

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT-8 – Informação e Tecnologia

KNOWLEDGE GRAPH: O ELO ENTRE O CONHECIMENTO ENCICLOPÉDICO COLABORATIVO E A WEB PRAGMÁTICA

Gian Carlo Decarli (Universidade Estadual de Londrina – UEL)

Silvana Drumond Monteiro (Universidade Estadual de Londrina – UEL)

KNOWLEDGE GRAPH: THE ELO BETWEEN COLLABORATIVE ENCYCLOPEDIA KNOWLEDGE AND THE PRAGMATIC WEB

Modalidade da Apresentação: Comunicação Oral

Resumo: A Wikipédia é uma enciclopédia *on-line*, colaborativa, multilíngue e de licença livre. Seu *modus operandi* retrata a essência da *Web Social*, dimensão que pôs fim ao modelo linear de produção e consumo de informações na rede, legado da *Web 1.0*. Entretanto, a descentralização e a liberdade produtiva de conteúdo resultaram em problemas de recuperação de dados nesse ambiente. A *Web Semântica*, sua sucessora, emergiu com o propósito de estruturar dados para que fizessem sentido para humanos e agentes computacionais. Nessa dimensão, tecnologias como as *Knowledge Bases* ligam dados correlacionados de diferentes páginas *Web* com o objetivo de gerar sentido/significado sobre determinadas temáticas. A *Web Pragmática* agrega o legado estrutural das dimensões anteriores para explorar o significado a partir da percepção do usuário em múltiplos contextos. Além disso, tecnologias como o *Knowledge Graph* auxiliam os mecanismos a melhor estruturar e exibir resultados em suas *Search Engine Results Pages*. Ao observar essa conjuntura tecnológica, esta pesquisa objetivou compreender o papel do *Knowledge Graph* em relação ao conhecimento enciclopédico colaborativo da Wikipédia e a *Web Pragmática* por intermédio do Mecanismo de Busca Google. De natureza básica, a pesquisa adotou um delineamento qualitativo para realizar o acompanhamento do termo elencado “Tim Berners-Lee”, escolhido por sua representatividade em relação ao tema pesquisado. O percurso metodológico analisou esse termo sob a condição de verbete da Wikipédia, seguida de sua transformação em entidade no Wikidata e posterior conversão em *query* no Mecanismo de Busca Google. Os resultados apontaram que o *Knowledge Graph* é um agente de ligação do conhecimento enciclopédico colaborativo da Wikipédia com a concepção de uma *Web Pragmática*, mediada pelo Mecanismo de Busca Google.

Palavras-Chave: Wikipédia; *Knowledge Graph*; *Web Pragmática*; *Knowledge Base*; Mecanismos de busca.

Abstract: Wikipedia is an online encyclopedia, multilingual, collaborative and free license. Their *modus operandi* portrays the essence of the Social Web, which put an end to the linear model of production and consumption of information on the net, legacy of Web 1.0. However, decentralization and freedom of content production have resulted in data recovery problems in this environment. The Semantic Web, your successor, emerged with the purpose of structuring data so that make sense for humans and computer agents. In this dimension, technologies such as Knowledge Bases connect data correlated from different Web pages with the goal of generating meaning/significance on certain topics. The Pragmatic Web aggregates the legacy structure of previous dimensions to explore the meaning from the perception of the user in multiple contexts. In addition, technologies such as Knowledge Graph assist mechanisms to better structure and display results in their Search Engine Results Pages. When looking at this technological climate, this research aimed at understanding the role of Knowledge Graph in relation to encyclopedic knowledge from Wikipedia collaborative and Pragmatic Web through search engine Google. Of basic nature, the research adopted a qualitative design to follow the listed term "Tim Berners-Lee", chosen by your representation on the subject researched. The methodological path analyzed this term under the condition of entry in the Wikipedia, followed by your transformation into entity in the Wikidata and subsequent conversion in query in the Google Search Engine. The results showed that the Knowledge Graph is a liaison officer of the encyclopedic knowledge from Wikipedia with collaborative design of a Pragmatic Web, mediated by the Google Search Engine.

Keywords: Wikipedia; Knowledge Graph; Pragmatic Web; Knowledge Base; Search engines.

1 INTRODUÇÃO

Desde seu surgimento, readequações estruturais marcaram o desenvolvimento da *Web*, porém seu propósito inicial de interoperabilidade transpassou por esses diferentes períodos que, são identificados como: *Web* Sintática, *Web* Semântica e *Web* Pragmática (VECHIATO; VIDOTTI, 2014; MONTEIRO, 2012).

A *Web* Sintática abrange as duas primeiras camadas¹ da *World Wide Web* (*WWW* ou *Web*); são elas: *Web* 1.0 e *Web* 2.0 ou Social. A terminologia “Sintática” advém da forma como a informação é indexada e recuperada, por considerar apenas a estrutura das palavras e não o significado a elas agregado (MONTEIRO, 2015; SANTAELLA, 2012). 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 (GÁRCIA ARETIO, 2014).

Originada em 2001, a Wikipédia representa o propósito da *Web* Social. Em seu *website* pessoas podem colaborar para a 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 (MILLER, 2010). Assim, o conhecimento de mundo de cada indivíduo corrobora para tornar a Wikipédia uma base do conhecimento enciclopédico colaborativo na *Web*.

A *Web* Semântica não se atenta tanto ao modo como palavras, frases e expressões são formadas, mas vislumbra o significado por elas produzido para que seja passível de interpretação por humanos e agentes computacionais (MONTEIRO, 2015; SANTAELLA, 2012). Nesse sentido, as chamadas *Knowledge Bases* (*KB*), são exemplos de iniciativas correspondentes à dimensão da *Web* Semântica, pois estruturam em seu âmago o conhecimento disperso na rede, mediante conexões semânticas entre entidades (UYAR; ALIYU, 2015).

A *Web* Pragmática, por sua vez, busca agregar o legado estrutural de suas antecessoras com o propósito de explorar o significado segundo a perspectiva do usuário, ou seja, considera

¹ Nesta investigação, o termo “camada(s)” foi empregado para estabelecer nuances entre a *Web* 1.0 e 2.0, porém, entende-se que estas “camadas” não possuem fronteiras rigidamente delimitadas dentro da dimensão Sintática, sendo, por vezes, abstratas.

que a produção do sentido se dá a partir do sujeito em múltiplos contextos (WEIGAND; PASCHKE, 2012).

Além de resultados, os Mecanismos de Busca (MBs) 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. Assim, tecnologias como o *Knowledge Graph* (KG), empregadas para aprimorar os resultados exibidos pelos MBs com informações advindas de KBs, corroboram para a estruturação e exibição de informações fatuais sobre determinados temas nas *Search Engine Results Pages* (SERPs) destes mecanismos (ROSPOCHER et al., 2016).

A observação dessa conjuntura tecnológica presente no Ciberespaço fez emergir a inquietação para elaborar esta pesquisa, mais precisamente, em buscar responder o seguinte questionamento: qual o papel do *Knowledge Graph* do Google em relação ao uso do conhecimento enciclopédico colaborativo da Wikipédia sob a concepção de uma *Web Pragmática*?

2 WEB SINTÁTICA, SEMÂNTICA E PRAGMÁTICA

Diferentes adjetivações nomearam a *Web* de acordo com o modo como as informações são produzidas, organizadas, disseminadas e recuperadas em seu âmago. Nesta pesquisa, essas fases da *Web* são interpretadas na perspectiva de Vecchiato 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).

Na perspectiva de Monteiro (2015), a *Web* é Sintática por se relacionar com as dimensões da linguagem, fato que a torna legível por máquinas, mas compreensível apenas por humanos. Ao vinculá-la ao contexto evolutivo da *Web*, a nomenclatura “Sintática” identifica as duas primeiras camadas da *Web* e corresponde ao modo como a informação era indexada e recuperada, por considerar apenas aspectos referentes à formação lógica de palavras e frases, sem necessariamente atribuir-lhes significado, reservando a humanos essa tarefa.

A *Web Social*, segunda camada da *Web Sintática*, evidencia-se por seu caráter social, cooperativo e participativo. Fuchs et al. (2010) consideram que o caráter social desta camada 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 foi nas tecnologias 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 nesse contexto. O desenvolvimento e a implementação dessas tecnologias na *Web Social* 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 (ALESSO; SMITH, 2006). Isto é possível porque o *XML* possui metadados que, segundo Yu (2007, p. 10, tradução nossa), são definidos como “[...] dados sobre dados [...]” e descrevem de forma sistemática (estruturada) recursos informacionais, fato que viabiliza a interpretação dessas informações por agentes computacionais. 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).

Outras duas importantes tecnologias iniciadas na *Web 2.0* que contribuíram significativamente para a expansão do ideal social são os *weblogs (blogs)* e *wikis*. Com a proposta de editar conteúdo sem necessariamente possuir um conhecimento técnico, essas tecnologias foram disseminadas 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).

Os *blogs* conectam pessoas com interesses em comum, resultando assim em comunidades *on-line*. 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). Nesse sentido, a Wikipédia parece ser um exemplo representativo da aplicabilidade dos princípios *wiki*.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

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). Berners-Lee, Hendler e Lassila (2001) se referem à *Web Semântica* como uma extensão da anterior, em que a informação tem seu significado bem definido para que pessoas e computadores possam trabalhar em cooperação.

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). 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 *Uniform Resource Identifier (URI)* interligadas (BERNERS-LEE; HENDLER; LASSILA, 2001). Para Bax (2012) o *RDF* é uma linguagem, dentre outros padrões propostos pela *World Wide Web Consortium*, que possibilita a representação do conhecimento na *Web Semântica* e, conseqüentemente, a interpretação por máquinas, por meio de uma *Semântica Formal*.

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*. Entretanto, no contexto *Web*, e em consonância com Bax (2012) identificou-se uma similaridade com os pressupostos da *Semântica Formal*, pois faz a união entre sentido e referência para a criação do significado, haja vista que o lógico precisa de evidências do mundo palpável para gerar o sentido. (OLIVEIRA, 2006). 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*.

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 dimensão. Essas tecnologias da *Web Semântica* permitem a criação e armazenamento de dados, a construção de vocabulários e de

regras para o tratamento dessas informações, entre as quais se destacam: *Rule Interchange Format*; *Uniform Resource Identifier*; *Internationalized Resource Identifier*; *Resource Description Framework Schema*; *Web Ontology Language*; *Simple Protocol and RDF Query Language* (W3C, 2017).

Já a *Web Pragmática*, segundo Monteiro (2015, p. 3), essa dimensão também pode ser compreendida como: “[...] *Web 4.0*, *Web Ubíqua*, *Web Social Semântica* e *Web Semiótica*.” A origem desta 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).

Os aspectos conceituais que envolvem a *Web Semântica* (lógica de dados) e a *Web Pragmática* (perspectiva humana) não podem ser segregados, mas trabalhados em conjunto para a formação de uma nova dimensão *Web* (PIETARINEN, 2003). 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, na qual agentes computacionais e humanos compartilham da mesma linguagem.

Singh (2002, 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.” Dessa forma, a *Web Pragmática* parece ser a junção da *Web Social* com a *Web Semântica*, regida por uma 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:

[...] 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.

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.

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 WIKIPÉDIA E AS KNOWLEDGE BASES

Formalmente, a Wikipédia foi iniciada em 15 de janeiro de 2001. Desde então, a quantidade de leitores, colaboradores e artigos aumentou progressivamente. 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). A quantidade extraordinária de artigos produzidos e armazenados nas diferentes edições da Wikipédia torna-a uma grande base do conhecimento enciclopédico colaborativo na *Web*.

Entretanto, está limitada a buscas por palavras-chave, além de apresentar inconsistências decorrentes da duplicidade de conteúdo entre suas diferentes edições (MORSEY et al., 2009). Desse modo, tecnologias como as *Knowledge Bases (KBs)* auxiliam a Wikipédia a tornar seu conteúdo cada vez mais acessível e disponível ao restante da *Web*.

Segundo Sterckx et al. (2016, p. 1, tradução nossa), *KBs* são “[...] bancos de dados contendo milhões de fatos sobre pessoas, organizações, eventos, produtos [...]”. Também podem ser consideradas bases de dados de entidades, nas quais as entidades correspondem a informações disponíveis na *Web* sobre objetos e conceitos que, possuem propriedades e relacionamentos em comum com outras entidades (UYAR; ALIYU, 2015).

Iniciado em 2012, o Wikidata contava com um contingente pequeno e restrito de editores que podiam 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 (VRANDECIC; KRÖTZSCH, 2014).

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 1.0*” para a publicação mais flexível de seus dados (WIKIDATA, 2017a). No Wikidata as entidades são compreendidas como “*Items*”, 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, 2017a). Atualmente, o Wikidata passa por um processo de migração de conteúdo, não do seu, mas recepcionando os dados advindos do Freebase. 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).

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* Semântica. Além disso, possibilitam, por meio de padronizações e interoperabilidade, o entendimento entre humanos e agentes computacionais no Ciberespaço.

4 MECANISMOS DE BUSCA E O KNOWLEDGE GRAPH

Os Mecanismos de Busca (MBs) se tornaram cada vez mais requisitados 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 MBs foi alcançada apenas recentemente com a evolução dessas ferramentas, sendo o *Knowledge Graph* (KG) uma das tecnologias que complementam seu funcionamento.

Nesse sentido, o *KG* é um importante recurso para aplicações *Web* e *mobile* que utilizam Inteligência Artificial para processos de recuperação de informações. Pode ser interpretado como um grafo multirrelacional composto 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 MBs 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 aprimorar os resultados fornecidos pelos *MBs* com base em informações estruturadas em *KBs* como Freebase, DBpedia, Yago e Wikidata (ROSPOCHER et al., 2016).

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.” Os *MBs* são 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). Assim, 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 (SINGHAL, 2012).

Portanto, os *KGs* são a exibição gráfica/textual de informações semanticamente interligadas em *KBs*. Assim, têm por objetivo fazer com que os resultados apresentados nos *MBs* tenham mais sentido para além do significado, isto é, buscam envolver o contexto e perfil de uso do sujeito informacional na apresentação dos resultados.

5 PERCURSO METODOLÓGICO

De natureza básica, esta investigação não propôs o desenvolvimento de soluções aplicáveis como produtos dos fenômenos aqui estudados. Sem a pretensão de quantificar o grandioso volume informacional existente na *Web*, esta pesquisa adotou a abordagem qualitativa para compreender o papel do *Knowledge Graph* do Google entre o conhecimento enciclopédico colaborativo da Wikipédia e a concepção de uma *Web Pragmática*. Quanto ao delineamento, optou-se pelo documental, visto a necessidade de seu emprego para o estreitamento de laços entre *corpus* teórico (revisão de literatura) e metodológico (Wikipédia, Wikidata e o *KG* do Google).

Em virtude de sua representatividade no que tange o tema *Web*, “Tim Berners-Lee” foi elencado como o termo a ser analisado. Feito isso, a Wikipédia foi classificada mediante consulta realizada a lista² de enciclopédias *on-line* disponibilizada em seu próprio domínio. Assim, buscou-se por enciclopédias que fossem multilíngues, gerais, colaborativas e que

² https://en.wikipedia.org/wiki/List_of_online_encyclopedias

realizassem conexões com *KBs*. A *Knowledge Base* analisada foi selecionada com base nas pesquisas de Färber et al. (2015), Vrandecic e Krötzsch (2014) e Tanon et al. (2016), nas quais, o Wikidata foi evidenciado em virtude de sua representatividade em comparação à outras *KBs*. Nessa fase, houve a validação do ciclo de análise por meio do *website* do BabelNet³, no qual foram verificadas as fontes de sua rede semântica, apontando a existência de conteúdo advindo da Wikipédia e Wikidata para o termo elencado.

O MB Google foi selecionado devido ao seu alto índice de utilização, comprovado por meio de consultas realizadas ao *website* NetMarketShare⁴. As etapas de análise realizadas na Wikipédia, Wikidata e *KG* do Google foram necessárias para compreender a origem e trajetória da informação até seu aproveitamento pelo *KG* do Google.

5 ANÁLISE E APRESENTAÇÃO 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 Wikipédia⁵, deu-se início ao processo de análise desta investigação. Posteriormente, o Wikidata também foi acessado no intento de compreender como o conhecimento enciclopédico colaborativo extraído da Wikipédia era estruturado em seu âmago. A última etapa do processo de análise investigou se o *KG* do Google utilizava o conhecimento enciclopédico extraído da Wikipédia e enriquecido semanticamente pelo Wikidata na estruturação de resultados de sua SERP. Assim, as próximas subseções apresentam as especificidades de cada uma das etapas de análise desta pesquisa.

5.1 Wikipédia

Na Wikipédia, vislumbrou-se o processo criativo que envolve a produção do conhecimento enciclopédico colaborativo, em especial, sobre o verbete referente a “Tim Berners-Lee”. Para tanto, o termo elencado foi inserido na caixa de busca do MB nativo da Wikipédia, o qual apontou duas opções de pesquisa por meio do “*auto-suggest*”, são elas: “Tim Berners-Lee” e “*containing... Tim Berners-Lee*”. A primeira alternativa redireciona para a página do verbete referente a “Tim Berners-Lee”; a segunda corresponde a quantidade de ocorrências encontradas sobre o termo em diferentes artigos da Wikipédia, no caso, foram

³ <http://babelnet.org/>


⁴ <https://www.netmarketshare.com/>

⁵ https://en.wikipedia.org/wiki/Main_Page

⁶ São sugestões ou previsões de pesquisa baseadas nas atividades dos usuários (SULLIVAN, 2012).

recuperadas 2.153 evidências. 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.

Ao acessar o verbete sobre “Tim Berners-Lee”, o usuário se depara com seis 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 são responsáveis pela exibição do conteúdo do verbete. A guia “*Talk*” redireciona para uma página de discussão pertencente ao artigo, destinada apenas para adequações específicas do verbete. 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.

Na guia “*Talk*” também há 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*”⁷ que prima por manter um padrão linguístico consensual entre os membros da comunidade. Um verbete também pode alcançar um determinado *status* a partir de classificações realizadas pela própria comunidade. O artigo de “Tim Berners-Lee” possui o selo de qualidade “ *C-Class*” e consta na lista “*Vital articles/Expanded/People*”⁸ que elenca os artigos mais significativos da Wikipédia para que sirvam de modelo para os demais. O restante das informações presentes na página “*Talk*”, referem-se principalmente a notas e avisos sobre atividades que foram e que ainda devem ser cumpridas pela comunidade para a manutenção do artigo.

Ao acessar a página da guia “*View source*”, um aviso incomum na maioria das outras páginas de verbetes da Wikipédia alerta que o artigo de “Tim Berners-Lee” é semi-protegido e tem edição restrita aos membros da comunidade. 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 ataques de vandalismo realizados ao conteúdo do artigo.

A guia “*View history*” é responsável por exibir 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 ter sido criado no ano de 2001 pelo

⁷ https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style#National_varieties_of_English

⁸ https://en.wikipedia.org/wiki/Wikipedia:Vital_articles/Expanded/People

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

23 a 27 de outubro de 2017 – Marília – SP

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 uma sigla em negrito.

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.

De volta à guia “*Article*”, algumas considerações a respeito da organização e apresentação do verbete de “Tim Berners-Lee” puderam ser tecidas. Assim, seu conteúdo está disposto ao longo de nove seções de assuntos, apoiadas em 69 referências, além de um “*Infobox*” que apresenta informações pontuais sobre os temas abordados em cada seção, conforme apontado na Figura 1.

Figura 1: *Infobox* do verbete “Tim Berners-Lee” na Wikipédia

Born: Timothy John Berners-Lee/8 June 1955 (age 62)/London/England/UK
Other names: TimBL - TBL
Alma mater: The Queen's College, Oxford (BA)
Occupation: Computer scientist
Spouse(s): Rosemary Leith (m. 2014)/Nancy Carlson (m. 1990; div. 2011)
Children: 2
Parent(s): Conway Berners-Lee/Mary Lee Woods
Awards: Turing Award (2016)/Queen Elizabeth Prize (2013)/OM (2007)/KBE (2004)/etc.
Website: www.w3.org/People/Berners-Lee
Institutions: World Wide Web Consortium/University of Oxford/University of Southampton/etc.

Born	Timothy John Berners-Lee 8 June 1955 (age 62) ^[1] London, England, UK
Other names	TimBL TBL
Alma mater	The Queen's College, Oxford (BA)
Occupation	Computer scientist
Spouse(s)	Rosemary Leith (m. 2014) Nancy Carlson (m. 1990; div. 2011)
Children	2
Parent(s)	Conway Berners-Lee Mary Lee Woods
Awards	Turing Award (2016) Queen Elizabeth Prize (2013) OM (2007) KBE (2004) FRS (2001) ^[2] FREng (2001) FRSA (2001) DFBCS (1999) See full list of honours
Website	www.w3.org/People/Berners-Lee# Scientific career
Institutions	World Wide Web Consortium University of Oxford University of Southampton Passey MIT

Fonte: adaptado de Wikipédia – 2017.

Dessa maneira, ao analisar o verbete sobre “Tim Berners-Lee” na Wikipédia foi possível constatar que há uma preocupação crescente da comunidade com o estabelecimento de padrões e políticas de produção e manutenção do conteúdo existente.

5.2 Wikidata

Em continuidade ao procedimento de análise desta investigação, esta seção se destinou a analisar como o Wikidata estrutura as informações acerca do termo “Tim Berners-Lee” advindo da Wikipédia. Acessar e compreender o modo de funcionamento do Wikidata mostrou-se uma tarefa simples por conta da familiarização com o padrão de interface da Wikipédia. No Wikidata, o verbete torna-se uma entidade que pode ser compreendida como “*Item*”, que corresponde a um objeto, conceito ou evento do mundo real, seguido de propriedades/declarações e valores (WIKIDATA, 2017a).

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 ao evento Technology, Entertainment and Design⁹, nos anos de 2009, 2010 e 2014. A opção “*containing...*” identificou 374 correspondências existentes no 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 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 *Internet Protocol* (IP) do computador em que a edição foi realizada. Desde então, foram realizadas mais de 500 edições nesta entidade. Além disso, outra semelhança corresponde à proteção das páginas (*semi-protected*) para casos de vandalismo de conteúdo.

A página correspondente ao “*Item*” de “Tim Berners-Lee¹⁰” no Wikidata parece ter sua interface destinada de fato ao usuário final, com informações bem organizadas e claras. Os

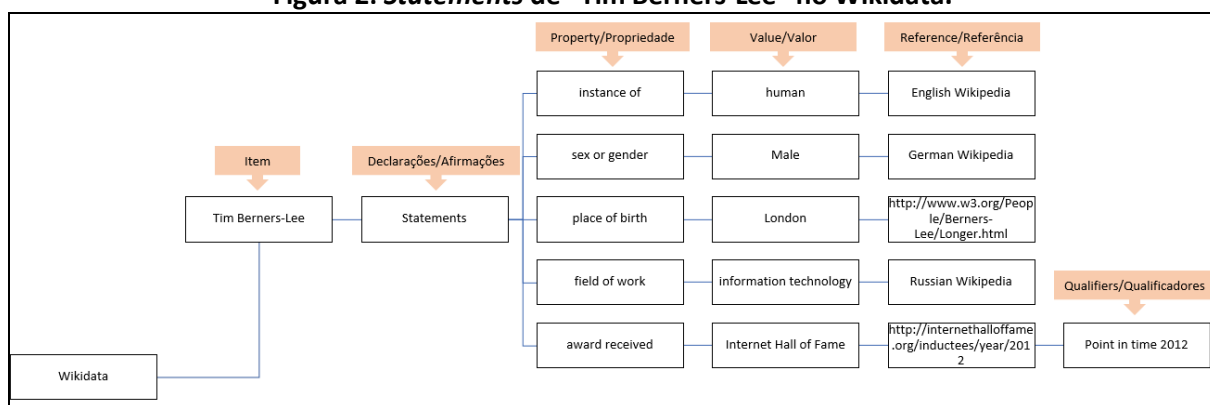
⁹ <https://www.ted.com/>

¹⁰ <https://www.wikidata.org/wiki/Q80>

“*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”.

As informações existentes em “Q80” advêm de diferentes fontes, como *hyperlinks* para *websites* externos, além da Wikipédia, 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, mas é possível inferir com base na verossimilhança de conteúdo um possível trajeto, conforme apresentado na Figura 2.

Figura 2: Statements de “Tim Berners-Lee” no Wikidata.



Fonte: os próprios autores com base em Wikidata – 2017b.

Ao observar a Figura 2, o modo como as declarações são organizadas remetem à perspectiva de Berners-Lee, Hendler e Lassila (2001), na qual o *RDF* usa o *XML* para atribuir propriedades e valores a um sujeito mediante *URI* interligadas. Além disso, esta padronização corrobora com a visão de Bax (2012) a respeito da Semântica Formal utilizada para a representação do conhecimento na *Web Semântica*.

5.2 Google Knowledge Graph

Nesta seção, o ciclo de análise cumpre sua última etapa ao investigar o modo como o *KG* do Google emprega o conhecimento originado na Wikipédia e semanticamente estruturado no Wikidata para a exibição de resultados sob a concepção de uma *Web Pragmática*. Logo, o termo “Tim Berners-Lee” foi inserido na caixa de busca do MB Google, que indicou quatro opções de pesquisa por meio do “*auto-suggest*”, são elas: “tim berners-lee”, “tim berners-lee net worth”, “tim berners-lee quotes” e “tim berners-lee invention”. Entretanto, o enfoque está apenas no desdobramento do termo exato “Tim Berners-Lee” e no

KG a ele vinculado. Assim, a Figura 3 apresenta um comparativo entre este KG do Google e as informações presentes na Wikipédia e Wikidata.

Figura 3: Comparativo entre Google Knowledge Graph, Wikipédia e Wikidata

The figure illustrates the relationship between three knowledge sources for Tim Berners-Lee. On the left is the Google Knowledge Graph (KG), which provides a structured overview of his life and achievements. In the center is the Wikipedia article, which offers a more detailed narrative. On the right is Wikidata, which provides structured data and multilingual labels. Green arrows show how the KG's structured data is mirrored in the Wikipedia article and Wikidata. For example, the KG's 'Born' field (June 8, 1955) corresponds to the Wikipedia's birth information and Wikidata's birth date. The KG's 'Awards' field lists numerous honors, which are also detailed in the Wikipedia article. The KG's 'Quotes' field includes a famous quote about time and dreams, which is also featured in the Wikiquote section of Wikidata. The KG's 'Profiles' field shows a link to his Twitter account, which is also listed in Wikidata's 'Twitter username' field. Finally, the KG's 'People also search for' field lists other notable figures like Vint Cerf and Steve Jobs, which is also reflected in the Wikipedia article's 'People also search for' section.

Fonte: os próprios autores.

Ao observar a Figura 3, foi possível compreender a trajetória do conhecimento acerca de “Tim Berners-Lee”. Desse modo, identificou-se que o texto inicial do KG do Google é semelhante ao apresentado no início do verbete sobre “Tim Berners-Lee” na Wikipédia, em seguida, o KG aponta algumas informações pontuais por meio dos campos “Born”, “Awards”, “Education”, “Spouse” e “Children”. Apesar de em alguns casos a nomenclatura do campo ser diferente das existentes na Wikipédia, os valores são equivalentes aos encontrados no “Infobox” do verbete.

Na subdivisão “Quotes” do KG, o texto da primeira citação corresponde ao campo “Wikiquote” existente no “Item” sobre “Tim Berners-Lee” no Wikidata. Além disso, a seção “Profiles” apresenta uma conexão com o Twitter¹¹ de “Tim Berners-Lee”, informação interconectada por meio do campo “Identifier” do Wikidata, denominado como “Twitter username”. A última parte do KG do Google apresenta a opção “People also search also” que

¹¹ https://twitter.com/timberners_lee

corroborar com a visão de Santaella (2012, p. 41) a respeito da forma como os resultados são apresentados na *Web Pragmática*, utilizando “[...] características dos grupos sociais e políticos de que o usuário participa, ou pelo perfil de pesquisa [...]” .

Por fim, o elemento “*Feedback*” do *KG* do Google permite que o usuário, ao selecioná-lo, possa editar as informações que o compõem para avaliar/alterar os dados e enviá-los a 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 Vaidhyanathan (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 evidenciaram a presença da *Web Pragmática* no funcionamento do MB, em especial, na estrutura do *KG*.

6 CONSIDERAÇÕES FINAIS

Esta investigação buscou compreender o papel do *KG* em relação ao uso do conhecimento enciclopédico colaborativo da Wikipédia no MB Google sob a concepção de uma *Web Pragmática*. Para tanto, com base neste objetivo, o termo “Tim Berners-Lee” foi elencado para a execução de um procedimento de monitoramento iniciado na Wikipédia, que perpassou pelo Wikidata e se finalizou no MB Google.

A aplicação do procedimento de análise supracitado trouxe a possibilidade de realizar inferências a respeito do modo como o *KG* do Google emprega o conhecimento enciclopédico colaborativo produzido na Wikipédia para proporcionar aos seus usuários uma experiência próxima da concepção de uma *Web Pragmática*. Assim, o termo elencado para análise passou por um processo de monitoramento que descreveu sua trajetória pela Wikipédia, Wikidata e o *KG* do Google.

O acesso ao histórico de versionamento do verbete indicou que o conhecimento enciclopédico da Wikipédia pode ser taxado como colaborativo, visto que o desenvolvimento do verbete 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.

No Wikidata o verbete é convertido em “Item”, “Propriedade e/ou Valores”. A interface simples promove a compreensão de seu funcionamento sem a necessidade de acesso a ferramentas externas, provando ser uma opção viável de KB colaborativa. Em seu âmago, o verbete sobre “Tim Berners-Lee” é enriquecido semanticamente para mais tarde se tornar um resultado da *SERP* do Google.

O KG do Google possui traços da *Web Pragmática* em seu funcionamento, visto que apresentou informações pontuais advindas da Wikipédia e do Wikidata em sua estrutura, isto é, colocou recursos semânticos em contextos pragmáticos. Além disso, também possui a opção “*Feedback*”, um indicativo da prática da colaboração em prol de um conhecimento pautado no consenso. Portanto, o conhecimento enciclopédico colaborativo advindo da Wikipédia e semanticamente enriquecido no Wikidata é utilizado pelo KG do Google na apresentação de resultados sob a concepção de uma *Web Pragmática*. Assim, o KG do Google exerce o papel de elo entre o conhecimento enciclopédico colaborativo da Wikipédia e a concepção de uma *Web Pragmática*, com resultados sensíveis ao contexto e apresentados sobre uma concepção humana do significado.

Para contribuições futuras, sugere-se novos estudos com a inclusão de outras *Knowledge Bases* que não puderam ser abordadas nesta investigação e também uma análise comparativa entre outros MB.

REFERÊNCIAS

ALESSO, H. P.; SMITH, C. **Thinking on the web**: Berners-Lee, Gödel, and Turing. New Jersey: Wiley-Interscience, 2006. 260 p.

ARENAS, M. et al. Faceted search over RDF-based knowledge graphs. **Web Semantics: Science, Services and Agents on the World Wide Web**, v. 37, n. 1, Koblenz, mar. 2016.

BAX, M. P. A evolução da *Web* rumo à *Web Semântica*. **PRISMA.COM**, Porto, n. 19, p. 1-27, 2012.

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.

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.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

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.

GÁRCIA ARETIO, L. Web 2.0 vs web 1.0. **Contextos Universitarios Mediados**, v.1, n. 14, jan. 2014.

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

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

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

MONTEIRO, S. D. 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.

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

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.

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

MURUGESAN, S. Understanding Web 2.0. **IEEE Computer Society**, Washington D.C., v. 9, n. 4, p. 34-41, jun./ago. 2007.

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.

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

ROSPOCHER, M. et al. Building event-centric knowledge graphs from news. **Web Semantics: Science, Services and Agents on the World Wide Web**, Koblenz, v. 37, n. 1, mar. 2016.

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

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: 19 jul. 2017.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

SINGH, M. 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, 2002.

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.

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.

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: 19 jul. 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.

UYAR, A.; ALIYU, F. M. Evaluating search features of Google Knowledge Graph and Bing Satori: entity types, list searches and query interfaces. **Online Information Review**, n.1, v. 39, p. 1-26, jan. 2015.

VECHIATO, F. L.; VIDOTTI, S. A. B. G. Encontrabilidade da informação: atributos e recomendações para ambientes informacionais digitais. **Informação e Tecnologia (ITEC)**, Marília/João Pessoa, n. 1, v. 2, p. 42-58, jul./dez. 2014.

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. In: PROCEEDINGS OF THE TWENTY-EIGHTH AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, 28., 2014, Québec. **Anais...** Palo Alto: AAAI Press, 2014. p. 1112-1119.

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: 19 jul. 2017a.

WIKIDATA. **Tim Berners-Lee**. Disponível em: <<https://www.wikidata.org/wiki/Q80>>. Acesso em: 19 jul. 2017b.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

WIKIPEDIA. **Tim Berners-Lee**. Disponível em: <https://en.wikipedia.org/wiki/Tim_Berners-Lee>. Acesso em: 19 jul. 2017.

W3C. **Semantic Web**. 2017. Disponível em:
<<https://www.w3.org/standards/semanticweb/>>. Acesso em: 19 jul. 2017.

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