

MODELOS DE REPRESENTAÇÃO SEMÂNTICA NA ERA DO BIG DATA

Models of semantic representation in the age of Big Data

Janailton Lopes Sousa (1) Paulo George Miranda Martins (2) Rogerio Aparecido Sá Ramalho (3)

Universidade Federal de São Carlos, janailtonlopes20@gmail.com (1) pgeorgem22@gmail.com (2) ramalho@ufscar.br (3)

Resumo

O termo Big Data se refere ao grande volume de dados produzidos e disponibilizados em ambientes digitais. Ao longo dos últimos anos novos modelos de representação têm sido propostos no intuito de aperfeiçoar as formas de representação de informações em ambientes digitais. O presente trabalho está vinculado a um projeto de pesquisa em andamento, financiado pelas agências FAPESP e CNPq, e possui como objetivo analisar os princípios que fundamentam o Big Data e sua relação com os novos padrões de representação *Resource Description Framework* (RDF); *Simple Knowledge Organization System* (SKOS) e *Ontology Web Language* (OWL). A pesquisa possui caráter teórico e abordagem qualitativa, pois busca apresentar características voltadas à descrição, compreensão e explicação das relações do Big Data com os novos modelos de representação. A partir do levantamento teórico realizado, foi verificado que os modelos de representação analisados contribuem para interligar grandes volumes de dados sem perder o contexto no qual são originados, favorecendo um melhor entendimento do Big Data e os novos paradigmas de representação em ambientes digitais.

Palavras-chave: Sistemas de Organização do Conhecimento. RDF. SKOS. OWL. Big Data.

1 Introdução

O desenvolvimento das tecnologias de comunicação e informação realçaram o modo de produção e processamento de informações e, o uso de novos padrões que suportam essa tendência são adotados como um meio de modelar os dados para representá-los em ambientes dinâmicos, como a *Web*. Ao efetuar buscas pela internet é possível se deparar com grandes volumes de informações e dados produzidos por seres humanos e máquinas, no entanto, seu processamento é executado por mecanismos automáticos com o auxílio de softwares ou padrões inteligentes.

A análise desses dados tornou-se um desafio, pois o volume tornava-se superior à velocidade de processamento. Tecnologias como *MapReduce*, que trabalha o processamento de forma independente surgiu como uma alternativa para grandes instituições de tecnologia que precisavam lidar com um grande volume de dados. Outra tecnologia desenvolvida para lidar na análise de dados foi o *Hadoop*, uma arquitetura de computação escalável, na forma de *clusters*, com nós de processamento

Abstract

The term Big Data refers to the large volume of data produced and made available in digital environments. Over the last few years, new models of representation have been proposed to improve the forms of representation of information in digital environments. The present work is linked to an ongoing research project, funded by FAPESP and CNPq agencies, and aims to analyze the principles underlying Big Data and its relationship with the new Resource Description Framework (RDF) representation patterns; Simple Knowledge Organization System (SKOS) and Ontology Web Language (OWL). The research has a theoretical character and a qualitative approach, as it seeks to present characteristics aimed at describing, understanding and explaining the relationships between Big Data and the new models of representation. From the theoretical survey carried out, it was verified that the representation models analyzed contribute to the interconnection of large volumes of data without losing the context in which they originated, favoring a better understanding of Big Data and the new paradigms of representation in digital environments.

Keywords: Knowledge Organization Systems. RDF. SKOS. OWL. Big Data.

mento que trabalham em conjunto e, seu processamento proporcionou o surgimento de tecnologias como o *Apache Spark*, um *framework* para computação distribuída para análise unificada de processamento de dados em larga escala.

Tais tecnologias surgem como respostas ao nível de velocidade necessárias para analisar o grande volume de dados intitulado *Big Data*. O volume desses dados foi visto como primeiro desafio, pois envolvia um passo fundamental antes de qualquer análise, ou seja, o armazenamento, no qual a evolução das tecnologias de armazenamento foi essencial para alocar dados oriundos dos mais diversos dispositivos.

O armazenamento e a velocidade de processamento permeiam o cerne da discussão sobre *Big Data*, no entanto, dentre os 3Vs (três vês) essenciais que orientam essa vertente de dados, o volume, velocidade e variedade são amplamente conhecidos na literatura. É este último que revela um desafio oriundo da própria natureza da *Web*, que nasce em uma perspectiva sintática e paulatinamente está migrando para uma *Web*

Semântica, isto é, um fato transversal com o *Big Data*, pois o que dificulta essa transição semântica é o fato de existir inúmeros formatos de dados que não podem ser processados adequadamente.

A variedade de formatos e padrões que foram criadas ao longo tempo oferecem a diversidade de dados que possuímos atualmente e, a padronização desses dados oferecem uma alternativa para encontrar os *links* semânticos que unem esse grande volume de dados. Nesse sentido, esta pesquisa discute os princípios que fundamentam o *Big Data* sob a perspectiva de análise dos modelos de representação da informação na *Web* como: o *Resource Description Framework* (RDF), o *Simple Knowledge Organization System* (SKOS) e a *Ontologie Web Language* (OWL).

O objetivo deste trabalho é apresentar o contexto de uso destas tecnologias e como elas são utilizadas para representar as informações no ambiente *Web* no contexto do *Big Data*. O RDF é um padrão que se fundamenta na tripla: sujeito, predicado e objeto na qual é possível atribuir semântica aos dados representados e sua estruturação agrega outras triplas que passam a compor um ecossistema com dados vinculados.

A partir dessa concepção o RDF oferece subsídios para a construção semântica da *Web* por meio de dados interligados que podem ser implementados em sistemas de recuperação de informações com grande volume e variedade de dados. Ademais, por ser um formato desenvolvido a partir da linguagem *eXtensible Markup Language* (XML) possibilita o trabalho em conjunto com outras variantes como a *Web Ontology Language* (OWL) e o *Simple Knowledge Organization System* (SKOS), formatos estes que tornam possíveis a semântica entre os mais distintos conceitos, prática desejada pela *Web* para sistematização e padronização na representação de dados e informações nesse ambiente, abordados a seguir.

2 Resource Description Framework (RDF)

A partir da criação de modelos de dados com grande capacidade semântica, o ambiente *Web* constitui-se como uma das mais ricas fontes de informação, e com isso há um esforço no sentido de converter a *Web* em uma *Web* Semântica, na qual o *Resource Description Framework* (RDF) tornou-se a linguagem escolhida para modelagem, interconexão e fusão de dados (Bischof et al., 2011; Ramalho, 2015).

Desenvolvida em 1997 pelos editores Ora Lassila e Ralph Swick, o RDF foi projetado para ser um mecanismo para trabalhar com metadados que promovem o intercâmbio de informações entre processos automatizados sem perda de significado (Alesso e Smith, 2005; Furgeri, 2006).

Considerada a primeira linguagem que busca ir além das representações sintáticas, RDF permite a formali-

zação de relacionamentos agregando informações semânticas e contextuais que fundamentam a representação semântica entre diversos recursos informacionais, permitindo uma inferência de conteúdo informacional com maior eficiência na interoperabilidade de informações na *Web* (Kontopoulos; Kravari; Bassiliades, 2007), promovendo a ampla comunicação entre diversos sistemas na internet.

Grácio (2002, p. 50) destaca que, o RDF é a “[...] base para o processamento de metadados e tem como principal objetivo proporcionar interoperabilidade entre aplicações que trocam informações eletrônicas da *Web*”. Ou seja, RDF é uma estrutura utilizada para representar e organizar informações na *Web*, cuja sintaxe define um modelo de dados capaz de ligar todas as linguagens baseadas em RDF e suas especificações.

Para Ramalho (2006), a aplicação do RDF em ambientes digitais possibilita adicionar semântica a um recurso informacional independente de sua estrutura, permitindo a reutilização de metadados estruturados e/ou esquemas de metadados.

Nesse cenário, o RDF é um *framework* projetado para utilizar a infraestrutura da WWW aproveitando o armazenamento de informações, a capacidade de comunicação e de processamento de metadados por computadores, e utiliza a linguagem *eXtensible Markup Language* (XML) como base para a representação de recursos.

O RDF e a XML são modelos de representação que se complementam, pois se o RDF é um modelo de metadados que define o domínio de endereços por referência e o relacionamento entre si e diversos outros recursos, a XML é a estrutura semântica que torna esses relacionamentos possíveis.

A criação da *Web* de dados é constituída com a descrição de documentos através das relações existentes entre os links e as coisas arbitrárias descritas em RDF, no qual o *Uniform Resource Identifier* (URI) identifica qualquer tipo de objeto ou conceito descrito em uma linguagem compatível com o RDF.

O RDF explora a infraestrutura da *Web* usando uma forma especial de URI, o *International Resource Identifier* (IRI) identifica um recurso e o *Uniform Resource Locator* (URL) é uma forma de IRI que é utilizado para nomear endereços da *Web*. Dessa maneira, os recursos e os predicados em RDF identificados em URI, possibilitam uma maneira global e única de nomear itens na *Web* (Ferreira; Santos, 2013).

Para Martins e Ramalho (2016), dada a grande capacidade de padronizar declarações de vocabulários, a estrutura de RDF permite definir elementos de metadados específicos às necessidades dos desenvolvedores de sistemas de informação, possibilitando criar relacionamentos formais com qualquer parte de um documento contextualizando as relações já existentes.

O uso de um modelo como o RDF melhora o acesso à grande infinidade de informações disponíveis na *Web* uma vez que sua arquitetura é necessária para suportar a descrição de metadados disponibilizados nesse ambiente (Tauberer, 2014, tradução livre).

Com a criação das boas práticas de trabalho no âmbito da *Web* para o movimento de dados interligados (Marcondes, 2012), cujo objetivo é potencializar a interoperabilidade de informações em ambientes digitais a níveis mais elevados, o modelo RDF é a linguagem que melhor representa e formaliza essa realidade.

Ao utilizar os *namespaces* para realizar as inferências de recursos já existentes no ambiente *Web*, o uso do RDF aumenta a diversidade de termos utilizados na representação, já que sua estrutura é bastante flexível e sem restrições quanto à utilização de conceitos, e os sistemas computacionais podem distinguir se os relacionamentos existentes, entre recursos, representam um termo ou uma definição (Beckett et al., 2014).

Para Berners-Lee (2009, tradução livre), a representação de informações não se trata apenas de colocar dados na *Web*, mas sim de fazer links para que uma pessoa ou computador possa explorar de forma mais abrangente a *web* de dados que, com a interligação de dados em ambientes digitais, é possível encontrar qualquer outro dado relacionado a partir de um já vinculado previamente.

Com isso, a inovação tecnológica desenvolvida ao longo dos anos possibilitou um avanço substancial nos processos de representação de recursos informacionais no ambiente *Web* ao permitir a criação de modelos que potencializam relacionamentos semânticos mais complexos.

3 Simple Knowledge Organization System (SKOS)

O *Simple Knowledge Organization System* (SKOS) foi resultado do desejo de expressar tesouros em RDF dos desenvolvedores Phil Cross, Dan Brickley e Traugott Koch que transformaram os resultados deste estudo em uma proposta publicada em conjunto, pelo *Institute for Learning and Research Technology* (ILRT), no Reino Unido, e a *Lund University Library Netlab*, na Suécia, para codificar um conjunto de relacionamentos de tesouro usando um esquema RDF (Baker et al, 2013). Tal proposta está inserida no Projeto Europeu *Language Independent Metadata Browsing of European Resource* (LIMBER), de 1999-2001, no qual foi definido um vocabulário baseado em "conceitos" rotulados por termos em vários idiomas.-

O SKOS possui uma característica muito comum com os *Knowledge Organization Systems* (KOS), pois com ele é possível criar tesouros, sistemas de classificação, listas de cabeçalhos de assunto, taxonomias, terminologias, glossários (Isaac e Summers, 2009). Este mode-

lo de representação de informações na *Web* não tem como objetivo substituir os modelos tradicionais de organização do conhecimento, como vocabulários controlados em seu contexto original de uso, e sim favorecer uma maior reutilização e interoperabilidade entre os vocabulários existentes. (Ramalho, 2015).

Nesse sentido, as novas formas de representação da informação no ambiente *Web* modificaram a forma de buscar informações, por isso emerge a necessidade do profissional da informação desenvolver habilidades que superem esta necessidade e crie novas demandas. Catarino e Souza (2012) destacam que, além de repensar e discutir, será necessário escrever os vocabulários controlados existentes em linguagens propostas pelo W3C como: RDF, SKOS, OWL e RIF.

Nesse sentido, o SKOS é um sistema simples de representação do conhecimento, escrito em RDF, no qual é possível trabalhar com os insumos da Ciência da Informação. Entender as teorias da Ciência da Informação e as unidades básicas de descrição expressas em XML e RDF são a base para compreender o SKOS e sua aplicação nesse contexto. Catarino, Cervantes e Andrade (2015) ressaltam que, pelo fato de estar em RDF isso torna os esquemas de conceitos expressos neste modelo passíveis de serem lidos por agentes inteligentes conforme as recomendações da W3C.

Segundo Maculan (2015), o SKOS possui uma extensão *eXtension for Labels* (XL) que possibilita o uso de propriedades da aplicação RDF, criando expressões de relações que podem ser interpretadas (legíveis) por máquinas, auxiliando na interoperabilidade entre diferentes vocabulários. Ramalho (2015) explica que os elementos em SKOS podem ser categorizados como: conceitos identificados a partir de URIs; propriedades, que possuem como principal função descrever formalmente atributos dos conceitos, e os relacionamentos, que incorporam semântica aos esquemas, a partir do estabelecimento de diferentes categorias, hierarquias e conceitos

Nesta perspectiva, os SKOS originam-se nos fundamentos do *Knowledge Organization System* (KOS), concebendo o conceito como unidade básica do conhecimento (Dahlberg, 1978). Neste sentido o SKOS caracteriza-se como resultado dos KOSs como corroboram Catarino;(2014); Lara (2015) e Ramalho (2015). Portanto, o SKOS projeta uma relação dos KOSs no contexto *Web*, observando os padrões escritos em RDF/XML para esquematização conceitual e semântica. No que diz respeito ao KOS, os mesmos "[...] são de domínio dos profissionais da informação, no entanto, a modelagem destes, seguindo o padrão da Web Semântica com base no RDF, aparentemente não tem sido difundido o suficiente na formação dos profissionais" (Catarino, 2014, p.18).

Neste sentido, o SKOS envolve a apropriação dos fundamentos da Biblioteconomia e Ciência da Informação

como substrato das tecnologias emergentes, no que diz respeito à classificação, representação e recuperação da informação no ambiente *Web*, as relações semânticas propõem uma ordem conceitual e categórica dentro de uma linguagem.

Em agosto de 2009, o SKOS foi adotado pela W3C como padrão para dar suporte à representação, uso e interoperabilidade de diferentes tipos de vocabulários controlados, como tesouros, taxonomias ou esquemas de classificação (Zoghلامي; Kerhervé e Gerbé, 2011). Como um modelo desenvolvido a partir de padrões interoperáveis e reconhecidos, o SKOS representa a importação dos tradicionais sistemas de classificação e representação do conhecimento para a aplicação de recursos tecnológicos padronizados e nativos do ambiente *Web*, como é o caso do XML, para modelar a descrição de artefatos. No caso do RDF é possível fazer uma modelagem semântica por meio das triplas e o SKOS, a modelagem conceitual.

4 Web Ontology Language (OWL)

Desenvolvida pelo W3C, a *Web Ontology Language* (OWL) é uma linguagem da *Web Semântica* projetada para representar semanticamente um conhecimento rico e complexo sobre diversos recursos e as relações entre eles, e suportar a inferência entre bases de dados (Gao, 2011; World Wide Web Consortium, 2012).

No âmbito da *Web semântica* as ontologias em OWL são fundamentais para a criação e desenvolvimento dos processos de inferência informacional, já que necessitam de uma semântica bem definida para potencializar o intercâmbio de dados em ambientes digitais (Pastor Sánchez, 2012).

Considerada mais expressiva e complexa para a criação de ontologias, a OWL fornece um amplo conjunto de modelagem que permite a um desenvolvedor de ontologias especificar mais ricamente as classes em termos das condições necessárias e suficientes para a adesão a essas classes (Gibbins e Shadbolt, 2009; Silva et al., 2009; Pastor Sánchez, 2012).

Na prática, a OWL proporciona uma maior interpretação de conteúdos informacionais disponíveis na *Web* do que o suportado pelo XML, RDF e RDF *Schema*, uma vez que fornece um vocabulário adicional com uma semântica formal, representando de forma mais ampla o significado dos termos e as relações existentes entre eles.

De acordo com o World Wide Web Consortium (2012), a segunda versão da recomendação da OWL não difere muito do original, uma vez que a função central de compatibilidade entre RDF/XML, o papel de outras regras de sintaxe e os relacionamentos semânticos não foram alterados.

Com isso, uma sintaxe concreta é necessária para armazenar e intercambiar uma ontologia OWL entre

ferramentas e aplicativos, cuja sintaxe de intercâmbio primária é o RDF/XML, que é de fato a única sintaxe que deve ser suportada por todas as ferramentas dessa linguagem (World Wide Web Consortium, 2012).

Freitas Jr. e Jacyntho (2016, p. 53) salientam que a criação de uma ontologia tem por objetivo relacionar nomes a itens representados, no qual a escolha da mesma “[...] deve ser adequada ao objetivo que se espera alcançar na questão de atribuir significados e o reconhecimento dos mesmos por sistemas que irão interpretar as triplas RDF produzidas por uma aplicação”.

Dada as constantes atualizações da linguagem OWL, Ramalho (2006) destaca que as contribuições da área de Ciência da Informação podem ser significativas a partir de estudos interdisciplinares que permitam estruturar e organizar os recursos informacionais desenvolvidos para os ambientes digitais.

Nessa perspectiva, o uso dessas linguagens para a criação de novas linguagens, modelos de representação, vocabulários e ontologias constituem ferramentas tecnológicas valiosas para a representação e recuperação de informações em um determinado domínio, pois com elas é possível aumentar a capacidade semântica dos dados representados na *Web* permitindo que dados e informações sejam reutilizados de forma mais ampla nesse ambiente.

As ontologias conectam conceitos pertencentes a determinada área para efetivar uma busca significativa. Os vocabulários beneficiam-se de padrões que facilitam a busca e recuperação da informação em determinado domínio. Para Freitas Júnior e Jacyntho (2016) uma ontologia é criada de forma a associar nomes a relações entre entidades e aquilo que representam em um universo definido. Por isso, ontologias também são conhecidas por vocabulários.

Nesse cenário, apesar da discussão que envolve *Web Semântica*, *Ontologias*, *LinkedData*, *Bigdata*, *Data Lakes*, etc, estarem no auge das conferências, palestras, treinamentos e atividades voltadas ao profissional da informação, este trabalho volta-se para o elemento comum dessas nuances, o RDF, e apesar dos avanços e usos que o envolvem, a perspectiva da Ciência da Informação precisa atentar para esta forma de representação da informação na web, que atualmente é explorada por diversas instituições, como a própria LC.

A criação de ontologias e catálogos de metadados podem ajudar na descoberta, navegação, exploração e interpretação de lagos de dados heterogêneos e podem simplificar a interpretação, elevar a qualidade dos dados e simplificar a integração de vários conjuntos de dados, para dar suporte semântico à web (Mayer et al, 2018)

Desde a publicação do artigo *The Semantic Web* de Tim Berners-Lee, havia uma proposta de extensão da

Web atual dando um significado bem definido à informação, habilitando computadores e pessoas a trabalharem em cooperação (Berners-Lee; Hendler e Lassila, 2001). No entanto, a resposta não estava apenas em criar ontologias, mas em compreender o estado atual da *web* e como seria possível aplicar essa transição da *web* sintática para *web* semântica, considerando os diversos fenômenos que são subsequentes a essa proposta como o *BigData*, *Linked Data* e recentemente os *Data Lakes*. Isto tem envolvido a Ciência da Informação no elenco da contribuição dos sistemas de organização do conhecimento, classificação e recuperação da informação.

5 Big Data e sua variedade

O *Big Data* é caracterizado pelo grande volume de dados que está distribuído no ambiente digital. Os *browsers*, por exemplo, indexam páginas *web* diariamente e para isso necessitam identificar campos, valores, descritores, *tags*, termos correlatos ou até mesmo sentenças para gerar um resultado de busca consistente. Os avanços ocorridos nas tecnologias de comunicação e informação realçaram o modo de produção e processamento de informações, o uso de novos padrões que suportam essa tendência é adotado como um meio de modelar os dados para representá-los em ambientes dinâmicos como a *web*.

O *Big Data* causa grande impacto na *web*, pois a quantidade de dados produzidos nesse ambiente é tão grande que segundo o UNCTAD (2017), estima-se que só o tráfego global de protocolo da Internet (IP) aumente em um ano 23% durante o período 2014–2019, o equivalente a 142 milhões de pessoas projetando streaming de vídeos HD na Internet simultaneamente, todo dia, todos os dias, até 2019. Até lá, o tráfego global da Internet será 66 vezes o volume de todo o tráfego global da Internet em 2005.

Estes números colossais são redimensionados por notações de armazenamento que outrora não se pensara, pois até então disquetes conseguiam armazenar arquivos no limite de poucos megabytes. Atualmente, quando se fala em *Big Data* é possível deduzir quantidade de dados que podem ultrapassar 1 Geopbyte. Neste sentido, Ribeiro (2014) aponta em seu estudo que só o mercado brasileiro alcançará a marca de 1.6 Zettabytes (1.600 Exabytes) em 2020 conforme as informações apresentadas pela empresa EMC, instituição especializada em armazenamento de dados.

Diante desses números, que são multiplicados em escala global, é possível estabelecer um ponto de referência no universo do *Big Data*, os Vs de Volume e Velocidade são vencidos por outro fator, o terceiro V, de Variedade. Neste espaço os dados podem ser estruturados, semiestruturados ou não estruturados, além de serem provenientes dos mais diversos dispositivos eletrônicos. Dos principais desafios originalmente associados ao *Big Data*, ou seja, volume, velocidade e variedade, o

aspecto Variedade é o menos abordado pelas arquiteturas analíticas padrão. (Mayer et al, 2018).

Em seus estudos sobre gerenciamento da variedade de dados Mayer et al (2018) destacam que a falta de semântica entre os dados, que considerem sua origem, tempo, metadados de descrição padronizado, entre outros fatores, tem proporcionado o surgimento de Lago de Dados (*Data Lakes*), que são repositórios com dados de todos os tipos. Quando não tratados corretamente, um *Data Lake* pode se tornar um “Pântano de Dados” (*Data Swamp*). Portanto, o valor dos metadados e das anotações semânticas está rapidamente se tornando uma consideração-chave, em relação à concepção e manutenção de lagos de dados (Mayer et al, 2018).

O RDF é uma das proposições apresentadas por Mayer et al (2018) como abordagem flexível para representar entidades de dados e suas relações, e gráficos de conhecimento derivados de corpora muito grandes de dados não estruturados, facilitando a análise e a descoberta de conhecimento. Neste sentido a proposição de modelos semânticos no contexto *Big Data* surge como alternativa para contextualizar esses dados diversos. Alternativas como o *eXtensible Markup Language* (XML) já tinha sido apresentada por Wilder-James (2012) para uma possível padronização de armazenamento, devido à característica interoperável, fácil, porém como uma sintaxe extremamente coesa.

Não é por coincidência que o RDF seja uma variação da Linguagem de Marcação XML, pois é possível preservar as características do arquivo XML incorporando a semântica do RDF por meio da tripla: sujeito, predicado e objeto. Os metadados semânticos podem ajudar a descrever e gerenciar a variedade de estrutura, proveniência, visibilidade e uso permitido. (Mayer et al, 2018). Neste sentido os modelos semânticos de origem XML/RDF surgem como alternativa para lidar com a variedade de dados existente no *Big Data*.

5 Conclusão

Por descender do XML, o RDF permite compilar novos modelos que melhor atendam às necessidades de representação e processamento de dados a nível semântico, como o SKOS, um modelo de esquematização de conceitos que pode ser utilizado no ambiente *web* por meio de URIs associados a um determinado conceito. A perspectiva semântica permanece e amplia o modo de representação simples para o mais complexo, no caso das ontologias, que no ambiente *Web* utiliza o formato OWL.

Estes modelos são utilizados como fundamentos para *Web Semântica* e propiciam um alto nível de processamento. Esta pesquisa enfatiza esses modelos que surgem como meio de vincular grandes volumes de dados sem perder o contexto no qual são originados. Portanto, são partes que contribuem para o entendi-

mento do fenômeno Big Data e os novos paradigmas de processamento e representação no contexto da web semântica.

O *Big Data* é um fenômeno contínuo no ambiente digital, pois os dados fluem de todos os cantos da sociedade, de medidores elétricos, câmeras de segurança, registros de chamadas, cliques on-line, atualizações de status mídia e reações postadas (como “curtir”) Unctad (2017). Isso põe em pauta os limites éticos desse fenômeno que utiliza diferentes modos para contextualizar esses dados a interesses econômicos. Em contrapartida, também existem benefícios que podem ser extraídos com esse tipo de tecnologia, por exemplo, na África Subsaariana, onde grandes conjuntos de dados sobre as características do solo são extraídas para ajudar a determinar a quantidade ideal de fertilizante e aumentar a produtividade. Unctad (2017).

Portanto, é necessário refletir holisticamente sobre esse fenômeno considerando os impactos que isso pode causar na sociedade, todavia oferecendo subsídios que possam reduzir sua negatividade. Os modelos semânticos são instrumentos oportunos para melhorar o tratamento da diversidade de dados por meio dos conhecimentos da Ciência da Informação, que possui fundamentos essenciais para o tratamento de dados e sua transformação em conhecimentos que podem impactar positivamente a sociedade.

Referências

- Alesso, H. P.; Smith, C. F (2005). *Developing Semantic Web services*. Massachusetts: A K Peters, Ltd, 2005.
- Baker, T. et al. (2013). Key Choices in the Design of Simple Knowledge Organization System (SKOS) // *Journal of Web Semantics* 20 (2013). <http://www.cs.vu.nl/~guus/papers/Baker13a.pdf>. (2018-04-15).
- Beckett, D. et al (2014). RDF 1.1 Turtle: terse RDF triple language. (2014). <https://www.w3.org/TR/turtle/>. (2018-04-15).
- Berners-Lee, T (2009). *Linked Data*. (2009). <https://www.w3.org/DesignIssues/LinkedData.html>. (2018-04-15).
- Berners-Lee, T.; Hendler, J.; Lassila, O (2001). The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. // *Scientific American*. (2001). https://www.sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feaure%20Article_%20The%20Semantic%20Web_%20May%202001.pdf. (2018-04-15).
- Bischof, S. et al. (2011). Mapping between RDF and XML with XSPARQL. (2011). <http://link.springer.com/article/10.1007/s13740-012-0008-7>. (2018-04-15).
- Catarino, M. E (2014). Simple Knowledge Organization System: construindo sistemas de organização do conhecimento no contexto da Web Semântica // *Informação & Tecnologia (ITEC)* 1: 1 (jan./jun, 2014.) 17-28.
- Catarino, M. E.; Cervantes, B. M. N.; Andrade, I. A. de (2015). A representação temática no contexto da web semântica // *Inf. & Soc.:Est.* 25:3 (set./dez. 2015) 105-116.
- Catarino, M. E.; Souza, T. B. de (2012). A representação descritiva no contexto da web semântica // *TransInformação* 24:2 (maio/ago. 2012) 77-90.
- Dahlberg, I (1978). Teoria do conceito // *Ci. Inf.* 7:2 (1978) 101-107.
- Ferreira, J. A.; Santos, P. C. L. V. A. C. (2013). O modelo de dados resource description framework (rdf) e o seu papel na descrição de recursos. // *Informação & Sociedade: Estudos* 23: 2 (2013) 13-23. <http://basessibi.c3sl.ufpr.br/brapci/v/a/13566>. (2018-04-15).
- Freitas Junior, N.; Jacynoto, M. D. de A (2016). Um protótipo Linked data para catalogação semântica de publicações // *Perspectivas em Ciência da Informação* 21:4 (out./dez. 2016) 48-65. <http://dx.doi.org/10.1590/1981-5344/2664>. (2018-04-15).
- Furgeri, S. (2006). O papel das linguagens de marcação para a Ciência da Informação // *TransInformação* 18:3 (set./dez. 2006). 225-239. <http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/670>. (2018-04-15).
- Gao, J (2011). *Linked Data based Enterprise Content Management*. United Kingdom: University of Southampton. (2011). https://www.academia.edu/5048590/Linked_Data_Based_Enterprise_Content_Management_System. (2018-04-15).
- Gibbins, N.; Shadbolt, N (2009). *Resource Description Framework*. (2009). https://www.academia.edu/2821151/Resource_Description_Framework_RDF_. (2018-04-15).
- Grácio, J. C. A (2002). Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade. Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, (2002). <http://hdl.handle.net/11449/93722>. (2018-04-15).
- Isaac, A.; Summers, Ed (2009). *SKOS Simple Knowledge Organization System Primer: W3C Working Group Note*. (18 August 2009). <https://www.w3.org/TR/2009/>. (2018-04-15).
- Kontopoulos, E.; Kravari, K.; Bassiliades, N (2007). Object-Oriented modeling of RDF schema ontologies, (2007). https://www.academia.edu/16254938/ObjectOriented_Modeling_of_RDF_Schema_Ontologies. (2018-04-15).
- Lara, M. L. G. de. (2015). Propostas de tipologias de KOS: uma análise das referências de formas dominantes de organização do conhecimento // *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação* 20: 1 (fevereiro 2015) 89-107.
- Maculan, B. C. M. dos S (2015). *Estudo e aplicação de metodologia para reengenharia de tesouro: remodelagem do THESAGRO*. Belo Horizonte: Universidade Federal de Minas Gerais, Escola de Ciência da Informação, 2015.
- Marcondes, C. H (2012). “Linked data” – dados interligados - e interoperabilidade entre arquivos, bibliotecas e museus na web // *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação* 17: 34 (maio. /ago., 2012) 171-192. <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17n34p171>. (2018-04-15).
- Martins, P. G. M.; Ramalho, R. A. S (2016). Evolução das linguagens de marcação: um estudo preliminar a partir do enfoque da área de Ciência da Informação. // *3th Encontro Internacional Dados, Tecnologia e Informação. Perspectivas e interdisciplinaridades em Ciência da Informação*, Marília: UNESP, 2016. 733-745. <https://goo.gl/Lec61v>. (2018-04-15).
- Mayer, W et al. (2018). *Variety Management for Big Data*. // Hoppe T., Humm B., Reibold A. (eds) *Semantic Applications*. Springer Vieweg, Berlin, Heidelberg, 2018. 47-62.
- Pastor Sánchez, J. A (2012). *Tecnologías de la web semántica*. Barcelona: Editorial UOC, 2012.

- Ramalho, R. A. S (2015). Análise do Modelo de Dados SKOS: Sistema de Organização do Conhecimento Simples para a Web // *Informação & Tecnologia (ITEC)*. 2:1: (Jan. / Jul. 2015) 66-79. <http://periodicos.ufpb.br/ojs/index.php/itec/article/view/25995/14680>. (2018-04-15).
- Ramalho, R. A. S (2006). *Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação*. Marília: Universidade Estadual Paulista, Marília, 2006. <http://hdl.handle.net/11449/93709>. (2018-04-15).
- Ribeiro, C. J. S. (2014). Big Data: os novos desafios para o profissional da informação // *Informação & Tecnologia (ITEC)* 1:1 (jan./jun., 2014) 96-105.
- Silva, D. L da. et al. (2009). Ontologias e unified modeling language: uma abordagem para representação de domínios de conhecimento // *DataGramaZero* 10:5 (2009). <http://www.brapi.inf.br/index.php/article/view/0000008256/aa38ddb84d6b891d2f6cc49d6b363aaf>. (2018-04-15).
- Tauberer, J (2014). What is RDF and what is it good for? (2014). <https://github.com/JoshData/rdfabout/blob/gh-pages/intro-to-rdf.md>. (2018-04-15).
- Unctad (2017). *Information Economy Report 2017: Digitalization, Trade and Development*. Switzerland, United Nations, 2017.
- Wilder-James, E (2012). What is big data? O'Reilly Radar. (2012). <http://radar.oreilly.com/2012/01/what-is-big-data.html>. (2018-04-15).
- World Wide Web Consortium (2015). OWL 2: Web Ontology Language. (2012). <https://www.w3.org/TR/2012/REC-owl2-overview-20121211/#ref-owl-2-manchester-syntax>. (2018-04-15).
- Zoghliami, K.; Kerhervé, B.; Gerbé, O (2011). Using a SKOS engine to create, share and transfer terminology data sets. // *Proceedings of 7th International Conference on Signal Image Technology & Internet-Based Systems*, 2011, California: IEEE Computing Society, 2011. 46-53.

Copyright: © 2018, Sousa, Martins e Ramalho. This is an open-access article distributed under the terms of the Creative Commons CC Attribution-ShareAlike (CC BY-SA), which permits use, distribution, and reproduction in any medium, under the identical terms, and provided the original author and source are credited.

Received:2018-08-30. Accepted: 2018-09-03