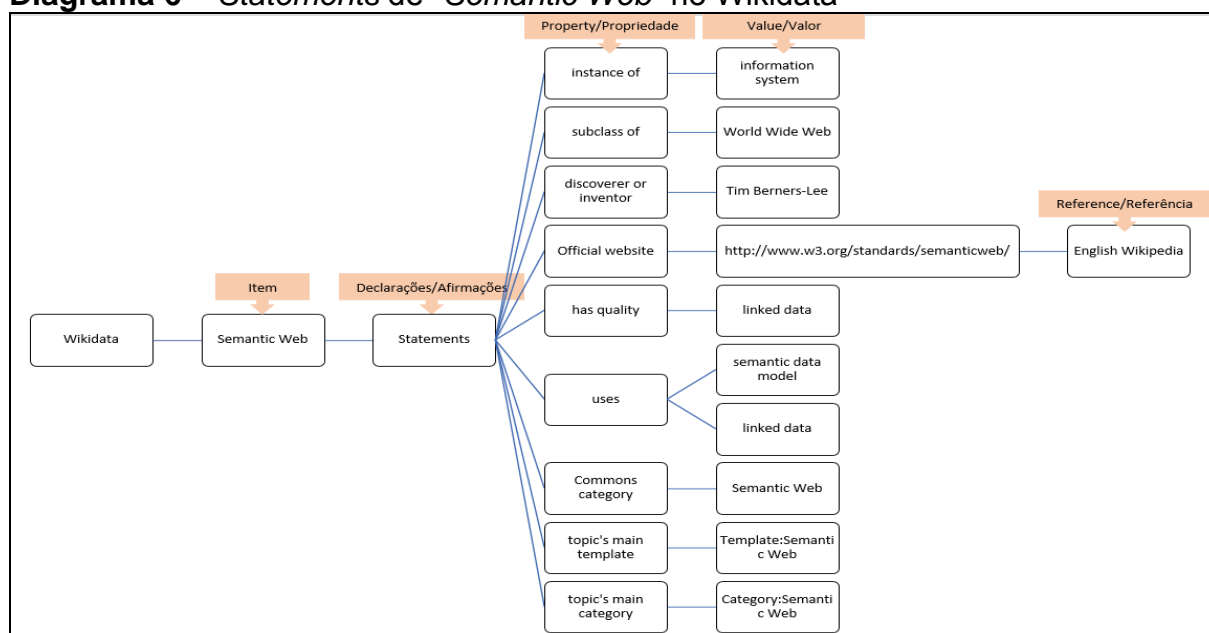
Na sequência, a página da entidade “*Semantic Web*” foi acessada para a análise de suas declarações. Em sua composição há nove propriedades e onze valores, além de 42 entradas para outros idiomas. Curiosamente, os valores dos campos “*Description*” e “*Aliases*” para o idioma português do Brasil estavam vazios. Diante disso, este pesquisador se prontificou a preencher o campo “*Description*” com a seguinte informação: “É uma extensão da *Web* para facilitar o intercâmbio de dados na rede.”, esta frase foi escrita com base na entrada existente em língua inglesa. Apesar de não haver limitação de caracteres para a quantidade

de informação inserida no campo, se buscou seguir o padrão adotado em outros idiomas, o de descrições breves, minimalistas e informais.

Desenvolvida pelo usuário “Gewild”, a página sobre a entidade “*Semantic Web*” de codificação “*Unique identifiers*” equivalente a “Q54837” passou por 85 edições desde sua criação, no ano de 2012. O número de edições realizadas é inferior aos alcançados pelos termos anteriormente analisados, o que talvez ocorra em razão de se tratar de um conceito de base tecnológica, ainda muito recente e, conseqüentemente, pouco difundido a nível mundial. Esta entidade ainda é composta por nove propriedades e onze valores e todas suas declarações puderam ser representadas no Diagrama 6, em virtude de seu tamanho reduzido.

Em nenhuma das propriedades há valores referenciados ao seu respectivo verbete da Wikipédia, a única referência utilizada, a conecta com *website* da W3C, indicado por uma propriedade como o *website* oficial desta entidade.

Diagrama 6 – Statements de “*Semantic Web*” no Wikidata



Fonte: o próprio autor com base em Wikidata (2017d).

Dessa forma, os procedimentos de análise realizados nas *Knowledge Bases*: DBpedia e Wikidata foram concluídos. Constatou-se que o funcionamento dessas *KBs* está de acordo com o que a literatura havia indicado por meio de Sterckx et al. (2016), Bizer et al. (2009) e Biega, Kuzey e Suchanek (2013), principalmente sobre a forma como o conhecimento enciclopédico colaborativo

advindo da Wikipédia é semanticamente estruturado mediante a extração de dados semiestruturados dos verbetes.

Apesar das duas *KBs* serem de acesso aberto e colaborativas, o nível de conhecimento técnico exigido para contribuir com a expansão das conexões semânticas das entidades na DBpedia parece ser mais elevado do que o requisitado no Wikidata. Fato este, que possivelmente desencoraje o ingresso de novos membros na comunidade DBpedia ou restrinja-se a um número menor de colaboradores. Entretanto, iniciativas como a RelFinder e OpenLink Virtuoso auxiliaram na compreensão sobre o funcionamento da DBpedia e como se dão os relacionamentos entre entidades. Além disso, as tecnologias da *Web Semântica* estão presentes na maioria dos processos que envolvem a interoperabilidade desses dados no âmago dessas *KBs*, seja para a padronização, extração e/ou publicação de informações.

A seção seguinte contempla a terceira etapa de análise desta investigação, e investigou se os principais mecanismos de busca da atualidade utilizam o conhecimento enciclopédico colaborativo monitorado até aqui para aprimorar os resultados de busca apresentados aos seus usuários.

5.3 BING, GOOGLE E YAHOO

Nesta seção, o ciclo que envolveu a análise dos termos “Tim Berners-Lee”, “England” e “*Semantic Web*” completa-se na busca. Em um breve resgate sobre os procedimentos realizados até o momento, a primeira etapa deste ciclo visou atender ao **OBJETIVO ESPECÍFICO A** por meio da descrição das atividades que envolvem a criação, edição e manutenção desses termos sob a condição de verbetes no ambiente da Wikipédia. Na fase subsequente, em cumprimento ao **OBJETIVO ESPECÍFICO B**, esses mesmos termos foram submetidos à análise nas *Knowledge Bases*: DBpedia e Wikidata para vislumbrar a forma como verbetes são convertidos em entidades através de processos de estruturação de informações.

Dessa maneira, esta seção está designada a contemplar o **OBJETIVO ESPECÍFICO C**, que investigou se os principais MBs da atualidade (Bing, Google e Yahoo!) utilizam o conhecimento enciclopédico colaborativo advindo da Wikipédia e estruturado nas *KBs*: DBpedia e Wikidata. Todos os MBs tiveram suas configurações de região e idioma alteradas para o idioma inglês (Estados Unidos), para que

houvessem correspondências, visto que nas etapas anteriores Wikipédia, DBpedia e Wikidata pertenciam a este mesmo padrão.

Para uma recuperação de informações pautada o mais próximo possível da imparcialidade, elencou-se o navegador Vivaldi⁴⁹ para a realização das buscas nos MBs. Seu uso foi necessário, visto que havia a possibilidade de outros navegadores, pertencentes às mesmas corporações responsáveis pelo desenvolvimento dos MBs, influenciarem os resultados apresentados por mecanismos concorrentes. Por exemplo, ao realizar uma pesquisa pelos termos elencados no MB “Bing” por intermédio do navegador “Google Chrome” ou o contrário, no MB “Google” através do *browser* “Internet Explorer” os resultados poderiam vir a serem recuperados e ordenados de maneiras diferentes. Assim, para evitar possíveis conflitos de interesses ou divergências nos resultados, o navegador Vivaldi foi designado como uma opção “neutra” para esta investigação.

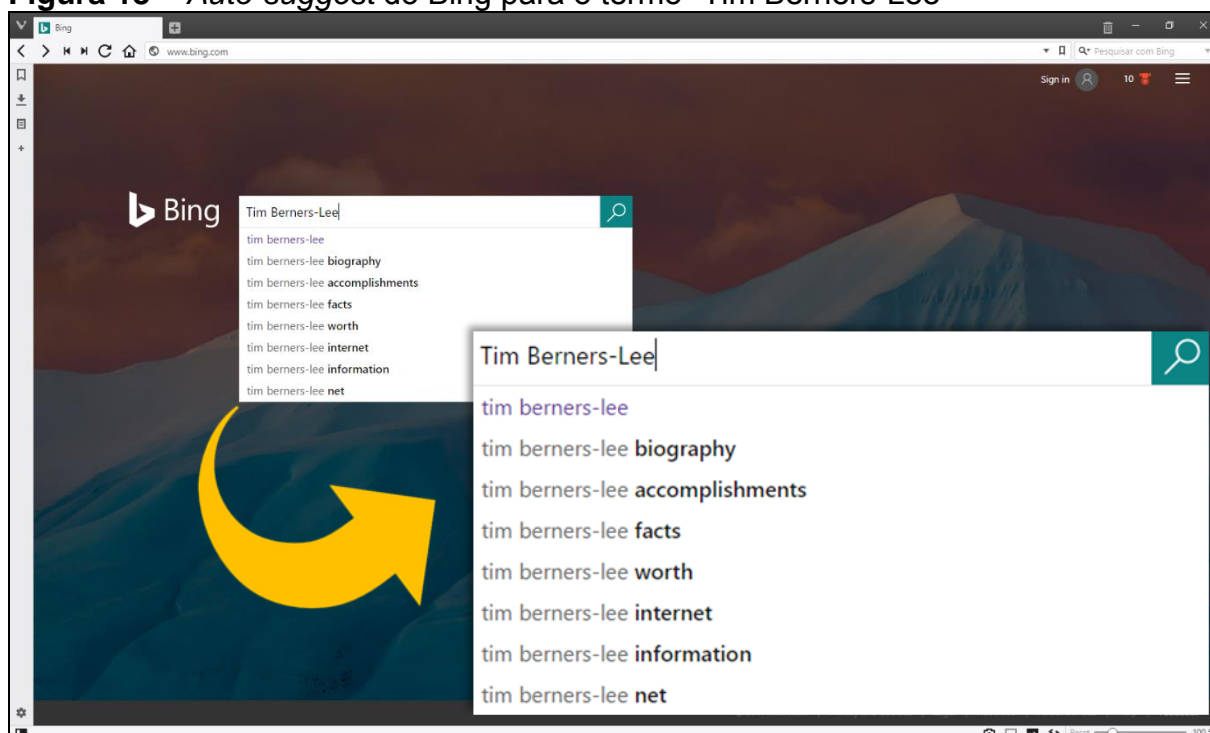
Além disso, o histórico de buscas e as preferências do navegador e dos MBs analisados foram reconfiguradas ao fim da análise de cada termo, para que não houvesse tentativas de perfilamento que pudessem influenciar os resultados.

5.3.1 Bing

Com base nas configurações anteriormente estabelecidas, o procedimento de análise foi iniciado a partir do acesso realizado à página do mecanismo Bing⁵⁰. Dessa forma, o termo “Tim Berners-Lee” foi o primeiro a ser inserido em sua caixa de busca, ativando imediatamente o *auto-suggest* que indicou oito opções de pesquisas complementares ao termo pretendido, conforme apresentado na Figura 18.

⁴⁹ <https://vivaldi.com/?lang=en>

⁵⁰ <http://www.bing.com/?FORM=Z9FD1>

Figura 18 – Auto-suggest do Bing para o termo “Tim Berners-Lee”

Fonte: adaptado de Bing (2017a).

Ao comparar as indicações realizadas pelo *auto-suggest* do Bing com as que foram exibidas nos mecanismos nativos da Wikipédia (Figura 5) e do Wikidata (página 100), observou-se que não há correspondências entre as pesquisas sugeridas nas três plataformas para o termo de “Tim Berners-Lee”, além, é claro, da primeira que, comumente, corresponde ao próprio termo digitado. As oito sugestões de pesquisa tiveram suas *Search Engine Results Pages* (SERPs) acessadas individualmente com o propósito de verificar semelhanças e divergências entre seus resultados. Feito isso, elas foram compiladas na Figura 19 a fim de traçar este comparativo de forma simplificada.

Figura 19 – Mosaico de SERPs do Bing para o termo “Tim Berners-Lee”

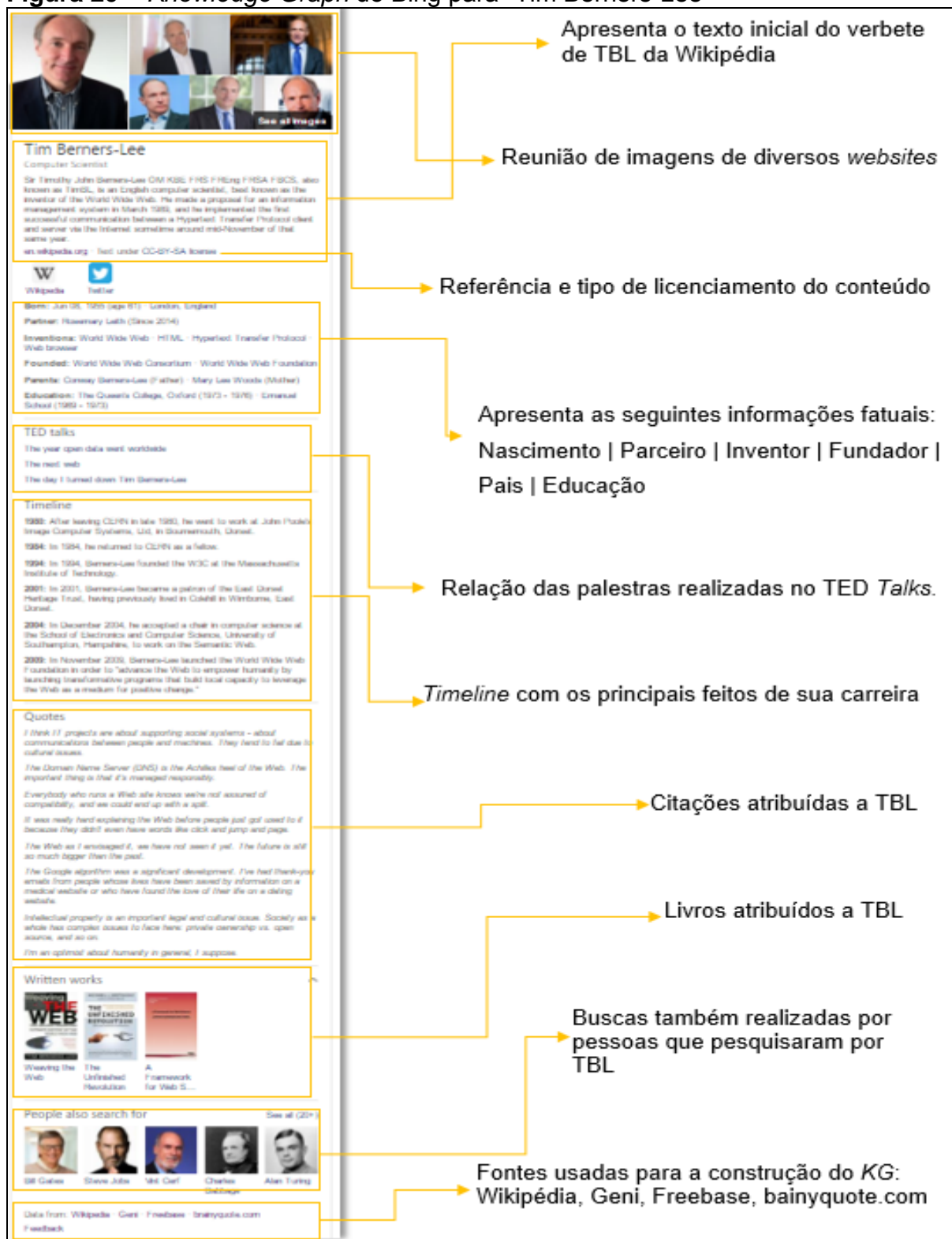


Fonte: adaptado de Bing (2017a).

O comparativo foi iniciado a partir da localização de, pelo menos, uma referência entre os resultados listados na SERP das oito requisições, mediante um *hyperlink* que pudesse redirecionar para o artigo de “Tim Berners-Lee” na Wikipédia. Nesse sentido, todas as SERPs continham o redirecionamento para o verbete que, geralmente, era apresentado entre os três primeiros resultados; apenas na SERP “5” o verbete foi listado como o último item da página. Além disso, nas SERPs “5” e “7”, houve a presença de um recurso denominado como “*answer box*” ou caixa de resposta, no qual o MB apresenta o conteúdo de determinado *website* na sua própria SERP, como se fosse uma resposta rápida à requisição realizada, sem que seja necessário acessar diretamente a fonte de informação.

Localizado no canto direito de todas as SERPs analisadas, o *Knowledge Graph* apresenta o mesmo padrão de exibição de dados para todas as requisições e sua composição assemelha-se ao *infobox* da Wikipédia, conforme apresentado na Figura 20.

Figura 20 – Knowledge Graph do Bing para “Tim Berners-Lee”



Fonte: adaptado de Bing (2017a).

O KG possui seções que o relacionam com uma série de outras informações, como: a “TED talks” onde constam suas palestras, a “Timeline” em que ficam datados importantes marcos de sua história, a “Quotes” que registra suas

principais citações, a “*Written works*” que indica suas principais obras e a “*People also search for*” que o relaciona com outros entusiastas e cientistas do ramo da tecnologia, com base no registro de pesquisas realizadas pelos usuários, conforme a Figura 20.

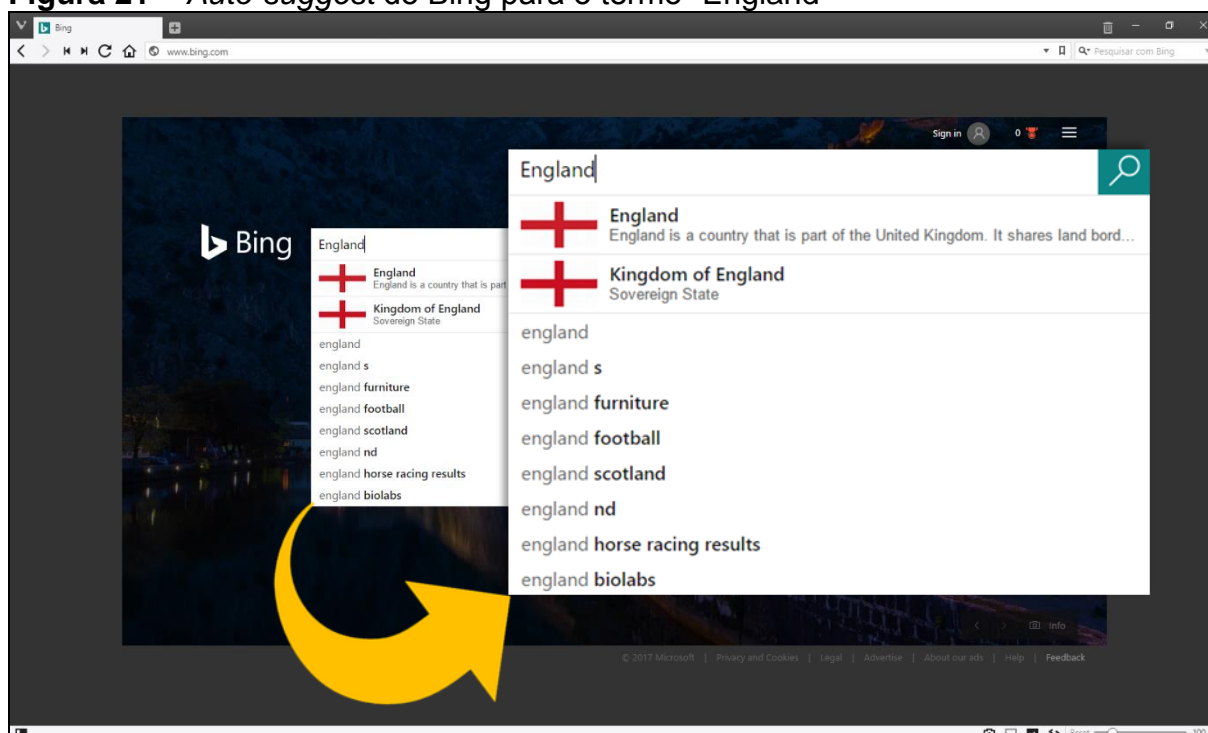
No topo da estrutura do *KG* são exibidas imagens advindas de variados endereços da *Web*, mais abaixo, após a identificação da entidade “Tim Berners-Lee (*Computer scientist*)”, há um campo destinado à apresentação de um breve resumo sobre o item pesquisado. As informações que o compõem são extraídas de seu verbete na Wikipédia e evidenciam o uso do conhecimento enciclopédico colaborativo pelo MB Bing. Esses acontecimentos respondem positivamente ao **OBJETIVO ESPECÍFICO C**, que buscou investigar se os MB utilizam o conhecimento enciclopédico colaborativo que perpassou por Wikipédia, DBpedia e/ou Wikidata na apresentação de seus resultados.

Não foi possível identificar com precisão quais dados têm origem nas *KBs*, mas a última seção do *KG* sobre “Tim Berners-Lee” indica a utilização de referências como Wikipédia e Freebase. Porém, o *hyperlink* referente à fonte “Freebase” redireciona para uma página⁵¹ da Google que informa sobre a desativação da *KB*, com a possibilidade de *download* de seu legado. Fato este que corrobora com a afirmação de Tanon et al. (2016) sobre o processo de migração de dados do Freebase para o Wikidata. A maioria dos *hyperlinks* presentes nos campos de descrição das informações fatuais sobre a entidade corresponde a “*queries*” prontas para redirecionar o usuário às SERPs correspondentes.

Em continuidade à análise no Bing, o termo “England” foi inserido na caixa de busca do mecanismo que indicou, por meio do *auto-suggest*, dez requisições complementares à pesquisa principal, conforme exposto na Figura 21. Nesta requisição, o *auto-suggest* apresentou opções que continham, além da descrição textual, duas imagens vinculadas à busca.

⁵¹ <https://developers.google.com/freebase/>

Figura 21 – *Auto-suggest* do Bing para o termo “England”



Fonte: adaptado de Bing (2017b).

As duas primeiras requisições, além do adendo das imagens, também são uma forma de desambiguação apresentada no próprio *auto-suggest* do Bing, simplificando o processo de recuperação de informações acerca da temática requisitada. As dez sugestões de pesquisas complementares tiveram suas SERPs acessadas para que uma análise comparativa entre os elementos apresentados nos resultados pudesse ser realizada, conforme a Figura 22.

Dessa maneira, as SERPs “2, 5, 6, 7, 8, 9 e 10” não correspondem aos resultados para “England” especificamente, mas aos assuntos correlatos como a seleção nacional de futebol e empresas com sede no país. As SERPs “1” e “3” retornaram um padrão de organização dos resultados semelhante, visto que a única diferença na requisição consiste no modo como a palavra é digitada, uma em letras maiúsculas e a outra em letras minúsculas. Além disso, elas foram as únicas que retornaram o *hyperlink* específico para acesso ao verbete “England” da Wikipédia.

Alguns itens foram usados para identificar tipos de informações existentes nas SERPs, assim, a letra “W”, na cor azul, representa a existência de *hyperlinks* para a Wikipédia sobre o tema, enquanto as letras “KG”, na cor vermelha, indica a presença do *Knowledge Graph* como elemento constituinte da SERP.

Figura 22 – Mosaico de SERPs do Bing para o termo “England”



Fonte: adaptado de Bing (2017b).

Entretanto, validar a existência ou não de *hyperlinks* de redirecionamento para os verbetes aqui analisados não configura um indicativo de uso do conhecimento enciclopédico colaborativo, nem mesmo assegurar que os *hiperlinks* influenciam para a instauração da *Web Pragmática*. Assim, algumas tecnologias aplicadas ao contexto de apresentação dos resultados nas SERPs

puderam ser consideradas como indícios de uso desse conhecimento, a exemplo, o *Knowledge Graph*.

Nesse sentido, as SERPs “1”, “3” e “4” foram as únicas que exibiram o *KG*, apresentado na Figura 23, em suas recuperações, apesar das variações de escrita do termo nas três requisições.

Figura 23 – *Knowledge Graph* do Bing para “England”

The image shows a Bing search result for 'England'. At the top is a map of the United Kingdom with England highlighted. Below the map is a section titled 'England' with a red cross icon and a brief Wikipedia-style description. Underneath are key statistics: Population (53.01 million), Calling code (44), Area (50,348 sq miles), GDP (\$2 trillion USD), Capital (London), and Currency (UK £). The 'Destinations' section features five image thumbnails for London, Birmingham, Manchester, Liverpool, and York. The 'Points of interest' section shows five thumbnails for Stonehenge, Big Ben, Buckingham Palace, River Thames, and London Eye. The 'People also search for' section displays five thumbnails for David Cameron, Elizabeth I of England, Prince Charles, William Shakespeare, and Boudica. At the bottom, there is a 'Data from Wikipedia · Freebase' section and a 'Feedback' link. Three yellow arrows point from text annotations on the right to specific parts of the interface: the first points to the Wikipedia description, the second points to the Destinations and Points of interest sections, and the third points to the bottom section containing the data source and feedback link.

Informações extraídas do verbete “England” na Wikipédia

Nas seções “Destinations”, “Points of interest” e “People also search for” são indicadas sugestões de pesquisas relacionadas ao termo “England”

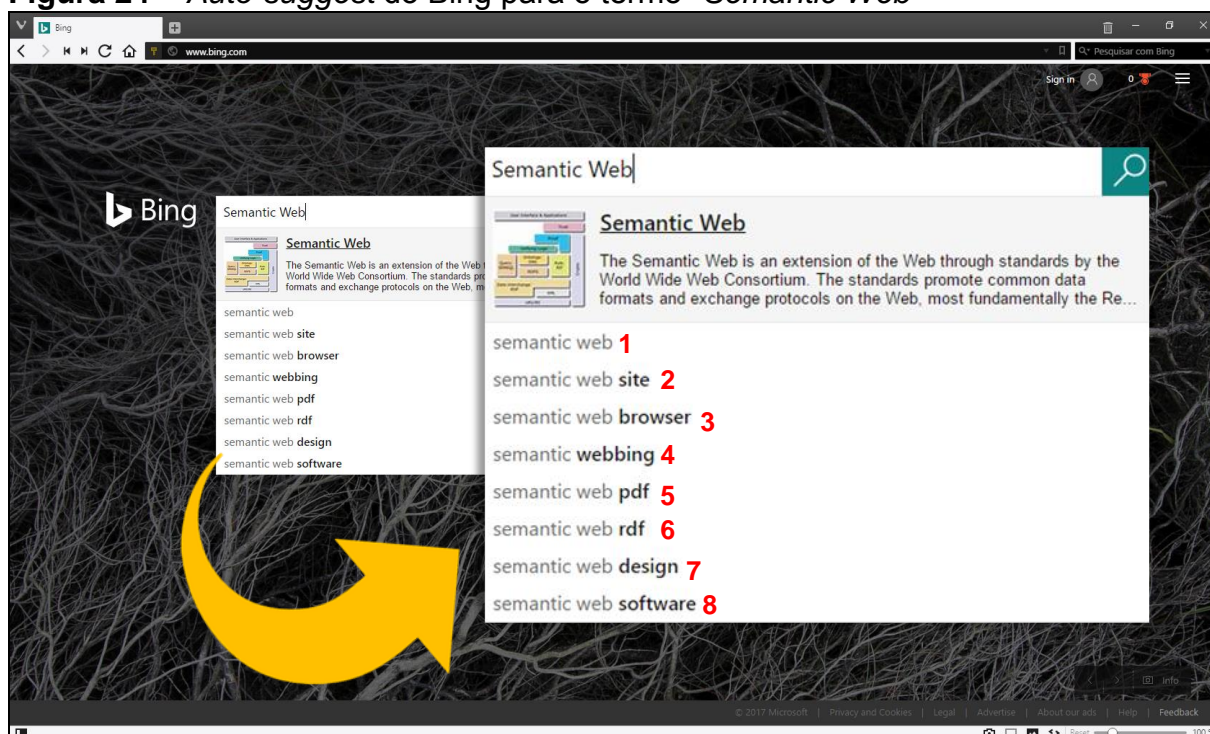
Indicação das fontes utilizadas para o constructo do *KG* e também a opção “Feedback”

Fonte: adaptado de Bing (2017b).

A informação inicial sobre a entidade, apresentada no topo do *KG*, é idêntica ao do texto existente no verbete da Wikipédia. Assim, não há como identificar de maneira segura se essas informações foram obtidas diretamente da Wikipédia ou por meio de propriedades e declarações de *Knowledge Bases*. Entretanto, infere-se que o conteúdo inicial do *KG* é alimentado diretamente pelo respectivo verbete da Wikipédia, visto que Rospocher et al. (2016) apontam que apenas os relacionamentos e informações fatuais são obtidos de *KBs*.

O termo “*Semantic Web*” foi o último elemento investigado no MB Bing. Para tanto, navegador e motor de busca foram novamente reconfigurados para assegurar ao máximo a imparcialidade dos resultados. Feito isso, o termo “*Semantic Web*” foi inserido na caixa de busca do mecanismo e retornou oito sugestões de pesquisas, conforme exposto na Figura 24.

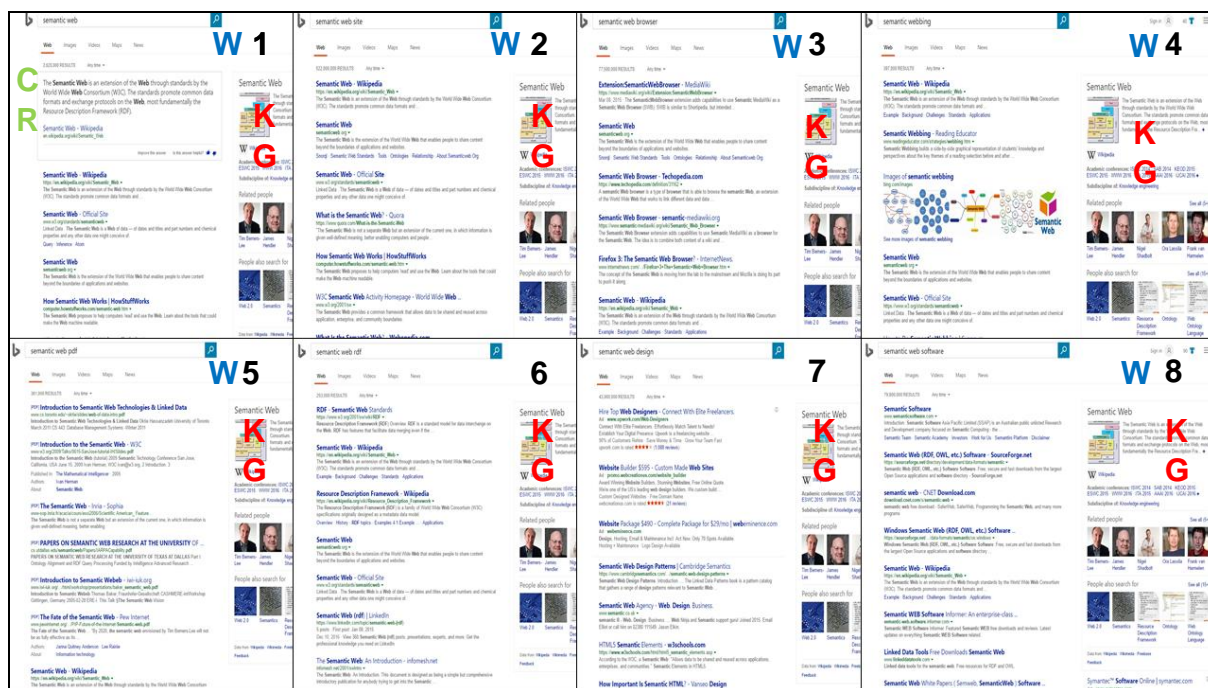
Figura 24 – *Auto-suggest* do Bing para o termo “*Semantic Web*”



Fonte: adaptado de Bing (2017c).

A primeira sugestão de busca, indicada pelo *auto-suggest*, se apresentou da mesma maneira que havia ocorrido para o termo “*England*”, com atributos visuais e textuais. As oito requisições complementares foram acessadas para que as SERPs fossem submetidas à uma análise comparativa, conforme a Figura 25.

Figura 25 – Mosaico de SERPs do Bing para o termo “Semantic Web”



Fonte: adaptado de Bing (2017c).

Por conta da quantidade de SERPs acessadas, a resolução do mosaico ficou comprometida ao compilá-las, fato que não prejudicou a análise por ter sido usado como elemento de apoio para a discussão. Apenas as SERPs “6” e “7” não continham referências, entre a lista de resultados, para o artigo da Wikipédia. Na SERP “1” se identificou a “*answer box*”, que trouxe o texto inicial do verbete “*Semantic Web*” da Wikipédia, indicando a edição em inglês como referência.

Após uma averiguação acerca do funcionamento do recurso “*answer box*”, foi constatado que este refere-se a um projeto colaborativo intitulado como “*Bing Distill*”⁵² e visa conectar as perguntas realizadas ao Bing às caixas de respostas que serão exibidas entre os primeiros resultados da SERP. Para contribuir, os usuários devem possuir uma conta Microsoft⁵³ que possibilitará o acesso, a edição e a criação de respostas concisas, fatuais e escritas corretamente, além da necessidade de indicar sempre fontes confiáveis. Assim, a “*answer box*” pode ser compreendida como uma melhoria na exibição dos resultados, vinculada ao funcionamento do KG (BUTZBACH, 2017).

⁵² <http://www.bing.com/distill/questions?filter=Unpublished>

⁵³ <https://www.microsoft.com/pt-br/>

Todas as SERPs da Figura 25 tiveram a ocorrência do KG e sua estrutura de organização de informações foi apresentada de maneira idêntica em todos os casos, conforme exposto na Figura 26.

Figura 26 – Knowledge Graph do Bing para “Semantic Web”

Semantic Web

The Semantic Web is an extension of the Web through standards by the World Wide Web Consortium. The standards promote common data formats and exchange protocols on the Web, most fundamentally the Resource Description Fra... +

Wikipedia

Academic conferences: ISWC 2014 · SAB 2014 · KEOD 2015 · ESWC 2015 · WWW 2016 · ITA 2015 · AAI 2016 · IJCAI 2016 +

Subdiscipline of: Knowledge engineering

Related people See all (5+)

Tim Berners-Lee James Hendler Ora Lassila Nigel Shadbolt Frank van Harmelen

People also search for See all (15+)

Web 2.0 Semantics Resource Description Framework Ontology Web Ontology Language

Data from: Wikipedia · Wikimedia · Freebase

Feedback

Annotations:

- Texto inicial idêntico ao do verbete na Wikipédia, referenciado logo abaixo
- Hyperlinks com queries sobre eventos acerca da temática, além de estar vinculada como subdisciplina da Engenharia do Conhecimento
- Pessoas envolvidas com a temática
- Pesquisas sugeridas com base nas buscas de outros usuários
- Fontes utilizadas: Wikipédia, Wikimedia e Freebase
- Feedback

Fonte: adaptado do Bing (2017c).

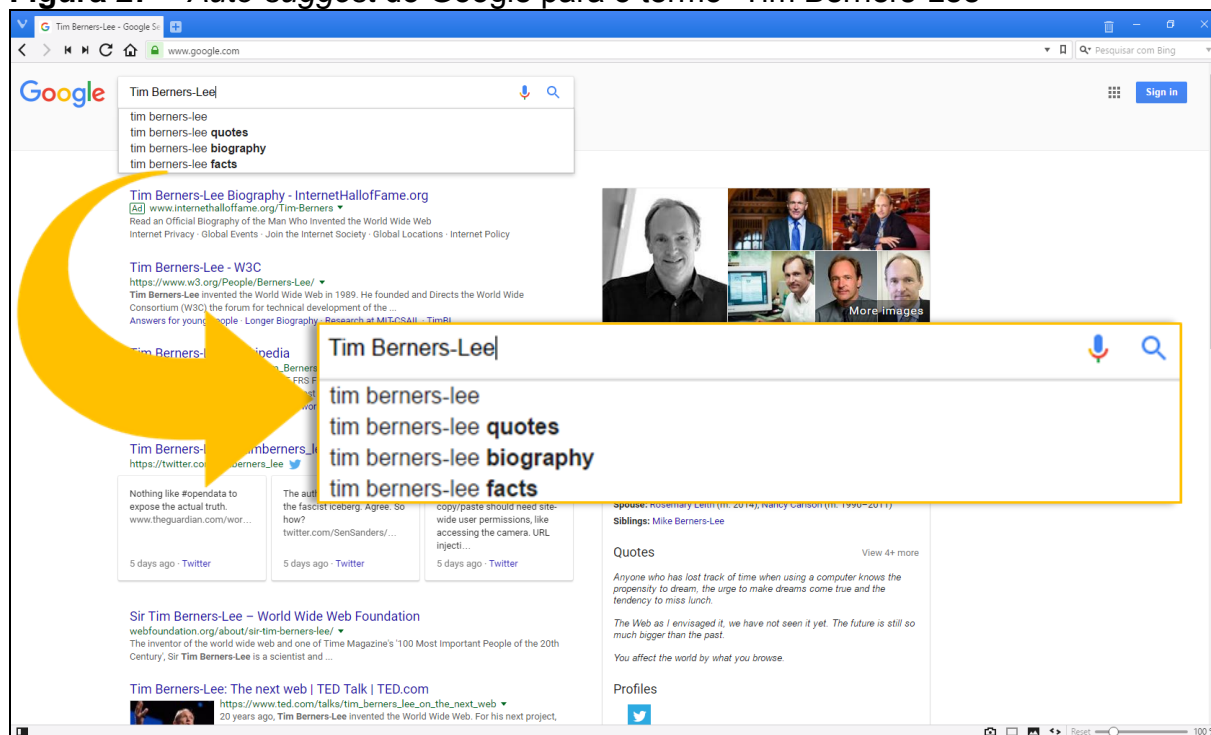
Novamente, o texto inicial do KG correspondeu exatamente ao conteúdo do verbete existente na Wikipédia. Entre as fontes utilizadas por este KG, houve a presença de uma nova referência, a “Wikimedia”, que, após ter seu *hyperlink* verificado, foi possível indicar que se refere à imagem presente no início do KG. A estrutura do KG é encerrada com a opção “Feedback”, na qual o usuário pode enviar críticas e sugestões para a melhoria do KG, porém sem a possibilidade de edição direta do conteúdo, visto que as informações são enviadas para avaliação, sob responsabilidade do Bing, e somente depois podem vir a compor o KG. Em

continuidade ao procedimento de análise desta dissertação, o próximo mecanismo investigado foi o Google, conforme abordado na seção subsequente.

5.3.2 Google

Nesta segunda fase, o MB Google foi analisado com base nos mesmos parâmetros realizados na seção anterior. Desse modo, após a reconfiguração do navegador e do mecanismo de pesquisa, o termo “Tim Berners-Lee” foi inserido na caixa de busca, que indicou, por meio do *auto-suggest*, quatro possibilidades complementares de buscas, conforme apresenta a Figura 27.

Figura 27 – Auto-suggest do Google para o termo “Tim Berners-Lee”



Fonte: adaptado de Google (2017a).

No Google, as sugestões são realizadas pelo MB simultaneamente ao redirecionamento para a SERP. Assim, o usuário não necessita selecionar uma opção de requisição do *auto-suggest* para ser redirecionado, pois conforme o termo é inserido, a SERP é carregada e recarregada até o término da digitação. Esse funcionamento corresponde ao emprego da função *auto-complete* em consonância com a *auto-suggest* que, segundo Butzbach (2017), é um complemento propiciado pelos avanços no KG do Google e permite a realização de pesquisas preditivas que

intentam concluir a pesquisa automaticamente para o usuário. Assim, a Figura 28 reuniu as SERPs derivadas dessas funções para que fosse possível analisá-las de forma comparativa.

Figura 28 - Mosaico de SERPs do Google para o termo “Tim Berners-Lee”

The figure displays four distinct Google search result pages for the query "Tim Berners-Lee":

- 1. All:** Shows a general overview with a Knowledge Graph card, a list of search results, and social media links.
- 2. Quotes:** Features a Knowledge Graph card, a "Quotes" section with a quote about dreaming, and search results for quotes.
- 3. Biography:** Includes a Knowledge Graph card, a "Biography" section with a detailed bio, and search results for biographies.
- 4. Facts:** Contains a Knowledge Graph card, a "Facts" section with several bullet points, and search results for facts.

Fonte: adaptado de Google (2017a).

Em comparação às informações apresentadas no MB Bing para o mesmo termo, os resultados do Google mostraram uma organização e uma ordenação diferentes para cada SERP, além de divergências estruturais entre os KGs vinculados a cada requisição. As SERPs “1” e “3” foram as únicas que listaram resultados na primeira página referentes ao verbete de “Tim Berners-Lee” na Wikipédia. Nas SERPs “2” e “4”, houve a presença de “*answer box*” que, segundo Butzbach (2017), são comumente exibidas quando a temática da requisição é reconhecida pelo KG. Assim, o MB apresenta uma resposta à requisição em sua própria SERP no formato de caixa e inclui a referência do *website* utilizado como base logo abaixo da resposta.

Sobre o KG, apenas as SERPs “1” e “4” possuem informações e estrutura equivalentes deste elemento. Na SERP “2” a diferença para as duas anteriores está na ausência do campo “Quotes”, designado à apresentação das principais falas atribuídas a “Tim Berners-Lee”. No KG da SERP “3” não constam as seções “Quotes”, “Profiles” e “People also search for”, além de adicionar mais dois

campos fatuais, não existentes nos outros: “Books” e “Influenced by”. Dessa maneira, infere-se que essa divergência aconteça em virtude da mudança de entidade realizada pelo mecanismo, pois, para as SERPs “1”, “2” e “4”, trata-se de “Tim Berners-Lee (*Computer scientist*)” e, na SERP “3”, a entidade de “Tim Berners-Lee” está vinculada como “Tim Berners-Lee (*Biography*)”.

Além disso, o conhecimento enciclopédico colaborativo se fez presente como parte integrante da breve descrição sobre a entidade no KG, o qual utiliza informações extraídas do texto inicial do verbete de “Tim Berners-Lee” na Wikipédia, conforme exposto na Figura 29.

Figura 29 - Knowledge Graph do Google para “Tim Berners-Lee”

The image shows a screenshot of the Google Knowledge Graph for Tim Berners-Lee. The interface includes a grid of images at the top, a name and profession section, a Wikipedia-style text snippet, a list of biographical facts (born, awards, education, spouse, siblings), a quotes section, a social media profile (Twitter), and a 'People also search for' section with small portraits of Vint Cerf, Robert Cailliau, Robert E. Kahn, Ted Nelson, and Steve Jobs. A 'Feedback' button is at the bottom right. Yellow arrows point from text annotations on the right to specific elements in the Knowledge Graph.

Identificação da entidade e possível ponto de divergência entre o KG da SERP “3” para os demais.

Texto extraído do verbete de TBL na Wikipédia.

Informações fatuais sobre TBL, nas quais os *hyperlinks* redirecionam para *queries* sobre o valor de cada campo

Principais frases de autoria de TBL

Perfil em ferramenta de rede social

Relação de buscas realizadas pelos usuários que também pesquisam por TBL

Informa ao Google possíveis erros no KG

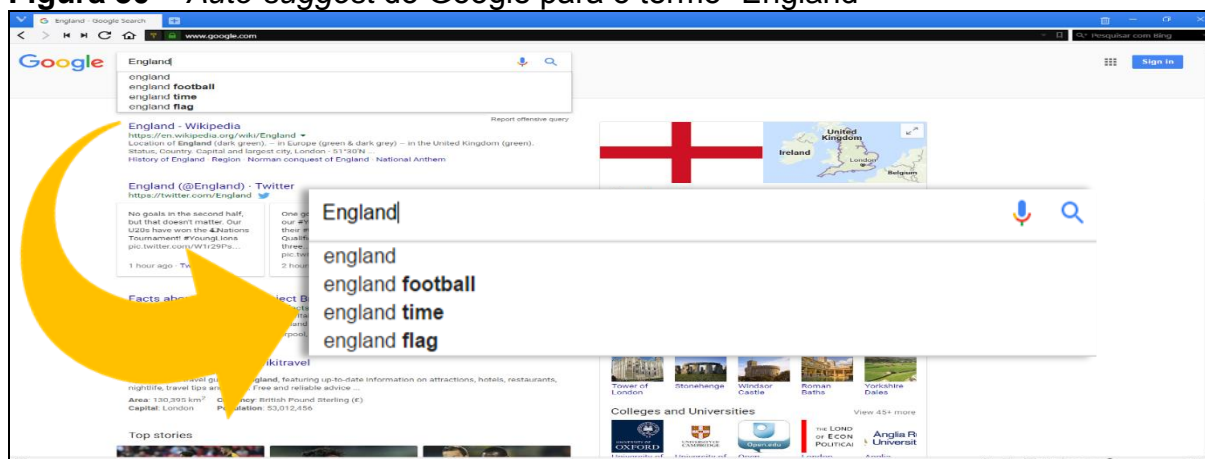
Fonte: adaptado de Google (2017a).

Onde há os campos que descrevem informações fatuais sobre “Tim Berners-Lee”, os *hyperlinks* são compreendidos como valores que redirecionam para a página de seus relacionamentos por meio de requisições prontas para cada item, isto é, os valores que constam na entidade de “Tim Berners-Lee” são compostos por outras SERPs. Em “Quotes” são apresentadas as frases mais emblemáticas atreladas a “Tim Berners-Lee”, porém não há indicação da fonte das quais foram extraídas. O KG também vincula os perfis de ferramentas de redes sociais de “Tim Berners-Lee”, além de apresentar uma relação de pesquisas realizadas por outros usuários, que também haviam buscado por “Tim Berners-Lee”.

O elemento “Feedback” do KG de “Tim Berners-Lee” permite que o usuário, ao selecioná-lo, possa editar as informações que o compõem e enviar para avaliação do Google. Ao verificar que há disparidade nas informações e realizar melhorias, o indivíduo contribui para a formação de um conhecimento consensual na rede (SCHOOP; MOOR; DIETZ, 2006). Esse tipo de ação também corrobora com a visão de Vaidhyathan (2011, p. 74 *apud* MONTEIRO, 2015, p. 10-11) de que, com base no pragmatismo norte-americano, o funcionamento do Google se aproxima a um processo de obtenção da “verdade” resultante de “[...] um processo de experimentação, descoberta, *feedback* e consenso.” Nesse sentido, esses fatores evidenciam a presença da *Web Pragmática* no funcionamento do MB.

Após reconfigurar navegador e MB, o termo “England” foi inserido na caixa de busca do Google para a continuidade do procedimento de análise. Como esperado, o MB indicou quatro sugestões de pesquisas complementares para o termo “England”, expostas na Figura 30.

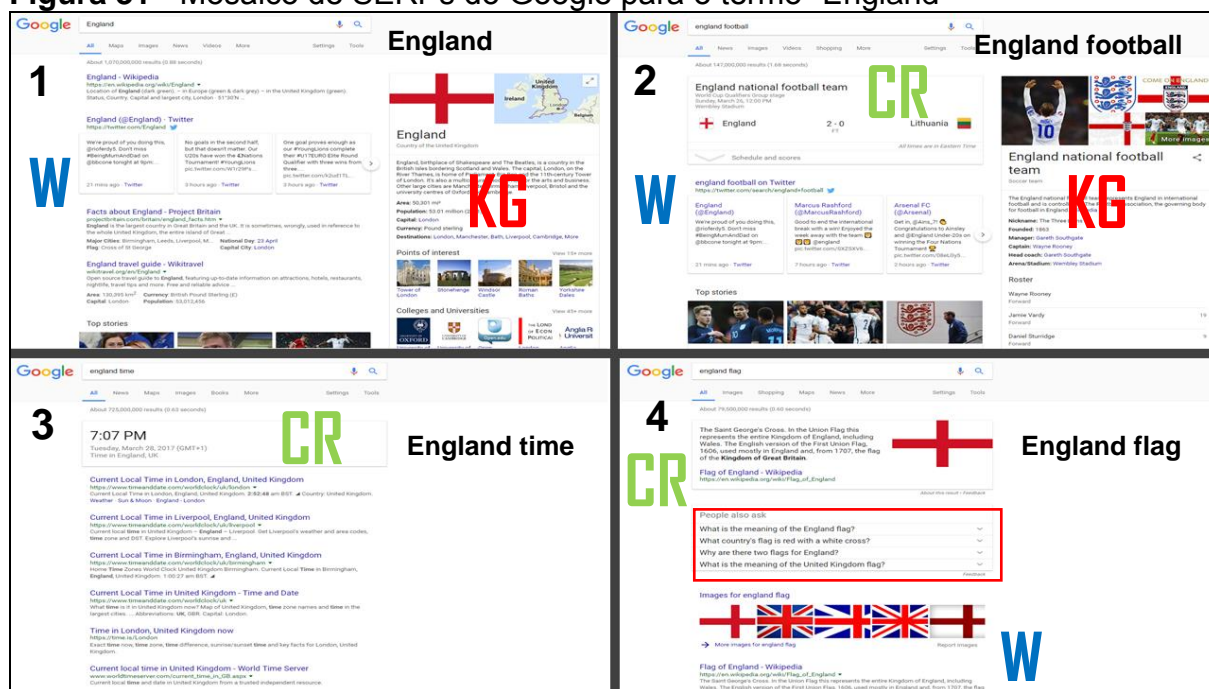
Figura 30 – *Auto-suggest* do Google para o termo “England”



Fonte: adaptado de Google (2017b)

Desse modo, as requisições sugeridas pelos mecanismos foram acessadas para que suas SERPs pudessem ser analisadas comparativamente, como se apresenta na Figura 31.

Figura 31 - Mosaico de SERPs do Google para o termo “England”



Fonte: adaptado de Google (2017b).

Nas SERPs “3” e “4” não há KG na apresentação dos resultados, mas, juntamente com a SERP “2”, houve a identificação do elemento “*answer box*”, que possivelmente é exibido por se tratarem de respostas a requisições simples, como horário e bandeira da Inglaterra. Ainda, na SERP “4” um componente incomum foi identificado e está destacado na cor vermelha na Figura 31. Esse elemento é intitulado como “*People also ask*” e está logo abaixo da “*answer box*”. Sua função é apresentar questionamentos e respostas relacionadas à requisição, desenvolvidos com base em pesquisas realizadas por demais usuários, uma forma de FAQ⁵⁴ a respeito do tema que circunda a requisição.

Foram localizadas, entre os resultados, referências a verbetes da Wikipédia nas SERPs “1”, “2” e “4”, porém apenas a primeira SERP apresentou o artigo correspondente ao termo “England”. Dessa maneira, o KG da SERP “1” está exposto na Figura 32 com o objetivo de vislumbrar se havia outras ocorrências do conhecimento enciclopédico colaborativo em sua estrutura.

⁵⁴ Frequently Asked Questions (FAQ)

Figura 32 – Knowledge Graph do Google para “England”

Entidade identificadora do KG

Resumo

Informações fatuais sobre a entidade

Informações turísticas

Instituições de ensino

Feedback

Instituições de ensino

Fonte: adaptado de Google (2017b).

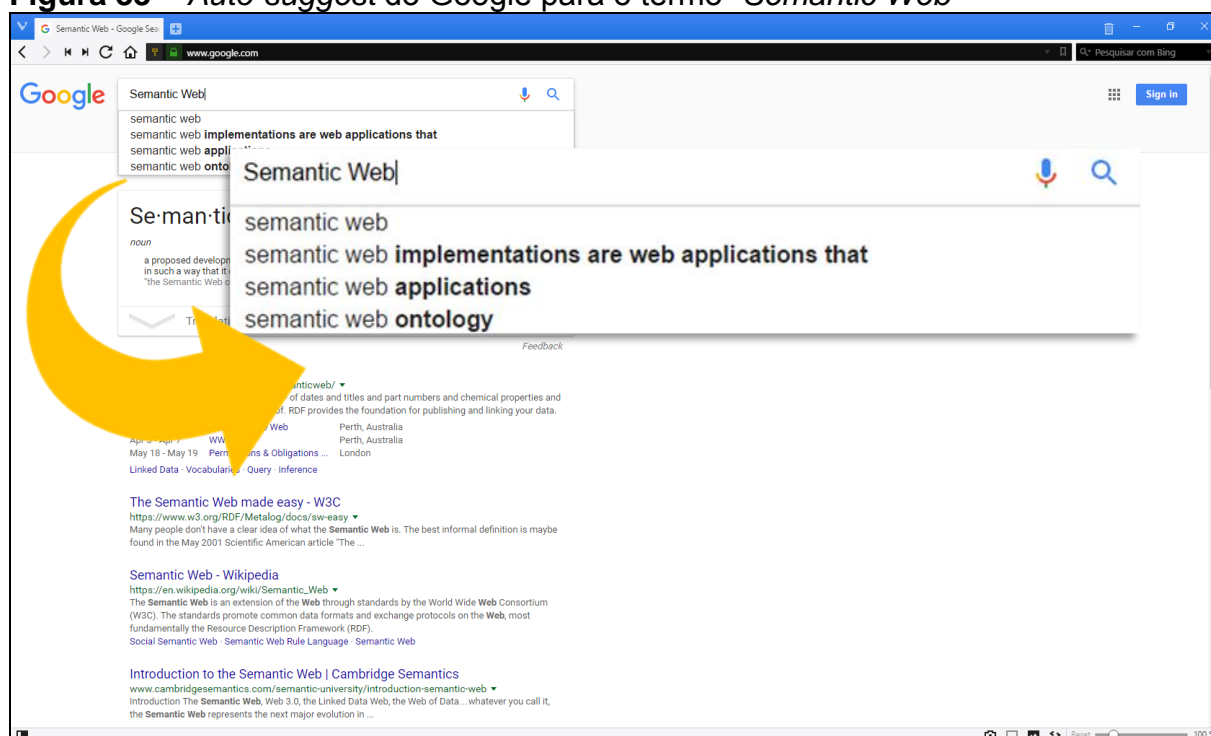
O resumo apresentado no início do KG sobre “England”, no qual se esperava encontrar evidências de uso do conhecimento enciclopédico colaborativo

em sua formação, mostrou não utilizar informações advindas de Wikipédia, DBpedia e Wikidata em sua confecção, visto que o texto foi analisado e comparado com o conteúdo existente nesses três ambientes e não retornou nenhuma correspondência.

Nesse contexto, infere-se que o MB da Google esteja utilizando informações advindas de outros locais da *Web* ou até mesmo dados remanescentes da Freebase, conforme indicou Tanon et al. (2016). A falta de fontes também abrange todas as outras seções do *KG*, e os *hyperlinks* existentes em sua estrutura redirecionam apenas para outras SERPs derivadas de requisições prontas do próprio MB.

Em continuidade ao procedimento de análise no mecanismo Google, o termo “*Semantic Web*” foi inserido em sua caixa de busca, que retornou quatro sugestões de pesquisa, conforme indicado na Figura 33.

Figura 33 – *Auto-suggest* do Google para o termo “*Semantic Web*”

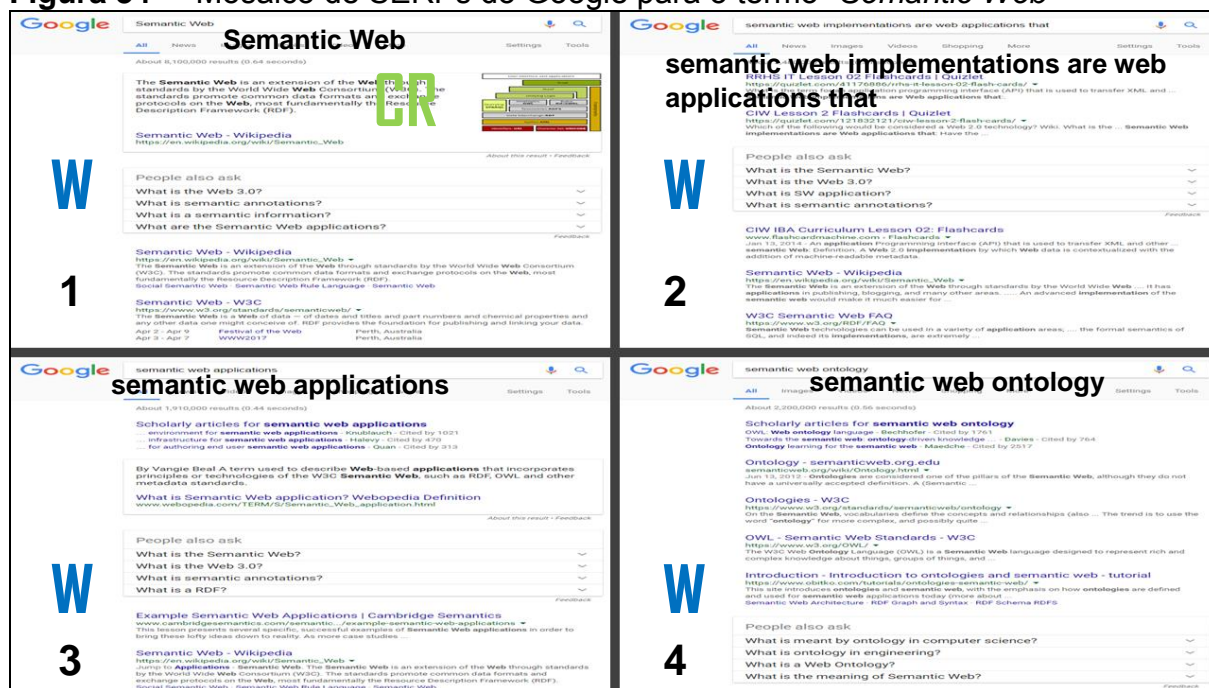


Fonte: adaptado de Google (2017c).

Dessa maneira, a requisição principal e suas respectivas indicações do *auto-suggest* foram compiladas na Figura 34 para uma análise comparativa entre SERPs. Assim, as SERPs “1”, “2” e “3” apresentaram um *hyperlink* para acesso ao

artigo sobre “*Semantic Web*” na Wikipédia. Na SERP “4” há redirecionamento para a Wikipédia, mas referente a outros verbetes.

Figura 34 – Mosaico de SERPs do Google para o termo “*Semantic Web*”



Fonte: adaptado de Google (2017c).

Somente na SERP “1” há a presença de uma “*answer box*”, que apresenta brevemente o conteúdo do artigo da Wikipédia no formato de resposta no próprio mecanismo. Apenas um elemento esteve presente em todas as SERPs investigadas; trata-se novamente do “*People also ask*” que lista possíveis perguntas e respostas relacionadas à temática da requisição. Além disso, a análise deste último termo revelou a inexistência de *KG*, tecnologia utilizada até então como forte indício da utilização do conhecimento enciclopédico colaborativo pelos mecanismos de busca.

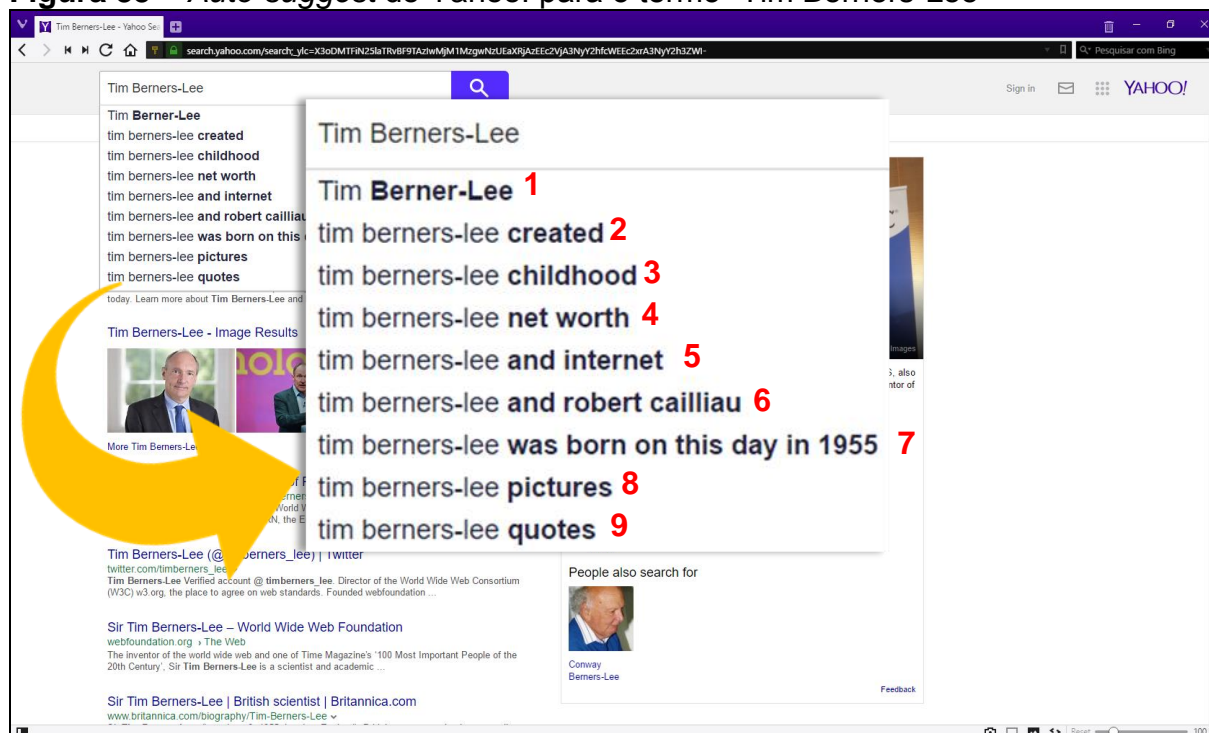
Por outro lado, a “*answer box*” apresentada na SERP “1” fez a conexão necessária com o respectivo verbete na Wikipédia ao exibir uma breve descrição do conceito em formato de resposta, pautando-se em informações advindas do artigo sobre “*Semantic Web*” na Wikipédia. Além disso, por estar alocada na primeira posição de resultados da SERP, ela pode ser considerada como uma opção mais relevante a ser exibida ao usuário na concepção do algoritmo do Google (SILVA; SANTOS; FERNEDA, 2013). Na próxima seção, os mesmos termos

aqui abordados, serão analisados sob a perspectiva de funcionamento do último MB deste ciclo de análise, o Yahoo!.

5.3.3 Yahoo!

Nesta última fase, o MB Yahoo! foi analisado com base nos mesmos pressupostos de avaliação aplicados aos seus antecessores: Bing e Google. No navegador “Vivaldi”, o histórico de pesquisa foi removido e as configurações de busca e exibição de resultados alteradas para o idioma inglês (Estados Unidos). O primeiro termo inserido na caixa de busca foi o de “Tim Berners-Lee”, que retornou, por meio do *auto-suggest*, nove opções de requisições complementares ao tema, conforme exposto na Figura 35.

Figura 35 – *Auto-suggest* do Yahoo! para o termo “Tim Berners-Lee”



Fonte: adaptado de Yahoo (2017a).

As opções indicadas pelo *auto-suggest* do Yahoo! tiveram suas SERPs acessadas com a pretensão de investigá-las comparativamente, em especial, aos indícios de uso do conhecimento enciclopédico colaborativo na apresentação de seus resultados. Nesse sentido, somente a SERP “4” não apresentou referências ao verbete de “Tim Berners-Lee” da Wikipédia e nas SERPs

“5, 6 e 7” o *KG* não se fez presente. Por outro lado, as SERPs “1, 2, 3, 4, 8 e 9” incluíram o mesmo *KG* de “Tim Berners-Lee” para requisições de pesquisa diferentes, conforme a Figura 36.

Figura 36 – Mosaico de SERPs do Yahoo para o termo “Tim Berners-Lee”



Fonte: adaptado de Yahoo (2017a).

Minimalista, este *KG* exibe poucas informações se comparado aos apresentados por Bing e Google para o mesmo termo. Porém, o modo de funcionamento é equivalente, pois o texto inicial, exibido na parte superior de sua estrutura, é o mesmo existente no verbete da Wikipédia e também utilizado pelos outros dois MBs avaliados anteriormente.

Figura 37 – Knowledge Graph do Yahoo para “Tim Berners-Lee”



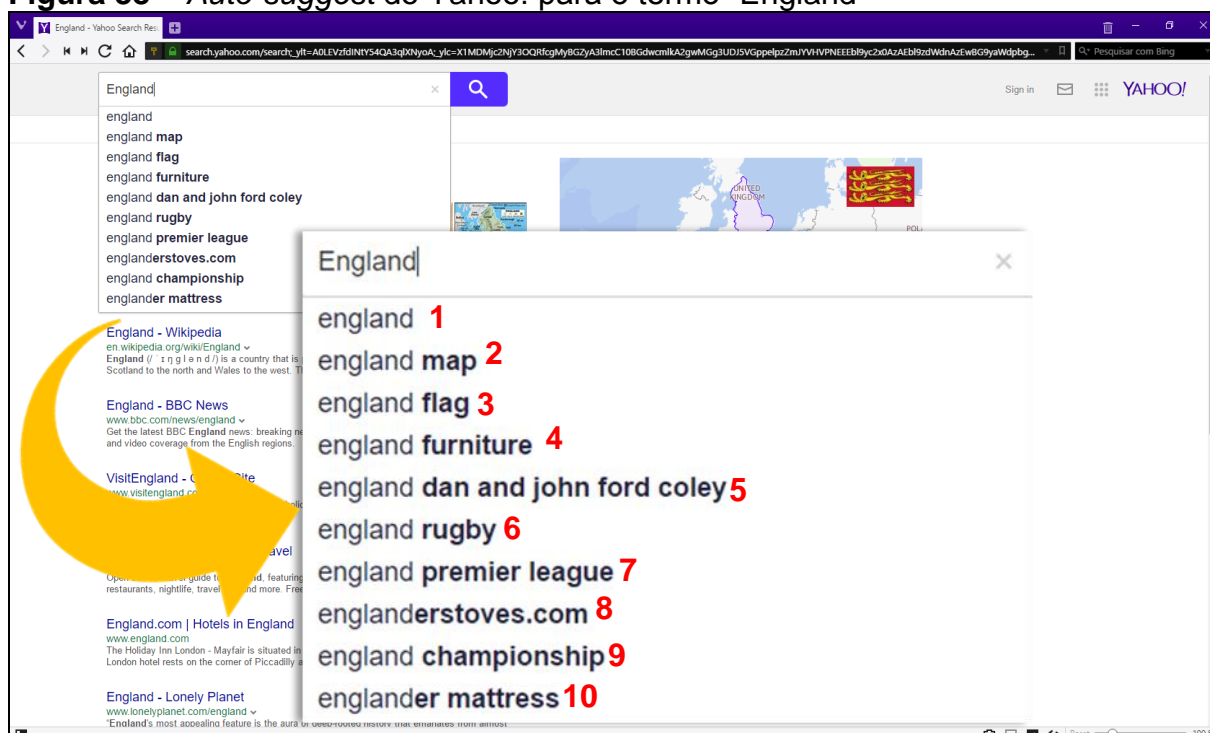
Fonte: adaptado de Yahoo (2017a).

As descrições existentes em campos como “*Spouse*” e “*Parents*”, apresentadas no formato de *hyperlinks*, tratam-se de *queries* prontas a respeito da própria descrição e, ao selecioná-la, o usuário é redirecionado para a SERP correspondente ao assunto. A penúltima seção do *KG* está destinada a listar ferramentas de redes sociais nas quais “Tim Berners-Lee” possui perfis, sendo que, neste caso, consta apenas a conta do “Twitter”⁵⁵. A opção “*Feedback*” permite que os usuários do MB indiquem melhorias e correções sobre o conteúdo apresentado no *KG*, além disso, essas alterações são classificadas, validadas e complementadas pela própria comunidade *on-line* do Yahoo!.

Após reconfigurar navegador e MB, o termo “England” foi inserido na caixa de busca do Yahoo! em continuidade ao procedimento de análise neste mecanismo. Assim, dez opções de pesquisas complementares ao termo “England” foram indicadas pelo *auto-suggest* do Yahoo!, conforme apontado na Figura 38.

⁵⁵ <https://twitter.com/?lang=en>

Figura 38 – Auto-suggest do Yahoo! para o termo “England”



Fonte: adaptado de Yahoo (2017b).

Em acesso às SERPs originadas das sugestões de busca do Yahoo!, buscou-se averiguar a existência de resultados que estivessem vinculados ao conhecimento enciclopédico colaborativo referente ao verbete “England” da Wikipédia. Essas SERPs foram compiladas na Figura 39 em conjunto com algumas marcações para identificação de elementos que utilizaram este conhecimento, como *KG* e resultados da própria lista.

Figura 39 – Mosaico de SERPs do Yahoo! para o termo “England”

The figure displays a grid of 10 search engine result pages (SERPs) for the term "England" on Yahoo!. The pages are numbered 1 to 10. SERP 1 shows a map of England and a Wikipedia entry. SERP 2 shows a map of England and a Wikipedia entry. SERP 3 shows a map of England and a Wikipedia entry. SERP 4 shows a map of England and a Wikipedia entry. SERP 5 shows a map of England and a Wikipedia entry. SERP 6 shows a map of England and a Wikipedia entry. SERP 7 shows a map of England and a Wikipedia entry. SERP 8 shows a map of England and a Wikipedia entry. SERP 9 shows a map of England and a Wikipedia entry. SERP 10 shows a map of England and a Wikipedia entry. A red "KG" watermark is visible on several pages, indicating knowledge graph information.

Fonte: adaptado de Yahoo (2017b).

Ao comparar os resultados obtidos no Yahoo! com os mecanismos anteriores (Bing e Google) para o mesmo termo, observa-se uma queda na quantidade de correspondências com o verbete “England” da Wikipédia nas dez SERPs analisadas. Nesse sentido, apenas a SERP “1”, referente ao respectivo termo, apresentou em sua lista de resultados uma referência ao artigo da Wikipédia. Em algumas destas SERPs também constavam resultados para a Wikipédia, mas referentes aos outros verbetes.

Somente as SERPs “1”, “6” e “7” apresentaram o KG na composição de seus resultados, porém a única que retratou a temática relativa ao termo “England” foi a SERP “1”, derivada do termo exato. Na sequência, a estrutura do KG foi evidenciada por meio da Figura 40, na qual a primeira seção textual (resumo) correspondeu ao conteúdo extraído do verbete “England” da Wikipédia. Além disso, as informações fatuais, comumente apresentadas no formato de *hyperlinks* como *queries*, neste KG, constam apenas como informações textuais sobre a Inglaterra.

Figura 40 – Knowledge Graph do Yahoo! para “England”



Fonte: adaptado de Yahoo (2017b).

Em outra seção do *KG* foram identificados vínculos com ferramentas de redes sociais, como Wikipédia e “Tripadvisor”⁵⁶. Nesta última, o usuário é redirecionado para uma página da ferramenta, na qual existem melhores destinos, hotéis, restaurantes, entre outras recomendações com base em classificações sociais de seus membros. Há também a guia “*Related places*” que associa o termo pesquisado “England” com localizações semanticamente correspondentes. O último item do *KG* credita a origem da imagem utilizada, que compõe o topo de sua estrutura.

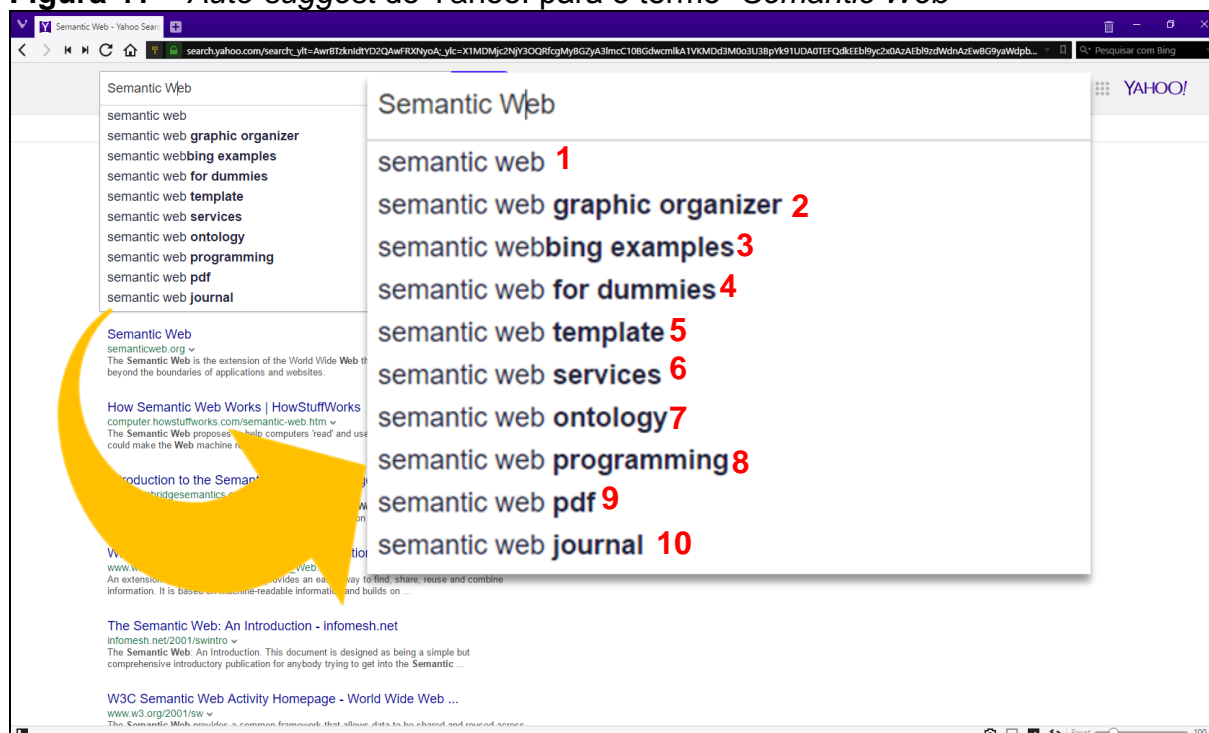
Na parte inferior da Figura 40, há um espaço destinado à exibição de anúncios relacionados ao contexto do termo pesquisado. Ao selecioná-lo, o usuário é redirecionado para a página do anunciante para que possa fazer reserva de vagas em hotéis. Essas funções vão ao encontro com o exposto por Sullivan (2012), a

⁵⁶ <https://www.tripadvisor.com.br/Tourism-g186217-England-Vacations.html>

respeito dos *KGs* serem destinados aos fatos e não às ações, pois se fossem, seria possível reservar o hotel diretamente da SERP do MB.

O último termo analisado neste MB foi “*Semantic Web*”. Após sua inserção na caixa de busca do Yahoo! foram indicadas dez sugestões de pesquisas associadas a este termo, conforme apresentado na Figura 41.

Figura 41 – *Auto-suggest* do Yahoo! para o termo “*Semantic Web*”



Fonte: adaptado de Yahoo (2017c).

As SERPs originadas das sugestões de busca da Figura 41 foram acessadas para uma análise comparativa entre o resultado principal “*Semantic Web*” e os outros nove desmembramentos, conforme exposto na Figura 42. As SERPs “1”, “3”, “8” e “9” foram as únicas que continham, em suas listas de resultados, *hyperlinks* para o respectivo verbete na Wikipédia.

Na SERP “1”, correspondente ao termo “*Semantic Web*”, o primeiro resultado apresentado refere-se, especificamente, ao verbete existente na Wikipédia. Porém, em nenhuma das SERPs analisadas houve a ocorrência de “*answer box*”, *KG* ou qualquer outro elemento que apontasse um uso diferente do conhecimento enciclopédico colaborativo para além da composição na lista de resultados.

Figura 42 – Mosaico de SERPs do Yahoo! para o termo “*Semantic Web*”



Fonte: adaptado de Yahoo (2017c).

Ao finalizar a investigação sobre o termo “*Semantic Web*” no Mecanismo de Busca Yahoo!, o ciclo de análise desta dissertação também foi encerrado. Dessa forma, os termos “Tim Berners-Lee”, “England” e “*Semantic Web*” tiveram sua transição monitorada, desde sua concepção como verbetes da Wikipédia até sua conversão em entidades, propriedades e valores em *Knowledge Bases* como DBpedia e Wikidata. Além da consequente aplicação em algumas SERPs dos mecanismos Bing, Google e Yahoo!.

No âmbito dos MBs, funções como “*auto-suggest*” e “*auto-complete*” auxiliam o usuário na formulação da requisição de pesquisa por meio da indicação de assuntos correlacionados ao termo principal. Ao analisar as pesquisas complementares indicadas por estas funções nos mecanismos Bing, Google e Yahoo! para os três termos, notou-se que poucas sugestões coincidem entre eles. Talvez, esta não correspondência evidencie a inexistência de *Knowledge Bases* para a execução dessas funções específicas. Por conseguinte, ainda que Schwartz (2015) aponte que Bing e Yahoo! utilizam o mesmo motor de busca, não foram encontradas similaridades de funcionamento entre esses mecanismos.

As SERPs derivadas das sugestões de pesquisa dos mecanismos também foram analisadas, ainda que não fossem peças centrais desta investigação. Observá-las auxiliou no processo de identificação de uso do conhecimento

enciclopédico colaborativo pelos MB. Nesse sentido, os mosaicos apresentados para cada termo investigado visaram traçar comparativamente divergências e semelhanças entre a SERP principal e as das sugestões de pesquisa indicadas pelos mecanismos.

A presença de elementos adicionais à comum lista de resultados apresentada pelas SERPs como “*answer box*”, “*People also ask*” e *Knowledge Graph*, se mostraram recursos importantes para a exibição do conhecimento enciclopédico colaborativo nos MBs. Entre esses componentes, o *KG* foi o que melhor retratou essa aplicação, visto que, na maioria dos casos, exibia no topo de sua estrutura informações textuais advindas dos verbetes da Wikipédia, além de acrescentar ao final as fontes consultadas para a confecção do *KG*. Apenas o termo “*Semantic Web*” não foi representado por um *KG* nos mecanismos Google e Yahoo!.

Entre os *KGs* analisados, os exibidos pelo Bing foram os que apresentaram estruturas mais robustas, referentes à quantidade e organização das informações por eles exibidas. O Google, por sua vez, manteve apenas informações essenciais, sendo diferente do Yahoo!, que expôs *KGs* minimalistas com poucas informações descritivas sobre as entidades investigadas.

O *KG* foi importante para evidenciar as influências do conhecimento enciclopédico colaborativo no desenvolvimento da *Web Pragmática*. Isso se deu pela utilização das informações advindas da Wikipédia e, para alguns casos, também das *Knowledge Bases*, haja vista que o *KG* possibilitou que recursos semânticos (informações da Wikipédia enriquecidas semanticamente pelas *KBs*) fossem colocados em contextos pragmáticos.

6 CONSIDERAÇÕES FINAIS

Esta dissertação buscou analisar o conhecimento enciclopédico colaborativo e suas influências no desenvolvimento da *Web Pragmática*. Aparentemente distintos, esses fenômenos puderam ser compreendidos a partir de incursões teóricas realizadas em outras áreas do conhecimento mediadas pela interdisciplinaridade inerente à Ciência da Informação.

Esses fundamentos teóricos abordaram desde a formação do significado na virtualidade do indivíduo até a assimilação do sentido por máquinas no Ciberespaço. Assim, com base no objetivo principal desta investigação, três termos semanticamente interligados (Tim Berners-Lee, England e *Semantic Web*) foram elencados para a execução de um procedimento de rastreio iniciado na Wikipédia, que perpassou pelas *Knowledge Bases*: DBpedia e Wikidata e foi finalizado nos mecanismos de busca Bing, Google e Yahoo!

A aplicação do procedimento de análise supracitado trouxe a possibilidade de inferir as influências do conhecimento enciclopédico colaborativo no desenvolvimento da *Web Pragmática*. Isso se deu a partir de uma perspectiva holística, na qual esse conhecimento criado na Wikipédia e estruturado nas *Knowledge Bases* era possivelmente utilizado pelos Mecanismos de Busca.

Nesse contexto, os três termos tiveram mais que seu percurso descrito na Wikipédia; foram submetidos a análises individuais em torno de sua construção. Ao monitorá-los, o *modus operandi* da Wikipédia, conseqüentemente, foi evidenciado. O acesso ao histórico de versionamento dos verbetes indicou que o conhecimento enciclopédico da Wikipédia pode ser taxado como colaborativo, visto que o desenvolvimento dos verbetes perpassa por diferentes usuários, que contribuem voluntariamente, ao longo de anos, com a formatação e a inserção de textos, além de também controlar todas essas modificações.

Apesar dos rumores que envolvem a confiabilidade acerca do conteúdo publicado nos artigos da Wikipédia, os verbetes analisados revelaram pertencer a um esquema minucioso de classificação de qualidade e representatividade dentro da comunidade *on-line*. Ao analisar a forma como o conteúdo dos verbetes é categorizado em seções temáticas ou em tabelas e *infoboxes* com informações pontuais, mostra-se também a preocupação da

comunidade em constantemente produzir e manter informações de maneira mais estruturada possível.

Além disso, a Wikipédia revelou possuir princípios constitutivos equivalentes aos que permeavam o desenvolvimento de projetos enciclopedistas historicamente reconhecidos. Entre eles, destaca-se a busca pela totalidade do conhecimento, a colaboração dos envolvidos e o fato de ser prospectiva, isto é, sempre será uma obra em aberto.

No âmago da DBpedia, os três verbetes eram simultaneamente interpretados como entidades, propriedades e/ou valores. Os procedimentos que envolvem a colaboração em torno do enriquecimento semântico dessas entidades não se mostraram receptivos a usuários desprovidos de conhecimento técnico sobre o funcionamento de tecnologias da *Web Semântica*. Iniciativas como o OpenLink Virtuoso e RelFinder revelaram-se como alternativas acessíveis para a compreensão do funcionamento da DBpedia. Dessa maneira, as informações semiestruturadas pela comunidade Wikipédia são extraídas por agentes computacionais ou manualmente e disponibilizadas no formato de declarações na DBpedia, fato que permite uma máquina interpretar não apenas a palavra, mas o significado que a circunda por meio de informações expressas pelas tecnologias da *Web Semântica*.

No Wikidata os verbetes são convertidos em itens, propriedades e/ou valores. Ao compará-lo com a DBpedia, o grande diferencial está na facilidade com a qual se pode colaborar para a estruturação de informações sobre as entidades, visto que o próprio pesquisador conseguiu deixar sua contribuição para a entidade sobre o termo "*Semantic Web*". A interface simples promove a compreensão de seu funcionamento sem a necessidade de acesso a ferramentas externas, provando ser uma opção viável de *Knowledge Base* colaborativa.

Os verbetes convertidos em entidades nas *Knowledge Bases* mais tarde se tornariam resultados nos Mecanismos de Busca. Desse modo, as SERPs de Bing, Google e Yahoo! foram exploradas no intuito de verificar se o conhecimento enciclopédico colaborativo era empregado na apresentação dos resultados. A investigação revelou o funcionamento singular de cada mecanismo para a indicação de requisições complementares e na apresentação e ordenação de seus resultados. Os termos designados a pessoa e local comumente apresentaram elementos complementares à lista de resultados nas SERPs, se comparados aos do conceito.

As influências do conhecimento enciclopédico colaborativo foram manifestadas no momento da análise dos Mecanismos de Busca. Assim, funções como “*auto-suggest*”, “*auto-complete*”, “*answer box*” e seções como “*People also search for*” e “*People also ask*” mostraram a preocupação dos mecanismos na apresentação de resultados sensíveis ao contexto com base em uma perspectiva mais humana. Nesse sentido, o *Knowledge Graph* também possui traços da *Web Pragmática* em seu funcionamento, visto que na maioria dos casos apresentou informações pontuais advindas da Wikipédia e de *Knowledge Bases* em sua estrutura, isto é, colocou recursos semânticos em contextos pragmáticos. Além disso, os *Knowledge Graphs* que continham a opção “*Feedback*” se tornaram um indicativo da prática da colaboração em prol de um conhecimento pautado no consenso.

Portanto, o conhecimento enciclopédico colaborativo advindo da Wikipédia e semanticamente estruturado por *Knowledge Bases* é utilizado pelos principais Mecanismos de Busca da atualidade no estabelecimento de uma *Web Pragmática*. Assim, suas influências são positivas para esse desenvolvimento, ao passo que seu uso por tecnologias como o *Knowledge Graph* viabiliza a instauração da *Web Pragmática* com base em resultados sensíveis ao contexto, apresentados por máquinas sobre uma concepção humana do significado.

Ao longo desta dissertação algumas limitações de pesquisa foram levantadas, uma delas, foi a inviabilidade técnica descrita na seção “4 Percurso Metodológico” que impossibilitou a análise das *Knowledge Bases* Yago e OpenCyc. Além disso, a segunda, ateuve-se ao fato dos Mecanismos de Busca serem proprietários, circunstância que permitiu realizar apenas inferências sobre seu funcionamento.

Para contribuições futuras, sugere-se a conversão destas limitações em possibilidades de novos estudos como a inclusão das *Knowledge Bases* que não puderam ser abordadas nesta investigação e também a análise de Mecanismos de Busca que sejam *open-source*, a exemplo, o YaCy⁵⁷.

⁵⁷ <http://yacy.net/en/index.html>

REFERÊNCIAS

- ABBENHAUS, C. Web 2.0. In: FORUM INFORMATIK 2006 IN DRESDEN - INFORMATIK FÜR MENSCHEN!, n. 6, 2006, Dresden. **Anais...** Dresden: Informatik-Spektrum, 2006. p. 447-471.
- AGHAEI, S.; NEMATBAKHSH, M. A.; FARSANI, H. K. Evolution of the world wide web: from web 1.0 to web 4.0. **International Journal of Web & Semantic Technology**, Isfahan, v. 3, n. 1, p. 1-10, jan. 2012.
- ALESSO, H. P.; SMITH, C. **Thinking on the web**: Berners-Lee, Gödel, and Turing. New Jersey: Wiley-Interscience, 2006. 260 p.
- ALLAN, K. Dictionaries and encyclopedias: relationship. **Elsevier**, Australia, v. 3, n. 1. p. 573-577, 2006.
- ANDERSON, J. J. **Wikipedia**: the company and its founders. Mineapolis: ABDO Group, 2011.
- ARENAS, M. et al. Faceted search over RDF-based knowledge graphs. **Web Semantics: Science, Services and Agents on the World Wide Web**, Koblenz, 2016.
- AYERS, P.; MATTHEWS, C.; YATES, B. **How Wikipedia works**. San Francisco: No Starch Press, 2008. 303 p.
- BAEZA-YATES; R.; RIBEIRO-NETO, B. **Modern information retrieval**. New York: ACM Press, 1999. 499 p.
- BARANAUSKAS, M.; MARTINS, M.; VALENTE, J. **Codesign de redes digitais: tecnologia e educação a serviço da inclusão social**. Porto Alegre: Penso, 2013.
- BARRETT, D. J. **MediaWiki**. Sebastopol: O'Reilly Media, 2009. 359 p.
- BAUER, M. W.; GASKELL, G. (Org.). **Pesquisa qualitativa com texto, imagem e som: um manual prático**. Petrópolis: Vozes, 2002.
- BAX, M. P. A evolução da *Web* rumo à *Web Semântica*. **PRISMA.COM**, Porto, n. 19, p. 1-27, 2012.
- BERNERS-LEE, T. **Information management**: a proposal. 1989.
- _____. **The World Wide Web**: past, presente and future. 1996.

BERNERS-LEE, T.; FIELDING, R. T.; MASINTER, L. **Uniform Resource Identifier (URI): generic syntax**. Washington: The Internet Society, 2005.

BERNERS-LEE, T.; FISCHETTI, M. **Weaving the Web: the original design and ultimate destiny of the World Wide Web by its inventor**. New York: HarperCollins, 2000.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, maio. 2001.

BIEGA, J.; KUZEY, E.; SUCHANEK, F. Inside YAGO2s: a transparent information extraction architecture. In: PROCEEDINGS IN THE DEMO FALL, n.1, 2013, Silicon Valley. **Anais...Silicon Valley, DEMO FALL**, 2013.

BING. **Tim Berners-Lee**. Disponível em: <<http://www.bing.com/search?q=Tim+Berners-Lee&qs=n&form=QBLH&sp=-1&pq=tim+berners-lee&sc=8-15&sk=&cvid=BE9980832F974B3B9A22CEAF2DB7D68E>>. Acesso em: 24 mar. 2017a.

BING. **England**. Disponível em: <<http://www.bing.com/search?q=England&qs=n&form=QBRE&sp=-1&pq=england&sc=8-15&sk=&cvid=6B19298279414B6CA6491050B007F3AA>>. Acesso em: 24 mar. 2017b.

BING. **Semantic web**. Disponível em: <<http://www.bing.com/search?q=Semantic+Web&qs=n&form=QBRE&sp=-1&pq=semantic+web&sc=4-12&sk=&cvid=1A23A1E53DBF43079F0282EA0194B7AE>>. Acesso em: 24 mar. 2017c.

BIZER, C. et al. DBpedia - a crystallization point for the Web of data. **Elsevier**, Berlim, 2009.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data – the story so far. In: HEATH, T.; HEPP, M.; BIZER, C. (eds.). **International Journal on Semantic Web and Information Systems (IJSWIS)**, Special Issue on Linked Data, 2009.

BLÜHDORN, H. A relação entre pragmática, semântica e gramática. **Rev. Est. Ling.**, v. 6, n. 2, p. 5-43, jul./dez. 1997.

BOSAK, J.; BRAY, T. XML and the second-generation Web: the combination of hypertext and a global Internet started a revolution, a new ingredient, XML, is poised to finish the job. **Scientific American**, mayo 1999.

BOULOS, M. N. K.; WHEELERT, S. The emerging web 2.0 social software: an enabling suite of sociable Technologies in health and health care education. **Health Information and Libraries Journal**, Richmond, v. 24, n. 1, p. 2-23, mar. 2007.

BOT. In: Wikipedia: the free encyclopedia. 2017. Disponível em: <<https://en.wikipedia.org/wiki/Wikipedia:Bots>>. Acesso em: 17 mar. 2017.

BUTZBACH, A. **6 ways get content googles knowledge graph results**. 2017. Disponível em: <<https://www.brafton.com/blog/6-ways-get-content-googles-knowledge-graph-results/>>. Acesso em: 15 mar. 2017.

BREITMAN, K. **Web semântica: a internet do futuro**. Rio de Janeiro: LTC, 2010.

CAVAZZA, M.; ZWEIGENBAUM, P. Lexical semantics: dictionary or encyclopedia. In: SAINTDIZIER, P.; VIEGAS, E. (Eds.). **Computational lexical semantics, studies in natural language processing**. Cambridge: Cambridge University Press, 1995. p. 336-347.

CHATFIELD, T. B. **The complete guide to wikis: how to setup, use, and benefit from wikis for teachers, business, professionals, families, and friends**. Florida: Atlantic, 2009.

CHOUDHURY, N. World Wide Web and its journey from web 1.0 to web 4.0. **International Journal of Computer Science and Information Technologies**, v. 5, n. 6, 2014.

CONNAWAY, L. S.; POWELL, R. R. **Basic research methods for librarians**. 5. ed. Santa Barbara: Libraries Unlimited, 2010. 370 p.

CORMODE, G.; KRISHNAMURTHY, B. Key differences between web 1.0 and web 2.0. **First Monday**, Miami, v. 13, n. 6, jun. 2008.

CROFT, W.; CRUSE, D. A. **Cognitive linguistics**. New York: Cambridge University Press, 2004.

CROFT, W.; METZLER, D.; STROHMAN, T. **Search engines: information retrieval in practice**. New Jersey: Pearson Education, 2015.

DBPEDIA. **DBpedia community**. Disponível em: <<http://wiki.dbpedia.org/about/dbpedia-community>>. Acesso em: 05 abr. 2016.
DBPEDIA. **Tim Berners-Lee**. Disponível em: <http://dbpedia.org/page/Tim_Berners-Lee>. Acesso em: 27 mar. 2017a.

DBPEDIA. **England**. Disponível em: <<http://dbpedia.org/page/England>>. Acesso em: 27 mar. 2017b.

DBPEDIA. **Semantic web**. Disponível em: <http://dbpedia.org/page/Semantic_Web>. Acesso em: 27 mar. 2017c.

DOMINGUE, J.; FENSEL, D.; HENDLER, J. **Handbook of semantic web technologies**. Berlin: Springer, 2011.

ECO, U. **From the tree to the labyrinth: historical studies on the sign and interpretation**. Cambridge: Harvard University Press, 2014.

FÄRBER, M.; ELL, B.; MENNE, C.; RETTINGER, A. A comparative survey of DBpedia, Freebase, Opencyc, Wikidata, and YAGO. **IOS Press**, v. 1, n. 5, p. 1-26, 2015.

FACHIN, G. Recuperação inteligente da informação e ontologies: um levantamento na área de Ciência da Informação. **Biblos**, Rio Grande, v. 23, n. 1, p. 259-283, 2009.

FOAF. In: Wikipedia: the free encyclopedia. 2017. Disponível em: <[https://en.wikipedia.org/wiki/FOAF_\(ontology\)](https://en.wikipedia.org/wiki/FOAF_(ontology))>. Acesso em: 23 mar. 2017.

FUCHS, C. et al. Theoretical foundations of the Web: cognition, communication, and co-operation. towards an understanding of Web 1.0, 2.0, 3.0. **Future Internet**, Basel, v. 2, n.1, p.41-59, fev. 2010.

GEERAERTS, D. **Cognitive linguistics: basic readings**. Berlin: Mouton de Gruyter, 2006.

GOOGLE. **Tim Berners-Lee**. Disponível em: <https://www.google.com.br/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=Tim+Berners-Lee&*>>. Acesso em: 25 mar. 2017a.

GOOGLE. **England**. Disponível em: <https://www.google.com.br/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=England&*>>. Acesso em: 25 mar. 2017b.

GOOGLE. **Semantic Web**. Disponível em: <https://www.google.com.br/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=Semantic+Web&*>>. Acesso em: 25 mar. 2017c.
GRIMMELMANN, J. The Structure of Search Engine Law. **Law 93 Iowa L. Rev.:** [s. l.], n. 1, 2007.

HAIMAN, J. Dictionaries and encyclopedias. **Lingua**: [s. l.], v. 50, n. 4, p.329-357, 1980.

HEATH, T.; BIZER, C. **Linked data**: evolving the web into a global data space. San Rafael: Morgan & Claypool, 2011. 136 p.

HOLEWA, H. New paradigms: a collaborative web-based research tool. In: TATNALL, A. (Org.). **Web technologies**: concepts, methodologies, tools, and applications. New York: Information Science Reference, 2010. p. 670-680.

HO KIM, J. Cibernética, ciborgue e ciberespaço: notas sobre as origens da cibernética e sua reinvenção cultural. **Horizontes Antropológicos**, Porto Alegre, v. 10, n. 21, p. 199-219, jan./jun., 2004.

ILARI, R. Semântica e pragmática: duas formas de descrever e explicar fenômenos da significação. **Rev. Est. Ling.**, Belo Horizonte, v. 9, n. 1, p. 109-162, jan./jun. 2000.

INFOBOX. In: Wikipedia: the free encyclopedia. 2017. Disponível em: <<https://en.wikipedia.org/wiki/Infobox>>. Acesso em: 17 mar. 2017.

IRI. In: **Wikipédia: a enciclopédia livre**. Disponível em: <https://en.wikipedia.org/wiki/Internationalized_Resource_Identifier>. Acesso em: 12 mar. 2017.

JEMIELNIAK, D. **Common knowledge?: an ethnography of Wikipedia**. Stanford: Stanford University Press, 2014. 293 p.

JENTZSCH, A. et al. **DBpedia Mapping Language**. 2010. Disponível em: <http://jens-lehmann.org/files/2012/program_el_dbpedia_live.pdf>. Acesso em: 22 mar. 2017.

KHARE, R.; RIFKIN, A. XML: a door to automated web applications. **IEEE Internet Computing**, Los Angeles, v.1, n. 4, jul./ago. 1997.

KIEFER, F. Linguistic, conceptual and encyclopedic knowledge: some implications for lexicography. In: PROCEEDINGS OF THE 3RD EURALEX INTERNATIONAL CONGRESS, n. 3, 1988, Budapest. **Anais...** Budapest: Akadémiai Kiadó, 1988. p. 1-10.

KNIGHTS, M. Web 2.0. **Communications Engineer**, Stevenage, v. 5, n. 1, fev./mar. 2007.

KOCH, I.; ELIAS, V. **Ler e compreender**: os sentidos do texto. São Paulo: Contexto, 2008.

KONTOKOSTAS, D. et al. Internationalization of Linked Data: The case of the Greek DBpedia edition. **Web Semantics**, v. 15, n. 1, jan. 2012.

LANDIS, C. A web of meaning: linked open data resources on the web. **C&RL News**, Middletown, v. 1, n. 1, out. 2014.

LAVRENKO, V. **A generative theory of relevance**. Berlin: Springer, 2009.

LAZZARIN, F.; AZEVEDO NETTO, C.; SOUSA, M. Informação, memória e ciberespaço: considerações preliminares no campo da Ciência da Informação no Brasil. **Transinformação**, Campinas, n. 27, v.1, p. 21-30, jan./abr., 2015.

LEDFORD, J. L. **SEO: Search Engine Optimization – Bible**. Indianapolis: Wiley Publishing Inc, 2009. 389 p.

LEGG, C. Ontologies on the Semantic Web. **Annual Review of Information Science and Technology**, [S. l.], v. 41, n. 1, jan. 2008.

LEHMANN, J. et al. DBpedia: a large-scale, multilingual knowledge base extracted from Wikipedia. **Semantic Web Journal**, v. 6, n. 2, p. 167-195, 2015.

LEUF, B.; CUNNINGHAM, W. **The wiki way: quick collaboration on the web**. New Jersey: Addison-Wesley, 2001.

LÉVY, P. **Cibercultura**. Editora 34, 1999.

LEWANDOWSKI, D. New perspectives on web search engine research. In: _____. **Web search engine research**. 4. ed. Bingley: Emerald Group Publishing Limited, 2012. p. 1-16.

LIH, A. **The Wikipedia revolution: how a bunch of nobodies created the world's greatest encyclopedia**. New York: Hachette Books, 2009.

MAHDISOLTANI, F.; BIEGA, J.; SUCHANEK, F. YAGO3: a Knowledge base from multilingual wikipedias. In: PROCEEDINGS OF THE 7th CONFERENCE ON INNOVATIVE DATA SYSTEMS RESEARCH, n. 7, 2015, Asilomar. **Anais...** Asilomar: CIDR, 2015.

MARCONDES, C. Linguagem e documento: fundamentos evolutivos e culturais da Ciência da Informação. **Perspectivas em Ciência da Informação**, v. 15, n. 2, p. 2-21, maio/ago. 2010.

MEIJS, W.; VOSSEN, P. In: so many words: knowledge as a lexical phenomenon. In: PROCEEDINGS OF THE FIRST SIGLEX WORKSHOP ON LEXICAL SEMANTICS

AND KNOWLEDGE REPRESENTATION, n. 1, 1991, London. **Anais...** London: Springer, 1991. p. 137-153.

MIKA, P. **Social networks and the semantic Web**. New York: Springer, 2007.

MIHALCEA, R.; CSOMAI, A. Wikify!: linking documents to encyclopedic knowledge. In: CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, n. 16, 2007, Lisbon. **Anais...** New York: ACM, 2007. p. 233-241.

MILLER, M. **Sams teach yourself Wikipedia in 10 minutes**. New Jersey: Pearson Education, 2010. 209 p.

MONTEIRO, S. D. O ciberespaço e os mecanismos de busca: novas máquinas semióticas. **Ci. Inf.**, Brasília, v. 35, n. 1, p. 31-38, jan./abr. 2006.

_____. A dobra semiótica e os agenciamentos maquínicos: por uma ontologia das tecnologias de informação e comunicação. In: CERVANTES, B. M. N. (Org.). **Horizontes da organização da informação e do conhecimento**. Londrina: EDUEL, 2012. p. 63-96.

_____. O ciberespaço: o termo, a definição e o conceito. **DataGramZero**, Rio de Janeiro, v. 8, n. 3, jun. 2007.

_____. Knowledge Graph e a significação: novos agenciamentos semióticos dos índices contemporâneos. In: XVI ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, n. 16, 2015, João Pessoa. **Anais...** João Pessoa: XVI ENANCIB, 2015.

MONTEIRO, S. D.; ABREU, J. O pós-moderno e a organização do conhecimento no ciberespaço: agenciamentos maquínicos. **DataGramZero**, Rio de Janeiro, v. 10, n. 6, dez. 2009.

MONTEIRO, S. D.; FIDENCIO, M. As dobras semióticas do ciberespaço: da web visível à invisível. **Transinformação**, Campinas, v. 25, n.1, p. 35-46, jan./abr., 2013.

MOOR, A. de; Patterns of the pragmatic web. In: PROCEEDINGS OF THE 13th INTERNATIONAL CONFERENCE ON CONCEPTUAL STRUCTURES, n. 13, Kassel. **Anais...** Berlin: Springer Verlag, 2005.

MORAES, M. A forma e o infinito, de Diderot a Baudelaire. **Teresa Revista de Literatura Brasileira**, São Paulo, n. 12/13, 2013.

MORRIS, C. **Fundamentos da teoria dos signos**. Eldorado: São Paulo, 1976.

MORSEY, M. et al. DBpedia an the live extraction of structured data from Wikipedia. 2012.

MURPHY, M. L. **Semantic relations and the lexicon**: antonymy, synonymy, and other paradigms. Cambridge University Press, Cambridge, 2003.

MURUGESAN, S. Understanding Web 2.0. **IEEE Computer Society**, Washington D.C., p. 34-41, jun./ago. 2007

NETMARKETSHARE. **Market share reports**. 2016. Disponível em: <<https://netmarketshare.com/>>. Acesso em: 30 out. 2016.

FAZENDA. **NF-e**. 2017. Disponível em: <<http://www.nfe.fazenda.gov.br/portal/principal.aspx>>. Acesso em: 16 mar. 2017.

OLIVEIRA, G.; ARAÚJO, W. Usar ou não usar – qual a relevância das meta tags na recuperação da informação pelos mecanismos de busca? **Biblionline**, João Pessoa, v. 8, n. 1, p. 60-77, 2012.

OLIVEIRA, R. P. de. Semântica. In: MUSSALIN, F.; BENTES, A. C. (Orgs.). **Introdução à linguística**: domínios e fronteiras. São Paulo: Cortez, 2006. p. 17-46.

ORLANDI, E. P. **O que é linguística**. São Paulo: Brasiliense, 2009.

PASCHKE, A. et al. Rule responder: ruleML-based agents for distributed collaboration on the pragmatic web. In: PROCEEDINGS INTERNATIONAL CONFERENCE ON THE PRAGMATIC WEB, n. 2, Tilburg. **Anais...** Tilburg: ACM, 2007.

PERINI, M. Sobre língua, linguagem e linguística: uma entrevista com Mário A Perini. **REVEL**, v. 8, n. 14, 2010.

PIETARINEN, Ahti-Veikko. The semantic + pragmatic web = semiotic web. In: PROCEEDINGS OF INTERNATIONAL CONFERENCE WWW/INTERNET, n. 1, Las Vegas. **Anais...** Las Vegas: DBLP, 2003.

PINTO, J. Pragmática. In: MUSSALIN, F.; BENTES, A. C. (Orgs.). **Introdução à linguística**: domínios e fronteiras. São Paulo: Cortez, 2006.

POMBO, O. O hipertexto como limite da ideia de enciclopédia. In: POMBO, O.; GUERREIRO, A.; ALEXANDRE, A. F. **Enciclopédia e Hipertexto**. Lisboa, Edições Duarte Reis, 2006.

_____. Leibniz and the encyclopaedic project. In: ACHEVERRIA, A; ROLDAN, J (eds.). ACTAS DEL CONGRESO INTERNACIONAL "CIENCIA, TECNOLOGÍA Y

BIEN COMÚN: LA ACTUALIDAD DE LEIBNIZ”, nº 2, Valencia. **Anais...** Valencia: Univ. Polytechnic de Valencia, 2002.

PRAGMATISMO. In: Wikipedia: a enciclopédia livre. Disponível em: <<https://pt.wikipedia.org/wiki/Pragmatismo>>. Acesso em: 16 mar. 2017.

QIAN, R. Understand Your World with Bing. 2013. Disponível em: <<http://blogs.bing.com/search/2013/03/21/understand-your-world-with-bing/>>. Acesso em: 16 mar. 2017

RELFINDER. **Interactive relationship discovery in RDF data**. Disponível em: <<http://www.visualdataweb.org/relfinder/relfinder.php>>. Acesso em: 27 mar. 2017.

ROSCOCHER, M. et al. Building event-centric knowledge graphs from news. **Web Semantics: Science, Services and Agents on the World Wide Web**, Koblenz, 2016.

SANTAELLA, L. A tecnocultura atual e suas tendências futuras. **Signo y Pensamiento**, Bogotá, v. 30, n. 60, p. 30-43, jan./jun. 2012.

_____. **Navegar no ciberespaço: o perfil cognitivo do leitor imersivo**. São Paulo, Paulus, 2004.

SANTAREM SEGUNDO, J.; SOUZA, J.; CONEGLIAN, C. Web semântica: introdução a recursos de visualização de dados em formato gráfico. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015, João Pessoa. **Anais...** João Pessoa: XVI ENANCIB, 2015.

SANTOS, P.; VIDOTTI, S. Perspectivismo e tecnologias de informação e comunicação: acréscimos à Ciência da Informação? **DataGramZero**, Rio de Janeiro, v. 10, n. 3, jun. 2009.

SÁ RAMALHO, R.; VIDOTTI, S.; FUJITA, M. Web semântica: uma investigação sob o olhar da Ciência da Informação. **DataGramZero**, Rio de Janeiro, v. 8, n.6, dez. 2007.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspec. Ci. Inf.**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

_____. Relevance: a review of and a framework for the thinking on the notion in Information Science. **Journal of the American Society for Information Science**, [s. l.], p. 321-343, nov./dez. 1975.

_____. Research on relevance in Information Science: a historical perspective. In: CARBO, T.; HAHN, T. B (Ed.). *International Perspectives on the History of Information Science and Technology. PROCEEDINGS OF THE ASIS&T – PRE-CONFERENCE ON THE HISTORY OF ASIS&T AND INFORMATION SCIENCE AND TECHNOLOGY*, n. 75, 2012, Medford. **Anais...** Medford: ASIS&T, out. 2012.

_____. Why is relevance still the basic notion in Information Science? In: PEHAR, F.; SCHÖGL, C.; WOLFF, C. (Ed.). *Re:inventing Information Science in the Networked Society. PROCEEDINGS OF THE 14th INTERNATIONAL SYMPOSIUM ON INFORMATION SCIENCE*, n. 14, 2015, Zadar. **Anais...** Zadar: ISI, 2015.

SCHOOP, M.; MOOR, A.; DIETZ, J. L. G. The pragmatic web: a manifesto. **Communications of the ACH**, v. 49, n. 5, p. 75-76, maio 2006.

SERRA, P. Informação e sentido: notas para uma abordagem problemática do conceito de informação. **BOCC-LABCOM**, Lisboa, p. 1-15, 1999.

SHENOY, A.; PRABHU, A. **Introducing SEO**: your quick-start guide to effective SEO practices. New York: Apress, 2016.

SINGH, M. The pragmatic web. **IEEE Internet Computing**, Los Angeles, p. 4-5, may./june, 2002a.

_____. The pragmatic web: preliminar thoughts. In: *PROCEEDINGS OF THE NSF-ONTOWEB WORKSHOP ON DATABASE AND INFORMATION SYSTEMS RESEARCH FOR SEMANTIC WEB AND ENTERPRISES*, n. 3, 2002, Dawsonville. **Anais...** Dawsonville: NSF-ONTOWEB, 2002b.

SILVA, S. da. Reflexões sobre web 1.0, web 2.0 e web semântica. **Sinergia**, São Paulo, v. 11, n. 2, p. 129-135, jul./dez. 2010.

SILVA, R.; SANTOS, P.; FERNEDA, E. Modelos de recuperação de informação e *Web semântica: a questão da relevância*. **Informação & Informação**, Londrina, v. 18, n. 3, p. 27-44, set./dez. 2013.

SIMBIOSE. In: Wikipédia: a enciclopédia livre. 2017. Disponível em: <<https://pt.wikipedia.org/wiki/Simbiose>>. Acesso em: 16 mar. 2017.

SINGHAL, A. **Introducing the Knowledge Graph**: things, not strings. 2012. Disponível em: <<https://googleblog.blogspot.com.br/2012/05/introducing-knowledge-graph-things-not.html>>. Acesso em: 16 mar. 2017.

SHERMAN, C.; PRICE, G. **The invisible web**: uncovering information sources search engines can't see. New Jersey: Information Today Inc., 2001.

STERCKX, L. et al. Knowledge base population using semantic label propagation. **Elsevier Science Publishers B. V.**, Amsterdam, v. 108, n. 1, p. 1-13, set. 2016.

SUCHANEK, F. The YAGO knowledge base. In: PROCEEDINGS OF PARIS BIG DATA, n. 1, 2016, Paris. **Anais...** Paris: TÉLÉCOM PARISTECH, 2016.

SULLIVAN, D. **Google launches knowledge graph to provide answers, not just links.** 2012. Disponível em: <<http://searchengineland.com/google-launches-knowledge-graph-121585>>. Acesso em: 15 mar. 2017.

SULLIVAN, D. **How Google instant's autocomplete suggestions work.** 2011. Disponível em: <<http://searchengineland.com/how-google-instant-autocomplete-suggestions-work-62592>>. Acesso em: 17 mar. 2017.

SCHWARTZ, B. Yahoo Has Their Own Knowledge Graph, Not Without Their Own Embarrassing Issues. 2015. Disponível em: <<http://searchengineland.com/yahoo-has-their-own-knowledge-graph-not-without-their-own-embarrassing-issues-218538>>. Acesso em: 17 mar. 2017.

TANON, T. et al. From Freebase to Wikidata: the great migration. PROCEEDINGS THE INTERNATIONAL WORLD WIDE WEB CONFERENCE COMMITTEE (IW3C2), n. 3, 2016, Montreal. **Anais...** Montreal: IW3C2, 2016.

VIOTTI, E. **Introdução aos estudos linguísticos.** Florianópolis: UFSC, 2008.

VRANDECIC, D.; KRÖTZSCH, M. Wikidata: a free collaborative knowledgebase. **COMMUNICATIONS OF THE ACM**, New York, v. 57, n. 10, out. 2014.

WANG, Z. et al. Knowledge graph embedding by translating on hyperplanes. **AAAI**, Palo Alto, 2014.

WEB FOUNDATION. **History of the Web.** 2017. Disponível em: <<http://webfoundation.org/about/vision/history-of-the-web/>>. Acesso em: 07 mar. 2017.

WEIGAND, H.; PASCHKE, A. The pragmatic web putting rules in context. In: BIKAKIS, A.; GIURCA, A. (eds.). Rules on the web: research and applications. PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM RULEML, n. 6, 2012, Montpellier. **Anais...** Montpellier: Springer-Verlag Berlin Heidelberg, 2012

WIKIDATA. **Main page.** Disponível em: <https://www.wikidata.org/wiki/Wikidata:Main_Page>. Acesso em: 24 mar. 2017a.

WIKIDATA. **Tim Berners-Lee.** Disponível em: <<https://www.wikidata.org/wiki/Q80>>. Acesso em: 24 mar. 2017b.

WIKIDATA. **England**. Disponível em: <<https://www.wikidata.org/wiki/Q21>>. Acesso em: 24 mar. 2017c.

WIKIDATA. **Semantic web**. Disponível em: <<https://www.wikidata.org/wiki/Q54837>>. Acesso em: 24 mar. 2017d.

WIKIPEDIA. **Main page**. Disponível em: <https://en.wikipedia.org/wiki/Main_Page>. Acesso em: 05 jan. 2017a.

WIKIPEDIA. **Tim Berners-Lee**. Disponível em: <https://en.wikipedia.org/wiki/Tim_Berners-Lee>. Acesso em: 27 mar. 2017b.

WIKIPEDIA. **Vital articles**. Disponível em: <https://en.wikipedia.org/wiki/Wikipedia:Vital_articles>. Acesso em 27 mar. 2017c.

WIKIPEDIA. **England**. Disponível em: <<https://en.wikipedia.org/wiki/England>>. Acesso em: 27 mar. 2017d.

WIKIPEDIA. **Semantic web**. Disponível em: <https://en.wikipedia.org/wiki/Semantic_Web>. Acesso em: 27 mar. 2017e.

WIKIPROJECT INTERNET. In: Wikipedia: the free encyclopedia. 2017. Disponível em: <https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Internet>. Acesso em: 17 mar. 2017.

WITTER, G. Pesquisa bibliográfica, pesquisa documental e busca de informação. **Estudos de Psicologia**, v. 7, n. 1, p. 5-30, jan./jul. 1990.

WOODS, D.; THOENY, P. **Wikis for dummies**. Indianapolis: Willey Publishing, 2007.

W3C. **RIF Primer**. 2. ed. 2012. Disponível em: <<https://www.w3.org/2005/rules/wiki/Primer>>. Acesso em: 12 mar.2017.

W3SCHOOLS. **XML**: introduction. Disponível em: <<http://www.w3schools.com/xml/default.asp>>. Acesso em: 18 fev. 2016.

W3C. **Layer Cake**. 2013a. Disponível em: <<https://www.w3.org/2007/03/layerCake.png>>. Acesso em: 12 mar. 2017.

W3C. **RIF FAQ**. 2013b. Disponível em: <https://www.w3.org/2005/rules/wiki/index.php?title=RIF_FAQ&action=history>. Acesso em: 12 mar. 2013.

W3C. **SPARQL**. 2013c. Disponível em: <<https://www.w3.org/TR/rdf-sparql-query/>>. Acesso em: 13 mar. 2013.

W3C. **Semantic Web**. 2017a. Disponível em: <<https://www.w3.org/standards/semanticweb/>>. Acesso em: 12 mar. 2017.

W3C. **Linked Data**. 2017b. Disponível em: <<https://www.w3.org/standards/semanticweb/data>>. Acesso em: 13 mar. 2017.

YAHOO. **Tim Berners-Lee**. Disponível em: <https://search.yahoo.com/search;_ylc=X3oDMTFiN25laTRvBF9TAzlwMjM1MzgwNzUEaXRjAzEEc2VjA3NyY2hfcWEEc2xrA3NyY2h3ZWI-?p=Tim+berners-Lee&fr=yfp-t&fp=1&toggle=1&cop=mss&ei=UTF-8>. Acesso em: 24 mar. 2017a.

YAHOO. **England**. Disponível em: <https://search.yahoo.com/search;_ylt=A0LEVwhVX.RYSc0AbaFXNyoA;_ylc=X1MDMjc2NjY3OQRfcgMyBGZyA3lmcC10BGdwcmlkA1RxaldyUU56UUM2c081VXBWeEhoM0EEbl9yc2x0AzAEbl9zdWdnAzkEb3JpZ2luA3NIYXJjaC55YWhvby5jb20EcG9zAzAEcHFzdHIDBHBxc3RybAMwBHFzdHJsAzEcXVlcnkDRW5nbGFuZAR0X3N0bXADMtQ5MTM2MTcxNw--?p=England&fr2=sb-top&fr=yfp-t&fp=1>. Acesso em: 24 mar. 2017b.

YAHOO. **Semantic web**. Disponível em: <https://search.yahoo.com/search;_ylt=A0LEV0.0X.RYWf8A82IXNyoA;_ylc=X1MDMjc2NjY3OQRfcgMyBGZyA3lmcC10BGdwcmlkA0pEWnk2TU53UktpYUo2dXRwVnplIUUEEbl9yc2x0AzAEbl9zdWdnAzEwBG9yaWdpbgNzZWYy2gueWFob28uY29tBHBvcwMwBHBxc3RyAwRwcXN0cmwDMARxc3RybAMxNARxdWVyeQNTZW1hbnRpYyUyMFdIYgR0X3N0bXADMtQ5MTM2MTczNQ--?p=Semantic+Web&fr2=sb-top&fr=yfp-t&fp=1>. Acesso em: 24 mar. 2017c.

YU, L. **Introduction to the semantic web and semantic web services**. New York: Chapman & Hall/CRC, 2007.