

Guilhermina de Melo Terra
(Organizadora)

Biblioteconomia e os Ambientes de Informação

Atena Editora
2019

AVANÇO DAS NOVAS TECNOLOGIAS E USO EM NÚVENS APLICÁVEIS ÀS BIBLIOTECAS

Marcos Luiz Mucheroni
José Fernando Modesto da Silva

RESUMO: A produção de conteúdos digitais tem emergido novas mudanças na produção e armazenamento de dados, com a consequente geração de dados ocorrem sobre esta influência modificou-se alguns usos e aplicações em bibliotecas. Neste cenário, nota-se em bibliotecas de relevância interna, um papel de destaque como referência no uso dos recursos computacionais no desenvolvimento de seus Serviços de Informação ao usuário. Entre estas, com destacada evidência, estão a instituição *Online Computer Library Center* (OCLC). O objetivo deste tópico é destacar o desenvolvimento de serviços em nuvens e sua aplicação no ambiente da Web Semântica. Em termos metodológicos, desenvolve-se um texto descritivo, que a partir de uma revisão de literatura, descreve-se projetos e conceitos que abordam estas novas tecnologias.

PALAVRAS-CHAVE: Dados Ligados; Web Semântica; Mecanismos de Busca; VIAF; OCLC

ABSTRACT: The production of digital content has emerged new changes in the production and storage of data, with the consequent generation of data occurring on this influence has modified some uses and applications in

libraries. In this scenario, we note in libraries of internal relevance, a prominent role as reference in the use of computational resources in the development of its Information Services to the user. Among these, with outstanding evidence, are the *Online Computer Library Center* (OCLC). The purpose of this topic is to highlight the development of cloud services and their application in the Semantic Web environment. In methodological terms, a descriptive text is developed, that from a literature review, describes projects and concepts that approach these new technologies.

KEYWORDS: Linked Data; Semantic Web; Search Engines; VIAF; OCLC

1 | INTRODUÇÃO

A crescente produção de conteúdos digitais, com a consequente geração de um volume maior de dados ocorrem sob forte influência tecnológica, o que requer processos de aperfeiçoamento e armazenamento que tratem este volume. Com isto tornou-se necessário o desenvolvimento de tecnologias e investimentos na criação e aprimoramento de ferramentas que envolvam os processos de captura e de validação dos dados, para posteriores: análise, armazenamento e acesso permanente. Estas são situações que afetaram

as finalidades das bibliotecas no que se refere à organização e ao tratamento das informações, mas não como finalidade principal que é o de armazenar.

O processo de curadoria digital dos dados deve abranger uma série de atividades que se inicia pela determinação das corretas estruturas de dados que precisam ser mapeados em diferentes arquivos. Para este fim, é preciso incluir esquemas de metadados com objetivo de prover longevidade e integração dos dados com os instrumentos de seus processos e análises. Sem a adoção de metadados explícitos, a interpretação dos dados torna-se apenas implícitas. A curadoria digital tem como foco a seleção, preservação, manutenção, coleção e arquivamento dos dados digitais; sem curadoria, certamente, os dados podem ser perdidos no universo digital (HEY, TANSLEY e TOLLE, 2011). Dados perdidos e não estruturados impossibilitam a sua coleta, recuperação e integração com outros dados, em um ambiente de Web Semântica. Assim, é possível questionar sob qual ambiente se fará isto?

Neste cenário, em especial, nota-se que muitas bibliotecas, de abrangência global, podem ser visualizadas como modelo ou exemplo de ambientes *casos* por possuir seus principais serviços desenvolvidos sob influências tecnológicas. Estas bibliotecas, também se encontram apoiadas, com destacada evidência, pela OCLC (*Online Computer Library Center*), uma entidade criada em 1978, como uma cooperativa sem fins lucrativos, com a finalidade de compartilhar tecnologias, pesquisas e ideias inovadoras no ambiente das bibliotecas. No cenário mundial torna-se uma importante referência para o desenvolvimento de produtos, serviços e processos baseados em tecnologias para bibliotecas (GODBY, WANG, MIXER, 2015). Além de referência para desenvolvimento de ambientes semânticos em serviços bibliotecários.

Deve-se entender “nuvens” como plataformas e serviços de armazenamento disponíveis em larga escala com acesso via Web, que é uma plataforma sobre a internet, onde o armazenamento deve ser permanente (por exemplo, usando dispositivos de armazenamento tipo *storage*, que garante este aspecto) e onde os serviços devem estar disponíveis de forma amigável aos usuários.

Neste contexto, o objetivo deste artigo é destacar o desenvolvimento de serviços em nuvens e, ainda, a sua aplicação no ambiente da Web Semântica. Além de também ressaltar os projetos bibliotecários envolvidos com o ambiente semântico.

Ademais destaca-se a importância de desenvolver uma complexa teoria para o uso de “software robôs” - é importante notar que uso destes robôs também chamados de agentes inteligentes estão em pleno desenvolvimento - que lhes permitam vasculhar a Web por meio do conceito de *Linked Data* (Dados Ligados), capazes de apoiar a tomada de decisões precisas e confiáveis, e como esse conceito afeta os serviços bibliotecários.

Em termos metodológicos, desenvolve-se um texto descritivo que, baseado em revisão de literatura, descreve projetos e conceitos que abordam os termos tratados.

O *linked data* traz, como conceito, a preocupação central de que dados não estejam dispersos de maneira a favorecer tanto a sua difícil recuperação, quanto a

sua capacidade de validação da veracidade, autoria e demais estruturas que tornem a informação durável e confiável.

2 | A OCLC E OS PROJETOS PARA WEB SEMÂNTICA

A OCLC participa ativamente do Consórcio Mundial da Web (*World Wide Web Consortium - W3C*), e em diversos projetos e ações relacionadas ao desenvolvimento de sistemas que usam a tecnologia da Web Semântica. Cita-se como exemplo: *Schema Bib Extend Community Group* (importante para decidir as definições indexadas no *Schema.org*); o *VIAF (Virtual International Authority File)*; e o *FAST (Faceted Application of Subject Headings)*. Ademais, a instituição promove recomendações orientadas para organizações como a Google, o Yahoo, a Microsoft, e o Yandex, bem como, a Wikipedia.

Deve-se perceber que junto ao Wikipedia, em função do *linked data*, apareceu a base de dados intitulada DBPedia, a qual traz referência fundamental a emergência de uma Web Semântica estruturada e com *linked data* (PALETTA, MUCHERONI, 2014).

2.1 O Grupo Extendido da Comunidade Bib

O projeto do *Schema Bib Extend Community Group* é formado por um grupo de trabalho que tem a missão de discutir e preparar propostas que ampliem o esquema “*schema.org*” de maneira a melhor representar a marcação e o compartilhamento da informação bibliográfica. A OCLC desempenha papel ativo na criação e nas atividades do grupo cujo desenvolvimento de propostas envolve implementar vocabulário para *Linked Data* (OCLC, 2016).

O uso do *Schema.org*, na descrição de recursos bibliográficos, é uma maneira de compartilhar dados estruturados sobre esses recursos na Web. O *Schema* é um vocabulário genérico, projetado para descrever a maioria das “coisas” encontráveis na Web e, para tanto, alguns procedimentos na nomeação de tipos e propriedades têm de ser feitas, quando comparado com um vocabulário específico de um único domínio.

Por exemplo, no ambiente bibliográfico espera-se que os livros tenham uma propriedade chamada “título”, enquanto que, no *Schema.org*, usa-se o termo “nome”. Além disso, há procedimentos na granularidade de detalhes que podem ser descritos com o *Schema*, em comparação com a riqueza dos vocabulários bibliográficos específicos. Esse domínio dos detalhes bibliográficos específicos iria complicar o carácter genérico adotado pelo *Schema*.

Apesar disto, o *Schema* desenvolve um processo de descrição dos recursos bibliográficos em detalhe suficiente para ser compreendido pelos mecanismos de busca e da Web em geral. O *Schema.org* é um vocabulário que também pode ser representado em RDF (*Resource Description Framework*), por todas as sinalizações RDF, como RDF/XML (W3C, 2015).

Muitos ambientes já utilizam esta forma de descrição da Web Semântica. Como exemplo ilustrativo, apresenta-se na figura 1 um modelo geral de relacionamentos de uma coleção de páginas wiki (um conjunto de páginas interligadas, que podem ser acessada e o seu conteúdo editado), que demonstram a marcação proposta para uma obra em vários tipos.

A finalidade do exemplo da figura 1 é mostrar o uso de vocabulário e os exemplos de marcação. O exemplo se refere a um livro, com inclusão de link para uma descrição conceitual da obra da qual é um exemplo, além de link para uma descrição do mesmo livro na base de dados bibliográfica do *WorldCat.org*.

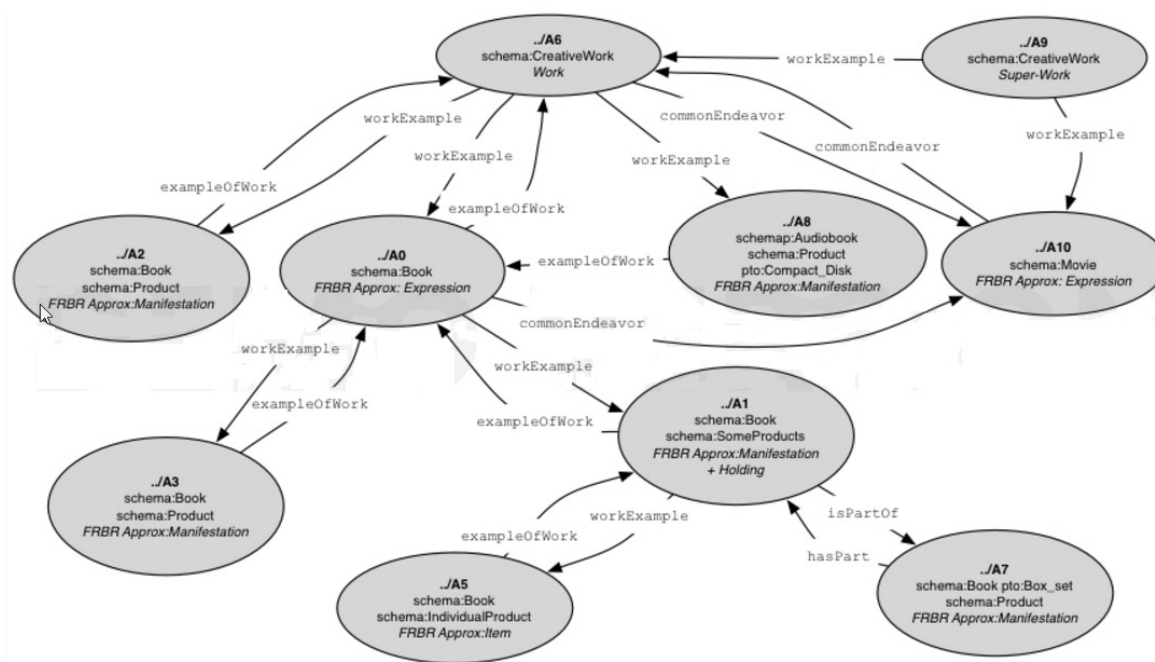


Figura 01 – Modelagem de Relacionamento de Recursos Bibliográfico no *Schema.org*.

Fonte: W3C, 2014

O exemplo da figura 1 foi extraído da proposta desenvolvida pelo *Schema Bib Extend Community Group*.

O modelo é uma estrutura parcial de visualização das relações de dados. Permite observar que os recursos bibliográficos não podem ser descritos detalhadamente, tanto quanto desejado na implementação do catálogo bibliográfico tradicional. Isto é por razões de clareza, concisão, etc.

Trata-se ainda de proposta com um protótipo experimental construído sobre um ambiente semântico.

O *Schema* adota, em seus processos de estruturação e ligação dos dados, o VIAF que a partir de sua criação, em 2012, evoluiu para se tornar um serviço OCLC de importante referência para identificação e localização de autores e dos processos de controle de autoridade desenvolvidos no fluxo de trabalho de catalogação realizados em bibliotecas.

É um projeto que combina vários arquivos de nome de autoridade em um único serviço de nome de autoridade.

Ao longo do tempo, o número de agências bibliográficas participantes tem crescido, chegando já a mais de 200, em mais de 100 países diferentes, espera-se através das bibliotecas nacionais chegar a todos os países que possuem bibliotecas ou sistemas de referências universais, onde as publicações são catalogadas.

Em acréscimo aos objetos do presente trabalho, destaca-se os estudos de Mucheroni, Santos (2018) que abordam as ligações entre os sistemas de autoridade do VIAF e as ontologias SPAR, explorando a adoção de ontologias e *linked data* para melhorar e tornar mais automáticas a elaboração das referências e registros bibliográficos, sob o conceito da Web Semântica.

Os colaboradores do VIAF incluem bibliotecas nacionais, agências culturais e outras instituições de todo o mundo. Atualmente, 62 instituições de mais de 34 países estão registradas como colaboradores no VIAF. A governança do VIAF é realizada por meio do Conselho, que fornece orientação sobre as políticas, práticas e operação do projeto. O Conselho é formado por representantes das seguintes agências: Biblioteca e Arquivos do Canadá, Biblioteca Nacional de Espanha, Biblioteka Narodowa, Biblioteca do Congresso, Bibliothèque nationale de France, Deutsche Nationalbibliothek (OCLC, 2019).

A seguir se descreve os modelos do VIAF e FAST, contextualizada na compreensão da web semântica, estruturada sob exemplos aplicados.

2.2 VIAF – *Virtual International Authority File*

O serviço internacional da OCLC foi projetado para acesso compartilhado, e uniformizado aos catálogos de autoridade criados pelas principais agências bibliográficas do mundo (embora limitado nesta etapa experimental para consulta aos nomes de pessoas). Sua concepção foi pensada como um dos componentes essenciais da Web Semântica, e que permite apresentar os nomes de pessoas na forma alfabética e de língua do usuário.

A longo prazo, o VIAF pretende incluir os registros de autoridade de muitas outras bibliotecas de maneira a compor um serviço global, e livremente disponível na rede. Mediante o relacionamento dos diferentes nomes de uma mesma pessoa ou organização deve tornar o VIAF em um meio apropriado para que uma ampla comunidade de bibliotecas e de agências bibliográficas possam redimensionar seus dados bibliográficos para atender os usuários que o utilizam em diferentes línguas.

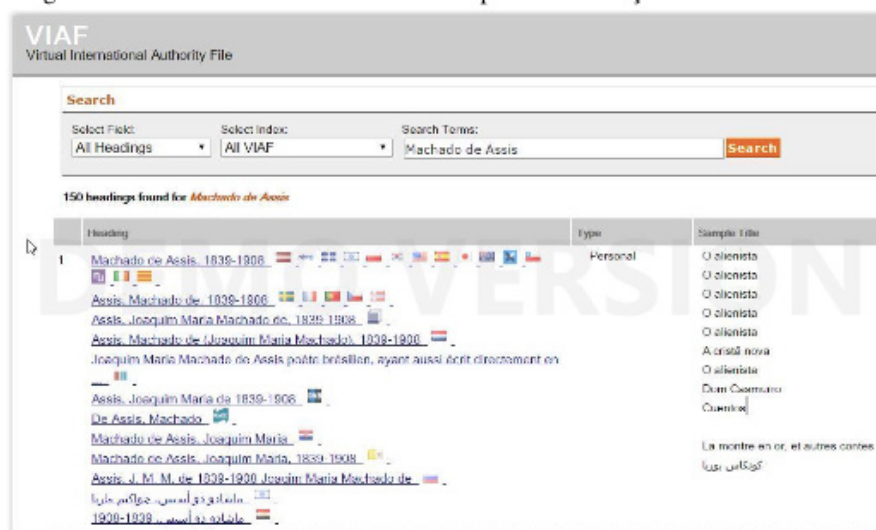


Figura 02 – Interface de Consulta do VIAF para identificação de Nome do Autor.

Fonte: OCLC. VIAF, 2016.

A OCLC para viabilização do VIAF, desenvolveu um software especial de comparação de nomes, construído com base no *WorldCat Identities* (<http://worldcat.org/identities>) que é um serviço que fornece informações resumidas existente no WorldCat que é um banco de dados de informações bibliográficas dos mais abrangentes, do mundo, sobre coleções de bibliotecas.

O *WorldCat Identities* conta, atualmente, com cerca de 30 milhões de nomes, incluindo nomes de pessoas, organizações e personagens fictícios. No exemplo de consulta constante da figura 2, utilizou-se o nome de “Machado de Assis”, sendo listado as várias formas do nome adotado pelas bibliotecas participantes do projeto VIAF.

O VIAF se baseia no conceito do modelo entidade-relacionamento preconizado pelo FRBR (Requisitos Funcionais para Registros Bibliográficos) destinado ao agrupamento e representação semântica de registros bibliográficos. A figura 2 ilustra as características do VIAF.

Os critérios de comparação do VIAF incluem o exame dos nomes e sua associação com as obras registras em múltiplos catálogos bibliográficos nacionais, produzido pelas diversas agências bibliográficas espalhadas pelo mundo e o *WorldCat* da OCLC. Os registros de autoridade do VIAF se constroem a partir das informações extraídas durante o processo de comparação de dados e inclui elementos dos registros fonte de autoridade e bibliográfico. Adere a aplicação do *Linked Data*, aplicado ao registro bibliográfico. O VIAF é um projeto modelar de adoção do conceito do *Linked Data* aplicável à nuvem.

2.3 FAST – *Faceted Application of Subject Headings*

O FAST é uma lista de aplicações terminológicas facetadas por assunto, surgida em 1998, derivada do cabeçalho de assunto utilizado pela *Library of Congress* e que, inicialmente, visou descrever tematicamente recursos da Web em esquema simples

de metadados. Para tanto, adotou o padrão *Dublin Core* para operar sob oito facetas distintas: tópicos, geográfico, nome pessoal, nome da entidade coletiva, forma, cronológico, título (como sujeito e nome de grupo). Cada faceta pode ser mapeada por um elemento específico do padrão *Dublin Core* (QYIANG, 2008).

O objetivo de adaptar a Lista de Cabeçalhos de Assunto da *Library of Congress* - LCSH, com uma sintaxe simplificada, para criar o FAST foi para aproveitar o rico vocabulário de LCSH e, ao mesmo tempo, criar um esquema de fácil compreensão, controle, aplicação e uso. Dada a sua concepção de origem, o esquema mantém compatibilidade com LCSH e qualquer conjunto válido de cabeçalhos do LCSH podem ser convertidos aos elementos do FAST. As dificuldades de uso do LCSH também se encontram no FAST, porém há novas possibilidades de aplicação impulsionadas pelo crescimento da web. Aspectos que forçam mudanças nos sistemas de controle bibliográfico de maneira a torná-los mais fáceis de utilizar.

A interface de pesquisa do FAST denominada *searchFAST* é amigável. Na figura 3, ilustra-se sua interface de busca, ao consultar o termo “Machado de Assis”.

The screenshot shows the searchFAST interface. At the top, there is a search bar with the text 'machado de assis' and a search button. Below the search bar, there are two main sections: 'FAST TERMS' and 'TERM DETAILS'. The 'FAST TERMS' section displays search results for 'machado de assis', showing a list of terms with their respective facets and uses. The 'TERM DETAILS' section is currently empty.

Heading	Facet	Uses
Machado de Assis, 1839-1908	person	1270
Dom Casimiro (Machado de Assis)	file	73
Memórias póstumas de Brás Cubas (Machado de Assis)	file	49
Quincas Borba (Machado de Assis)	file	31
Memorial de Aires (Machado de Assis)	file	9
Esaú e Jacó (Machado de Assis)	file	8
Alenista (Machado de Assis)	file	8
Jair Garcia (Machado de Assis)	file	1

Figura 03 – Interface de Consulta do FAST

Fonte: OCLC. searchFAST, 2016.

Outro aspecto a destacar no FAST é a sua disponibilidade como *Linked Data*, aproveitando a abordagem de dados publicados e que aumenta a utilidade da informação na Web. Principalmente ao estabelecer referências de busca para pessoas (exemplo da Figura3), lugares, coisas, etc., de maneira mais consistente e passível de ligação entre domínios.

3 | A WEB SEMÂNTICA E O USO DE NOVAS TECNOLOGIAS DE NUVENS EM BIBLIOTECAS

Considerando-se os avanços tecnológicos da última década, e ao texto publicado por Berners-Lee e outros autores em 2001, delineando os objetivos, que ainda eram ideais, sobre ferramentas e um novo conceito de Web - a Web Semântica. No mencionado artigo, os autores manifestavam-se:

“A Web Semântica irá trazer estrutura ao conteúdo significativo de páginas Web, criando um ambiente onde agentes de software de roaming a partir de uma página para outra podem facilmente realizar tarefas sofisticadas para os usuários. ... não é uma Web separada, mas uma extensão do atual, na qual é dada significação à informação, permitindo que computadores e pessoas possam trabalhar em cooperação. Os primeiros passos para tecer a Web Semântica a partir da estrutura Web existente já está em andamento. Num futuro próximo, estes desenvolvimentos darão início a uma nova funcionalidade significativa como as máquinas se tornam muito mais capazes de processar e “entender” os dados que eles simplesmente exibem no presente” (BERNERS-LEE, HENDLER E LASSILA, 2001).

É importante entender que os autores ainda descreviam um cenário hipotético, baseado em chamadas telefônicas e que, só hoje, podem ser pensadas como aplicativos operados em smartphones, reduzindo o volume de dispositivos próximos ao seu “agente” humano que podia buscar a uma chamada urgente para um médico, seguida de logísticas complexas na marcação de consultas e prescrições destinadas a um parente idoso. Na atualidade este processo, pode ser feito por meio de um WhatsApp, mas ainda com mediação de um agente humano. A indagação é se o crescimento das tecnologias e o uso das nuvens podem mudar isto?

Em um futuro próximo, estes detalhes poderão ser gerenciados automaticamente por “software robôs” (*softbots*), rodando em uma Web mais inteligente, na identificação de nomes, endereços, e avaliações de prestadores de serviços orientados a área da saúde e às companhias de seguros, por exemplo.

Evoluções tecnológicas a partir de agentes inteligentes dependerá, em grande parte, do desenvolvimento do *Linked Data* na Web e da existência de serviços mais estruturados e complexos para o tratamento de dados em nuvens.

A adoção de novas tecnologias, surgidas a partir da linguagem XML (*Extensible Markup Language*), transformando-se em gramáticas RDF (*Resource Description Framework*) e depois em ontologias - taxonomia de conceitos com atributos e relações que proporcionam um vocabulário consensual para definir redes semânticas formadas por unidades de informação inter-relacionadas – e que especificariam as regras lógicas para que os agentes de software reconheçam e classifiquem cada conceitos (RODRÍGUEZ PEROJO e RONDA LEÓN, 2005).

São aspectos que geram estímulos ao desenvolvimento de uma inovação aberta a partir dos recursos de informação de bibliotecas, arquivos, museus e outras instituições culturais. A Web se converte em um espaço global de dados e, neste sentido, as bibliotecas participam publicando seus dados por meio do uso das tecnologias RDF,

além das licenças abertas, para potencializar serviços inovadores desenvolvidos por terceiros, reutilizando seus dados. O *Linked Data* tem um efeito fundamental nos arquivos, bibliotecas e museus que não é outro que a de ampliar o âmbito de aplicação dos dados que geram o chamado LAM (*Library, Archives and Museum*), e sobre tudo, o sentido de ampliar os pontos de vista e de análises da informação (SAORÍN, 2012).

A Web Semântica pode ser compreendida como uma extensão da Web existente, porém contemplando a informação que adquire significado mediante o uso de metadados para prover uma descrição e categorização semântica de seu conteúdo e possibilitar uma razoável automatização. De maneira geral, pode-se dizer que um metadado é um dado que se encarrega de manter no registro seu significado, contexto ou propósito enquanto um objeto de informação, de forma a poder descobrir, entender, extrair e administrar esse objeto. Ademais, os esquemas de metadados, e o W3C ao desenvolver o RDF, viabilizaram uma infraestrutura para descrição de recursos na web que é mais que um formato de metainformação, é um “metamodelo de metadados” que permite codificar distintos esquemas de metadados e, ademais, criar outros vocabulários específicos por meio de uma linguagem de descrição RDF (RDF Schema, 2014).

A declaração original de princípios para *Linked Data*, é discutida em detalhes técnicos no tópico seguinte, interessa agora saber como uma rede de declarações estruturadas, ou um “conjunto de dados expressos usando um vocabulário” pode ter um modelo que identifique entidades como “coisas” que as pessoas comuns entendem (BERNERS-LEE, 2006).

De maneira simplificada, os dados ligados referem-se a “coisas” que apontam para outras “coisas”. Assim, os princípios de *Linked Data* ou Dados Ligados também estipulam a necessidade de recursos como a autoria, no exemplo de dados associado ao dramaturgo inglês do século XVI, “William Shakespeare” para que seja confiável qualquer ligação persistente e acessível na Web.

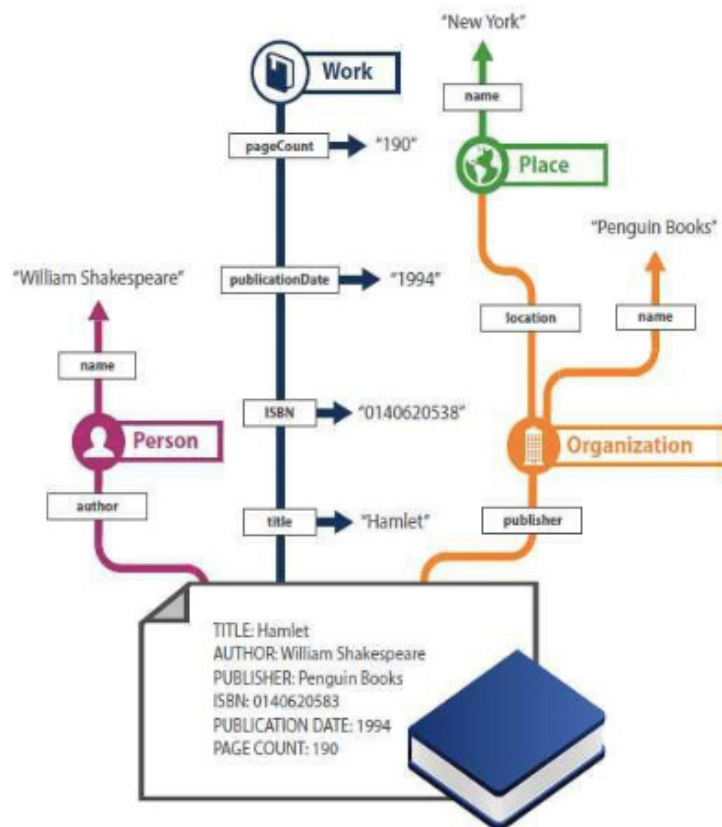


Figura 4 – Descrição detalhada de um registro em um grafo

Fonte: Adaptação de Godby, Wang, Mixer, 2015.

No exemplo acima, o autor “William Shakespeare” mencionado na descrição da obra

“Hamlet” apresentado na Figura 4, está corretamente identificado como o nome do autor e ícone literário Inglês, e não como o nome de algum animal ou de alguma banda de rock, e representa o conhecimento sobre o autor de Hamlet que perdura para além de uma única indicação em um documento individual qualquer.

As enciclopédias e outras referências podem satisfazer esta necessidade na Web de Documentos, mas estruturadas de tal forma que só os humanos possam lê-las e as buscas em infinitos arquivos podem ser demoradas e redundantes se não forem auxiliadas por máquina.

Deve-se ressaltar que, a introdução de novos instrumentos de tratamento dos dados como a *Resource Description e Access* (para catalogação bibliográfica), o XML e as tecnologias da Web Semântica em geral, viabilizam a ideia de criar ligações entre os dados que permitam a compreensão de seu significado por computadores, tende a se expandir em popularidade. Neste momento, a informação encontrada em ambos os ambientes: catálogos bibliográficos e a Web é geralmente legível e compreensível, mas não otimizadas para a compreensão total pela máquina.

É assim que uma representação na Web Semântica pode ser compreendida por alguém que deseja fazer uma busca confiável, mas com um modelo computacional de busca de “coisas” que os seres humanos se lembram e reconheçam como sendo

verdadeiras e significativas. Em outras palavras, as convenções da Web Semântica para descrever entidades devem ter um efeito que estabeleçam ligações claras entre um texto a alguma “coisa” no mundo.

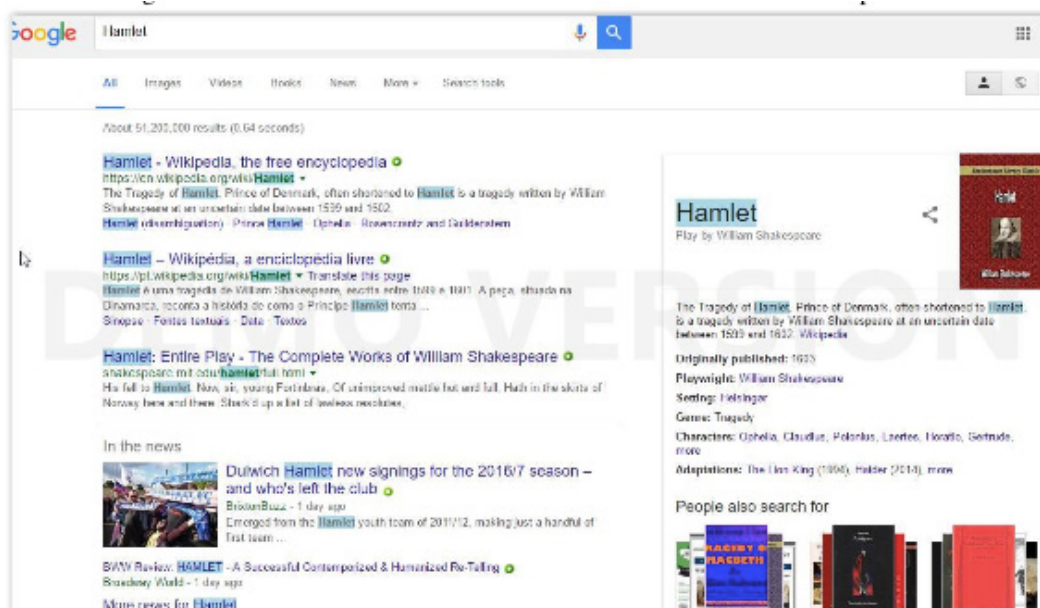


Figura 5 – Busca estruturada e não estruturada sobre Hamlet de Shakespeare.

Fonte: Godby; Wang; Mixer, 2015

Os grafos anunciados como sendo utilizado pelo Google, as vezes chamado de “ficha do conhecimento” ou “diagrama de conhecimento”, mostra ser diferente de uma lista de documentos. Traz informações relacionadas ao termo pesquisado diretamente na página de resultados da busca, como dados demográficos de cidades, datas de aniversário de pintores ou informações sobre alguma obra ou escritor. Em suma, o sistema fornece informações estruturadas e detalhadas sobre um determinado tema, além de uma lista de links para outros sites. O objetivo é fornecer aos usuários as informações necessárias para responder às suas dúvidas sem que tenham de navegar para outros sites.

Em relação ao exemplo da Figura 4 anterior, se referem a entidade do mundo real “Hamlet”, e não apenas uma relação de texto procurada ao “pé da letra” pelo mecanismo de busca do Google, usando grafos ou não. Não estabelece, por exemplo, uma relação de autoria típica encontrada no projeto de *Linked Data* do VIAF. Ele é construído sob um processo de mineração de fatos, em fontes de dados, como Freebase, Wikipedia e o CIA World Factbook.

A inscrição apresentando apenas a informação mais confiável que pode ser através de algoritmos de descoberta.

Como Tim Berners-Lee e seus colaboradores disseram na citação mencionada neste texto, a Web Semântica não é separada a partir da Web de Documentos, mas é sim, uma extensão dela. Evidência de sua coexistência é apresentada na Figura 5,

trata-se de uma captura de tela de uma pesquisa realizada no Google para o termo “Hamlet”.

À esquerda é mostrada uma lista ordenada de documentos, produzidos pela correspondência encadeada de textos nos quais o usuário consulta as versões atualizadas de períodos variados.

Os algoritmos de recuperação de informação codificam tais informações sem nenhuma compreensão do mundo real no qual os textos são subentendidos, enquanto à direita está um display construído a partir de um modelo estruturado do Conhecimento.

Não é questão de demonizar ou idolatrar o mecanismo de busca do Google, ele tem uma função delimitada de busca, não oferece uma estrutura que a Web Semântica poderá ter daqui para frente. Pode-se observar estes conceitos com buscas estruturadas e não estruturadas de conhecimento, conforme apresentado na Figura 5.

A partir de 2012, o Google já tinha gerado registros de Conhecimento para 500 milhões de entidades, que são enriquecidas com outros 3,5 milhões de indicações extraídas do *Knowledge Graph* (PATEL, 2016).

Um sistema experimental atualiza o *Knowledge Graph* montada em uma rede de entidades e relacionamentos de entradas variadas e utiliza algoritmos de aprendizagem para fazer avaliações de qualidade e remover duplicatas, mesclar indicações sobre uma determinada entidade de diferentes fontes, e fazer inferências que podem ser traduzidas em mais declarações processáveis pelas máquinas (DONG, 2014).

4 | CONSIDERAÇÕES FINAIS

A modelagem de dados bibliográficos em uma forma que os mecanismos Web entendam podem colaborar para maior visibilidade das bibliotecas. Espaços provedores de conteúdo ou de dados, nos quais as pesquisas por informação estão agora possíveis de começarem a emergir em bibliotecas, com o armazenamento em nuvens.

Neste contexto, é que se descreve as contribuições da OCLC com a aplicação do conceito de *Linked Data* em nuvem, e que produzem modelos e conjuntos de dados em RDF de variados recursos, amplamente referenciados e publicados pela comunidade de bibliotecas, incluindo os serviços: WorldCat, VIAF, e FAST, além de outros dados codificados no formato *Schema.org*. Formato aliás, que subsidia os principais mecanismos de busca para a exibição estruturada de dados descritivos.

As experiências da OCLC dedicadas aos dados ligados em nuvem serão benéficas para bibliotecários, arquivistas, cientistas da informação e demais profissionais interessados em modelar descrições bibliográficas como ligações de dados, na Web Semântica.

Assim, se a biblioteca passa a utilizar dados vinculados para apresentar os seus recursos informacionais na Web, uma pesquisa de um livro, por exemplo, pode

apresentar um gráfico semelhante ao mostrado no *Knowledge Graph* do Google, contemplando informações sobre o livro e sua disponibilidade de consulta. Links para o autor e o editor e outras informações sobre o livro seriam disponibilizadas por uma interface gráfica. As bibliotecas já coletam e gerenciam tais informações, agora as possibilidades descritas possibilitam que elas tornem mais úteis essas informações na Web, para os usuários.

Esse é, aliás, um dos passos futuros para o qual caminha a Biblioteconomia e a Ciência da Informação.

REFERÊNCIAS

BERNERS-LEE, T. Linked Data. **World Wide Web Consortium**, 27 July 2006. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 02 abr. 2016.

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

DONG, X. L.; GABRILOVICH, E.; HEITS, G.; HORN, W.; LAO, N.; MURPHY, K.; SUN, S.; STROHMANN, T.; ZHANG, W. Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion. In: 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. **Proceedings**. New York: Association of Computing Machinery, 2014. Disponível em: <<https://goo.gl/6MJSk9>>. Acesso em: 02 abr. 2016.

GODBY, C. J.; WANG, S.; MIXER, J. K. **Library Liked Data in the Cloud: OCLC's Experiments with New Models of Resource Description**. Indiana, EUA: Morgan & Claypool, 2015.

HENDLER, J. Emerging Web 3.0. **Computer**, Jan. 2011, p. 111-113.

HEY, T.; TANSLEY, S.; TOLLE, K. **O quarto paradigma: descobertas científicas na era da eScience**. São Paulo: Oficina de textos, 2011.

MUCHERONI, M. L., SANTOS, E. VIAF and OpenCitations: cooperative work as a strategy for information organization in the linked data era. In: 15th INTERNATIONAL ISKO CONFERENCE. Portugal, July 9 – 11, 2018. **Proceedings**. Porto (Portugal): Universidade do Porto, 2018. Disponível em: http://www.iskoiberico.org/wp-content/uploads/2017/03/ISKO_2018_Programme_Final.pdf. Acesso em: 5 fev. 2019.

MURPHY, B. Virtual International Authority File service transitions to OCLC: contributing institutions continue to shape direction through VIAF Council. **OCLC Releases**, 2012. Disponível em: <<http://goo.gl/r8atqu>>. Acesso em: 10 jan. 2016.

OCLC. **Community**. Disponível em: <<http://goo.gl/5ymP1a>>. Acesso em: 30 mar. 2016.

OCLC. **SearchFAST**. Disponível em: <<http://fast.oclc.org/searchfast/>>. Acesso em: 30 mar. 2016.

OCLC. **VIAF: The Virtual International Authority File**. Disponível em: <<https://viaf.org/>>. Acesso em: 05 fev. 2019.

PALETTA, F. C., MUCHERONI, M. L. O desenvolvimento da WEB 3.0: Linked Data e DBPEDIA, **PRISMA**, n. 25, Porto: 2014, p. 73-90.

PATEL, N. **O Guia do Iniciante ao Knowledge Graph do Google**. 12 de janeiro de 2016. Disponível

em: <<http://goo.gl/LUMEMI>>. Acesso em 03 maio 2016.

QIANG, Jin. Is FAST the Right Direction for a New System of Subject Cataloging and Metadata? **Cataloging & Classification Quarterly**, vol.45, n. 3, p. 91–110, 2008.

RDF Schema. **W3C Recommendation** 25 February 2014. Disponível em: <<https://www.w3.org/TR/rdf-schema/>>. Acesso em: 02 abr. 2016.

RODRÍGUEZ PEROJO, K.; RONDA LEÓN, R. Web semántica: un nuevo enfoque para la organización y recuperación de información en el web. **ACIMED**, Habana, vol.13, n.6, nov.-dic., 2005. Disponível em: <<http://goo.gl/dh4J0u>>. Acesso em: 20 jan. 2016.

SAORÍN, T. Cómo linked open data impactará en las bibliotecas a través de la innovación abierta. **Anuario ThinkEPI**, vol. 6, p. 288-292, 2012.

W3C. **Examples Library**, 2014. Disponível em: <<https://goo.gl/3BYIQ3>>. Acesso em: 02 abr. 2016.

W3C. **Schema Bib Extend Community Group**, 2015. Disponível em: <<https://www.w3.org/community/schemabibex/>>. Acesso em: 02 abr. 2016.