



XVI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB)
ISSN 2177-3688

GT 8 – Informação e Tecnologia
Comunicação Oral

WEB SEMANTICA: INTRODUÇÃO A RECURSOS DE VISUALIZAÇÃO DE DADOS EM FORMATO GRÁFICO¹

SEMANTIC WEB: INTRODUCTION DATA VIEW RESOURCES IN GRAPHIC FORMAT

Jose Eduardo Santarem Segundo, USP
santarem@usp.br

Jessica Oliveira de Souza, UNESP
osz.jessica@gmail.com

Caio Saraiva Coneglian, UNESP
caio.coneglian@gmail.com

Resumo: A Web Semântica tem provocado mudanças na estrutura e ligação entre dados disponibilizados na Internet, a fim de permitir que estes dados atendam de forma mais adequada às necessidades dos usuários. Grande parte dos estudos está diretamente ligada à estrutura e armazenamento dos dados em formato semântico, entretanto, posteriormente a este processo, a apresentação destas informações para os usuários é deficitária. Esta pesquisa tem como objetivo analisar novos meios de apresentação das informações de aplicações que utilizam a Web Semântica como base. Para tal, realizaram-se procedimentos exploratórios para analisar tais tecnologias. Ao final, verificou-se há poucos recursos disponíveis para apresentação de dados em formato gráfico de ambientes baseados em tecnologias da Web Semântica. Identificou-se a necessidade de investimento e desenvolvimento de novas ferramentas e plataformas que tornem mais interativas as possibilidades de acesso aos dados disponibilizados em formato semântico.

Palavras-chave: Web Semântica, Visualização da Informação, Aplicações Semânticas, Acesso a Dados

Abstract: The Semantic Web has caused changes in the structure and connection between data published on the Internet in order to allow these data more adequately meet the needs of users. Most studies is directly linked to the structure and storage of data in semantic format, however, later in this process, the presentation of information to users is in deficit. This research aims to analyze new ways

¹ O conteúdo textual deste artigo, os nomes e e-mails foram extraídos dos metadados informados e são de total responsabilidade dos autores do trabalho.

of presenting the application information using the Semantic Web as a base. There were exploratory procedures for analyzing such technologies. At the end, we found a few resources available for data presentation in graphical format based environments Semantic Web technologies. It identified the need for investment and development of new tools and platforms making it more interactive possibilities of access to data available in semantic format.

Keywords: Semantic Web, Information Visualization, Semantic Applications, Data Access.

1 INTRODUÇÃO

Representar as informações, usando estruturas que contribuam para a construção de novos conhecimentos do usuário é fundamental para qualquer processo, foi assim desde os documentos até as redes, tema do Enancib 2015.

A Web Semântica tem provocado mudanças profundas na Web, quando se trata da disponibilização e processamento dos dados, levando em consideração o domínio na qual as informações e o usuário estão inseridos, sendo possível que a Web trabalhe auxiliando o usuário em suas tarefas.

Desde seu princípio em 2001, a Web Semântica tem sido moldada e implementada por diversas organizações e padronizada pelo W3C (World Wide Web Consortium). Inúmeras iniciativas têm mostrado as possibilidades que podem ser realizadas com o conjunto de conceitos e tecnologias baseados na Web Semântica. Tais iniciativas estão sendo cada vez mais utilizadas, mostrando que a proposta da Web Semântica, é algo real e aplicável.

A proposta da Web Semântica tem como fundamento a ideia de maximizar o uso e integrar o grande volume de informações disponibilizados na Web para que possam ser recuperadas e utilizadas por meio de agentes automatizados. Este processo só é possível por intermédio do uso de tecnologias que definem uma estrutura de representação dos documentos, de forma que agentes computacionais sejam capazes de “compreender” os dados, beneficiando a utilização das informações pelos usuários.

Segundo Heath (2012), o Linked Data é um conjunto de melhores práticas para publicação e conexão de dados estruturados na Web, permitindo estabelecer links entre itens de diferentes fontes de dados para formar um único espaço de dados global.

O conteúdo disponibilizado no Linked Data é publicado através de datasets, que são as bases de dados que compõe a estrutura global de informações. Os dados dos datasets estão ligados entre si e também estão ligados com os dados de outros datasets, formando um grande emaranhado de ligações. A quantidade de dados publicados vem crescendo exponencialmente, no último levantamento realizado, no ano de 2014, possuíam mais de 570 *datasets* vinculados ao projeto *Linked Data* (CYGANIAK; JENTZSC, 2014).

O grande conjunto de dados interligados que vem aumentando de forma rápida e consistente, baseados em tecnologias e conceitos da Web Semântica, estão, em sua grande maioria, ligados a contextos e domínios específicos de informação, aproximando os usuários de suas necessidades informacionais.

O uso de interfaces mais adequadas e recursos gráficos podem auxiliar na demonstração das informações que são relevantes ao usuário. Estes recursos conseguem tornar mais eficiente o processo de acesso e recuperação da informação.

As imagens digitais têm provocado melhoras ao acesso intelectual, pois elas podem agregar elementos que contribuem tanto para indexação quanto recuperação dos documentos (WATERS, 1997). Além disso, as imagens passam a ser cada vez mais utilizadas como disseminadoras de informações, pela facilidade como as imagens podem ser compartilhadas e disponibilizadas na Web (SIMIONATO, 2012).

Neste sentido, a área de Visualização da Informação tem tido um papel fundamental com esta evolução das interfaces gráficas dos computadores, para que estas tecnologias possam agregar no processo de transmissão dos conhecimentos. Possibilitando uma melhora na interação do usuário com as informações que estão sendo disponibilizadas a ele.

As duas questões tratadas anteriormente, Visualização da Informação e Web Semântica, estão diretamente interligadas. É fundamental que os sistemas que utilizam conceitos e tecnologias da Web Semântica possam apresentar os dados, permitindo que o usuário consiga visualizar as informações ali disponíveis. Sendo possível disponibilizar para o usuário todas as estruturas e relações existentes de um determinado recurso.

Existe uma grande preocupação em adaptar as informações para a Web Semântica, para que a interação entre o usuário e os computadores seja aperfeiçoada, tornando a Web mais interativa. Entretanto, posteriormente a disponibilização destes dados, existe dificuldades em consumi-los e principalmente de apresentá-los em um formato adequado aos usuários. Algumas iniciativas têm criado novos meios para a apresentação das informações que estão inseridos dentro do contexto da Web Semântica, onde o usuário consegue ter uma visão mais ampla dos dados, auxiliando na construção de novos conhecimentos.

Dadas as dificuldades em apresentar os dados baseados na Web Semântica aos usuários, e tendo como parâmetro que a Ciência da Informação está na área de Ciências Sociais Aplicadas, o objetivo principal desta pesquisa é identificar e analisar ferramentas ou plataformas de apresentação gráfica de informações que utilizam a Web Semântica como centro de seus processos. O formato gráfico é privilegiado para a visualização dos dados, a fim de identificar alternativas mais adequadas de apresentação e consumo destes dados aos usuários. A

metodologia utilizada neste trabalho foi uma pesquisa bibliográfica com posterior procedimento exploratório perante as tecnologias. O procedimento exploratório foi dividido em duas partes: (1) identificação de ferramentas disponíveis para disponibilização de dados em formato gráfico; (2) análise de integração entre datasets e ferramentas através de procedimentos de consultas baseados em SPARQL (Simple Protocol and RDF Query Language).

A organização do trabalho aconteceu da seguinte maneira: tecnologias e componentes da web semântica são discutidos na seção 2. Tipos e formas de visualização de informações semânticas na web são abordados na seção 3. Na seção 4 são analisadas ferramentas de apresentação de informações para web semântica seguida pelas considerações finais.

2 WEB SEMÂNTICA

A web quando criada em 1989, tinha como objetivo permitir o fácil acesso, intercâmbio e recuperação de informações e acabou crescendo exponencialmente, totalmente descentralizado. Disponibilizou-se uma grande quantidade de documentos, preparados somente para a leitura humana (SOUZA; ALVARENGA, 2004).

A ação de inserir documentos somente para a leitura humana foi contrário a proposta inicial da Web, que dizia que além dos seres humanos, as máquinas também deveriam conseguir realizar a leitura e trabalhar com os dados contidos na Web. Tim Bernes-Lee, então, propôs o modelo da Web Semântica em 2001, que definia como esta poderia ajudar o usuário a fazer tarefas cotidianas, usando agentes computacionais que auxiliariam o usuário. Assim, as informações que o usuário necessitava estariam todas interligadas, e de uma maneira interoperável.

É necessário entender que a Web Semântica, não é uma nova Web, ou uma Web separada, mas sim uma extensão da atual, que tem como função oferecer uma maior interação entre o usuário e os computadores (BERNERS-LEE, 2001). O W3C descreve a Web Semântica como a Web dos dados onde, por meio de padrões e tecnologias específicas, máquinas conseguem atribuir significado semântico aos documentos.

Com o uso das tecnologias da Web Semântica, é possível construir aplicações semânticas. Estas aplicações permitem que os agentes realizem tarefas mais úteis para os usuários, ou seja, consigam auxiliar o ser humano nas tarefas básicas de seu dia-a-dia, estabelecendo inferências sobre os dados, e desta maneira, construindo um ambiente colaborativo de verdade. (W3C, 2015).

2.1 TECNOLOGIAS DA WEB SEMÂNTICA

Após a proposta inicial, a Web Semântica vem se difundindo, e vários conceitos,

tecnologias e funcionalidades foram agregados a esta proposta, com o intuito de construir ambientes semânticos. Entre essas tecnologias, destacam-se o RDF (Resource Description Framework), XML (eXtensible Markup Language), OWL (Web Ontology Language), SPARQL e vários outros conceitos que são descritos pelo W3C. Todas estas tecnologias tornam possível a materialização da Web Semântica, e estão ligadas ao processo de armazenamento e construção das informações (SANTAREM SEGUNDO, 2014).

Um destes conceitos são as URIs (Universal Resource Identifier). Uma URI é um identificador único para um recurso da Web, como imagens, documentos, arquivos. As URIs são fundamentais dentro da Web Semântica, pois é necessário que um recurso não seja confundido. Na linguagem humana muitas vezes uma palavra pode ter diversos significados, e isto pode causar muitos problemas aos sistemas, assim, para resolver este problema deve-se usar as URIs diferentes para cada conceito (CYGANIAK; JENTZSC, 2014).

A Web Semântica define padrões necessários para definir relações entre os dados, a tecnologia adotada pelo W3C é o RDF, que é considerado como o alicerce da Web Semântica. Este é o padrão para o relacionamento dos dados na Web. Sendo possível fazer várias ligações e inferências utilizando os meios disponibilizados neste padrão (DONG et al., 2014).

O RDF é constituído de três objetos básicos: recursos, propriedades e declarações. Um recurso é uma informação que pode ser identificada por uma URI. Propriedades são as informações que representam as características do recurso, ou seja, são os atributos que permitem distinguir um recurso de outro ou que descrevem o relacionamento entre recursos. A declaração é a constituição da informação completa, que compreende um recurso com suas propriedades e valores para as propriedades (SANTAREM SEGUNDO; VIDOTTI, 2011).

O protocolo SPARQL também se apresenta como recurso importante na aplicação das técnicas da Web Semântica. O SPARQL é um protocolo que permite a realização de consultas a dados publicados em RDF e OWL, e possibilita realizar recuperação de informações de uma maneira semântica, pois funciona levando em consideração as relações que existem entre os dados e os axiomas existentes nas ontologias. É possível utilizar *queries* do SPARQL para verificar as relações existentes e inferir novas relações.

Para representar as estruturas e relações dos conhecimentos são utilizadas as ontologias. As ontologias, no contexto da Web, são descrições formais de relações entre termos e conceitos (HEATH, 2012), sendo estruturas de representação, que conseguem expressar informações semanticamente, devido a possuir relações, propriedades e restrições que possibilitam a realização de inferências a respeito dos dados representados pela ontologia.

Para o uso como tecnologia da Web Semântica, entendemos as ontologias como

artefatos computacionais que descrevem um domínio do conhecimento estruturado através de classes, propriedades, relações, restrições, axiomas e instâncias.

Para a construção e aplicação de ontologias, o W3C indica a linguagem OWL (W3C, 2004).

Os elementos da linguagem OWL são essencialmente classes, propriedades e indivíduos. As classes são representações concretas de um conceito ou de uma entidade. Os indivíduos são os objetos de um domínio. E as propriedades são relações entre as classes e atributos destas classes. Possibilitando também inserir restrições e realizar inferências sobre as relações que existem.

Todas estas tecnologias que tornam possível a materialização da Web Semântica trabalham como base dos sistemas de significados, permitindo a estes realizarem inferências e relacionamentos. Estes conceitos e tecnologias também podem trabalhar em conjunto modelos de apresentação e visualização da informação, através das interligações semânticas.

3 VISUALIZAÇÃO DA INFORMAÇÃO

O avanço dos equipamentos de imagens e dos computadores tem permitido uma representação da informação cada vez mais rica e precisa. O uso de recursos gráficos permite que a apresentação da informação seja muito semelhante ao mundo real. A área de Visualização da Informação vem ganhando destaque com a evolução das interfaces gráficas dos computadores (DIAS; CARVALHO, 2007).

Tal área é definida por Freitas et al. (2001) como a área da ciência que tem como objetivo a representação gráfica possibilitando explorar a capacidade de percepção do usuário, e a partir disto, ser possível a interpretação e compreensão das informações, gerando por fim novos conhecimentos.

Os usuários têm hoje, uma grande quantidade de informação disponível, sendo que parte delas acaba se tornando irrelevantes. Este excesso de informações é uma das maiores preocupações nos mecanismos de busca e recuperação da informação. Portanto, é necessário que as informações sejam selecionadas e que as estruturas de Visualização de Informação apresentem dados que contenham informações com qualidade (DIAS; CARVALHO, 2007).

Dentro de sistemas de informação, apresentar dados com apelo visual gráfico possibilita uma aproximação do conjunto de informações com os usuários.

4 APLICAÇÃO DA VISUALIZAÇÃO DA INFORMAÇÃO EM APLICAÇÕES SEMÂNTICAS

Demonstrar através de recursos gráficos aos usuários todo o processo realizado pelas

aplicações construídas baseadas em conceitos e tecnologias da Web Semântica é um grande desafio. Isto porque, tais aplicações, realizam diversas associações, ligações e inferências para conseguir chegar aos resultados finais. A visualização dos dados destas aplicações também deve ser mais intuitiva, e seguir esta proposta da Web Semântica.

Nesta seção, são apresentados três meios onde aplicações semânticas se relacionam diretamente à Visualização da Informação, o Google Knowledge Graph, o Google Knowledge Vault e o uso de consultas SPARQL com o Google Charts. Explicam-se como estas aplicações funcionam e qual a sua forma de apresentação ou o modo como isto pode ser feito. As três ferramentas apresentadas são produtos da empresa Google, isto mostra que esta empresa se preocupa em apresentar soluções para visualização, de forma que tem sido, atualmente, as principais contribuições da área de visualização, quando se trata de Web Semântica.

Claramente identifica-se que grande parte das pesquisas e empresas tem se preocupado excessivamente com a estrutura dos dados, com pouco investimento na questão da visualização das informações.

4.1 GOOGLE KNOWLEDGE GRAPH

Em 2012, a Google criou um mecanismo chamado de Google Knowledge Graph (Mapa ou Grafo de Conhecimentos). Este mecanismo funciona em cima de um mapa de elementos com as diversas relações existentes. Ao fim de 2012 este mapeamento possuía uma quantidade aproximada de 570 milhões de entidades e 18 bilhões de fatos e conexões (BROWN, 2012).

Esta proposta surgiu com a ideia de inserir semântica ao processo de busca realizado pela Google. Na apresentação do projeto (SINGHAL, 2012) é destacada a necessidade dos buscadores irem além da busca por palavras chaves, ou textos, e conseguir realizar a busca por entidades, pois as pessoas quando realizam uma busca estão querendo saber sobre algo que existe no mundo, e não nos caracteres que um termo contém.

Este mapeamento realiza relações semânticas, assim como é feito em projetos como o DBPedia, onde são encontradas uma grande quantidade de elementos relacionados, sendo possível o usuário acessar e navegar através destas relações. Este mapa contém inúmeros elementos relacionados, com diversos tipos de ligações, como filiação, obra, entre outros.

O mapeamento é realizado utilizando base de dados de informações abertos da Web, destas bases destacam-se a Wikidata, Wikipedia e CIA World Factbook. Porém, o Knowledge Graph não utiliza apenas estas bases, podendo utilizar outras fontes de dados, pois os desenvolvedores do projeto estão focados em ter informações aprofundadas e abrangentes.

O Wikidata passou a ser utilizado no Knowledge Graph substituindo o Freebase

(FREEBASE, 2014). O Wikidata é um banco de dados livre, e nele são reunidas as informações estruturadas dos projetos da Wikimedia, entre eles o Wikipedia, Wikivoyage e Wikisources. Em julho de 2015 o Wikidata continha mais de 14 milhões de itens disponibilizados (WIKIDATA, 2015).

O CIA World Factbook é produzido pelos responsáveis políticos dos Estados Unidos e coordenado pela comunidade de inteligência estadunidense. O Factbook fornece informações sobre a história, pessoas, governo, economia, energia, geografia, comunicações, transporte, militares e questões transnacionais das 267 entidades mundiais (CENTRAL INTELLIGENCE AGENCY, 2015).

O Knowledge Graph inseriu elementos no processo de visualização dos resultados apresentados aos usuários. Atualmente para algumas buscas realizadas no Google, os resultados são apresentados contendo diversas estruturas, além da tradicional sequência de links apresentado pelos buscadores (GOOGLE, 2015).

Basicamente, foram inseridas três estruturas que estão relacionadas ao Knowledge Graph. Forma-se um painel na lateral, em buscas realizadas em computadores desktops, ou juntamente a lista de links resultantes da busca em ambientes *mobile*, como mostrado na Figura 1.

Além disso, em alguns momentos são apresentados um carrossel na parte superior da busca. Também é apresentado na parte inferior ao painel, um local de desambiguação, aparecendo outros significados para os mesmos termos pesquisados.

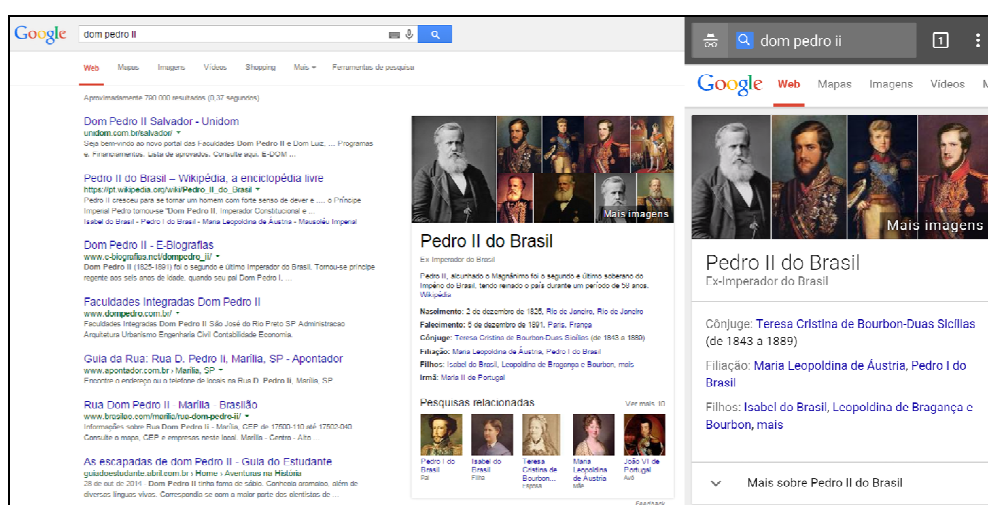


Figura 1: Telas mostrando o painel de informações

Fonte: próprio autor

A primeira estrutura, o painel, aparece associando a busca do usuário a algum objeto real do mundo. Mostrando resumo, datas, relacionamentos e outros tipos de relações, que este objeto apresenta, além de outras possibilidades.

O carrossel é apresentado para mostrar diversos objetos relacionados à busca do usuário. Por exemplo, se fizer pesquisa de museus em São Paulo, aparecerá um carrossel acima dos resultados orgânicos da busca, com os principais museus de São Paulo, como mostrado na Figura 2.

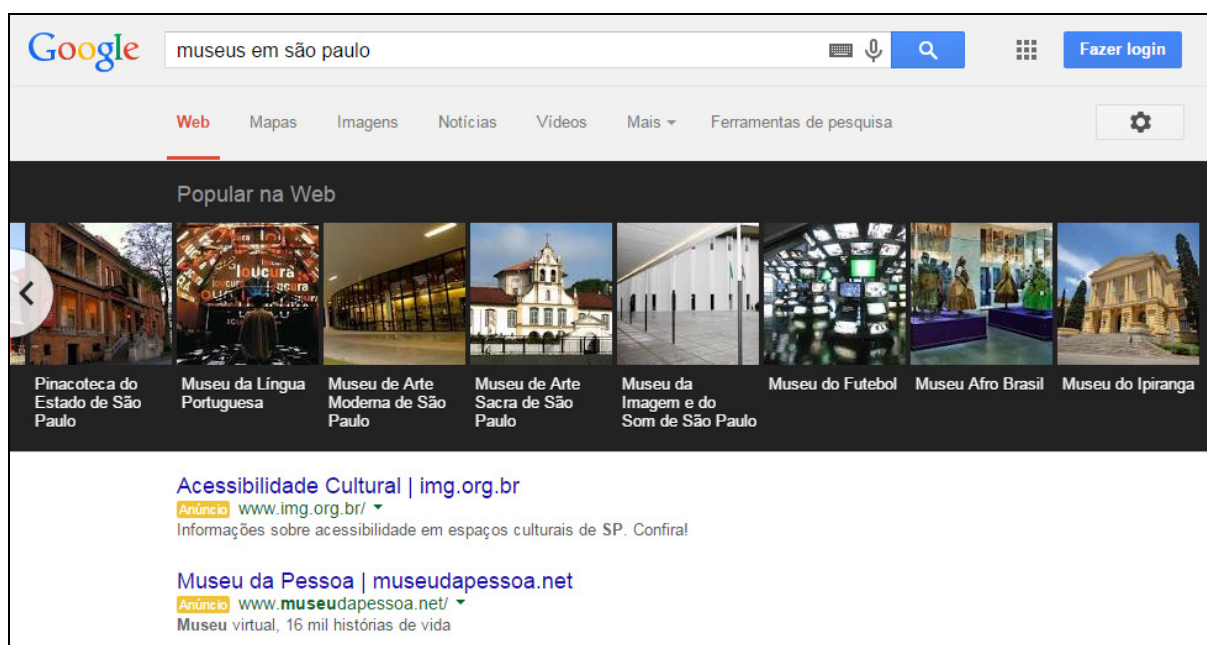


Figura 2: Carrossel mostrando os museus de São Paulo

Fonte: próprio autor

Outra possibilidade de aparecer o carrossel, é no momento que o usuário clica em algum item que aparece na parte de “Pesquisas Relacionadas”, no painel lateral. Este carrossel apresenta as pesquisas relacionadas, marcando em qual item está sendo apresentados os resultados, facilitando o acesso para o usuário clicar em outra pesquisa relacionada.

Por fim os links de desambiguação têm a função de apresentar links cujos os termos sejam iguais ao que o usuário pesquisou, mas com um significado diferente daquele que o buscador entendeu que eram as necessidades informacionais do usuário.

Singhal (2012) diz que este projeto melhora a busca de três modos, para desambiguar resultados; obter o melhor resumo do tema e; poder adquirir novos conhecimentos, indo além da pesquisa inicial.

A primeira função trata da desambiguação dos termos, pois uma palavra pode ter

significados diversos, por isso o Knowledge Graph coloca na página de resultados alguns links desambiguando o significado que aquele termo pode ter. Podendo o usuário, após realizar a busca, apontar em qual sentido ele colocou aquela expressão.

Já a segunda função trata da questão de obter o melhor resumo daquela busca que o usuário realizou. Este resumo contém informações relevantes sobre uma determinada coisa, seja ela pessoa, objeto, animal, entre outros. São colocadas informações específicas para cada tipo de coisa, por exemplo, um escritor será colocado informações referentes à suas obras, porém para um arquiteto será colocado informações a respeito de seus edifícios. Isto ocorre baseado nas estatísticas provenientes do banco de dados de busca, sendo possível verificar quais são as características mais relevantes buscadas sobre determinada pessoa ou objeto, podendo colocar apenas informações que de fato são relevantes àquele resumo.

Por fim, é destacada a questão de adquirir conhecimentos que não era esperado ao início da busca. Isto ocorre quando o usuário se depara com as conexões que o alvo de sua busca possui, é possível que ele adquira conhecimentos que ele não estava esperando. Por exemplo, que uma obra é de um determinado pintor, ou que um determinado escritor se inspirou nos nomes de seus familiares para os nomes das personagens de um livro ou de uma série de televisão.

O processo realizado por este sistema aproxima o usuário dos resultados obtidos, isto porque, quando se realiza uma pesquisa em um buscador tradicional, obtém-se como retorno milhões de páginas que contém algo relacionado à pesquisa. Um ser humano não seria capaz de analisar todos os resultados da pesquisa. Por mais que exista um algoritmo de relevância que irá posicionar resultados com maior probabilidade de atender as necessidades informacionais do usuário nas primeiras posições, é possível que o usuário não consiga aquilo que ele necessita de forma facilitada.

Utilizando-se dos recursos do Google Knowledge Graph, com a apresentação de uma estrutura visual baseado em processos semânticos, o usuário terá acesso a um formato diferenciado e baseado em visualização gráfica, portanto mais interativo e atrativo.

4.2 GOOGLE KNOWLEDGE VAULT

Na busca de encontrar meios de melhorar o processo de descoberta das informações, o Google oferece outro modelo para realizar a recuperação e o mapeamento das informações com o intuito de melhorar a sua busca. Em um artigo publicado, Dong et al. (2014) propõe uma base de conhecimento probabilístico que combina dados espalhados pela Web com informações prévias de repositórios de conhecimento existentes, chamado de Knowledge Vault (KV).

Este processo realiza uma combinação da grande quantidade de informações que estão espalhadas para a Web com dados estruturados, como os dados disponíveis em base de dados Wikidata e Freebase, por exemplo. O sistema, assim, não fica limitado apenas às informações que estão dentro desta base de dados, mas também, utiliza dados desestruturados, ou seja, no formato como a maioria das informações estão disponíveis na Web.

Esta iniciativa se mostra inovadora, pois é um meio de inserir semântica em informações desestruturadas que estão espalhadas pela Web. Há uma união entre dados desestruturados com dados estruturados, tornando diversos destes dados em conjuntos de estruturas RDF, que apresentam um sentido claro para as máquinas.

A extração dos dados ocorre de diversas maneiras dentro da Web. Utilizam-se como fontes: textos livres, árvores estruturadas do HTML, tabelas do HTML e as próprias anotações feitas pelos desenvolvedores das páginas.

Dong et al. (2014) descreve que o KV apresenta algumas características que o diferem de outros projetos. Primeiramente, é tratado da questão de combinar a extração de dados não estruturados com um conhecimento prévio, uma base de dados já existente, isto difere de outros projetos de bases de conhecimento automático. É possível corrigir tantos os erros do processo de extração quanto a falta de informações que existe em uma base de dados pronta, possibilitando a inferência de dados novos. Dessa forma, inserem-se, cada vez mais, informações que não estão dentro de bases de conhecimentos.

Também é tratado especificamente do tamanho que as bases de conhecimento possuem. O KV contém aproximadamente 1,6 bilhão de triplas, sendo que destas 324 milhões apresenta uma confiança de 0,7 ou mais e 271 milhões apresenta 0,9 ou mais de confiança, um número muito maior se comparado ao Dipe Dive, o maior sistema comparável ao KV, que contém 7 milhões de fatos confiáveis.

O KV funciona basicamente com três componentes: os extratores, os grafos anteriores baseados na base de conhecimento e a fusão do conhecimento, que basicamente verifica a probabilidade de uma tripla ser verdadeira.

Os extratores são algoritmos que realizam a extração de dados na Web. Basicamente, cada extrator fica responsável por recuperar dados de um determinado tipo. Os extratores do KV foram desenvolvidos para a extração de documentos de textos, árvores HTML, tabelas do HTML, páginas de anotações humanas, sendo que para cada um desses tipos, os algoritmos apresentam similaridades para conseguir realizar a extração conforme as diferenças de cada ambiente.

Os grafos das bases de conhecimento são necessários, pois os dados extraídos da Web

não são confiáveis. O uso destas bases auxilia a verificação da probabilidade de um dado estar correto, levando em consideração os dados já existentes nestas bases, sendo possível realizar a inferência de novas informações.

Por último, a questão da fusão dos dados é fundamental, pois ela irá realizar a união dos dados extraídos com as bases de dados com conhecimento anterior. São processos estatísticos e probabilísticos, que levam em consideração a qualidade dos dados, realizando novos relacionamentos entre os dados. Após o processo de fusão, existirão muitos novos fatos do que o existente nas bases de conhecimento, sendo que parte destes dados tem um alto índice de qualidade.

A pesquisa que traz a questão do KV é recente, 2014, e ainda não está sendo utilizada em algum produto ou aplicação. Ao que tudo indica, este tipo de processamento será o substituto para o Google Knowledge Graph, que tem como base repositórios de conhecimentos prontos. O uso do KV no motor de busca, ao invés do Knowledge Graph, expandiria significativamente o poder de realização de inferência e relações que este possui, conseguindo inserir cada vez mais semântica aos processos de buscas de um usuário.

Como não existem trabalhos que versam das implementações deste sistema, não é tratado como será realizado o processo de visualização das informações que o KV possui em sua base de dados. Porém, se for confirmada a questão deste sistema substituir o Knowledge Graph, é possível que a visualização destas relações, ocorra similarmente ao que acontece hoje com Knowledge Graph. O KV poderá agregar muito as estruturas visuais que são apresentadas atualmente no Google Knowledge Graph, pois a abrangência que os elementos contidos no KV é significativamente maior, portanto, expandiria a quantidade de buscas que aquela estrutura auxiliará ao usuário, enriquecendo cada vez mais o processo de visualização nas pesquisas realizadas em buscadores.

4.3 SPARQL E GOOGLE CHARTS

O protocolo SPARQL possibilita que diversos dados estruturados em RDF possam ser acessados, extraindo diversas relações, além de determinar números a respeito dos dados que estão sendo acessados.

O SPARQL possibilita extrair dos conjuntos de dados novas informações, pois com propriedades como o SUM, COUNT, é possível contabilizar a quantidade ou a soma de um determinado tipo de informação. Assim estas novas informações podem ser representadas e ter alguma relevância para o usuário.

Alguns sistemas construídos utilizando informações estruturadas abertas, disponibilizam

uma estrutura para realizar consultas SPARQL, chamado de SPARQL EndPoint. Um usuário qualquer consegue escrever uma consulta SPARQL e recuperar informações de uma base de dados aberta nesta estrutura. Um exemplo de uma base de dados aberta é o DBPedia, que é uma base que contém os dados do Wikipedia em formato RDF.

Os resultados apresentados a partir de uma consulta SPARQL são apresentados em formato de tabelas, como mostrada na figura 3. Nesta consulta é possível ver que é apresentado como resultado duas variáveis, no caso “nome_organizacao” e “qtd_pesquisador”, sendo que estas variáveis são interligadas, ou seja, tal organização tem um número de pesquisadores.

nome_organizacao	qtd_pesquisador
"Universidad Federal de Santa Catarina"^^<http://www.w3.org/2001/XMLSchema#string>	11
"Universidad de la Republica"^^<http://www.w3.org/2001/XMLSchema#string>	9
"Universidad Federal da Paraíba"^^<http://www.w3.org/2001/XMLSchema#string>	8
"Centro de Información y Gestión Tecnológica de Villa Clara"^^<http://www.w3.org/2001/XMLSchema#string>	5
"Fundação Oswaldo Cruz"^^<http://www.w3.org/2001/XMLSchema#string>	5
"Centro de Investigación y de Estudios Avanzados"^^<http://www.w3.org/2001/XMLSchema#string>	4
"Universidad de Pinar del Río"^^<http://www.w3.org/2001/XMLSchema#string>	4
"Universidade Federal de Minas Gerais"^^<http://www.w3.org/2001/XMLSchema#string>	4
"Universidade Federal de São Carlos"^^<http://www.w3.org/2001/XMLSchema#string>	4
"Universidad Federal do Estado do Rio de Janeiro"^^<http://www.w3.org/2001/XMLSchema#string>	4
"Universidad de Lleida"^^<http://www.w3.org/2001/XMLSchema#string>	3
"Universidad de Valencia"^^<http://www.w3.org/2001/XMLSchema#string>	3
"Universidade Estadual Paulista"^^<http://www.w3.org/2001/XMLSchema#string>	3
"Universidade Federal Fluminense"^^<http://www.w3.org/2001/XMLSchema#string>	3

Figura 3: Quantidade de pesquisadores por instituição

Fonte: próprio autor

Estas ferramentas podem ser utilizadas para diversos fins de análises das informações de uma determinada base de dados. Porém, apenas o uso de tabelas, pode limitar o processo de transformação desta tabela em gráficos. São necessárias operações manuais do usuário, onde este deverá realizar uma cópia dos dados apresentados pelo SPARQL EndPoint para uma ferramenta de planilhas, para criar gráficos, e realizar uma análise mais profunda do significado daqueles dados.

Nesse contexto, existe uma ferramenta que consegue transformar diretamente em gráficos os dados recuperados de um servidor de SPARQL EndPoint, o Google Charts (Gráficos).

O Google Charts, também chamada de Google Visualization API, é uma ferramenta da Google, criado pela própria empresa para a montagem dos mais variados tipos de gráficos. Esta ferramenta funciona diretamente em uma página HTML, que faz simplesmente uma referência ao seu uso. Para o uso desta API, utiliza-se o HTML juntamente com a linguagem JavaScript, que tem como característica a execução de tarefas, onde o usuário terá uma interação local, ou seja, que não precisa passar pelo servidor onde o serviço Web está hospedado. Para a utilização da ferramenta basta inserir alguns códigos em JavaScript configurando o funcionamento da

ferramenta, e inserir os dados do gráfico ou apontar para um local onde os dados podem ser acessados. O desenvolvedor da página conseguirá apontar qual será o tipo de gráfico, como pizza, coluna, linha, entre outros, além de configurações básicas, como o tamanho do gráfico, como serão os rótulos e a legenda, e as cores utilizadas (GOOGLE CHARTS, 2015).

Como dito anteriormente, é possível utilizar o Google Charts para transformar diretamente os dados de um servidor SPARQL EndPoint em gráficos. É necessário que seja feita a consulta ao servidor SPARQL diretamente na página HTML, e que o retorno do servidor SPARQL seja feito em um formato JSON. Os restantes das configurações são iguais a qualquer outra aplicação que faça uso desta ferramenta de gráficos.

Para obter os dados diretamente do servidor SPARQL e convertê-los no formato de uma tabela, é necessário indicar nos comandos uma URL que aponte o servidor SPARQL, além de conter as informações da consulta em si, ou seja, na própria URL aparecerão quais são as variáveis que serão retornadas, quais são as cláusulas da consulta, e todas as informações necessárias para poder ser realizada uma consulta SPARQL. Estando configurado o acesso ao servidor SPARQL, basta informar qual será o retorno da consulta que o servidor devolverá, e fazer a conversão deste formato, para o formato da API de Visualização do Google (DING, 2010).

Na figura 4, é possível ver um retorno que uma página HTML utilizando o Google Charts gera.

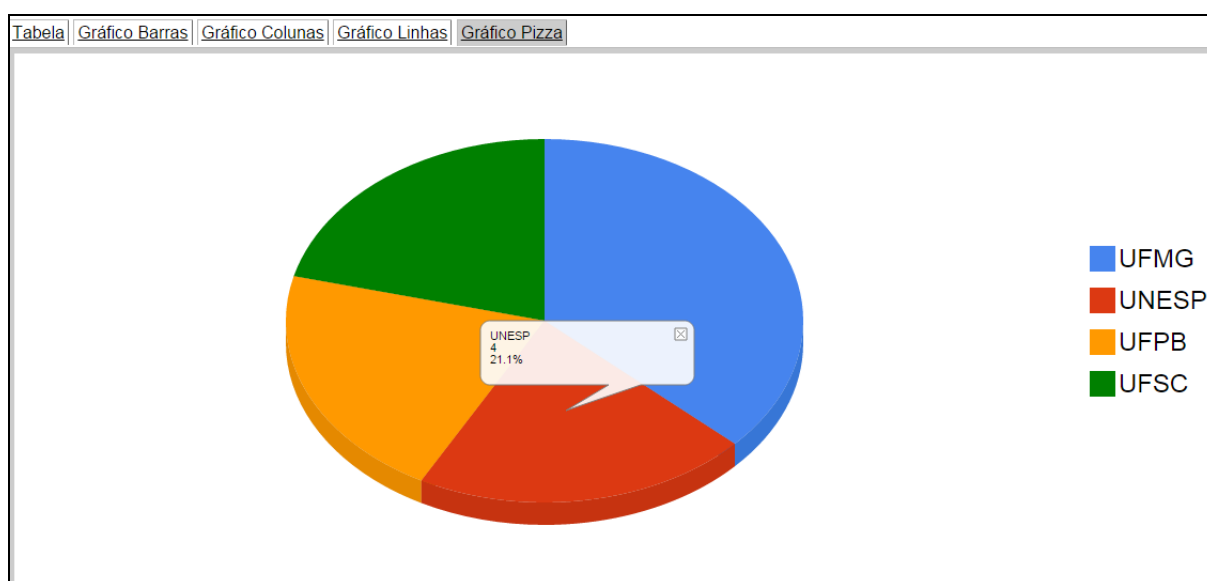


Figura 4: Página HTML com gráfico de pizza sendo apresentado

Fonte: próprio autor

Na figura 4 é possível ver que em uma mesma página HTML é possível inserir diversos tipos de gráficos. No caso, esta página contém quatro tipos, além de ser apresentada a tabela, dividida em abas. A navegação do usuário pode ser realizada através das diferentes apresentações dos gráficos, porém, independentemente do tipo de gráfico, os dados em si são os mesmos, apenas possibilitando que o usuário escolha o meio que melhor demonstre aquela informação. Outro ponto de interação que a ferramenta apresenta, diz respeito aos conteúdos dos dados. Quando o usuário clica nas barras do gráfico, é apresentado a ele mais informações sobre aquela estrutura visual.

A grande maioria das aplicações baseada nos conceitos e tecnologias da Web Semântica utiliza RDF e por consequência, tem a possibilidade de utilizar o SPARQL para realizar análises e descobertas em cima das informações que estão na base de dados das aplicações. Porém, os resultados obtidos de consultas SPARQL são tabelas, ou estruturas computacionais, e o uso de estruturas visuais, como os gráficos do Google Chart auxiliam significativamente no processo de visualização dos dados obtidos a partir das consultas realizadas.

5 CONSIDERAÇÕES FINAIS

A Web e a Internet estão presentes em quase todas as atividades do cotidiano. Nos últimos anos, as tecnologias que sustentam a Web Semântica têm sido elencadas para estruturar as informações publicadas na Web e os agentes computacionais, ferramentas que possam levar melhores resultados de pesquisa aos usuários.

Apesar de existirem e estarem cada vez mais em evidência os estudos relacionados à estrutura e organização da Web Semântica, ainda são tímidas e contidas as iniciativas que tratam o outro lado da Web Semântica, a informação que chega ao usuário. Percebe-se que há uma carência de tecnologias e produtos que se utilizem das bases que estão sendo construídas para levar informações de forma clara e facilitada ao usuário final. Essas iniciativas são ainda mais limitadas quando pensamos em apresentação gráfica de informações aos usuários.

Iniciativas como o Google Knowledge Graph e o Google Knowledge Vault, analisadas nesta pesquisa, constituem modelos interessantes do uso dos conceitos e tecnologias da Web Semântica, aplicados a um sistema de busca, no caso o Google, que é, hoje, o motor de buscas com maior número de usuários (NETMARKETSHARE, 2014). Estas tecnologias promovem a inovação e melhoraria dos processos realizados dentro da Web, transformando dados estruturados e em formato computacional em informações relevantes e aptas a serem consumidas pela comunidade em geral.

Outro ponto de destaque é o uso do SPARQL como intermediário entre os datasets

disponibilizados pelas várias comunidades que publicam dados em formato de Linked Data e a ferramenta Google Charts. O uso desta tecnologia gerando gráficos automatizados auxilia muito na visualização dos dados que são apresentados aos usuários. A apresentação das informações que contém dados estruturados com conteúdo relevante permite ao usuário visualizar mais intuitivamente, aproxima a informação do usuário, facilitando o seu entendimento.

Ferramentas e plataformas que pensem em novos formatos, interativos e gráficos, de apresentar os dados publicados com tecnologias da Web Semântica estão surgindo e, conseguem tornar a experiência do usuário neste tipo de ambiente algo bastante intuitivo, entretanto ainda são tímidas e poucas essas iniciativas.

É necessário que haja mais investimento e desenvolvimento de novas ferramentas e plataformas que tornem mais interativas as possibilidades de acesso aos dados disponibilizados em formato semântico, para que a Web Semântica realmente seja algo factível e acessível a comunidade civil.

REFERÊNCIAS

BERNERS-LEE, T. et al. The semantic web. *Scientific American*, v. 284, n. 5, p. 28-37, 2001.

BROWN, Aaron. **Obtenha respostas inteligentes em sete novas línguas**. 2012. Disponível em: <<http://googlebrasilblog.blogspot.com.br/2012/12/obtenha-respostas-inteligentes-em-sete.html>>. Acesso em: 12 ago. 2015.

CENTRAL INTELLIGENCE AGENCY. **The world factbook**. 2015. Disponível em: <<https://www.cia.gov/library/publications/the-world-factbook/>>. Acesso em: 1 ago. 2015.

CYGANIAK, R.; JENTZSCH, A. **The linking open data cloud diagram**. 2014. Disponível em: <<http://lod-cloud.net/>>. Acesso em: 13 ago. 2015.

DIAS, M. P.; CARVALHO, J. O. F. de. A Visualização da Informação e a sua contribuição para a Ciência da Informação. **DataGramaZero: Revista de Ciência da Informação**, v. 8, n. 5, out. 2007.

DING, L. **The data-gov wiki**. 2010. Disponível em: <http://iw.rpi.edu/wiki/How_to_render_SPARQL_results_using_Google_Visualization_API>. Acesso em: 13 ago. 2015.

DONG, X. et al. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. **Anais...** ACM, p. 601-610, 2014.

FREEBASE. **Freebase announcement**. 2014. Disponível em:

<<https://plus.google.com/109936836907132434202/posts/bu3z2wVqcQc>>. Acesso em: 13 ago. 2015.

FREITAS, C. M. D. S. et al. Introdução à visualização de informações. **Revista de informática teórica e aplicada**, v. 8, n. 2, p. 143-158, out. 2001.

GOOGLE. **The knowledge graph**. 2015. Disponível em: <<http://www.google.com/intl/ta/insidesearch/features/search/knowledge.html>>. Acesso em: 13 ago. 2015.

GOOGLE CHARTS. **Interactive charts for browsers and mobile devices**. 2015. Disponível em: <<https://developers.google.com/chart/>>. Acesso em: 12 ago. 2015.

HEATH, T. **Linked data-connect distributed data across the web**. 2012. Disponível em: <<http://linkeddata.org/>>. Acesso em: 14 ago. 2015.

NETMARKETSHARE. **Market share reports**. 2014. Disponível em <<http://www.netmarketshare.com/>>. Acesso em: 13 ago. 2015.

SANTAREM SEGUNDO, J. E.; VIDOTTI, S. A. B. G. Rede de tags para recuperação da informação no contexto da Representação Iterativa. **InCID: Revista de Ciência da Informação e Documentação**, v. 2, n.1, p. 86-109, 2011.

SANTAREM SEGUNDO, J. E. Web Semântica: introdução a recuperação de dados usando SPARQL. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO: além das nuvens, expandindo as fronteiras da Ciência da Informação, 2014. Belo Horizonte. **Anais...** Belo Horizonte: UFMG/ ECI, 2014. p. 3863-3882.

SIMIONATO, A. C. **Representação, acesso, uso e reuso da imagem digital**. 2012. 141 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2012.

SINGHAL, A. **Introducing the knowledge graph: things, not strings**. 2012. Disponível em: <<http://googleblog.blogspot.com.br/2012/05/introducing-knowledge-graph-things-not.html>>. Acesso em: 13 ago. 2015.

SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, v. 33, n. 1, p. 132-141, 2004.

WATERS, D. J. **Projeto conservação preventiva em bibliotecas e arquivos**. Do microfilme à imagem digital. Rio de Janeiro: Arquivo Nacional, 1997.

WIKIDATA. **Welcome to wikidata**. 2015. Disponível em: <https://www.wikidata.org/wiki/Wikidata:Main_Page>. Acesso em: 1 ago. 2015.

W3C. **OWL**. 2004. Disponível em: <<http://www.w3.org/TR/owl-features/>> acesso em 14 ago. 2014.

W3C. **Web semântica**. 2015. Disponível em <<http://www.w3c.br/Padroes/WebSemantica>> acesso em 14 ago. 2014.