

Web Semântica: a Web que se comunica com o homem

Carlos Roberto de Andrade¹
Danilo Luís Carlos Micali²

Resumo. O objetivo deste artigo é apresentar de forma teórica para os profissionais e usuários da internet as evoluções que vem ocorrendo na *Web* desde a sua criação. Muitos internautas não observam que Internet e *Web* são tecnologias distintas, mas ambas são tão importantes que os dois termos acabam se confundindo e virando sinônimos. A internet revolucionou os meios de comunicação, porém o seu sucesso só veio com a *Word Wide Web* ou simplesmente *Web*, uma ferramenta que auxilia na consulta e inserção de dados na rede. A *Web* se popularizou e muitas informações foram inseridas na Internet, o que ocasionou maior dificuldade em se encontrar conteúdos com qualidade, pois a máquina apenas apresenta as informações, já que não entende os significados dos conteúdos nas páginas da *Web*. Para resolver esse problema está sendo desenvolvida a *Web Semântica*, que é também denominada *Web 3.0*, com o objetivo de compreender a comunicação da linguagem humana. Os resultados obtidos demonstram que a *Web Semântica* já está sendo utilizada e que as máquinas começam a compreender os significados da informação, porém a *Web 3.0* ainda não vem a ser uma inteligência artificial.

Palavras-Chaves: *Web 3.0*; Internet; Informação; Comunicação.

Abstract. Semantic Web: the Web which communicates with the man. The purpose of this article is to present theoretically to professionals and Internet users the evolutions that have been occurring on the Web since its inception. Many netizens do not observe that Internet and Web are different technologies, but both are so important that the two terms end up confusing and becoming synonymous. The internet revolutionized the media, but its success came only with Word Wide Web or simply Web, a tool that assists in querying and inserting data on the network. The Web was popularized and many information was inserted on the Internet and therefore made it difficult to find contents with qualities, because the machine only presents the information and does not understand the meanings of the contents in the Web pages. To solve the problem is being developed the Semantic Web, which is also called Web 3.0, aiming to understand the communication of human language. The results have shown that the Semantic Web is being utilized and that the machines beginning to understanding the meanings of the information, but Web 3.0 is not yet an artificial intelligence.

Keywords: *Web 3.0*; Internet; Information; Communication.

1 Introdução

A internet é um meio de comunicação que surgiu no final da década de 60 se tornou tão natural na vida dos seres humanos que diariamente bilhões de pessoas usam essa ferramenta para se comunicar. Todas as informações adquiridas no decorrer do tempo pelos seres humanos

¹ Discente da Fatec Itu – carlosrobertodeandrade1986@gmail.com.

² Doutor em Estudos Literários pela UNESP-FCLAr e docente da Fatec Itu – dlcmicali@gmail.com.

estão disponíveis na internet e podem ser consultadas em qualquer parte do mundo. Atualmente há inúmeras ferramentas que auxiliam o usuário na troca da informação e na comunicação via internet como, por exemplo, *e-mail*, *Whatsapp*, *Skype*, dentre outras. Porém a grande ferramenta de acesso à internet que revolucionou o mundo tecnológico, criada no início da década de 90, foi a *Word Wide Web* que também é chamada de *Web*.

A *Web* facilitou muito a navegação na internet, pois através dela o usuário não precisa ter conhecimento técnico para buscar e inserir informação na rede mundial de computadores. Porém com o sucesso da *Web* alguns problemas surgiram, pois as informações se espalharam de forma tão desproporcional, que a navegação se tornou penosa, ou seja, ficou difícil encontrar conteúdos com qualidade e fontes confiáveis por causa do excesso de informações.

Para que os internautas consigam encontrar informações desejadas e com qualidade, é preciso que a máquina compreenda os significados das informações contidas dentro das páginas da *Web*, por isso surgiu o termo *Web Semântica*, também chamada de *Web 3.0*, que tem o objetivo de fazer os softwares entenderem e se comunicarem na linguagem humana.

Muitas dúvidas surgem quando se fala em *Web Semântica*, às quais este artigo procura responder: A *Web Semântica* é uma *Web Inteligente*? Ela será capaz de compreender o significado das informações? A *Web 3.0* já é uma realidade e existem aplicações em funcionamento? A máquina se comportará igual ao ser humano? A *Web Semântica* é uma inteligência artificial?

Este artigo apresentará o conceito de internet e *Web*, o significado das palavras na comunicação humana, os benefícios e estruturas da *Web 3.0*, assim como o funcionamento e os desafios para os profissionais e empresas de tecnologia da informação. Por fim, discorrerá sobre as aplicações da *Web Semântica* que já estão em funcionamento beneficiando os internautas.

2 Internet e Web

O sucesso da *Web* é tão grande que muitos usuários acabam considerando que o termo é sinônimo de internet (VITER, 2005). A Internet surgiu na década de 60 através do Departamento de Defesa Americano ARPA (*Advanced Research Project Agency*), com objetivo militares e a ideia era fazer os nortes americanos manterem as comunicações, caso acontecesse uma destruição nos meios convencionais de telecomunicações (DUMAS, s.d).

Após o período da guerra fria a internet também passou a ser utilizada por comunidades da área científica e acadêmica. No final da década de 80 um físico e cientista da computação chamado Timothy John Berners-Lee, percebeu que havia certa dificuldade de encontrar

informações na internet, pois era preciso ter conhecimentos técnicos para fazer a busca na rede (BALAN, 2006).

Berners-Lee criou o Protocolo de Transferência de Hipertexto (HTTP) para os computadores se comunicarem por meio da internet e também criou o Recurso Universal de Identificação (URI) para dar endereço único às informações na internet, e por fim desenvolveu o navegador a fim de facilitar a exibição das informações. Porém, Berners-Lee tentou vender a sua criação para Organização Europeia para a Pesquisa Nuclear (CERN), mas houve certa burocracia para aprovar o projeto e a venda não foi realizada. Então, junto com o colega Robert Cailliau, apresentou uma nova maneira de acessar a internet por meio de uma interface gráfica. Essa tecnologia foi denominada de *Word Wide Web* também chamada de *Web* (Ibidem).

A primeira Página da *Web* foi publicada no dia 6 de agosto de 1991, criada pelo próprio Tim Berners-Lee, que apenas explicava o conceito da *Web* e, segundo a ABC News, um ano depois foi publicada a primeira imagem na *Web* (G1,20012). Essa primeira *Web* era chamada de *Web 1.0* ou *Web* da Informação, pois ela apenas apresentava ao internauta as informações e o usuário precisava ter conhecimento técnico para inserir informações, quando o único meio de comunicação era por e-mail (POLLOCK, 2010).

Várias páginas da *Web* foram surgindo e não tinham padrão algum e as informações ficavam bagunçadas. Por isso foi preciso criar o consórcio de padronização da *Web* chamado de W3C, e um *site* criado nos padrões do W3C pode ser visualizado em qualquer idioma e sistema operacional espalhado pelo mundo (Ibidem).

A segunda geração *Web*, chamada de *Web 2.0* ou *Web* da Comunicação surgiu no início dos anos 2000, os *sites* deixaram de ser estáticos e os usuários começaram a se comunicar um com o outro. Nessa versão da *Web* surgiram as famosas redes sociais como *Orkut* e *Facebook*, pois os usuários começaram a inserir informações na *Web* e ganharam liberdade para expressar e atentar opiniões mesmo sem ter conhecimentos técnicos (Ibidem).

A mais nova geração da *Web* é chamada de *Web Semântica* ou *Web 3.0*, segundo o próprio Berners-Lee (2001): “A *Web Semântica* é uma extensão da *Web* atual, onde a informação possui um significado claro e bem definido, possibilitando uma melhor interação entre computadores e pessoas” (BREITMAN, 2005).

Tim Berners-Lee diz que o objetivo dessa ferramenta é de interligar o homem com a máquina, fazer com que o computador entenda e interprete a informação que a pessoa procura. Sendo assim, a busca na *Web Semântica* apresentará resultado mais eficiente, apresentando o que realmente é procurado, eliminando conteúdos equivocados que aparecem no meio da pesquisa nos motores de busca (POLLOCK, 2010).

2.1 Motores de Busca

Pollock (2010) afirma que um dos grandes sucessos da *Web* são justamente os motores de busca, que são softwares que ajudam a encontrar inúmeras respostas e conteúdos dentro da *Web*. Esses *softwares* denominados de robôs se comportam como os internautas e navegam por diversos *sites* no mundo, com finalidade de coletar informações e apresentar para o internauta.

Desde a criação da Internet e da *Web* inúmeros mecanismos de busca foram criados, porém o que mais se destaca no mundo atual é o mecanismo da Google, pois seu sistema é muito eficiente e traz resposta precisa. Segundo o documentário disponível no Youtube “História da Google 1995 – 2014”, que já fora apresentado no canal *Discovery Channel* no ano de 2014, a Google se destacou principalmente pelo algoritmo *PageRank*, que funciona sobre apontamento de *link* de uma página para outra, sugerindo que a página recebedora do *link* tem qualidade, e também por ser o primeiro buscador a inserir propagandas conforme a palavra-chave procurada pelo usuário e não de forma aleatória como os buscadores faziam antes do surgimento da Google.

Grande companhia de índices de busca possui milhares de computadores que usam softwares conhecidos como *spiders* ou robôs (ou apenas simples bots) para deter páginas *Web* e ler as informações que elas armazenam. Esses sistemas nem sempre pegam toda a informação em página ou de todas as páginas de um *site*, mas eles detêm uma quantia significativa de informações e usam complexos algoritmos, cálculos baseados em fórmulas complexas, para indexar essa informação (KENT, 2015, p. 10).

Cada motor de busca tem seu segredo para coletar dados, mas no geral eles são bem parecidos. Eles extraem os conteúdos dos *sites* seguindo os *links* e armazenando em seus servidores e junto com a data da última visita de cada *site*, porém não conseguem visualizar as imagens, vídeos e músicas e ignoram as partes gráficas de um *site* (SEBSAUVAGE, 2014). Segundo o autor, esses *softwares* analisam todos os textos e todas as palavras contidas em uma página *Web*, mas não interligam uma palavra com a outra, apenas procuram palavras idênticas ou parecidas com a pesquisa do usuário para depois apresentar em uma lista. Alguns buscadores manipulam a busca dando prioridade aos parceiros comerciais e os outros são obrigados a respeitar a regra de um determinado país e não apresentando determinado *site* em sua pesquisa. Mas o fato é que esses motores de busca são um sucesso para a *Web*, não é por acaso que segundo a revista Forbes (2016) a *Google* está entre as empresas mais valiosas do mundo.

3 Semântica

A Semântica vem da palavra grega *semantikós*, e é o “ramo da linguística que estuda o significado das palavras; estudo das relações entre os signos e os seus referentes” (PRIBERAM,

2017). Segundo Ilari e Geraldi (1995, p.6) “a semântica é a ciência que estuda a significação”, definindo-se como o estudo das palavras e frases, signos e expressões de um determinado contexto. De acordo com Garcia (2010), a linguagem ideal seria aquela em que uma palavra pudesse ter um significado preciso e único, mas isso é inexistente em qualquer língua, pois com o passar do tempo à palavra vai adquirindo novas significações.

Uma palavra só é compreensível quando ela ganha sentido em um contexto, pois sozinha não se sabe o que a palavra quer dizer. A Semântica classifica as palavras quanto à: Denotação, Conotação, Sinonímia, Antonímia, Homonímia, Paronímia e Polissemia.

A palavra “cobra” em seu sentido denotativo recebe o significado de “réptil ofídio”, já no sentido de conotação o seu significado pode ser “pessoa ruim”. O sentido de denotação é o sentido em que palavra está no significado original e literal, seu uso é mais comum, já a conotação é o sentido subjetivo e figurado que se atribui a uma palavra, que a diferencia do seu sentido original (VÍCOLA, 2011).

Também na semântica temos o conceito de sinonímia, que é a ligação entre duas ou mais palavras. O dicionário Novo Aurélio Século XXI define sinônimo como: “Palavra que tem o mesmo significado que outra ou outras, ou significado semelhante ou aproximado.”. Pode-se usar como exemplo as palavras “bonita” e “linda” ou “professor” e “docente”, que têm praticamente o mesmo significado. Ainda faz parte da semântica o conceito de antonímia, que se refere às palavras que expressam ideias contrárias, tais como as palavras “quente” e “frio” que possuem sentidos opostos.

No campo da semântica há palavras que pertencem ao conjunto da homonímia, que podem ser homógrafas, homófonas e perfeitas. As homônimas homógrafas são as palavras que possuem a mesma grafia, mas significado e pronúncia diferente (NORMACULTA, s.d.a). Exemplo: “Seu gosto, às vezes, é duvidoso. Eu gosto muito de viajar e conhecer o mundo³”. A primeira frase a palavra “gosto” é um substantivo, e passa a ideia de preferência ou personalidade própria. Já na segunda frase a palavra “gosto” é um verbo, e passa a ideia de que a pessoa adorar fazer algo, no caso viajar. Ambas as palavras apresentam a mesma grafia, mas com pronúncia diferente, pois na segunda oração a palavra “gosto” tem o primeiro “o” aberto, parecendo ser pronunciado duas vezes, tendo um som contínuo como se a pessoa estivesse falando “goosto”.

Já as homônimas homófonas são as palavras que têm significado e escrita diferentes, mas as pronúncias idênticas (NORMACULTA, s.d.a). É o caso das palavras “sessão”, que significa tempo/período e a palavra “seção”, que significa lugar/local/setor.

³ Exemplo retirado de página educacional da Internet: Disponível em: <<http://www.normaculta.com.br/palavras-homonimas>>. Acesso em: 07 ago. 2016.

E, por fim, as homônimas perfeitas que são palavras que possuem a mesma grafia e o mesmo som na pronúncia, mas o significado é diferente (NORMACULTA, s.d.a). Um exemplo seria a palavra “banco”, que pode significar um assento ou instituição financeira, assim como um verbo.

Quando uma palavra tem significado diferente, porém possui semelhança tanto na escrita quanto na pronúncia, ela é denominada de parônima (COMOESCREVE, 2012). As parônimas levam as pessoas a cometerem erros tanto na pronúncia quanto na grafia, como por exemplo, as palavras “absorver” que significa sorver, e “absolver” que significa perdoar.

Já o conceito de polissemia refere-se ao fato de a palavra adquirir outros significados no decorrer do tempo, sentidos figurados (NORMACULTA, s.d.b), a exemplo da palavra “cavalo” que significa um animal, e pode ganhar significado diferente como de “pessoa estúpida”.

E, por fim, na semântica, tem-se a ambiguidade que pode causar um duplo sentido na oração como, por exemplo, em um título de uma notícia do jornal folha de São Paulo em março de 2001, onde se diz: “Vaticano **admite** estupro de freiras por religiosos⁴”. Segundo Vícola (2013), a frase não está errada, mas pode gerar certa ambiguidade e acabar gerando uma dúvida em quem está lendo o título, pois o verbo “admite” significa tanto “reconhece, confessa” como “aceita, permite”. Para um leitor que domina o idioma português é clara a ideia de que o Vaticano reconhece o crime, mas para um leitor, no caso um estrangeiro que está aprendendo o português, isso seria um absurdo, pois ele poderia entender que o Vaticano permite esse crime. Deve-se ter cuidado com a ambiguidade principalmente no texto científico para não gerar dúvida, como aconteceu neste exemplo. Por outro lado, a ambiguidade pode ser desejada, não sendo vista como um problema em determinados contextos, como em propagandas e publicidade para vender produtos, sendo também utilizada em poesias e textos humorísticos.

4 Benefícios da Web Semântica

Pollock (2010) afirma que a Web Semântica foi desenvolvida para as máquinas, porém o principal beneficiado é o internauta que terá resultados muito mais eficientes em uma busca semântica, pois a máquina apresentará o resultado preciso eliminando os conteúdos equivocados que aparecem nas pesquisas.

No mundo dos negócios um dos grandes benefícios que a Web trouxe foi justamente permitir que o usuário realize uma compra online sem sair de casa. Segundo uma pesquisa feita em 2015 pelo Serviço de Proteção ao Crédito (SPC), o preço é um dos itens mais atrativos para

⁴ Exemplo retirado de página educacional da Internet: Disponível em: <<http://gramaticadoprofessordaniel.blogspot.com.br/2013/05/cuidado-com-ambiguidade.html>>. Acesso 01 set. 2016.

50% dos entrevistados. Isso acontece porque grande parte das lojas virtuais oferece preço menor que o das lojas físicas. Porém quando um usuário quer adquirir uma mercadoria por um preço menor ele precisa entrar em diversos *sites* para comparar os preços, visando à aquisição do produto no *site* que oferece o menor preço e ainda assim fazer um cadastro no *site*, caso seja a primeira compra do usuário nessa página.

Por meio das ferramentas da *Web Semântica* as compras se tornarão mais rápidas, pois assim que o usuário digitar o produto que deseja adquirir a máquina se encarregará de navegar pelos *sites* e trazer o produto com o menor preço (BREITMAN, 2005). O usuário tomará a decisão se deseja ou não adquirir o produto indicado pelo motor de busca. Outra vantagem será o fato de que o usuário não precisará mais ficar digitando diversos formulários de cadastro para realizar uma compra online em inúmeros *sites*, pois os dados pessoais do internauta ficarão em um único sistema para realizar diversas compras (POLLOCK, 2010).

Na busca semântica os motores de busca trabalharão com o conceito de polissemia, sinonímia, paronímia e homonímia, pois assim a busca semântica não analisará apenas as palavras-chaves e, sim, a linguística semântica humana, a *Web 3.0* será projetada para entender o contexto em que as palavras são usadas dentro de um conteúdo na *Web*, conforme disse Kassim e Rahmany(2009), citado por Reis et al (2012).

A *Web Semântica* também tem a intenção de ligar todas as redes sociais como, por exemplo, *Facebook*, *Twitter*, *LinkedIn*, *Whatsapp*, *Instagram* etc. Todos os usuários têm pelo menos uma conta em uma dessas redes sociais, e alguns têm em diversas contas em redes diferentes. A ideia da *Web 3.0* é fazer com que essas contas não fiquem separadas uma das outras, pois se o usuário é uma única pessoa por que as contas têm de ficar separadas? Na *Web Semântica* assim que o usuário fizer uma atualização em uma de suas redes sociais automaticamente as outras são atualizadas (POLLOCK, 2010).

Na *Web 3.0* os navegadores que os internautas usam para acessar a *Web* irão possuir inteligência e ajudarão o internauta na navegação, entendendo o conteúdo que o usuário estiver visualizando, podendo fazer indicações de conteúdos idênticos ou semelhantes referentes ao assunto que o internauta navegar. Na prática o navegador será um parceiro do usuário, pois está lendo e compreendendo os conteúdos e a preferência do usuário (Ibidem).

5 Estrutura da *Web Semântica*

De acordo com Breitman (2005), para melhorar a busca na *Web* e apresentar informações precisas e com qualidade, é necessário que cada página descreva qual o tipo de informação ela deseja transmitir, para isso é necessário acrescentar os metadados.

Pollock (2010) afirma que metadados são dados sobre dados, ou seja, informação utilizada para descrever outra informação. O consórcio W3C define que metadados são “informações na *Web* que podem ser compreendidas por máquinas”.

Apesar de não existir uma definição universal para o termo metadados, o tema ainda está aberto a discussões nas várias comunidades onde ele é utilizado. No entanto é importante notar que a utilização de metadados não é novidade e nem foi introduzida por pesquisadores da *Web Semântica*, pois se trata de um conceito que vem sendo aplicado há centenas de anos por bibliotecários, museólogos, arquivistas e editores (BREITMAN, 2005, p.16).

Um livro é um exemplo de algo que contém metadados, pois além do conteúdo que apresenta, traz informações que ajudam a identificá-lo como, por exemplo: autor, editora, ano da publicação, idioma de origem, categoria do livro, descrição do conteúdo etc. Neste sentido deverão ser as páginas na *Web 3.0*, ao invés de apenas serem identificadas por palavras-chaves pelos motores de busca, as páginas deverão ter uma descrição de conteúdo referente ao assunto, essa descrição é chamada de metadados (BREITMAN, 2005).

Um padrão de metadados criado em um workshop em 1994 na cidade americana de Dublin no estado de Ohio, foi o *Dublin Core*, pois esse padrão fornece uma série de elementos para descrever os dados como, por exemplo: título do documento da *Web*, assim uma breve descrição do que trata o documento, a data do documento assim como o criador, editor, formato etc. Também outros padrões de metadados foram criados como *Warwick*, que veio para suprir, completar o *Dublin Core*, inserindo outros elementos como, informações de segurança, autenticidade, assinaturas etc..(Ibidem)

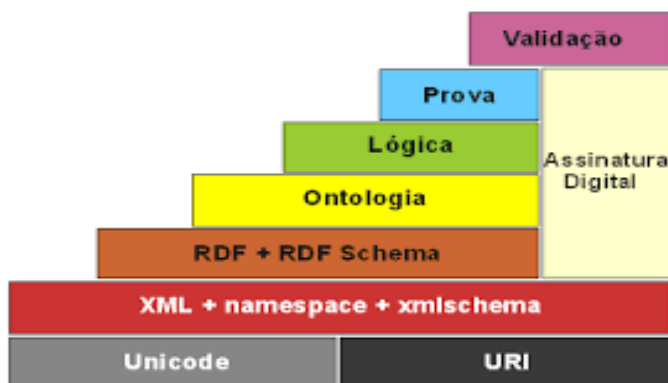
Os metadados contribuirão muito para a *Web Semântica*, mas eles vêm sofrendo diversas críticas por parte dos pesquisadores, principalmente porque não se pode confiar nas descrições das páginas *Web*, pois é de se esperar que alguns *sites* mintam sobre suas características, ou não as detalhem de forma correta. Outra questão levada em conta é que muitos donos de *sites* deixarão de colocar descrições de conteúdos visto que alguns usuários nem preenchem o campo assunto de um correio eletrônico. Apesar dos problemas os metadados são importantes para a *Web Semântica*, uma expressão norte-americana passa a ideia do que fazer com os metadados, essa expressão diz que “Não jogar fora o bebê junto com a água do banho”, ou seja, apesar dos problemas os metadados não devem ser descartados (Ibidem).

5.1 Camadas da *Web Semântica*

A *Web Semântica* já tem suas camadas definidas junto com os pesquisadores. O consorcio W3C tem estudado para melhorar o nível da *Web* atual. A ideia de Berners Lee é que

ao invés de criar uma nova arquitetura para a *Web*, que por consequência geraria uma reestruturação da Internet, a nova estrutura seria construída em cima da existente, pois ele acredita que é mais fácil chegar ao consenso fazendo modificações em cima da *Web* existente, do que conseguir apoio para fazer uma mudança drástica na *Web* (BREITMAN, 2005).

Figura 1- Estrutura da *Web* Semântica



Fonte: devmedia⁵.

A primeira camada corresponde à Unicode e URI, sendo a base da pirâmide da *Web* Semântica. O Unicode trabalha com todos os caracteres, e através dessa camada a máquina conectada na *Web* consegue apresentar qualquer página escrita indiferente ao idioma, apresentando texto, números ou símbolos (UNICODE.ORG, 2017). Já a URI é o alicerce da *Word Wide Web*, pois é a base para encontrar qualquer página ou conteúdo dentro da *Web*, pois se encarrega de dar endereço único para os recursos da *Web*, podendo ser um texto, uma página, música ou vídeo (CORRÊA, 2012). Sem o URI seria difícil de encontrar qualquer recurso disponível na internet (POLLOCK, 2010).

Um pouco mais para cima tem a camada XML, *namespace* e *xmlschema*. O *eXtended Markup Language* (XML) é uma linguagem para a formação de dados estruturados, e veio com objetivos de suprir a necessidade do HTML. Os recursos disponíveis na *Web* estão dentro de etiquetas denominadas de *Tags*, as quais têm a função de separar os conteúdos. Essas tags não são apresentadas para o usuário final, apenas as máquinas conseguem visualizar essas etiquetas e fazer a separação dos conteúdos, porém essas tags não têm significado algum com o conteúdo dentro dela (Ibidem).

O XML, além de fazer a separação de conteúdo, possibilita dar nome às *tags*, algo que o HTML⁶ não permite, possibilitando que o desenvolvedor dê nome que tenha sentido ao conteúdo que está dentro da Tag (Ibidem). Um exemplo que pode ser citado é que se um *site*

⁵ Imagem disponível em: <HTTP://arquivo.devmedia.com.br/REVISTAS/java/imagens/85/4/image001.jpg>. Acesso em: 07 set. 2016.

⁶ Significa *Hypertext Markup Language*, uma linguagem utilizada para desenvolver *sites*.

contém um documento com uma receita de bolo de laranja, pode-se abrir uma *tag* e colocar o nome de “receita de bolo de laranja”. Depois de se colocar o conteúdo sobre a receita para fazer um bolo de laranja pode-se fazer o fechamento da *tag* com o nome de “receita de bolo de laranja”. Isso ajudará o robô de busca na hora de procurar alguma receita de bolo de laranja disponível na *Web*, pois a informação estará etiquetada de forma clara para o motor de busca.

O *Namespaces* funciona parecido com um diretório de um sistema operacional como, por exemplo, o Windows, pois em um mesmo diretório não se podem colocar dois arquivos com o mesmo nome. Caso queira colocar arquivos com nomes idênticos em um mesmo diretório é necessário que o arquivo tenha uma característica específica, no caso, o formato. No Windows os arquivos com nomes iguais podem ser colocados no mesmo diretório caso tenham os formatos, por exemplo, “doc” e “pdf”. Já o *Namespaces* permite que as *tags* tenham uma característica específica podendo ser um prefixo ou uma particularidade, a função principal é evitar a ambiguidade de informação na página *Web* com *tags* com o mesmo nome, a exemplo do que acontece no sistema operacional Windows (Ibidem).

O *XML Schema*, também chamado de XSD (*XML Schema Definition*), é uma linguagem criada através do XML, porém ela trabalha com Tags pré-definidas, e indica qual informação deve ser preenchida dentro da *Tag*. Um exemplo que pode ser citado é que se uma *Tag* foi definida para ser uma data ou hora e o usuário inserir um texto, a informação não será validada e estará fora do padrão de busca (Ibidem).

O *Resource Description Framework* (DRF) é uma linguagem baseada em XML, sendo utilizada para descrever os metadados na *Web* Semântica. Utilizando o RDF pode-se descrever qualquer recurso na *Web*, podendo ser uma página *Web*, pessoas, lugares, livros etc... Para descrever o recurso os metadados utilizam uma tripla RDF, a qual é chamada de recurso, propriedade e declaração ou também denominado de sujeito, predicado e objeto. Para identificar e montar a tripla RDF é necessário fazer as perguntas: “Qual o recurso a descrever?”, “Quais as propriedades desse recurso?”, “Quais os valores dessas propriedades?” (Ibidem).

O recurso ou sujeito é tudo aquilo que se quer descrever na *Web* e que possui uma URI, no caso poderia ser, por exemplo, o livro “Romeu e Julieta” disponível para a venda online em qualquer *site* de comércio eletrônico. Esse livro possui propriedades ou predicados, podendo ser, o autor, ano de publicação, idioma e etc. As propriedades possuem um valor que é denominado declaração ou objeto. No caso da propriedade autor seu valor é “William Shakespeare”.

O RDFS ou *RDF Shema* tem a função de descrever a semântica na *Web*, pois ele modela de forma hierárquica as propriedades e recursos e relaciona com outras propriedades e recursos (Ibidem). Um exemplo que pode ser citado seria um texto em algum *site* da *Web* descrevendo

sobre o *Rottweiler*. O *Rottweiler* é uma classe que possui propriedades e declarações. O *Rottweiler* está relacionado com outra classe, no caso “Cachorro”, que possui propriedades e declarações. O Cachorro está relacionado com outra classe chamado “Animal”, e que também possui propriedades e declarações. Logo, se pode dizer para a máquina que *Rottweiler* é um cachorro e pertence a uma subclasse animal. Desse modo, quando um usuário quiser pesquisar sobre um animal, e perguntar para o motor de busca, qual animal tem determinada característica e digitar a característica do *Rottweiler*, o sistema apresentará de forma objetiva esse animal, no caso, o *Rottweiler*.

A Camada de Ontologia da *Web Semântica* tem a função de suprir a limitação da tecnologia RDF/RDFS, e para isso são acrescentados mais elementos para descrever as classes (as coisas), os relacionamentos entre as classes e as propriedades ou características que essas classes possuem. Para escrever a Ontologia na *Web* é recomendado pelo W3C o uso da linguagem OWL (*Web Ontology Language*), em português, Linguagem de Ontologia da *Web* (POLLOCK, 2010).

Ontologia são especificações formais e explícitas de conceitualizações compartilhadas. Ontologias são modelos conceituais que capturam e explicitam o vocabulário utilizado nas aplicações semânticas. Servem como base para garantir uma comunicação livre de ambiguidades. Ontologias serão a língua franca da *Web Semântica* (BREITMAN, 2005, p. 7).

Uma das possibilidades que a linguagem OWL permite é declarar que duas classes são idênticas, por exemplo, pode-se declarar que a classe professor é equivalente à classe docente, sendo assim todo indivíduo que é professor também é um docente e vice-versa. Também é possível declarar que duas classes possuem as mesmas características e propriedades, porém são diferentes como, por exemplo, ocorre na classe humano, pois nesta classe há subclasses de homem e mulher, ou seja, explica-se para a máquina que um humano pode ser homem ou mulher, apesar de possuírem características praticamente iguais (BREITMAN, 2005).

A camada Lógica tem a função de unificar todas as tecnologias das camadas anteriores como: XML, XSD, RDF, RDFS, OWL. Já a camada Prova tem a função de verificar se as aplicações da camada lógica estão funcionando corretamente com os agentes inteligentes (softwares de busca), ou seja, se as comunicações de todas as tecnologias estão sendo executadas conforme programada (POLLOCK, 2010).

As camadas Assinatura Digital e Confiabilidade têm a função de verificar se o conteúdo da informação é verídico. Sabe-se que qualquer indivíduo pode publicar alguma informação na *Web*, porém, na *Web Semântica* essa informação precisará ter uma assinatura digital que identifique de onde vem a fonte. Já a camada Confiabilidade tem a função de verificar se a fonte

de informação é verdadeira, para não aparecer resultados que não sejam verídicos em um motor de busca (Ibidem).

6 Desafios da Web Semântica

Segundo fonte do *site Internet Live Stats*, que apresenta em tempo real o uso da internet no mundo, em 2016 a *Web* atingiu a marca de mais de um bilhão de *sites* no mundo. Por causa dessa quantidade de páginas será um grande desafio estruturar a *Web* com a tecnologia da *Web* 3.0, visto que muitos desses *sites* não contêm a tecnologia da *Web* Semântica. Outro fator é que muitos profissionais de tecnologia da informação não têm o conhecimento técnico para reestruturar os *sites* ou não acham necessária a reestruturação por acreditarem que gastarão muito tempo para executar a mudança (ANDRADE, 2011).

Alguns cientistas dizem que a criação de ontologia é um grande desafio, pois na linguagem humana já é difícil se comunicar um com o outro, imaginem para o humano descrever de forma clara e objetiva o significado das informações contidas em uma página *Web* para as máquinas (Ibidem).

Já para os motores de busca o desafio é impedir que os desenvolvedores consigam burlar o sistema de busca, pois eles têm a intenção de fazer seus *sites* aparecerem no topo de uma pesquisa. O próprio Google já teve problema com o famoso algoritmo *PageRank*, quando muitos desenvolvedores criavam diversos *sites* apontados para outras páginas para dar voto de importância a *sites* que não tinham conteúdo algum; e também alguns desenvolvedores declaravam diversas palavras-chaves em seus *sites*, porém essas palavras também não tinham relacionamento algum com o conteúdo do *Web site* (GASPAR, 2015).

6.1 As mudanças nos Algoritmos da Google

A Google sempre faz pequenas mudanças diariamente em seus algoritmos a fim de deixar as buscas muito mais eficientes. Muitas dessas mudanças o usuário final acaba nem percebendo, mas quando uma mudança radical é feita, muitos desenvolvedores acabam ficando preocupados, pois não querem que seus *sites* desapareçam no sistema da Google (PATEL, 2015). Quando a Google realiza uma mudança radical não revela o segredo, mas deixa uma dica de como o sistema irá se comportar e como o desenvolvedor pode melhorar seu *site* (KENT, 2015).

No início de 2011 a Google lançou o algoritmo Google Panda, com a intenção de reduzir o *rank* de *site* com baixa qualidade, com pouco conteúdo ou conteúdo duplicado. A lógica desse algoritmo é que dois *sites* possuem um conteúdo idêntico. Não seria justo o *site* com conteúdo original aparecer em posicionamento inferior ao do que copiou o conteúdo. Neste caso, o Algoritmo Panda analisa que *site* postou o conteúdo primeiro, sendo o que possui o conteúdo original (Ibidem).

Em 2014 a Google lançou o algoritmo *Penguin*. O seu objetivo era punir o *site* que aponta para outro *site* com conteúdo diferente e baixa qualidade. Na prática significa que se um *site* quer fazer uma recomendação é melhor indicar uma página que possui qualidade no conteúdo. E também pune o *site* que encher de palavras-chaves, se essas palavras não possuem significado algum com o conteúdo do *site* (PATEL, 2015).

Em 2013 a Google lançou o algoritmo *Hummingbird*. Por meio dele o usuário pode encontrar a resposta enquanto ainda está digitando a frase. Nesse algoritmo a Google trabalha com o conceito de “Frases-chaves de cauda longa”. Assim, a pesquisa pode ser feita em uma linguagem natural (Ibidem).

Outro algoritmo lançado em 2014 pela Google é o *Pigeon*, que tem o objetivo de trabalhar com a localidade de um usuário conforme diz Sacoto (2015), ou seja, se um usuário estiver na cidade de Itu/SP e digitar apenas as palavras “faculdade”, esse algoritmo entende que o usuário está procurando alguma faculdade na cidade de Itu/SP, nisso o sistema apresentará nas primeiras colocações resultados com as faculdades da cidade de Itu.

O fato é que a Google está se preocupando em implantar a busca semântica em seu sistema, pois está tendo resultados eficientes e a máquina já está começando a entender, parcialmente, o que o usuário procura. Outros buscadores certamente irão implantar a busca semântica e o usuário final terá ótimos buscadores semânticos na Web 3.0.

7 Aplicações da Web Semântica

A Web Semântica já está em funcionamento e algumas aplicações já estão disponíveis e já podem ser utilizadas pelo usuário. Uma dessas aplicações é o *Skyscanner* que é um buscador de comprar passagens aéreas, reserva de hotéis e aluguel de carro. Através desse buscador o usuário pode comparar os preços das passagens aéreas de diversas companhias, apenas preenchendo a data e o destino da viagem, o buscador traz o resultado da companhia que possui voo naquela data e com um valor mais em conta. Isso economiza tempo para o usuário, pois ele não precisará ficar entrando em diversos *sites* de companhia aérea para comparar os preços e encontrar voo disponível para determinada data. Essa ferramenta, além do *site*, também está disponível para celulares como no sistema Androide, *Windows phone* e IOS (GOMES, 2012).

Figura 2 – Tela do inicial Skyscanner

The screenshot shows the Skyscanner homepage with a dark background. At the top, there are three tabs: 'Passagens aéreas' (selected), 'Hotéis', and 'Aluguel de carros'. Below the tabs, there are three radio buttons for trip type: 'Ida e volta' (selected), 'Só ida', and 'Várias cidades'. To the right is a 'Mapa' link. The search form includes fields for 'De' (origin) and 'Para' (destination). 'De' is set to 'São Paulo - Congonhas (CGH)' and 'Para' is 'Qualquer lugar'. There are checkboxes for 'Adicionar aeroportos próximos' for both origin and destination. Below these are date fields for 'Ida' (07/10/16) and 'Volta' (08/10/16). To the right is a dropdown for 'Classe de cabine e Passageiros' set to '1 adulto, Econômica'. There is a checkbox for 'Apenas voos sem conexão' and a large green 'Buscar' button.

Fonte: <http://www.skyscanner.com.br>.

Outra Ferramenta da Web Semântica é o *Wolfram Alpha*, cujo objetivo é trazer respostas eficientes apenas apresentando o resultado que o internauta deseja visualizar. O buscador *Wolfram Alpha* elimina os resultados que não interessam ao usuário, na prática esse motor de busca tenta ser curto e grosso. Essa ferramenta é gratuita e possui um grande banco de dados para responder perguntas sobre matemática, pessoas, música, esporte, geografia, entre outros assuntos. Sua vantagem é que não possui *link* patrocinado e a resposta fica no próprio *site*, porém esse buscador só funciona na grafia da língua inglesa (HECKE, 2014). A seguir são apresentadas algumas pesquisas feitas no motor de busca do *Wolfram Alpha*.

Figura 3 – Resultado da pesquisa sobre história no *Wolfram Alpha*

The screenshot shows the Wolfram Alpha search interface. The search bar contains the text 'Who was the president of Brazil in 1962'. Below the search bar, there are icons for input interpretation and a 'Result' section. The result is 'João Goulart (from 07.09.1961 to 02.04.1964)'. Below the result, there is a 'Basic information' table.

Basic information:	
official position	President
country	Brazil
start date	07.09.1961 (55 years 20 days ago)
end date	02.04.1964 (52 years 5 months 25 days ago)
duration of leadership	2 years 6 months 25 days

At the bottom, there is a 'Sequence' section with a 'More' button.

Fonte: <http://www.wolframalpha.com> (2016).

Figura 4 – Resultado da pesquisa sobre geografia no Wolfram Alpha

The screenshot shows the Wolfram Alpha interface. At the top, the search bar contains the text "Qual a cidade mais populosa do Brasil" and "What is the most populous city in Brazil". Below the search bar, there are navigation links: "Web Apps", "Examples", and "Random". A message states: "Assuming 'populous' is referring to a data query | Use as referring to cities instead". Under "Input interpretation:", the input is broken down into "most city" and "by city population in Brazil". The "Result" section, highlighted with a black box, shows "São Paulo (10.66 million people)". Below this, a table titled "Populations:" provides more data:

Category	Population	Notes
city population	10.66 million people	(country rank: 1 st) (2010 estimate)
metro area population	20.22 million people	(Sao Paulo metro area) (2007 estimate)

At the bottom right of the table is a "Show history" button.

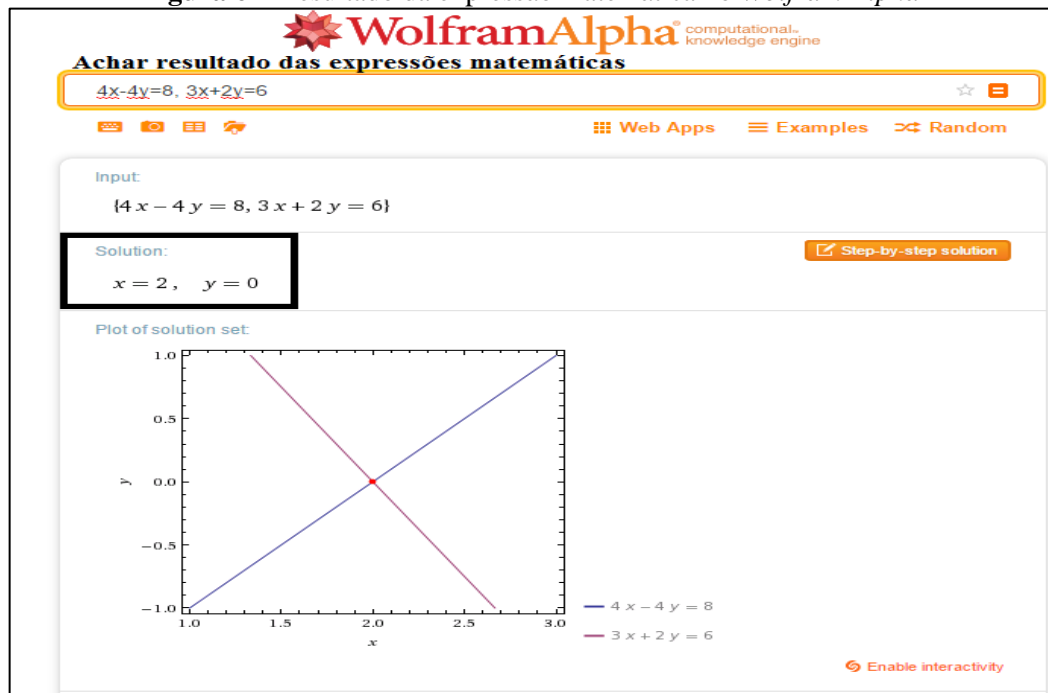
Fonte: <http://www.wolframalpha.com> (2016).

Figura 5 – Resultado da pesquisa sobre esporte no Wolfram Alpha

The screenshot shows the Wolfram Alpha interface. At the top, the search bar contains the text "Qual é o nome verdadeiro de Pelé" and "What is the real name of pelé". Below the search bar, there are navigation links: "Web Apps", "Examples", and "Random". A message states: "Assuming 'pelé' is a person | Use as a movie instead". Under "Input interpretation:", the input is broken down into "Pelé" and "full name". The "Result" section, highlighted with a black box, shows "Edson Arantes do Nascimento". At the bottom, there are links for "Sources", "Download page", and "POWERED BY THE WOLFRAM LANGUAGE". A footer section includes a "NEW: Have a question about using Wolfram|Alpha? Contact Pro Premium Expert Support" link and a "Give us your feedback" link.

Fonte: <http://www.wolframalpha.com> (2016)

Figura 6 – Resultado da expressão matemática no *Wolfram Alpha*



Fonte: <http://www.wolframalpha.com> (2016)

O objetivo da *Web Semântica* é trazer resposta clara e eliminar as demais informações que aparecem na pesquisa. Para isso, a ferramenta *Wolfram Alpha* é um avanço para a busca da informação na *Web*, onde responde exatamente ao que o usuário pergunta.

8 Considerações Finais

No decorrer deste artigo foram apresentadas as tecnologias disponíveis para a construção de uma *Web* estruturada. Pôde-se notar que a *Web Semântica* já vem fazendo parte da vida do internauta, e a busca por informação na *Web* realmente tem melhorado. Percebe-se essa evolução da *Web* desde a sua criação na década de 90. Também se nota uma evolução nos mecanismos de buscas, principalmente no mecanismo da Google, que vem tentando deixar a busca mais fácil por meio de uma linguagem natural.

Porém, a ferramenta em destaque da *Web Semântica* é justamente o motor de busca *Wolfram Alpha*, que tem por objetivo apresentar uma resposta única e precisa, exatamente aquilo que o internauta procura, eliminando conteúdos inúteis que aparecem no meio da busca.

Com base nesses fatos, pode-se concluir que os objetivos deste trabalho foram atingidos, pois a *Web Semântica* já vem sendo implantada e se mostra realmente capaz de entender a linguagem humana, entregando resposta clara e objetiva. Contudo, a Busca Semântica precisa ser aperfeiçoada. Os profissionais de TI precisam aplicar as tecnologias propostas pelo consórcio W3C nos *Web Sites*. Quando todas as páginas *Web* usarem a tecnologia da *Web 3.0*, serão mais visualizadas pelos motores de busca, o que trará benefício para os usuários.

Por fim, considera-se que apesar da *Web 3.0* ser mais eficaz que suas versões anteriores, ainda não é uma inteligência artificial, pois as informações ainda continuarão a ser inseridas pelos seres humanos. Entretanto, a quem diga que a *Web 3.0* é um passo para a construção da *Web 4.0*, quando a máquina chegará ao nível da inteligência artificial, se comportando igual ao ser humano. Mas isso é algo para o futuro, e não se pode ainda dizer como serão os mecanismos de funcionamento das próximas versões da *Web*.

9 Referências

- ANDRADE, Daniel Santos de. **Web Semântica: dificuldades e desafios**. Revista Tema, Campina Grande, v.11, n.16, jan./jun. 2011. Disponível em: <<http://revistatema.facisa.edu.br/index.php/revistatema/article/view/61/pdf>>. Acesso em: 10 set. 2016.
- BALAN, Willians Cerozzi. **Tim Berners-Lee, o pai da internet**. São Paulo: UMESP, 2006. Disponível em: <http://www.willians.pro.br/textos/tim_bernens_lee_o_pai_da_internet.pdf>. Acesso em: 27 fev. 2016.
- BREITMAN, Karin. **Web Semântica: A Internet do Futuro**. Rio de Janeiro, LTC, 2005.
- COMOESCREVE. **Palavras homônimas e parônimas**. Publicado 2012. Disponível em: <<http://www.comoescreve.com/2012/12/palavras-homonimas-e-paronimas.html>>. Acesso em: 25 set. 2016.
- CORRÊA, Eduardo. (2012). **Unicode: conceitos básicos**. Disponível em: <<http://www.devmedia.com.br/unicode-conceitos-basicos/25169>>. Acesso em: 07 set. 2016.
- DICIONÁRIO PRIBERAM. Priberam.pt. **Semântica**. Disponível em: <<https://www.priberam.pt/dlpo/sem%C3%A2ntica>>. Acesso em: 22 maio 2017.
- DUMAS, Véronique. **A origem da internet**. História Viva. (S.d). Disponível em: <http://www2.uol.com.br/historiaviva/reportagens/o_nascimento_da_internet.html>. Acesso em: 01 out. 2016.
- FERREIRA, Aurélio Buarque de Holanda. **Dicionário Eletrônico Aurélio Século XXI**. Rio de Janeiro: Editora Nova Fronteira e Lexikon Informática, 1999. Versão 3.0. 1 CD-ROM.
- FORBES, Revista. (2016). **50 maiores empresas do mundo em 2016**. Disponível em: <<http://www.forbes.com.br/listas/2016/05/50-maiores-empresas-do-mundo-em-2016>>. Acesso 09 jul. 2016.
- G1. (2012). **Primeiro site publicado na internet completa 21 anos**. Disponível em: <<http://g1.globo.com/tecnologia/noticia/2012/08/primeiro-site-publicado-na-internet-completa-21-anos-veja.html>>. Acesso em: 30 ago. 2016.
- GARCIA, Othon Moacyr. **Comunicação em Prosa Moderna**: aprenda a escrever, aprendendo a pensar. 27. ed. Rio de Janeiro: FVG, 2010.
- GASPAR, Heloisa. **Historia dos Buscadores – Dos diretórios ao surgimento do Google**. 2015. Disponível em: <<http://www.pwi.com.br/blog/historia-dos-buscadores-dos-diretorios-ao-surgimento-do-google>>. Acesso em: 07 set. 2016.

- GOMES, Felipe Morato. (2012). **Análise: Skyscanner à lupa.** Disponível em: <<http://www.almadeviajante.com/skyscanner-a-lupa-006104>>. Acesso em: 08 out. 2016
- HECKE, Caroline. **WolframAlpha.** Disponível em: <<http://www.baixaki.com.br/android/download/wolframalpha.htm>>. Acesso em: 08 out. 2016
- ILARI, Rodolfo; GERALDI, João Wanderley. **Semântica, Série Princípios.** São Paulo, Ática SA, 1995.
- INTERNET LIVE STATS. **Internet live Stats.** Disponível em: <<http://www.internetlivestats.com>>. Acesso em: 22 set. 2016.
- KENT, Peter. **Otimização para mecanismos de busca para leigos.** RJ: Alta Books, 2015.
- NORMACULTA. A. **Palavras homônimas.** Disponível em: <<http://www.normaculta.com.br/palavras-homonimas>>. Acesso em: 07 ago. 2016.
- NORMACULTA. B. **Polissemia.** Disponível em: <<http://www.normaculta.com.br/polissemia>>. Acesso em: 07 ago. 2016.
- PATEL, Neil. **O Guia Definitivo das Atualizações no Algoritmo Google.** 2015. Disponível em: <<http://neilpatel.com/br/2015/12/15/o-guia-definitivo-das-atualizacoes-no-algoritmo-do-google>>. Acesso em: 01 out. 2016
- POLLOCK, Jeffrey T. **Web semântica para leigos.** Rio de Janeiro, Alta Books, 2010.
- REIS, Júlio Cesar dos; BARANAUSKAS, Maria Cecilia C; BONACIN, Rodrigo. **Busca em Sistemas Sócio-Culturais Online: Desafios e Recomendações.** Campinas, 2012. 15p. Disponível em: <http://www.inf.pucminas.br/sbc2010/anais/pdf/semish/st01_04.pdf>. Acesso em: 15 ago. 2016.
- SACOTO, Gonçalo. **Google Pigeon: O impacto deste algoritmo no SEO.** 2015. Disponível em: < <http://www.marketingtecnologico.com/Artigo/google-pigeon-impacto-seo> >. Acesso em: 01 out. 2016.
- SEBSAUVAGE. (2014). **Comment ça marche un moteur de recherche.** Disponível em: < <http://sebsauvage.net/comprendre/recherche/index.html> >. Acesso em: 19 mar. 2016.
- SERVIÇO DE PROTEÇÃO AO CRÉDITO (SPC). **Índice de satisfação com compras na internet.** Disponível em: < <http://www.spcbrasil.org.br/imprensa/noticia/176> > Acesso em: 25 mar. 2016.
- UNICODE.ORG. **The Unicode Consortium.** Disponível em: <<http://www.unicode.org/consortium/consort.html>>. Acesso em: 22 maio 2017.
- VÍCOLA. Daniel Ricardo. **Sentido Literal X Sentido Figurado.** Publicado em 2011. Disponível em: < <http://www.gramaticadoprofessordaniel.blogspot.com.br/2011/09/sentido-literal-x-sentido-figurado.html> >. Acesso em: 30 ago. 2016.
- VÍCOLA. Daniel Ricardo. **Cuidado com a Ambiguidade.** Publicado em 2013. Disponível em: <<http://www.gramaticadoprofessordaniel.blogspot.com.br/2013/05/cuidado-com-ambiguidade.html>>. Acesso em: 13 ago. 2016.
- VITER, Luciana Nunes. (2005). **A diferença entre internet e World Wide Web.** Disponível em: <http://www.professorando.jex.com.br/informatica+educativa/a+diferenca+entre+internet+e+world+wide+Web>. Acesso em: 03 mar. 2016.
- YOUTUBE. **A história do Google - 1995 – 2014.** Produzido e exibido pelo Discovery Channel em 2014. Disponível em: < <https://www.youtube.com/watch?v=NK5dIOMwDq0> >. Acesso em: 11 maio 2016.