

# O desenvolvimento da WEB 3.0: Linked Data e DBPEDIA

*The development of the WEB 3.0:  
Linked Data and DBPEDIA*

---

**Francisco Carlos Paletta**

Universidade de São Paulo, SP, Brasil  
[fcpaletta@usp.br](mailto:fcpaletta@usp.br)

**Marcos Luiz Mucheroni**

Universidade de São Paulo, SP, Brasil  
[mucheroni.marcosl@usp.br](mailto:mucheroni.marcosl@usp.br)

## Resumo

O desenvolvimento inicial da Web como uma camada de textos sobre a internet tinha como objetivo apenas a publicação de conteúdos com maior facilidade, visibilidade e rapidez e era pensada para a comunidade científica. Porém com seu crescimento extraordinário criou-se a necessidade de sua crescente estruturação semântica, dando origem a esta construção através de linguagens de marcação de conteúdos como o XML e criação de ontologias. Com a Web 2.0 a participação do usuário tornou-se mais simples e efetiva, com a agregação de marcação pouco padronizada como *mashups* e *social tagging*. Embora discutida e sugerida em diversos trabalhos, a Web 3.0 começou a tornar-se realidade com o conceito de *Linked Data* e de tecnologias estruturadas que permitiram a inclusão de uma força de trabalho adicional e especializada atuantes na área da ciência da informação. Vários projetos então tornaram-se efetivos como o VIAF (Virtual Authority File) e o DBPedia desenvolvido a partir do Wikipédia, e explicitado neste trabalho

## Abstract

*The Web was originally developed for the primary purpose of sharing information and texts more easily, effectively and with a greater visibility. It was debated in the scientific community. However, its rapid growth required a semantic structure. The Semantic Web was thus designed, based on the use of markup languages such as the XML, and the creation of ontologies. The Web 2.0 allowed users to have a simpler and more effective interaction with each other, as well as the aggregation of a low level standardization, such as mashups and social tagging. Although the Web 3.0 has been suggested and discussed in several papers, it only became a reality with the concept of Linked Data and structured technologies, which will allow the inclusion of an additional and specialized workforce active in the Information Science field. A number of projects have then come into reality such as the VIAF (Virtual Authority File) and the DBpedia, based on Wikipedia and discussed in this paper.*

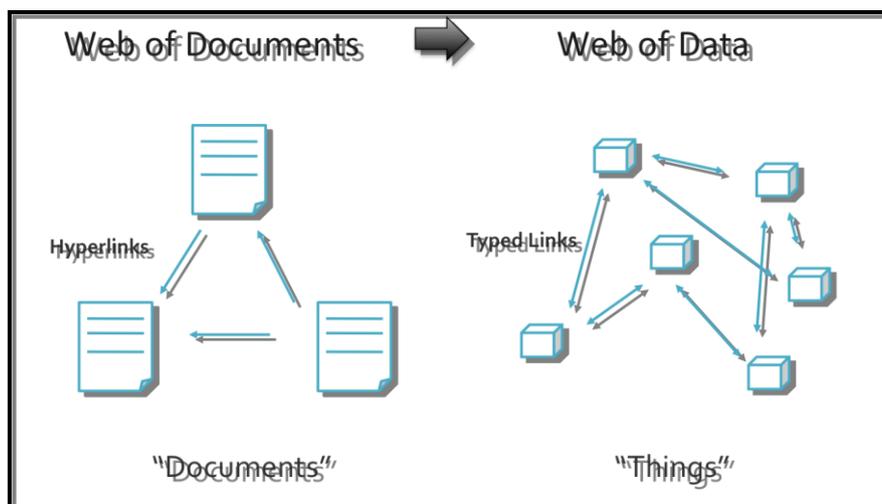
**Palavras-chave:** Web Semântica; *Linked Data*; **Keywords:** *Semantic Web*; *Linked Data*; *DBpedia*; *DBpedia*; Estudos de Usuários da informação; Gestão *User Study*; *Information Devices Management*; *Web 3.0* de Dispositivos de Informação; *Web 3.0*

## 1. Introdução

A *World Wide Web* continua a crescer e novas tecnologias, modos de interações, e aplicativos estão sendo desenvolvidos. Com base neste crescimento, as tecnologias da Web Semântica têm como objetivo proporcionar um espaço de informação semântica compartilhada, mudando qualitativamente as nossas experiências na Web. Com a consolidação das tecnologias da Web Semântica e estas permeando cada vez mais as áreas de desenvolvimento de aplicativos, novos desafios de pesquisa estão vindo à tona e alguns não resolvidos estão se tornando cada vez mais críticos. Estas questões incluem a criação e gerenciamento de conteúdos Web Semântica, viabilizar que as aplicações na Web Semântica se tornem mais robustas e escaláveis, organizar e integrar informações de diferentes fontes para novos usos, tornando a semântica explícita a fim de melhorar a experiência do usuário com as tecnologias de informação, e permitindo, assim, usar a riqueza de informações que está atualmente disponível em formato digital alterando de forma definitiva as atividades de busca da informação (ABERER, 2007).

Informalmente as pessoas costumam usar o termo "Internet" e "World Wide Web" (WWW) como sinônimos, mas isso é impreciso: a WWW é de fato apenas um dos muitos serviços prestados através da Internet. A característica distintiva da WWW é que esta é uma aplicação hipertexto, que explora a Internet para permitir a ligação cruzada de documentos em todo o mundo. O objetivo da Web Semântica é estabelecer normas através do qual as pessoas podem publicar documentos que consistem em dados, ou talvez uma mistura de dados e texto, permitindo que aplicações possam combinar dados de vários conjuntos de dados, assim como um usuário pode combinar informações de muitos documentos de texto, otimizando as ferramentas de busca da informação. Na Figura 1 representamos a transformação da Web dos Documentos para a Web dos Dados.

Figura 1 – **Web dos Documentos para Web de Dados**



Fonte: <http://projects.kmi.open.ac.uk/euclid>

O desenvolvimento inicial da Web (BERNES-LEE, 2001) propiciou maior disseminação e visibilidade dos conteúdos, porém o rápido crescimento dos dados tornou a Web Semântica uma necessidade e novas tecnologias: ontologias, bases semânticas, entre outras, foram desenvolvidas. A Web 2.0 trouxe uma maior participação do usuário, criaram-se comunidades virtuais, ferramentas de mídias sociais para o estabelecimento de redes, agregadores como mashups e social tagging, porém com pouca estruturação semântica.

A WEB participativa ou Web 2.0, consiste na participação ativa dos usuários na Internet, que criam conteúdos, adaptam e desenvolvem aplicações para uma grande variedade de campos. O conteúdo digital representa um fator importante na indústria das TICs – Tecnologia da Informação e Comunicação. A inovação tecnológica e a demanda dos novos usuários estão levando a novas formas de criação, distribuição e acesso ao conteúdo digital. A convergência nas aplicações, tais como convergência da bioinformática, da inteligência artificial, da biotecnologia, neurotecnologia, da robótica, da tecnologia da informação, entre outras tecnologias emergentes, fornecerão, provavelmente, maiores oportunidade e desafios para os pesquisadores que atuam na área da Ciência da Informação.

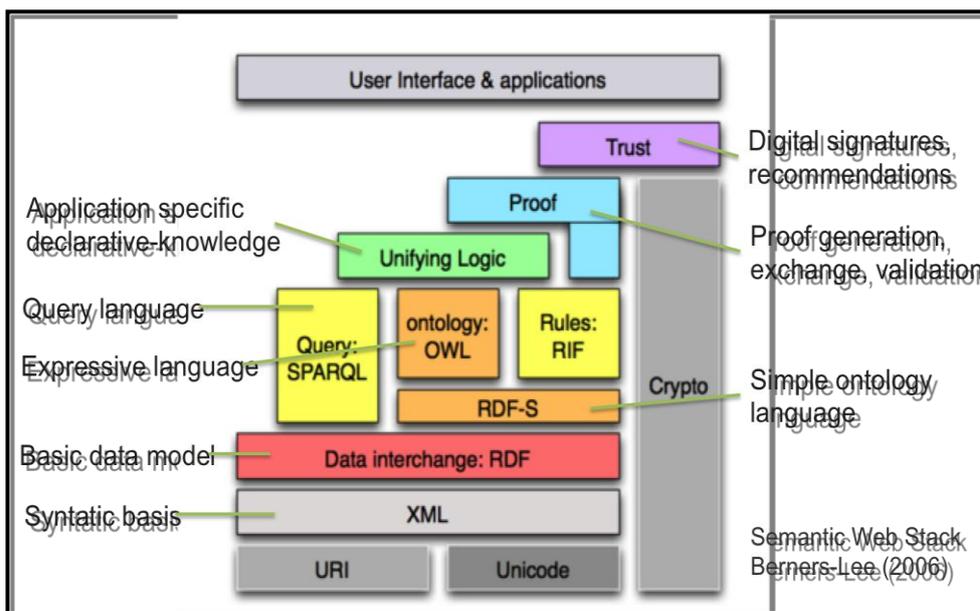
A Web 3.0 é uma das grandes propostas para o futuro da internet, a partir do conceito de Linked Data, com o uso de base de dados com estruturas semânticas e servidores SPARQL. Será ela que definitivamente organizará todas as informações que estejam na internet, proporcionará que o desenvolvimento de aplicativos baseados na Web sejam Open-Source e

viabilizará uma grande interatividade em diversas áreas da Web. Dentre as alternativas tecnológicas que fazem parte da evolução da Internet, o uso do celular como dispositivo de conectividade, tem uma forte colaboração na inclusão digital. Embora a Web 3.0 demore ainda alguns anos para chegar às organizações, os desenvolvedores de aplicativos já começam a oferecer serviços que farão parte deste futuro. A cada dia um número imenso de páginas da internet são criadas, com isto a busca por informações fica cada dia mais complexa em função da quantidade de informação disponibilizada ao usuário de forma não estruturada e desorganizada. A Web 3.0 virá com o objetivo de organizar estas informações para que os usuários tenham mais facilidade na busca da informação. A busca semântica, como são chamados os buscadores da Web 3.0, organiza as informações por assuntos determinados, algo bem mais complexo do que é usado atualmente nos processos de busca e apropriação da informação. A rede mundial de computadores ainda não consolidou a Web 2.0 e empresas já começam a desenvolver a Web 3.0 onde a busca semântica dividirá os resultados em categorias para que a experiência do usuário seja mais rápida e organizada do que as atuais na Web 2.0 (MATTOS, 2013).

## **2. O desenvolvimento da Web Semântica**

A concepção da Web Semântica surge delineada no artigo original de Tim Berners-Lee, Ora Lassila e James Hendler, publicado na *Scientific American* em 2001, e embora o termo Web 3.0 tenha sido cunhado por John Markoff, foi James Hendler (2009) que esclareceu que esta concepção depende ainda do desenvolvimento de tecnologias que ainda estão em processo. A Figura 2 é uma representação da arquitetura da Web Semântica proposta por Berners – Lee (GERBER, 2008):

**Figura 2 – Arquitetura da Web Semântica**



Durante a década de 90, Berners-Lee e colaboradores desenvolveram propostas para uma nova fase de desenvolvimento Web conhecida como a Web Semântica. Este conceito de longo alcance, primeiro divulgado em um artigo de 2001 na Scientific American, é parcialmente implementado no atual estágio de desenvolvimento Web, por vezes chamado de Web 3.0.

No artigo original da Web Semântica de 2001 se apontava “o caminho das ontologias, como caminho ‘natural’ para o desenvolvimento e agregar significado a informação na Web Semântica, com metodologias vindas da Inteligência Artificial, que no olhar de James Hendler (Web 3.0) passava por um ‘inverno’ criativo” (MUCHERONI, 2013).

Durante o processo percorrido pela Web Semântica, três ferramentas foram integradas e acabaram indicando um novo caminho: as ontologias que ajudam a construção de esquemas de organização simples do conhecimento chamadas SKOS (Simple Organization of Knowledge System), integradas em um banco de dados para consulta, usando a linguagem chamada SPARQL, e estas por sua vez, integradas naquilo que já era básico na Web Semântica, a sua linguagem de descrição, o RDF (Resource Description Framework) e em sua linguagem descritiva simples: o XML.

Este processo foi bem sucedido conforme esclarece Hendler, em função “das dificuldades para explicar a Web 3.0, ao contrário [das tecnologias] do browser original ou posteriores aos sistemas da Web 2.0, a tecnologia de Web Semântica tende a ser uma infraestrutura de

tecnologia” (HENDLER, 2009), e toda esta infraestrutura e seu desenvolvimento não são simples como foi a Web 2.0, onde usuários passaram a dispor de inúmeras ferramentas para seu auxílio que devem ser bem compreendidas a partir de seu uso em dois polos de desenvolvimento: empresas e instituições (bibliotecas, universidades, centros culturais, e unidades de informação).

A relação com a Ciência da Informação está no uso que o próprio Hendler aponta como "vocabulário controlado" que já existe há muito tempo, mas as tecnologias emergentes devem permitir que eles sejam mais facilmente integrados com os desenvolvimentos atuais da Web, uma vez que os primeiros foram desenvolvidos sem uma descrição para WEB (HENDLER, 2009), um RDF ou mesmo uma estrutura de aplicação XML mais simples.

Outro problema no desenvolvimento da Web 3.0 é o uso corporativo da Web Semântica e da arquitetura para apoiá-la. Surge o segundo aspecto destacado por Hendler, que é o desenvolvimento na comunidade acadêmica, que tem buscado numa lógica de raciocínio crescente, aplicativos *back-end* de grande escala, ontologias e novos sistemas SGBD (Bancos de Dados) que podem ajudar este desenvolvimento, agora com maior participação e integrando estes sistemas.

Mesmo antes do ambiente da Web, a internet e as redes de computadores já propiciavam a organização de usuários em listas de *e-mails*, desenvolvedores ou simples usuários em grupos de trabalhos (*Groupwares*) que se reuniam para executar uma determinada tarefa ou discutir o crescimento da Web, quando esta apontou para um caos geral.

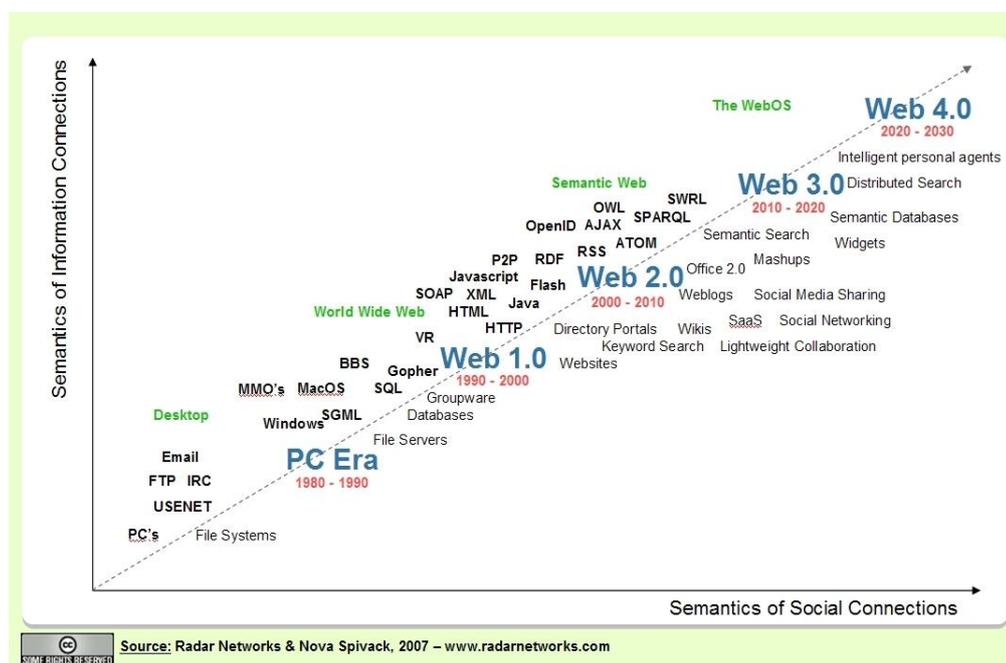
Pode-se apontar estas tecnologias desenvolvidas nesta etapa como linguagens de marcação XML (*eXtended Markup Language*), seus facilitadores para desenvolvimento de *sites Web*, as linguagens para *sites* dinâmicos: *Java JavaScript, Flash*, que, cada vez mais simples para os desenvolvedores, possibilitaram o desenvolvimento de portais e *web sites* e o surgimento das ferramentas sociais para *Weblogs* (ou *blogs*) até surgirem os primeiros CMS (*Content Managment System*) e ferramentas como AJAX, SOAP e P2P (*Peer to Peer*).

Juntas estas ferramentas, elementos de federação e agregadores: RSS, Atom e OpenID, que apesar de facilitadores, do ponto de vista da organização semântica dos dados, nem sempre auxiliaram sua construção, pois na busca por palavras e no uso de *mashups*, as marcações de metadados eram feitas aleatoriamente pelos usuários, e não auxiliam esta organização semântica do ambiente geral da Web, ainda que possam ajudar grupos isolados a determinar seus próprios metadados.

Se por um lado este movimento parecia criar ilhas de aplicações que pareciam confundir ainda mais o ambiente da Web global e tornar a tarefa de construção de uma Web Semântica ainda mais difícil, para não dizer impossível, a agregação de usuários não desenvolvedores através de milhares de *Web logs*, ferramentas de mídias sociais (incorretamente chamadas de Redes Sociais), criou uma rede de interesse geral na Web e de necessidades de organização destes ambientes.

Também a estruturação dos bancos de dados vai sofrer uma nova mutação, já que os primeiros ambientes SaaS (*Software as a Service*) apontam para bases semi-estruturadas, diferentemente dos Bancos Estruturados das linguagens tipo SQL, que justamente por isto são chamadas No-SQL. Proliferaram os *Blogs*, as ferramentas de *e-commerce*, as mídias sociais e com elas novas necessidades de estruturação e organização. A Figura 3 apresenta o desenvolvimento aproximado desde a Web 1.0 e futuras etapas de desenvolvimento da Web Semântica, incluindo uma nova etapa definida como Web 4.0.

**Figura 3 – O Futuro da Web Semântica**



A Web Semântica é uma Web de Dados - de dados, títulos, identificações, propriedades e quaisquer outros dados que se possa produzir. As tecnologias da Web Semântica (RDF, OWL, SKOS, SPARQL) fornecem um ambiente onde uma aplicação pode consultar esses dados e ou fazer inferências usando vocabulários controlados.

No entanto, para tornar a Web de Dados uma realidade, é importante ter esta enorme quantidade de dados na Web disponíveis em um formato padrão, acessíveis e gerenciáveis pelas ferramentas da Web Semântica. Além disso, não só a Web Semântica precisa de acesso aos dados, mas as relações entre os dados devem ser disponibilizados, para poder criar uma verdadeira Web de Dados (em oposição a uma mera coleção compartilhada de dados). Esta coleção de conjuntos de dados inter-relacionados na Web também pode ser referida como dados vinculados ou Linked Data..

Para alcançar e criar Linked Data, as tecnologias devem estar disponíveis para um formato comum (RDF), para fazer qualquer conversão ou acesso em tempo real para bancos de dados existentes (relacionais, XML, HTML). Também é importante ser capaz de estabelecer consultas terminais de configuração para acessar esses dados de forma mais conveniente. W3C fornece tecnologias (RDF, GRDDL, POWDER, RDFa , R2RML, RIF, SPARQL) que permitem acessar esses dados.

Deve-se observar que, do ponto de vista da busca, a Web 1.0 tinha possibilidade de uma busca por palavras e mecanismos de colaboração limitados (Linghtweight Colaboration), que a Web 2.0 potencializou a partir dos portais, dos Wikis e dos mecanismos de compartilhamento pelas mídias sociais. A partir de então a busca Semântica começa a se tornar possível, permitindo uma verdadeira colaboração – Web 3.0. Um ambiente com agentes inteligentes, feitos pela máquina e possibilitando uma maior interação homem-máquina nos levará a Web 4.0.

### **3. A Web 3.0 e *Linked Data***

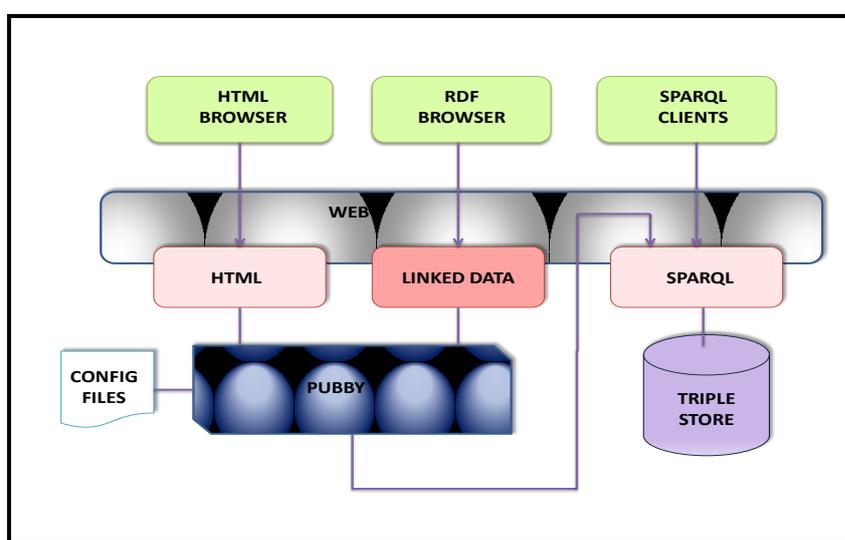
Surgem as primeiras bases semânticas, e o ambiente, aparentemente então complexo e de difícil organização, dá origem a um novo desenho para a Web Semântica, que é definido por Hendler, como o surgimento real da Web 3.0

Segundo Hendler, o melhor artigo na conferência de VLDB2007 foi sobre o uso de um SGBD para gerenciamento de dados da Web Semântica (ABADI, 2007), mas depois disto já surgiu

um banco de dados emergente, o SPARQL, feito justamente para a Web Semântica já com resultados práticos relevantes.

Estas tecnologias emergentes integradas na Web Semântica começam a produzir resultados, desde aplicações básicas, usando a descrição RDF (no âmbito da descrição de recursos), vincularem dados de vários sites da Web usando uma linguagem padrão SQL, a SPARQL que consulta RDF até ligações já prontas em XML ou ontologias em OWL (Cyganiak, 2014). A Figura 4 ilustra esta arquitetura:

**Figura 4 - Configuração de Servidor SPARQL Pubby Terminal**



Fonte: <http://wifo5-03.informatik.uni-mannheim.de/pubby> - adaptação dos autores

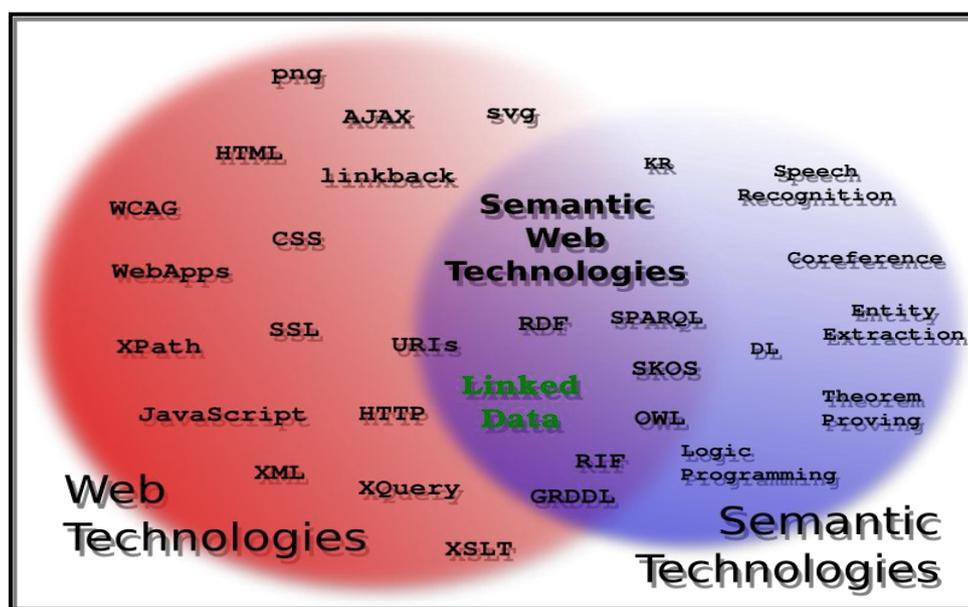
*Linked Data* está associado ao uso do *Resource Description Framework* (RDF) e *Hypertext Transfer Protocol* (HTTP) para publicar dados estruturados na Web e para conectar dados entre diferentes fontes de dados, permitindo efetivamente que dados em uma fonte de dados possam estar ligados a dados em outra fonte de dados. Os princípios de dados vinculados foram delineados pela primeira vez por Berners-Lee em 2006, e proporcionou ampla orientação sobre a qual os editores de dados começaram a perceber a Web de Dados. Esta orientação tem sido tema de pesquisas e desenvolvimento que utilizam as melhores práticas emergentes da comunidade *Linked Data*. O objetivo é criar ferramentas e diretrizes associadas para exposição, compartilhamento e interligação de recursos de dados na forma de *Linked Open Data*.

Um *Triple Store* é uma estrutura de banco de dados usada para armazenar e consultar dados

RDF. Recentemente, pode-se observar pelas publicações científicas que tem havido uma importante iniciativa de desenvolvimento em processamento de consultas, protocolos de acesso em tecnologias triple-store (dado composto por sujeito-predicado-objeto).

A ideia é oferecer um ambiente que seja simples ao usuário da informação em um contexto Web 3.0, a fim de disponibilizar seus recursos de dados e associá-los a outros recursos existentes, estimulando a publicação de dados na Web. Constituem funcionalidades esperadas deste ferramental mecanismos de importação, tratamento, inter-relacionamento com recursos na Web, anotação e referência a mecanismos de controle terminológico (ontologias). A figura 5 identifica *Linked Data* no contexto tecnológico:

**Figura 5 - *Linked Data* - Contexto Tecnológico**



Fonte: Multimedia and Linked Data

*Linked Data* ou Dados Vinculados descrevem um método de publicação de dados estruturados de modo que possam ser interligados e tornar-se mais útil. Ele se baseia em tecnologias Web, tais como HTTP, RDF e URIs, mas ao invés de usá-los para servir páginas web para usuários, estende-os a compartilhar informações de uma forma que pode ser lido automaticamente por computadores. Isso permite que os dados de diferentes fontes possam ser conectados e consultados (BIZER, 2009).

A *World Wide Web* permitiu a criação de um espaço de informação global que compreende

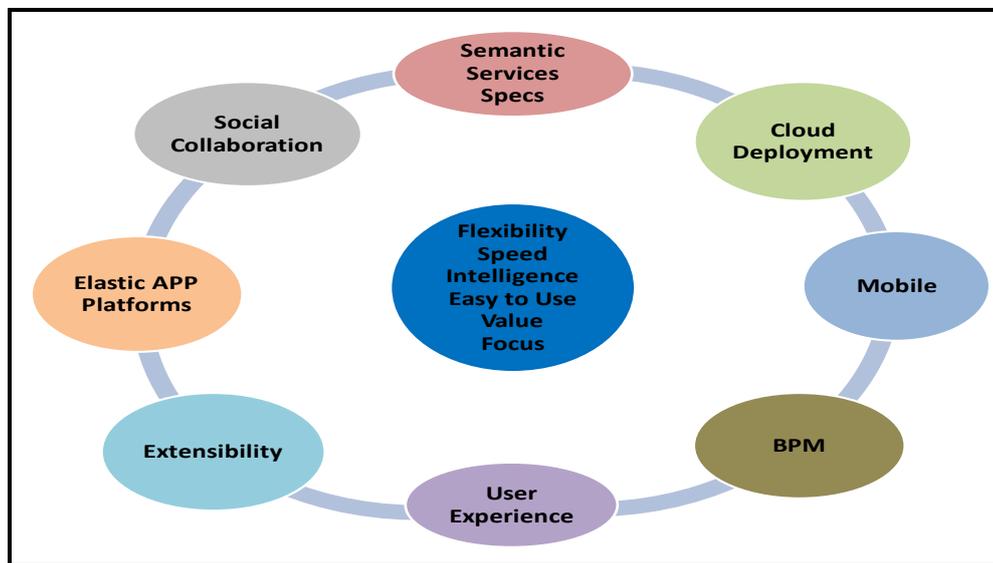
os documentos vinculados. À medida que a Web torna-se cada vez mais relevante nos processos de busca, apropriação e uso da informação, há um desejo crescente por acesso direto a dados brutos não disponíveis atualmente na Web ou vinculados em documentos de hipertexto. *Linked Data* fornece um paradigma de publicação em que não só os documentos, mas também os dados passam a ter relevância na Web, permitindo assim a extensão da Web com um espaço global de dados baseado em padrões abertos - a *Web of Data*.

A Web de Dados pode ser acessada usando navegadores *Linked Data*, da mesma forma que a Web tradicional de documentos é acessado usando navegadores HTML. No entanto, em vez de seguir ligações entre páginas HTML, navegadores *Linked Data* permitem que os usuários naveguem entre diferentes fontes de dados seguindo os *links* RDF. Isto permite que o usuário inicie o processo de busca com uma fonte de dados e, em seguida, mover-se através de uma potencial interminável de fontes de dados WEB, conectados por links RDF. Assim como o documento Web tradicional pode ser rastreado pelo seguinte *link* hipertexto, a Rede de Dados pode ser rastreada por links RDF. Trabalhando com os dados pesquisados, os motores de busca podem fornecer recursos de consulta sofisticados, semelhantes aos fornecidos pelos bancos de dados relacionais convencionais uma vez que os resultados da consulta são dados estruturados e não apenas *links* para páginas HTML, eles pode ser imediatamente processados, permitindo assim que uma nova classe de aplicações baseadas na Web de Dados. A "liga" que une os documentos na Web tradicional é a hiperligação entre páginas HTML. Na Web de Dados falamos de ligações RDF. Um *link* RDF afirma simplesmente que uma parte dos dados tem algum tipo de relação com o outro pedaço de dados.

Com o objetivo de desenvolver aplicações com base em tecnologias Web, os *Linked Data* servem como uma maneira útil para conectar dados previamente não relacionados e publicar dados estruturados na web. *Open-Linked Data* e aplicações da Web Semântica reúne novas teorias, resultados de pesquisas e estudos de casos que abrangem os desenvolvimentos recentes e as abordagens para aplicação de dados vinculados abertos e a rede semântica no contexto da ciência e sistemas de informação. A tecnologia está reformulando a proposta de valor no desenvolvimento de aplicativos com prioridade na consolidação, racionalização e arquiteturas baseadas na Web (HOPPERMANN, 2014). A Figura 6 apresenta as principais tendências de tecnologia que vão remodelar a forma que as equipes de desenvolvimento de aplicativos irão projetar, desenvolver, e selecionar aplicativos

de software na Web Semântica:

**Figure 6 – Tendências Tecnológicas - Desenvolvimento Web Semântica**



Fonte: Forrest Tesearch, Inc – adaptacao do autor

## 4. O Projeto DBpedia

Um dos projetos que demarcam o surgimento da Web 3.0 é o DBpedia, que permite extrair informações estruturadas na Wikipedia, tornando-as informações disponíveis na Web de modo semântico e estruturado. O tutorial *“How to Publish Linked Data on the Web”* (BIZER, et. al., 2007) apresenta como estruturar linked data a partir da Web Semântica e orienta o projeto de estruturação semântica da Web a partir da Wikipedia.

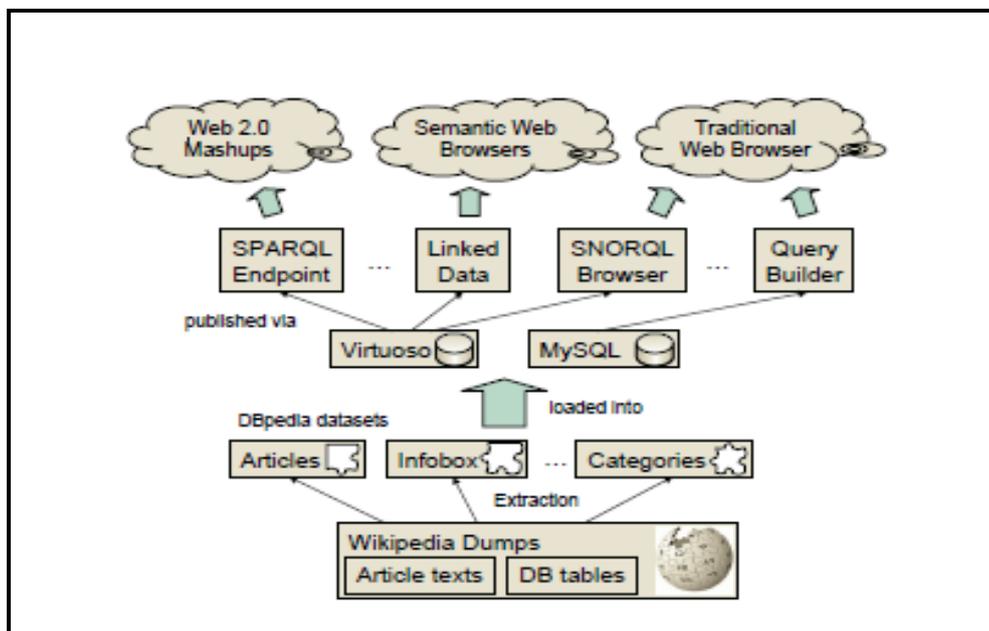
Esclarecem neste tutorial que dados ligados (*Linked Data*) é um método de publicação de dados RDF na Web que estão baseados no uso de identificadores URI (do tipo `http://URI`) como identificadores para recuperar as descrições de recursos. Conforme explicam os autores o DBpedia foi feito a partir do Wikipedia, uma vez que este ambiente apenas permite a pesquisa de texto completo, mas conforme se sabe este ambiente tem muitas propriedades difíceis para editar dados de forma colaborativa, além de ter dados contraditórios, convenções taxonômicas inconsistentes, erros e até mesmo *spam*. O projeto DBpedia é interessante por que faz esta conexão da Web 2.0 (que incorpora facilidades aos

usuários) com a Web 3.0, e neste sentido o projeto faz uma conexão forte destes dois desenvolvimentos da Web (BIZER, et. al., 2007):

- Desenvolvimento de uma estrutura de extração de informação, que converte o conteúdo Wikipedia para RDF. Os componentes básicos formam uma base sobre a qual ainda as pesquisas sobre a extração de informação, *clustering*, gerenciamento de incertezas e processamento de consultas possam ser realizadas.

- Usando a base de conteúdos da Wikipedia como um grande conjunto de dados RDF, multi-domínio, que pode ser usado numa variedade de aplicações da Web Semântica. Pode-se interligar o conjunto de dados DBpedia com outros conjuntos de dados abertos. Isto resulta numa Web de Dados contendo algo em torno de 2 bilhões de "triplas RDF" [em 2007], que podem ser acessados por meio de outros serviços da Web e ligado a outros sites. Uma indicação do alcance e escala da "nuvem" Linking Open Data é apresentado na Figura 7.

Figura 7 – **Visão Geral de Ferramentas DBpedia**



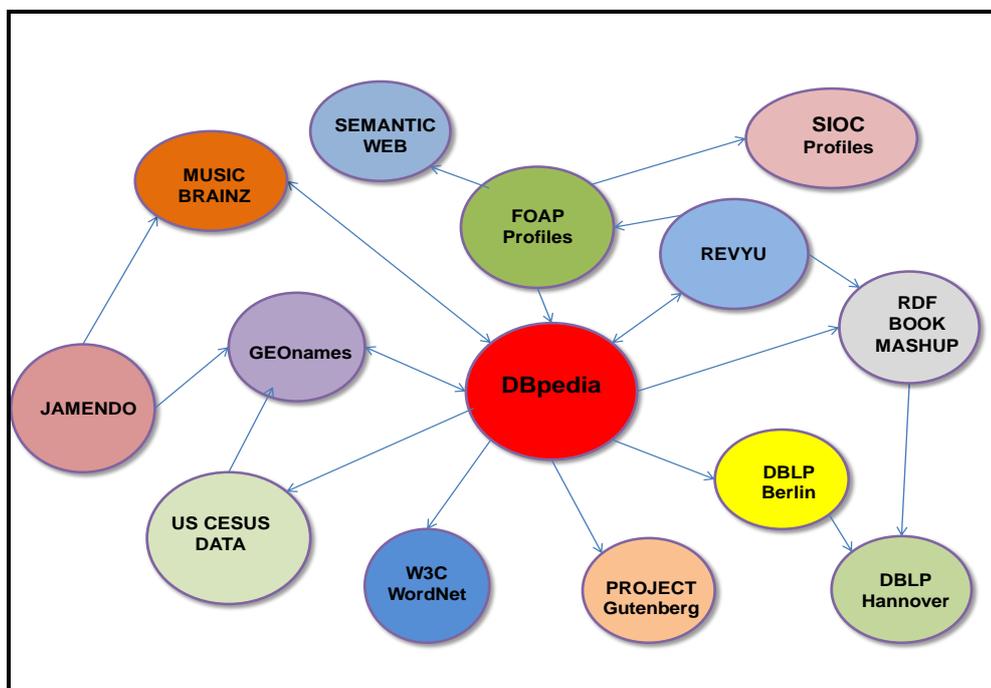
Fonte: Bizer, Richard Cyganiak, Heath, 2007

A Figura 7 mostra centros de interligação que são sites como DBpedia e Geonames. DBpedia

extrai "triples" RDF a partir de "infoboxes" comumente encontrados no lado direito de artigos da Wikipédia, e faz com que estes estejam disponíveis na Web em RDF a ser rastreado ou consultado com SPARQL. *Geonames* por sua vez, fornece descrições RDF de milhões de localizações geográficas em todo o mundo. DBpedia e *Geonames* fornecem URIs (e descrições RDF) para muitas coisas as quais se deseja consultar. Como esses URIs são reutilizados em outros conjuntos de dados, DBpedia e *Geonames* são hubs onde um número cada vez maior de conjuntos de dados são ligados, aumentando assim o potencial da rede. DBpedia é uma importante fonte de dados abertos na web e considerando a interligação do DBpedia com outras fontes de dados, esta poderá servir como um núcleo para a emergente Web de Dados.

Alem de projetos conhecidos, como FOAF *profiles* que indicam os usuários ligados as referencias do DBpedia, *Geonames*, *US Census Data*, os diversos DBLP, o histórico projeto *Gutenberg* e outros, novos diretórios como *Flicker Exporter* e *Audio-Scrobbler* poderão agregar usuários comuns e tornar os dados em torno do projeto DBpedia ainda mais atraentes e importantes para a reestruturação da Web Semântica (AUER, 2014). A Figura 8 apresenta uma visão geral sobre os conjuntos de dados que atualmente são interligados com DBpedia:

**Figura 8 - Linking Open Data em um Projeto DBpedia**



Fonte: DBpedia: A Nucleus for a Web of Open Data: adaptação dos autores

A fim de permitir aos usuários DBpedia a busca por novas informações, o conjunto de dados

DBpedia está interligado com várias outras fontes de dados na Web utilizando links RDF que permitem aos usuários da informação navegar a partir de dados dentro de uma fonte de dados para os dados relacionados dentro de outras fontes usando um navegador da Web Semântica. *Links* RDF podem fornecer sofisticados recursos de pesquisa e de consulta sobre os dados rastreados. O projeto DBpedia atua como um “linking hub” que permite a um conjunto de dados massivos e ontologias, como o *US Census*, *Geonames*, *MusicBrainz*, *the DBLP Bibliography*, *WordNet*, *Cyc* e muitos outros, interoperáveis na Web Semântica.

O futuro também é promissor para Bibliotecas Digitais e ambientes de organização e disseminação como centros culturais e ambientes educacionais, projetos como o VIAF (Base Virtual Internacional de Autoridade) está em plena expansão com a adesão de muitas Bibliotecas e Centros Informacionais em todo o mundo.

Não há dúvida que a Web 3.0 está em pleno desenvolvimento, e de sua estruturação dependerá o futuro da Web e a incorporação de mecanismos automáticos de *software*, que permitam agentes inteligentes trabalharem tornando a Web mais organizada e a informação que aí está disponível mais facilmente encontrada e podendo estabelecer critérios de relevância.

O futuro da Web Semântica depende de sua estruturação tecnológica como uma plataforma, e da incorporação do trabalho de instituições e profissionais da área da Ciência da Informação, capazes de auxiliar no processo de padronização e estabelecimento de critérios de relevância.

Os esforços do consórcio W3C e de instituições a ele agregadas podem dar uma direção mais clara para um pleno desenvolvimento da Web Semântica, em sua versão 3.0.

## 5. Conclusões

As TICs – Tecnologias da Informação e Comunicação, em particular a Internet, vem promovendo intensas e profundas modificações na sociedade e nas organizações. Fenômeno interessante nessa revolução, que a distingue de outras anteriores, é o fato de a mesma estar sendo incorporada de forma “natural” e quase automática pela sociedade, o que se observa não apenas quanto a tecnologia, mas também quanto aos serviços prestados, em particular aos associados a Web.

A Web inicialmente construída apenas para publicações abertas *online*, com seu rápido

crescimento criou a necessidade de sua estruturação semântica.

O projeto inicial apontado por Tim Berners-Lee e outros (2001) já indicava para uma Web Semântica com necessidades de estruturação, através de linguagens de marcação como XML e RDF, agentes inteligentes e ontologias, porém este cenário somente começa a se desenhar a partir dos dados ligados (*Linked Data*) e projeto como o DBPedia que estabelecem milhões de "triplas" – base de dados RDF - a partir de um banco de dados pronto.

Os artigos de James Hendler apontam para uma emergência e um novo desenho para a Web 3.0, entre eles o desenvolvimento da DBPedia, além de outros ambientes crescentes da Web Semântica como o projeto VIAF (*Virtual International Authority File*), mostram a realidade e viabilidade atual dos projetos Web 3.0.

O uso de triplas: URIs para denotar coisas, o endereço HTTP para dizer como elas podem ser encontradas e referenciadas e o uso de bases SPARQL feitas nos padrões RDF, estabelecem as ligações dos dados na Web, facilitando a sua publicação e busca semântica.

O uso de *Linked Data* tornou este ambiente possível e o uso da enciclopédia *on-line* Wikipedia criou rapidamente milhões de "triplas" de relações dadas (sujeito, predicado e objetos) efetivamente tornou-se possível a partir do uso de um banco de dados com uma estrutura conceitual pronta e uma estabelecendo a ligação semântica destes dados.

Assim como o DBPedia, outros projetos foram desenvolvidos, como o FOAF, que descreve a relação de pessoas, *Geonames*, com descrição RDF para mais de 7.5 milhões de relações entre características geográficas em todo mundo, o CKAN, registro de conteúdos e pacotes abertos fornecido pela *Open Knowledge Foundation* e muitos outros que estão sendo estruturados a partir do conceito de *Linked Data*.

No ambiente da Ciência da Informação, deve-se destacar o projeto VIAF, que a partir de Bibliotecas é operado pela prestigiosa OCLC-*Online Computer Library Center* e incorpora bibliotecas do mundo, realizando registro de autoria, conferindo a Web 3.0 além da busca semântica, uma maior precisão de autoridade sobre as informações.

O número de projetos e esforços agregados a Web 3.0 confere ao desenvolvimento da Web Semântica como uma realidade mas também prepara um futuro desenvolvimento para a Web com agentes inteligentes e mecanismos de *software* para auxílio na construção e organização do conhecimento. Fica evidente a necessidade de estudos e pesquisas que permitam entender o processo de busca, acesso, e apropriação da informação na Web 3.0.

## 6. Referências Bibliográficas

---

ABADI, D. J. et al. (2007). Scalable Semantic Web Data Management Using Vertical Partitioning. *Proc. 33 rd Conf. Very Large Data Bases, VLDB Endowment*, 2007, p. 411-422.

AUER, Soren; BIZER, Christian; KOBILAROV, Georgi; LEHMANN, Jens; CYGANIAK, Richard; IVES, Zachary (2014). *DBpedia: A Nucleus for a Web of Open Data*. Url: < <http://www.cis.upenn.edu/~zives/research/dbpedia.pdf> > ( Acesso: 24-11-2014).

ABERER, Karl; CHOI, Key Sun; NOY, Natasha; ALLEMANG, Dean; LEE, K.; NIXON, Lyndon; GOLBECK, Jennifer; MIKA, Peter; MAYNARD, Diana; MIZOGUCHI, Riichiro; SCHREIBER, Gus.; MAUROUX, Philippe Cudre (2007). *The Semantic Web*. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 6, ASIAN SEMANTIC WEB CONFERENCE, 2, Busan, Korea, 2007. Proceedings. Busan, Korea : ISWC/ASWC, 2007.

BERNERS-LEE, Tim (1989). *Information Management: a proposal*. Url: <http://www.nic.funet.fi/index/FUNET/history/mbase/w3c/proposal.html> (Acesso: 24-11-2014).

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora (2001). The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American Magazine*, May, 2001. Url: <http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html> (Acesso: 19-11-2014).

BIZER, Christian, CYGANIAK, Richard, and HEALTH, Tom (2007). *How to publish linked data on the web*. Url: <http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/> (Acesso:12-11-2014).

BIZER, Christian; HEATH, Tom; IDEHEN, Kingsley; BERNERS-LEE, Tim (2008). Linked Data on the Web (LDOW2008). In: *INTERNATIONAL WORLD WIDE WEB CONFERENCE*, 17, Beijing, China, 2008. Workshop. Beijing : ACM, 2008.

BIZER, Christian; LEHMANN, Jens; KOBILAROV, Georgi; AUER, Sören; BECKER, Christian; CYGANIAK, Richard; HELLMANN, Sebastian. (2009). DBpedia – A crystallization Point of Web of Data. Url: [http://jens-lehmann.org/files/2009/dbpedia\\_jws.pdf](http://jens-lehmann.org/files/2009/dbpedia_jws.pdf) (Acesso: 25-11-2014).

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, v.5, n.3,p.1-22. Url: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>

(Acesso: 24-11-2014).

CYGANIAK, Richard; BIZER, Chris. Pubby - A Linked Data Frontend for SPARQL Endpoints. Url: <http://wifo5-03.informatik.uni-mannheim.de/pubby/>. (Acesso: 19-11-2014).

GERBER, Aurona; VAN DER MERWE, Alta; BARNARD, Andries (2008). A functional Semantic Web Architecture. In: EUROPEAN SEMANTIC WEB CONFERENCE ON THE SEMANTIC WEB ESWC2008, 5, Tenerife, Canary Islands, Spain, 2008. Proceedings. Tenerife, Canary Islands, Spain, 2008. p.273-287. Url: <http://ksg.meraka.org.za/~agerber/Paper152.pdf> (Acesso: 24-11-2014).

HENDLER, Jim (2009). Web 3.0 Emerging. *Computer*, v.42,n.1 january 2009, p.111-113. Url: <http://dl.acm.org/citation.cfm?id=1512177> .(Acesso:24-11-2014).

HENDLER, James.(2010). Web 3.0: The Dawn of Semantic Search, *IEEE Computer*, january, 2010. Url: <http://ieeexplore.ieee.org/xpl/abstractKeywords.jsp?reload=true&arnumber=5398788> (Acesso: 25-11-2014).

HOPPERMANN , Jost; HAMERMAN, Paul D; LAWRIE, George (2013). The 10 Most Important Technology Trends in Business Application Architecture Today  
Url:  
[http://www.progress.com/~media/Progress/Documents/Pacific/Whitepaper/The\\_10\\_Most\\_Important\\_Tec.pdf](http://www.progress.com/~media/Progress/Documents/Pacific/Whitepaper/The_10_Most_Important_Tec.pdf)> (Acesso: 18-11-2014).

MARKOFF, John (2006). Entrepreneurs See a Web Guided by Common Sense. Url: [http://www.nytimes.com/2006/11/12/business/12web.html?\\_r=3&adxnnl=1&oref=slogin&adxnnlx=1385031867-PLsmlu8dAHscj5HcURWNug&](http://www.nytimes.com/2006/11/12/business/12web.html?_r=3&adxnnl=1&oref=slogin&adxnnlx=1385031867-PLsmlu8dAHscj5HcURWNug&) (Acesso: 18-11-2014).

MATTOS, Raynner Dourado. (2014) O Futuro da Web. Url:  
<http://www.designrdm.com/category/web-30/> .(Acesso: 20-11-2014).  
<http://corporate.canaltech.com.br/noticia/internet/Conheca-as-5-tendencias-para-o-futuro-na-web/>

MUCHERONI, Marcos L. (2014). Realidade e fantasias da Web 3.0. Url:  
[http://marcosmucheroni.pro.br/blog/?p=9663#.VHt9ekLF\\_fI](http://marcosmucheroni.pro.br/blog/?p=9663#.VHt9ekLF_fI)  
(Acesso: 26-11-2014).

MUCHERONI, Marcos L. (2014). Mais sobre Web 3.0. Url:  
[http://marcosmucheroni.pro.br/blog/?p=9704#.VHt9y0LF\\_fI](http://marcosmucheroni.pro.br/blog/?p=9704#.VHt9y0LF_fI)  
(Acesso: 26-11-2014).

W3C Recommendation (2004). *OWL Web Ontology Language: Use Cases and Requirements*. Url: <http://www.w3.org/TR/webont-req/>> (Acesso: 19-11-2014).