



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CIÊNCIA DA INFORMAÇÃO
GRADUAÇÃO EM BIBLIOTECONOMIA

LARISSA BACELAR PONTES LOPES

A tecnologia *big data*: da *web* às bibliotecas

Brasília/DF
2019

LARISSA BACELAR PONTES LOPES

A tecnologia *big data*: da *web* às bibliotecas

Monografia apresentada como pré-requisito para a obtenção do título de Bacharel em Biblioteconomia pela Faculdade de Ciência da Informação (FCI), da Universidade de Brasília (UnB).

Orientador: Professor Doutor Márcio Bezerra da Silva

Brasília/DF
2019

Dados Internacionais de Catalogação na Publicação (CIP)

L854t Lopes, Larissa Bacelar Pontes
 A tecnologia big data : da web às bibliotecas / Larissa
 Bacelar Pontes Lopes. – Brasília : [s.n.], 2019.
 93 f. : il.

 Monografia (Bacharelado em Biblioteconomia) –
 Universidade de Brasília, Faculdade de Ciência da
 Informação, Curso de Biblioteconomia, 2019.
 Orientador: Prof. Dr. Márcio Bezerra da Silva

 1. Pirâmide informacional. 2. Internet. 3. Big Data.
I. Título.

CDU 002:004



Título: A tecnologia *big data*: da *web* às bibliotecas.

Aluna: Larissa Bacelar Pontes Lopes.

Monografia apresentada à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Brasília, 10 de junho de 2019.

Márcio Bezerra da Silva - Orientador
Professor da Faculdade de Ciência da Informação (UnB)
Doutor em Ciência da Informação

Ailton Luiz Gonçalves Feitosa - Membro
Professor da Faculdade de Ciência da Informação (UnB)
Doutor em Ciência da Informação

Ernani Rufino dos Santos Júnior – Membro
Mestre em Ciência da Informação

*Dedico aos meus maiores apoiadores e
incentivadores: meus pais, Lúcio e Gisele.*

AGRADECIMENTOS

Primeiramente a Deus por direcionar a minha trajetória e me dar forças para perseguir meus objetivos.

Agradeço à minha família, especialmente aos meus pais que foram a minha sustentação durante todo o processo de elaboração da monografia. O amor e a fé que têm em mim, ajudam-me a perseverar diante dos obstáculos e desafios.

Ao meu professor e orientador Márcio Bezerra da Silva, além de valiosos conselhos e aprendizados, mostrou-se sempre paciente e acolhedor, apesar das minhas dificuldades.

Um agradecimento especial aos professores que passaram pela minha jornada na Universidade. Dentre muitos aprendizados, eu conheci docentes modelos que me inspiram e encorajam a ser uma boa profissional.

Não poderia deixar de citar os amigos que tive o prazer de conhecer e dividir momentos de alegria e de aflição durante os anos universitários, em especial: Ylkiane Lopes, Maria Fernanda Leite, Ana Luísa Morais e Tamiê Moraes. E sem esquecer dos amigos da vida que tanto me apoiaram e compartilharam momentos especiais: Ana Paula Araújo, Juliana Máximo, Bernardo Cavalcante, Vanessa Coelho e Leonardo Caruso.

Agradeço aos profissionais bibliotecários com os quais eu tive a oportunidade de aprender e amadurecer, tanto pessoal, quanto profissionalmente, nos estágios que realizei, sendo o primeiro no Tribunal de Justiça do Distrito Federal e Territórios (TJDFT), com Adelina Rodrigues, Dinaisa dos Santos, Georgia do Nascimento, Hosanna Rivetti, Jurema Assunção, Marcelo Hilario, Olimpio Antonio Filho, Raquel Bvuma e André Bertulho; e, por último, no Centro de Documentação e Informação (CEDI), na Câmara dos Deputados, com Raphael Cavalcante da Silva, Judite Martins, Maurinete dos Santos, Ernani Rufino, Ana Patrícia Rondon, Fabyola Madeira, Mariangela Oliveira e Corina Solino.

A todos, o meu muito obrigada.

“[...] certainly the fact that we're all connected, the fact that we've got this information space, does change the parameters. It changes the way people live and work. It changes things for good and for bad”

(Tim Berners-Lee)

RESUMO

Estudo que apresenta um cenário literário de uso da tecnologia *big data*, com destaque às bibliotecas. Engloba no referencial teórico discussão as temáticas: pirâmide informacional, baseada nos elementos dado, informação, conhecimento e sabedoria; Internet, abordando as *webs* 1.0, 2.0 e 3.0; e *big data*, ponderando características e exemplos de adoção. Caracteriza-se como um estudo exploratório, bibliográfico e qualitativo pela coleta de dados. Apresenta, como resultados, a identificação das características básicas da *big data*, representadas por volume, velocidade e variedade (3V); de formas de aplicação da tecnologia nos segmentos corporativo, governamental, científico e cotidiano social; de exemplos de uso da *big data* nas bibliotecas da *Washington University*, *Harvard University* e *Seattle Public*; e na elaboração de um quadro sugestivo de aplicações da tecnologia em setores de biblioteca segundo as características 3V, como volume-velocidade-variedade na administração e na circulação, volume-variedade no setor de desenvolvimento de coleções e valor-volume no setor de referências. Conclui-se que as características 3V, encaradas como básicas da tecnologia *big data* e complementadas pelas propriedades complexidade, veracidade e valor, podem auxiliar na tomada de decisão em setores das bibliotecas, como na administração, na circulação, no desenvolvimento de coleções e na referência, em um cenário de produção e compartilhamento de informações promovidos pelas tecnologias de informação e comunicação, como a Internet e em sentido exponencial.

Palavras-Chave: Pirâmide informacional. Internet. *Web*. *Big Data*. *Big Data* em bibliotecas.

ABSTRACT

Study that presents a literary scenario of use of the big data technology, with emphasis on libraries. Includes in the theoretical referential discussion the subjects: informational pyramid, based on the given elements, information, knowledge and wisdom; Internet, addressing web sites 1.0, 2.0, and 3.0; and big data, pondering characteristics and examples of adoption. Characterizes as an exploratory, bibliographic and qualitative study by data collection. Presents, as results, the identification of the basic characteristics of the big data, represented by volume, speed and variety (3V); of ways of applying technology in the corporate, governmental, scientific and everyday social segments; examples of use of the big data in the libraries of Washington University, Harvard University and Seattle Public; and in the elaboration of a suggestive picture of technology applications in library sectors according to 3V characteristics, such as volume-speed-variety in administration and circulation, volume-variety in the collections development sector and value-volume in the reference sector. Concludes that the 3V characteristics, seen as basic of the big data technology and complemented by the properties of complexity, veracity and value, can help in decision making in sectors of the libraries, such as administration, circulation, collections development and reference, in a scenario of production and sharing of information promoted by information and communication technologies, such as the Internet and in an exponential sense.

Keywords: Informational pyramid. Internet. Web. Big Data. Big Data in libraries.

LISTA DE ILUSTRAÇÕES

Figura 1 Pirâmide informacional.....	22
Figura 2 Página da Pousada Pirineus em 2002.	30
Figura 3 Diretório de pesquisa “Cadê?”.....	31
Figura 4 <i>Website</i> institucional da UnB em 1997.	32
Figura 5 Exemplo de <i>blog</i>	35
Figura 6 Exemplo de fórum.	36
Figura 7 Página da UnB no <i>Facebook</i>	37
Figura 8 Página da UnB no <i>Twitter</i>	38
Figura 9 Disposição de arquivos no <i>Dropbox</i>	39
Figura 10 <i>Google Drive</i> em dispositivos móveis.....	40
Figura 11 <i>Netflix</i> disponível em dispositivos diversos.....	41
Figura 12 Página inicial do canal da UnBTV no <i>Youtube</i>	42
Figura 13 Camadas da <i>web</i> semântica.	43
Figura 14 Resultado de busca no <i>Wolfram Alpha</i>	46
Figura 15 Serviço <i>Siri</i> sobre a previsão de chuva.	47
Figura 16 Quantidade de informações geradas em um dia.	49
Figura 17 Exemplos de variedade de formatos na <i>big data</i>	50
Figura 18 Informação gerada a partir de 2009 e estimativa para 2020.	51
Figura 19 <i>Google Now</i>	53
Figura 20 Sistema personalizado de recomendação no <i>Netflix</i>	54
Figura 21 Sistema personalizado de recomendação na <i>Amazon</i>	55
Figura 22 <i>The Large Synoptic Survey Telescope</i>	56
Figura 23 Sistema de recomendações da <i>Amazon</i>	61
Figura 24 Agenda de compromissos pessoais pelo <i>Google Now</i>	62
Figura 25 Recomendação do <i>Netflix</i>	62
Figura 26 Cinco modos que o <i>WalMart</i> usa a <i>big data</i>	64
Figura 27 Dados dos serviços prestados pela <i>FICO Falcon</i> em 2016.....	65
Figura 28 Aplicação da <i>big data</i> na <i>Gendarmerie Nationale</i> (França).	66
Figura 29 Câmera do LSST.....	67
Figura 30 Relatório anual da <i>Washington University Libraries</i> (2017/2018).....	69
Figura 31 Projeto <i>Geospatial Information Systems</i> (2017).	70
Figura 32 <i>Library Analytics Toolkit</i> desenvolvida por <i>Harvard Library Lab</i>	71
Figura 33 Painel da <i>Seattle Public Library</i> sobre empréstimos de obras.	72

LISTA DE QUADROS

Quadro 1 Caracterização da pesquisa	57
Quadro 2 Características básicas da <i>big data</i>	60
Quadro 3 Sugestão de aplicação da <i>big data</i> em bibliotecas	73

LISTA DE SIGLAS E ABREVIATURAS

3V	Volume, velocidade e variedade
ARPA	<i>Advanced Research Projects Agency</i>
ARPANET	<i>Advanced Research Projects Agency Network</i>
BBS	Banco de Dados
BD	<i>Balanced Scorecard</i>
BDTD	Biblioteca Digital de Dissertações e Teses
BRAPCI	Base de Dados Referencial de Artigos e Periódicos em Ciência da Informação
BSC	<i>Bulletin Board System</i>
CI	Ciência da Informação
EUA	Estados Unidos da América
HTML	<i>HyperText Markup Language</i>
HTTP	<i>HyperText Transfer Protocol</i>
IA	Inteligência Artificial
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
IDC	<i>International Data Corporation</i>
JPEG	<i>Joint Photographics Experts Group</i>
LSST	<i>Large Synoptic Survey Telescope</i>
MTML	<i>Mahanagar Telephone Mauritius Limited</i>
NS	<i>Namespace</i>
PB	<i>Petabytes</i>
PDF	<i>Portable Document Format</i>
RDF	<i>Resource Description Framework</i>
SGBD	Sistema de Gerenciamento de Base de Dados
SPARQL	<i>Protocol and RDF Query Language</i>
TB	<i>Terabytes</i>
TCC	Trabalho de Conclusão de Curso
TIC	Tecnologias de Informação e Comunicação
UnB	Universidade de Brasília
UNIX	<i>Uniplexed Information and Computing System</i>
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
W3C	<i>World Wide Web Consortium</i>
WWW	<i>World Wide Web</i>
XML	<i>eXtensible Markup Language</i>
XSD	<i>XML Schema Definition</i>
ZB	<i>Zettabytes</i>

SUMÁRIO

1 INTRODUÇÃO	15
1.1 PROBLEMAS	17
1.2 JUSTIFICATIVA	18
1.3 OBJETIVOS	20
1.3.1 Objetivo geral	20
1.3.2 Objetivos específicos	20
2 FUNDAMENTAÇÃO TEÓRICA	21
2.1 PIRÂMIDE INFORMACIONAL	22
2.2 INTERNET: da rede de computadores à <i>web</i>	27
2.2.1 <i>Web 1.0</i>	29
2.2.2 <i>Web 2.0</i>	32
2.2.3 <i>Web 3.0</i>	42
2.3 <i>BIG DATA</i> : conceitos e possibilidades de aplicação	48
3 METODOLOGIA	57
3.1 CARACTERIZAÇÃO DA PESQUISA	57
3.2 ETAPAS DA PESQUISA	59
4 RESULTADOS DA PESQUISA	60
5 CONSIDERAÇÕES FINAIS	75
REFERÊNCIAS	77

1 INTRODUÇÃO

No decorrer das décadas é possível perceber mudanças no funcionamento das sociedades, o que implica a dinâmica entre as pessoas, a forma como elas se comunicam. Entre os acontecimentos que promoveram essa mudança social, menciona-se o advento da escrita. Charaudeau (2018) aponta a escrita como uma ação que consiste no uso coordenado e ordenado de símbolos e signos. Complementando, Pinto (2006) expõe que a escrita possibilitou a preservação da memória humana a partir de registro de informações desde a Antiguidade, pois, como aponta Vygotsky (1998), uma vez registrados, poderiam ser acessados à posterioridade. Séculos à frente, a máquina de imprensa inventada por Gutemberg¹, aproximadamente em 1450, possibilitou uma forma mais rápida de produção de livros, propiciando que a circulação de informações em suportes físicos se tornasse mais facilmente disponível e para um público mais amplo (BRIGGS; BURKE, 2004). Tempos depois, em meados de 1870, com o surgimento do telefone por Bell², possibilitou-se a comunicação entre países, rompendo fronteiras territoriais (Ibid, 2004). Contudo, foi a partir dos adventos das tecnologias de informação e comunicação (TIC)³, especialmente representados pelo computador e pela Internet, respectivamente entre as décadas de 1940 e 1980 (Ibid, 2004), que as dinâmicas sociais sofreram significativas mudanças, diminuindo ainda mais as distâncias entre as nações e ampliando as formas de comunicação entre as pessoas (CASTELLS, 2003), o que potencializou, e continua expandindo, uma geração de intensa produção e consumo de informação, em menor tempo e em maior velocidade, a qualquer hora e em qualquer lugar.

Desde a Segunda Guerra Mundial, a Internet vem evoluindo, alterando a forma como as pessoas a usam, como interação entre si, enquanto ações formalizadas nas dinâmicas de disseminação, recuperação e acesso de informações. Neste sentido, a compreensão da evolução da Internet pode ser acompanhada pelas gerações da camada intitulada *World Wide Web* (WWW), que por sua vez, diante de características específicas, ajudam a entender as mudanças no comportamento das sociedades.

¹ Johannes Gensfleisch zur Laden zum Gutenberg.

² Alexander Graham Bell.

³ Expressão que “[...] refere-se à conjugação da tecnologia computacional ou informática com a tecnologia das telecomunicações e tem na Internet e mais particularmente na *World Wide Web* (WWW) a sua mais forte expressão” (MIRANDA, 2007, p. 43).

Percebida por Nonato (2015, p. 4) como “[...] um sistema de documentos de hipermídia que são interligados e executados na Internet”, a WWW tem a sua primeira geração caracterizada pela produção e disseminação de informação, principalmente no ramo dos negócios. Como o sentido era de comunicação, de publicitar pessoas, produtos e serviços, a *web* 1.0 (estática) apresentava *websites* de leitura, sem interações, ou seja, concedia “[...] interações de usuário limitadas ou contribuições de conteúdo e permitia apenas pesquisar informações e lê-las” (AGHAEI; NEMATBAKHS; FARSANI, 2012, p. 1, tradução nossa). Em relação a *web* 2.0 (colaborativa), “[...] caracteriza-se por potencializar as formas de publicação, compartilhamento e organização das informações [...]” (PRIMO, 2007, p. 2), alterando a dinâmica estabelecida na geração anterior no sentido de integração, participação e cooperação entre os usuários do *website*, como uma espécie de conhecimento coletivo, evoluindo do papel de espectador para a função de criador (KEEN, 2009). No que se refere a terceira geração, como uma consequência da anterior, idealizou-se uma *web* (semântica) em que fosse possível estruturar a enorme quantidade de dados produzida, relacionando-as com base em um sentido semântico e analisando-os de forma que a informação resultante possibilitasse uma tomada de decisão. Neste sentido, a *web* se apresenta em uma perspectiva de sabedoria que objetiva “[...] conectar, integrar e analisar dados provindos de fontes e formatos diversos para gerar novas informações” (AGHAEI; NEMATBAKHS; FARSANI, 2012, p. 5, tradução nossa). Os autores ainda afirmam que “a mais importante proposta da *web* semântica é fazer a *web* mecanicamente legível para máquinas e não apenas para humanos” (Ibid, p. 5, tradução nossa).

Ao passo em que a *web* foi evoluindo em gerações, a humanidade foi se construindo em uma forma de “sociedade em rede” a partir da popularização da Internet, que por sua vez caracteriza as transformações do modo de produção, disseminação e uso da informação na sociedade (CASTELLS, 2003), o que resulta na promoção da integração social e no aumento, em qualidade e quantidade, do nível de informação dos indivíduos (TAKAHASHI, 2000). Por outro lado, se a quantidade de informação é uma constante, a qualidade pode ser uma característica problemática, pois dependerá do quanto foram tratadas. Desta forma, desenha-se um ciclo que acompanha o nível de abstração/estruturação da informação, isto é, formado pelo dado, que se transforma em informação, que se torna conhecimento e que se consolida como sabedoria.

Diante do cenário em que o homem por natureza não consegue absorver tudo o que está à sua volta, ter informações específicas em suas mãos é fundamental para a correta tomada de decisão. É justamente nesta realidade que a *web 3.0* se mostra como uma geração ainda em desenvolvimento que permeia a sabedoria, ao estruturar e analisar uma grande quantidade de dados a partir de máquinas, acarretando em informações tratadas que ajudarão ao homem em sua tomada de decisão.

O prisma da sabedoria pode ocorrer a partir da análise e mineração de dados, que por sua vez conceitua-se como “[...] o processo não-trivial de identificar, em dados, padrões válidos, novos, potencialmente úteis e ultimamente compreensíveis” (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996, p. 40, tradução nossa). Contudo, ao avançar das gerações da *web*, a quantidade de dados cresce exponencialmente a cada dia na Internet, necessitando de uma forma de análise mais veloz e que assuma um maior volume, culminando na chamada *big data*. Para Taurion (2015), *big data* mostra-se como um conjunto de tecnologias, processos e práticas que permitem a análise de um grande volume e de uma grande variedade de dados que, em sua maioria, não eram acessíveis, em um espaço de tempo curto e passível de gerar valor de decisão.

Diante do contexto apresentado, o presente trabalho estrutura-se da seguinte maneira: o primeiro capítulo abrange a introdução, os problemas, a justificativa e os objetivos; o segundo capítulo apresenta a fundamentação teórica, abordando a pirâmide informacional, a evolução das *webs* e a tecnologia *big data*; o capítulo três delimita-se a metodologia do estudo; no quarto capítulo são apresentados os resultados da pesquisa, entre exemplos de uso da tecnologia *big data* em diversos segmentos da sociedade e, entre eles, a biblioteca (ensino); por fim, no capítulo cinco estão expostas as considerações finais e expectativas de estudos futuros.

1.1 PROBLEMAS

As mudanças globais que afetam a dinâmica social, transformando os indivíduos, as suas ideias, também alteram a mecânica e funções das organizações e das profissões ali realizadas. As bibliotecas, por exemplo, no decorrer da história humana, tiveram, e continua tendo, o seu papel discutido e modificado, o que conseqüentemente gera mudanças na atuação profissional do bibliotecário.

Se na Idade Média as bibliotecas eram meros depósitos (SILVA, 2006), com o desenvolvimento da indústria e a crescente urbanização, nos séculos XVIII e XIX, a biblioteca assumiu encargos sociais e educacionais (MUELLER, 1984). Neste caso, o bibliotecário já não tem apenas a atribuição única de guardar o acervo, mas de armazenar e disseminar a informação, pois o foco deixa de ser os suportes e passa para a sociedade, ou seja, para quem necessita de informação e a acessa. Contudo, algumas funções básicas da biblioteca “[...] não mudam – coleta, preservação, organização e difusão de informações – mas sim a maneira de desempenho e objetivos de seus serviços” (MULLER, 1984, p. 7).

Tanto o bibliotecário, quanto a biblioteca em si, assumem o papel de mediadores do conhecimento, se tornando agentes ativos às transformações sociais (BEIJO; VALOTO, 2014), ao buscarem “[...] conectar os indivíduos às bases de conhecimento e demais informações.” (CARVALHO; LIMA, 2017, p. 2). A referida inferência remete-se ao surgimento e a popularização de TIC no cotidiano das sociedades como um provedor na mudança de paradigma na Biblioteconomia, isto é, que deixava de ser uma técnica de organização (LE COADIC, 2004) e passa a se preocupar com a recuperação de informação. Se por um lado, essa preocupação está direcionada ao atendimento às necessidades informacionais dos usuários, como a disponibilização de dados de pesquisa, por outro, estar atento ao que vem sendo buscado pelo público nos catálogos pode contribuir na tomada de decisões sobre a gestão do acervo, o que pode influenciar na aquisição de obras, por exemplo.

Pelo fato de a biblioteca transcender o ambiente físico, oferecendo produtos e serviços digitais, ampliando o volume de informação produzida, armazenada e disponibilizada para recuperação e uso, cria-se um campo em que profissionais da informação, como bibliotecários e cientistas da informação, podem coletar, tratar e analisar dados com fins de tomada de decisão. Portanto, considerando a tecnologia *big data* nas bibliotecas, suscita-se a pergunta que fundamenta este trabalho: *Como se daria a adoção de big data nos serviços realizados e ofertados pelas bibliotecas?*

1.2 JUSTIFICATIVA

A biblioteca possui setores diversos, realizando ações de organização, tratamento e disseminação da informação. A partir da contemporânea revolução digital inaugurou-se uma nova fase, ou seja, oferecer produtos e serviços voltados às

necessidades do usuário e no formato digital, especialmente tirando proveito da Internet. Entretanto, não basta oferecer novos recursos sem que haja um acompanhamento, uma análise, com fins de futuras decisões, tanto sobre a continuidade dos produtos e dos serviços, pela atualização ou encerramento, assim como pela satisfação positiva, ou não, dos usuários. Dessa forma, a necessidade de avaliar a qualidade dos serviços e produtos oferecidos vai ao encontro de indicadores⁴ enquanto insumos de avaliações e assim subsidiar “[...] aos serviços de informação conhecer a adequação de seus serviços, seu rendimento e suas falhas” (ROZADOS, 2004, p. 13). Por outro lado, ainda que dados estatísticos não garantam a geração de indicadores (GEISLER, 2000), os valores apurados informam comportamentos, modo de operação e funcionamento da organização (ROZADOS, 2004), o que nas bibliotecas pode ser visualizado no setor de referência, por exemplo.

O setor de referência integra diferentes técnicas da Biblioteconomia, como: a) elaboração de tesauro; b) estudo de usuário; e c) bibliometria. Enquanto o tesauro permite o estabelecimento de relações entre termos e a padronização, possibilitando mensurar os termos mais pesquisados, e o estudo de usuários fornece informações acerca do comportamento informacional de seu público, fomentando reconhecer as necessidades do mesmo, a bibliometria caracteriza-se como um instrumento de gestão de informação registrada, em que os quantitativos relacionam-se às ações de empréstimo e devolução, e ainda possibilita identificar obras de maior relevância para a respectiva área do conhecimento (FIGUEIREDO, 1974).

Diante de suas características, constata-se que o serviço de referência é um setor que fornece dados quantitativos e qualitativos, que quando reunidos, periodicamente, resultam em um volume informacional que pode ser analisado e, conseqüentemente, gerar indicadores por meio de tecnologias como a mineração de dados e a análise de dados em grandes quantidades e em alta velocidade, a chamada *big data*.

Na iminência de que o bibliotecário deve manter-se em linha às TIC, a ideia não explorada em unidades de informação de Brasília de se usar a tecnologia *big data* nesses espaços foi sendo construída ao longo do curso de Biblioteconomia, da Universidade de Brasília (UnB), especialmente nas disciplinas “Planejamento e Elaboração de Base de Dados” (2016/1), “Redes de Informação e Transferência de

⁴ Indicador é “a especificação quantitativa e qualitativa para medir o atingimento de um objetivo” (BANCO INTERAMERICANO DE DESAROLLO, 1997, p. 6, tradução nossa).

Dados” (2017/1) e “Informática Documentária” (2017/2). Nas referidas disciplinas, ao debaterem a Internet como um ambiente onde o homem está inserido e que oferta aplicações diversas, como as redes sociais e os mapas interativos, gerando uma grande quantidade de dados e que é analisada em um curto espaço de tempo, como a indicação do melhor trajeto até o trabalho, nasceu o interesse em verificar a mencionada realidade de produção e análise no contexto da biblioteca, neste caso, a partir de uma investigação literária sobre características e práticas da tecnologia *big data*.

1.3 OBJETIVOS

1.3.1 Objetivo geral

- Apresentar um cenário literário de uso da tecnologia *big data* em bibliotecas.

1.3.2 Objetivos específicos

- Identificar características básicas de *big data*;
- Elencar formas de aplicação de *big data* no contexto geral;
- Apontar exemplos de uso de *big data* em bibliotecas;
- Exibir um quadro sugestivo de aplicações de *big data* em bibliotecas.

2 FUNDAMENTAÇÃO TEÓRICA

A busca pela verdade constitui-se como um objetivo universal da Ciência Moderna, que por sua vez molda-se, segundo Bachelard (2002), na tradição herdada da antiga Grécia, e modificada pelas culturas islâmica e europeia. Trata-se de um panorama de influências e modificações que constrói um ciclo em que a ciência vive, caracterizado pela constante ruptura de antigos modelos e tradições e, conseqüentemente, pela construção de novas verdades (Ibid, 2002).

Ao considerar o conhecimento como algo cumulativo (BACHELARD, 2002), a ciência faz uso desse aglomerado com fins de apropriar-se da realidade, explicar racional e objetivamente, como também prever resultados entre fenômenos. Esse conjunto, ou seja, o conhecimento, puro ou aplicado, somado a métodos, possibilita a compreensão de fenômenos, propor soluções e aplicá-los (JAPIASSU, 1983). Neste sentido, a ciência consiste na aplicação de métodos rigorosos com a finalidade de comprovar hipóteses pré-estabelecidas, ruptura ou confirmação de paradigmas (CANDIOTTO; BASTOS; CANDIOTTO, 2011), originando o que se pode chamar de pesquisa científica, qualificada por um método adotado e que, por sua vez, assegura confiabilidade nos resultados. Para tanto, existem etapas na pesquisa científica e entre elas encontra-se a fundamentação teórica, compreendida como uma fase primordial para explicitar os estudos sobre a temática abordada e sustentar as hipóteses levantadas. A fundamentação teórica se apresenta a “[...] partir de verdades garantidas para novas verdades” (SEVERINO, 2002, p. 83) e assim trazer a relevância dos conhecimentos e teorias com fins de solucionar ou explicar o problema proposto, (KÖCHE, 2001).

No caso do presente trabalho de conclusão de curso (TCC), a fundamentação teórica estrutura-se em discussões acerca da pirâmide informacional, apresentando conceitos de dado, informação, conhecimento e sabedoria; exposições sobre a evolução histórica da *web*, a partir das redes de computadores até alcançar a fase 3.0; e finaliza com explanações sobre a tecnologia *big data*, levantando conceitos, características e exemplos de aplicação.

2.1 PIRÂMIDE INFORMACIONAL

Na literatura científica é amplamente difundido a dificuldade em encontrar consenso em conceituar os termos “dado”, “informação” e “conhecimento”, “[...] que podem ser objetos de uma controvérsia de sentido [...]” (AQUINO, 2008, p. 80). O fato dos citados termos penetrarem em áreas diversas do conhecimento apenas intensifica a complexidade da situação. Contudo, evocando os processos de tomada de decisão, que por sua vez apresenta subsídios da “inteligência/sabedoria”⁵, enquanto uma extensão do conhecimento, ao pensar no conhecimento como uma aceitação e acúmulo nos recipientes da mente humana (BARRETO, 2002), faz-se necessário compreender o dado, a informação, o conhecimento e a sabedoria, minimamente, neste caso, usando como base a pirâmide informacional.

A pirâmide informacional é uma representação que discute “[...] a problemática da geração, da organização, da transferência e do aproveitamento da informação [...]” (URDANETA, 1992, p. 99, tradução nossa) a partir de uma relação de hierarquia por níveis, definida pelas variáveis “*qualidade vs. quantidade*” (Ibid, p. 100, tradução nossa). Conforme a sequência apresentada na figura um (1), enquanto a base (dados) possui um caráter mais quantitativo, no pico (inteligência) há um sentido qualitativo.

Figura 1: Pirâmide informacional.



Fonte: Urdaneta, 1992, p. 100.

⁵ No presente trabalho será adotada a expressão “sabedoria”, que significa “[...] conhecimento sobre conhecimento, saber qual conhecimento é relevante e qual conhecimento aplicar. [...] Sabedoria é singular, o que significa que não é quantificável como comparado a dado, à informação e ao conhecimento” (ALLEN, 2017, p. 16, tradução nossa)

Generalizando, os dados constituem, na área da Linguística, os “[...] registros icônicos, simbólicos (fonéticos ou numéricos) ou signos (linguísticos, lógicos ou matemáticos) por meio dos quais se representam fatos, conceitos ou instruções [...]” (URDANETA, 1992, p. 101, tradução nossa). Nos campos da Biblioteconomia e da Ciência da Informação (CI) é possível encontrar compreensões sobre dados como sendo “[...] elementos brutos, sem significado, desvinculados da realidade” (ANGELONI, 2003, p. 18). Trata-se de um elemento puramente matemático, armazenado e processado por computadores (SETZER, 1999), e sem a capacidade de “[...] transmitir uma mensagem ou representar algum conhecimento [...]” (MOREIRA; SANT’ANA; JORENTE, 2016, p. 73). Considerado o dado como uma unidade que promove outros elementos, Aquino (2008, p. 82) defende que:

São compreendidos como a matéria-prima a ser trabalhada e processada em pequenas ou grandes quantidades, através das atividades de coleta, envio, transmissão, processamento, armazenamento, feitura de “links” e escoamento dos dados codificados.

Considerando a premissa de que o dado constitui a matéria-prima da informação (ALBRECHT, 2004), correspondente ao segundo nível da pirâmide informacional (figura 1). De forma ampla, a informação refere-se ao “[...] ordenamento dos dados em função de obter um sentido cognitivamente relevante” (URDANETA, 1992, p. 101, tradução nossa), dotada de relevância e propósito (DRUCKER, 1989).

No contexto da Biblioteconomia e da CI, autores como Setzer (1999) afirmam que o principal atributo que separa os dois níveis da pirâmide informacional é o fato de a informação ser dotada de semântica, ou seja, significado ou contexto. Shannon e Weaver (1975) discutem a informação quanto a sua geração, neste caso, relacionada ao processo de comunicação⁶, o que exige a interação entre um emissor e um receptor para a transferência de informação segundo a Teoria Matemática da Comunicação⁷, desenvolvida pelos autores. Santos e Sant’Ana (2002) complementam ao exporem que essa transferência pode ser realizada por meio de conversação e de interação entre agentes, tanto humanos, quanto não humanos. Neste sentido, seria a transmissão de informação por uma conexão que envolve, tanto dois agentes não

⁶ De forma geral, comunicação “é a forma de interação humana realizada através do uso de signos” (BORDENAVE, 1984, p. 12).

⁷ Do inglês *The Mathematical Theory of Communication*.

humanos, como “[...] de um agente humano para um agente não humano dotado de alguma capacidade de tratamento desta informação [...]” (SANTOS; SANT’ANA, 2002, p. 4).

Considerando que a transmissão de informação pode ocorrer entre agentes não humanos, conseqüentemente também é considerável ponderar que ela seja representada e armazenada em computadores na forma de dados (SANTOS; SANT’ANA, 2002), ou seja, como “[...] unidades rigorosamente designáveis que se transformam em ‘bits’” (MORIN, 2000, p. 175), Trata-se de uma conversão/representação “[...] transformada pela máquina – como na formatação de um texto – mas não os seu significado, já que este depende de quem está entrando em contato com a informação” (SETZER, 1999, p. 2). Desta forma, o que atribui valor à informação, na visão da Administração, é o seu uso e que está diretamente ligado à “[...] capacidade de um indivíduo em retirar significado dessa informação [...]” (MARTINS, 2000, p. 5).

A informação adquire, nos mais diversos segmentos da sociedade, especialmente em uma contemporaneidade permeada pelas TIC, um status de “[...] recurso econômico e insumo produtivo [...]” (AQUINO, 2008, p. 83), o que “[...] requer análise, consenso de significado, e os indivíduos lhe atribuem sentidos e a transforma” (Ibid, p. 82) em conhecimento, que tanto pode ampliar a compreensão sobre determinada questão, como é possível construir uma nova interpretação sobre o que anteriormente era conhecido. Direcionando essa interpretação à pirâmide organizacional, o conhecimento refere-se ao terceiro patamar, considerado de forma ampla por Davenport e Prusak (2000, p. 5) da seguinte forma:

[...] um fluido misto de experiência, valores, informação contextual e perspectivas especializadas que fornece um enquadramento para avaliar e incorporar novas experiências e informação. É originado e é aplicado na mente dos indivíduos.

O conceito de Davenport e Prusak, que explora o conhecimento como um processo cognitivo e que pode alterar-se de acordo com o contexto ao qual está inserido, é comumente adotado como base discursiva nos campos da Biblioteconomia e da CI. Neste caso, o conhecimento constitui-se de um conjunto de “[...] estruturas informacionais que, ao se internalizarem, se integram a sistemas de relacionamentos simbólicos de mais alto nível e permanência” (URDANETA, 1992, p. 102, tradução

nossa), que por sua vez combinam a conjuntos de conceitos, valores e experiências de aprendizagens do próprio indivíduo. Ao considerá-lo desta forma, segundo os conceitos supracitados, o conhecimento gera-se a partir de processos internos do indivíduo, o que faz com que seja “[...] difícil de ser categorizado, localizado, transferido entre indivíduos ou grupos” (AQUINO, 2008, p. 83). Considerando-o por outro viés, diante da sua característica de ser algo subjetivo, o conhecimento relaciona-se à “[...] pragmática, isto é, [...] com alguma coisa existente no ‘mundo real’ do qual temos uma experiência direta” (SETZER, 1999, p. 3).

Como forma de viabilizar e compreender a possível transferência de informação, o conhecimento apresenta-se, especificamente, entre explícito e tácito. No caso do conhecimento explícito, refere-se ao “[...] conjunto de informações já explicitadas em algum suporte (livros, documentos, etc.) e que caracteriza o saber disponível sobre o tema específico” (MIRANDA, 1999, p. 287), facilmente transmitido entre os indivíduos por meio da linguagem formal, pois ainda que represente apenas uma pequena parte de todo o conhecimento (NONAKA; TAKEUCHI, 1997), “[...] pode ser formalmente articulada de maneira mais precisa. Ele pode ser mais facilmente codificado, documentado, transferido ou compartilhado” (SANTOS; SANT’ANA, 2002, p. 7). Quanto ao conhecimento tácito, “[...] é o conhecimento pessoal incorporado à experiência individual e envolve fatores intangíveis (crenças pessoais, valores e perspectivas) e é difícil ser articulado na linguagem formal” (DI CHIARA; ALCARÁ; TOMÁEL, 2005, p. 97). Contudo, pela possibilidade de transferência, infere-se que o conhecimento explícito tem ganhado papel mais importante, tanto economicamente, como um “[...] fator de produção na economia do conhecimento [...]” (SANTOS; SANT’ANA, 2002, p. 6), quanto nas organizações, como um fator competitivo.

Ao encarar o conhecimento como um recurso essencial para as competências individuais e inovações sociais, na verdade, fala-se aqui de um elemento fundamental para tomada de decisão. Em outras palavras-, evoca-se ao que se encontra no topo da pirâmide informacional, ou seja, a sabedoria, compreendida como “[...] ‘informação como oportunidade’, é dizer, de estruturas de conhecimento que, sendo contextualmente relevantes, permitem a intervenção vantajosa da realidade” (URDANETA, 1992, p. 103).

No escopo da Biblioteconomia e da CI também é possível encontrar uma discussão tênue entre o conhecimento e o saber, que por sua vez “[...] diz respeito ao grau mais profundo de apreensão e sistematização dos sentidos e aprendizados [...]” (MARTELETO, 2007, p. 9). Na essência da sabedoria, a autora defende que:

[...] *saber* [...] reúne os cabedais sociais, técnicos e cognitivos de instituições, pessoas e grupos, em grau de internalização e externalização que alia os conhecimentos às práticas e às experiências, formando, acervos, registros e documentos individuais e coletivos. (MARTELETO, 2007, p. 10, destaque do autor)

A sabedoria, expressão adotada como equivalente a inteligência (URDANETA, 1992), seria, então, a forma mais elevada de cognição, aliada à experiência e presente na totalidade dos modos e sistemas de conhecimento (MARTELETO, 2007), o que permite realizar uma tênue associação entre sabedoria e inteligência como expressões sinônimas. Diante dessa aproximação terminológica, Figueiredo (1999, p. 26) compreende a sabedoria como uma ação que se apresenta da seguinte maneira:

A capacidade de um indivíduo ou uma organização social [...] de adquirir nova informação ou conhecimento, fazer julgamentos, adaptá-la ao meio, desenvolver novos conceitos estratégicos e agir de maneira racional e eficiente, com base na ação recebida.

Tendo a sabedoria aliada à experiência, o seu uso resultaria em uma ação dependente do conjunto de informações adquiridas e conhecimentos gerados, o que, em uma organização, por exemplo, atinge o “nível mais avançado da hierarquia quando ela monitora ativamente seus sistemas para garantir que seu comportamento esteja de conformidade com o planejado” (IKEMATU, 2001, p. 1), de tal maneira que é possível “[...] detectar e diagnosticar qualquer comportamento anormal do sistema” (Ibid, p. 1) e, conseqüentemente, tomar uma decisão enquanto fruto da sabedoria.

Em suma, retomando a figura um (1), primeiramente ocorre a produção de dados, especialmente em tempos recentes, em quantidades exponenciais e desprovidas de organização e de significado; enquanto que as informações, formadas por dados organizados e associados, segundo contextos específicos, são providas de significado e valor. Afunilando a seqüência, os conceitos de conhecimento remetem a organização e processamento de informações e dados, que por sua vez resultam em aprendizados e compreensões acerca de determinada questão, assunto; avalizando

a sabedoria enquanto o nível superior da pirâmide informacional, formalizada por um conjunto de experiências e aprendizados e que representa o alcance máximo da cognição humana, capazes de gerar novas descobertas, bem como subsidiando a tomada de decisões.

O breve seguimento da pirâmide informacional pode ser identificado na Internet, especialmente quanto à evolução das fases da *web*. Aparentemente, as tecnologias da *web* vêm buscando aproximação da pirâmide informacional, delineando espaços digitais criados e usados apenas com fins de consumo/leitura, enquanto outros remetem a interações entre usuários, a produção e ao compartilhamento de conteúdos e a inferência de dados com fins de recomendações.

2.2 INTERNET: da rede de computadores à *web*

As inovações tecnológicas têm um papel significativo na evolução da forma de comunicação das sociedades, especialmente, nos últimos tempos, o que influencia diretamente na alteração da dinâmica social. Entre essas inovações destaca-se a Internet, cunhada em tempos de guerra, tendo as suas primeiras versões criadas durante a Segunda Guerra Mundial. De acordo com Squeff e Negri (2017), trata-se de um período marcado pelo grande número de pesquisas científicas voltadas à produção e fabricação de armas e tecnologias que possibilitassem vantagens em relação aos inimigos.

A Internet surgiu a partir de esforços conjuntos de centros governamentais e centros universitários (CASTELLS, 2003). Em outras palavras, “a Internet não teve origem no mundo dos negócios” (Ibid., p. 25), pois seria um investimento de alto custo e alto risco. No caso das iniciativas governamentais, os primórdios da Internet começaram como um projeto do setor da *Advanced Research Projects Agency* (ARPA), formado pelo Departamento de Defesa dos Estados Unidos, no final da década de 1950. O objetivo da agência era “[...] mobilizar recursos de pesquisa, particularmente do mundo universitário, com o objetivo de alcançar superioridade tecnológica militar em relação União Soviética [...]” (Ibid., p. 14). No viés acadêmico, os centros de pesquisas em universidades estadunidenses são os destaques, pois foram o aporte que garantiu a expansão global da Internet, especialmente por meio das iniciativas de redes autônomas, especialmente na década de 1990 (CAÑEDO ANDALIA, 2004).

Inicialmente, a *Advanced Research Projects Agency Network* (ARPANET) foi um esforço voltado para permitir a ligação de redes locais por meio de computadores com o objetivo de “[...] obter um nível maior de conexão, uma área mais ampla e uma rede mais rápida [...]” (CAÑEDO ANDALIA, 2004, p. 3), a partir de uma tecnologia revolucionária de comutação por pacote, inicialmente voltada para fins militares. O projeto forneceu uma rede descentralizada e flexível, o que possibilitou, inicialmente, a comunicação de quatro nós entre centros universitários da Califórnia e de Utah, no ano de 1969. Em 1971, já eram 15 nós espalhados entre centros universitários de pesquisa dos Estados Unidos. Com a referida tecnologia de comunicação criada, ao longo da década de 1970 foram feitos esforços para a criação e implementação de padrões que possibilitassem a comunicação entre redes (CASTELLS, 2003).

Em paralelo, outros esforços foram feitos por pessoas engajadas ao desenvolvimento tecnológico da época, como o *Bulletin Board System* (BBS), sendo um sistema de quadro de avisos, no fim da década de 1970, e o sistema operacional UNIX⁸, desenvolvido pelos Laboratórios *Bell* e disponibilizado para as universidades no ano de 1974. Outro desenvolvimento notável ocorreu na década de 1990, época em que a ARPANET já se encaminhava para a obsolescência. Nesse ano, um estudante desenvolveu o sistema operacional *Linux*, baseado no sistema UNIX. A ideia era que os usuários aperfeiçoassem o sistema e compartilhassem os resultados obtidos. A iniciativa resultou em um sistema operacional robusto, de código aberto (*open source*) e, conseqüentemente, colaborativo (ROSA, 2004).

O que permitiu a Internet alcançar um nível global foi o desenvolvimento da WWW, especialmente graças a Tim Berners-Lee ao realizar o feito que já havia sido proposto por Bush em 1945. Bush “[...] sugeriu um equipamento chamado Memex, cujo princípio de funcionamento era semelhante ao de um computador conectado à Internet” (FURGERI, 2006, p. 34). Uma das funções do equipamento era o “[...] de contribuir para a criação e distribuição do conhecimento adquirido em pesquisas, facilitando o compartilhamento e busca de informações” (Ibid., p. 34).

⁸ Expressão que surgiu como uma adaptação a *Uniplexed Information and Computing System* (UNICS).

[Berners-Lee] [...] definiu e implementou o software que permitia obter e acrescentar informação de e para qualquer computador conectado através da Internet: HTTP⁹, HTML¹⁰ e URI¹¹ (mais tarde chamado URL¹²). Em colaboração com Robert Cailliau, Berners-Lee construiu um programa navegador/editor em dezembro de 1990, e chamou esse sistema de hipertexto de *world wide web*, a rede mundial. (CASTELLS, 2003, p. 19)

A Guerra Fria forneceu o contexto de apoio governamental e popular para investir em pesquisas de inovação científica e tecnológica com fins dos Estados Unidos da América (EUA) estarem à frente da União Soviética, o que fomentou a Internet ganhar força para desenvolver-se. Se por um lado, o governo propiciou capital e liberdade, os centros universitários serviram de difusores da iniciativa, uma vez que a cultura da liberdade, por meio de potencial tecnológico, é uma forma de libertação e compartilhamento de ideias (CASTELLS, 2003). Neste caso, “o que sobressai destes relatos é que a Internet se desenvolveu um ambiente seguro, propiciado por recursos públicos e pesquisa orientada para a missão [...]” (Ibid, 2003, p. 26).

A evolução da Internet resultou nas diversas fases da *web*, que permite, em sua primeira geração, a inserção de dados em páginas estáticas apenas pelos detentores dos *websites*, com a possibilidade de recuperação e leitura. A partir da leitura pelo usuário, os dados ganham significado e um contexto que permitem gerar informação. Na sequência, a segunda fase *web* possibilita a transferência e o compartilhamento de informações entre usuários em ambiente digital. Em seguida, a terceira fase, a chamada *web 3.0*, possui a capacidade de geração de conhecimento, acarretando subsídios para a tomada de decisão.

2.2.1 Web 1.0

A geração da WWW iniciou-se com a *web 1.0*, de forma mais restrita, ou seja, o usuário desempenha o papel de consumidor da informação. Equivalente aos primórdios da WWW, a *web 1.0* é caracterizada pela presença de páginas estáticas, com capacidade limitada para qualquer interação entre os usuários e os conteúdos em si, possibilitando apenas a recuperação e a leitura de informações (AGHAEI;

⁹ *HyperText Transfer Protocol.*

¹⁰ *Mahanagar Telephone Mauritius Limited.*

¹¹ *Uniform Resource Identifier.*

¹² *Uniform Resource Locator.*

NEMATBAKHSH; FARSANI, 2012). Chamada de “*read only web*”¹³, a primeira versão começou como “[...] um lugar de informação para que empresas difundissem as informações delas para a população” (Ibid, 2012, p. 1, tradução nossa).

A Internet funcionava como um catálogo, em que as empresas apresentavam seus serviços e produtos. As pessoas, através de pesquisas, podiam recuperar resultados e contatar essas empresas (AGHAEI; NEMATBAHSH; FARSANI, 2012). Como um modelo unidirecional de páginas HTML¹⁴ e que não eram atualizadas com periodicidade definida, a meta dessa *web* era estabelecer presença/publicidade *on-line*. Desta forma, a *web 1.0* apresenta-se, basicamente, por duas limitações: as “[...] páginas só podem ser entendidas por humanos [...]” (CHOUDHURY, 2014, p. 8096, tradução nossa) e uma “ausência de representação dinâmica [...]” (Ibid, p. 8097, tradução nossa), como exemplificadas na figura dois (2), ao ilustrar um *website* intencionado em apresentar as suas informações sobre hospedagem, basicamente na forma textual, além de não ofertar qualquer modo de interação com o usuário.

Figura 2: Página da Pousada Pirineus em 2002.



0xx24 2225 1729 **pirineus@pousadapirineus.com.br**

O HOTEL | LADO DE FORA 1 | LADO DE FORA 2 | LADO DE DENTRO
APARTAMENTOS | CAFÉ DA MANHÃ | PISCINA | CAVE | MAPA | PREÇOS

Todos são do mesmo tamanho, aconchegantes, equipados com colchão de molas, TV colorida c/ controle remoto, frigobar, lareira, ventilador de teto e com vista para as montanhas.

O acesso é feito por escadas de pedra.

Os 1, 2 e 3 possuem cama tamanho padrão, e se diferenciam pela cor dos móveis patinados, que são o azul, verde e branco respectivamente.

Os apartamentos 4 e 5 possuem cama "queen size".

Como novidade, os 1, 2, 4 e 5 ganharam novas janelas que permitem a entrada de mais luz natural, permanecendo o apartamento 3 com seu estilo próprio que cativou um público fiel.

Para o seu conforto os apartamentos não possuem telefone. O nosso vai estar a sua disposição para ligações rápidas. Os celulares ATL funcionam bem na área externa. Ligações internacionais serão pagas a parte.

Fonte: Wayback Machine, 2014.

¹³ Web apenas para leitura.

¹⁴ HyperText Markup Language.

Ainda permeada pela falta de dinamismo, a web 1.0 também se caracteriza pelos diretórios, como o metabuscador chamado “Cadê?”, criado na década de 1990. O *website* permitia realizar pesquisas em páginas de hipertexto cadastradas no serviço, o que na época caracterizava um fator limitador nos resultados (GIORDAN, 1998). Outro atributo do “Cadê?” constituía a organização das páginas *web* em seções de temas genéricos, divididas em subseções, e ofertadas por “[...] endereços Web contendo informações relacionadas ao tema” (Ibid, p. 15). Conforme a figura três (3) é possível visualizar dezesseis seções de temas genéricos e, entre elas, a categoria “Ciência e Tecnologia”, a qual possui as subseções “Institutos” e “Publicações”.

Figura 3: Diretório de pesquisa “Cadê?”.



Fonte: Techtudo, 2015.

Ainda que caracterizada como a fase inicial da WWW, a *web* 1.0 ainda se faz presente, pois é o seio de muitos *websites*, especialmente pelo fato do HTML ser a linguagem de marcação básica, complementada de recursos das fases 2.0 e 3.0. Neste sentido, existem ambientes que assumem características da *web* 1.0 e que

evoluem conforme as fases da WWW, como, por exemplo, os *websites* institucionais, entre empresas, órgãos públicos, Instituições de ensino etc. A figura quatro (4) ilustra como a UnB, em 1997, dispunha informações da Instituição em forma de diretórios e *links*, aos quais direcionavam para páginas internas, basicamente textuais. Assim, é possível observar que o *layout* não era o foco, mas estar presente na WWW e disponibilizar conteúdos.

Figura 4: Website institucional da UnB em 1997.



Fonte: Wayback Machine, 2014.

Diferentemente da *web 1.0*, a necessidade de participação do usuário marca o desenvolvimento da próxima fase, em que se destacam a ascensão das redes sociais e, conseqüentemente, do compartilhamento de informação em formatos diversos entre os usuários que fazem parte dessas redes.

2.2.2 Web 2.0

A *web 2.0* é marcada pelo acréscimo de funcionalidades da WWW em relação à fase anterior, o que resultou na ampliação e rapidez na forma de comunicação entre os seres humanos, assim como no compartilhamento de informações entre os mesmos. A segunda geração foi definida por Dale Dougherty, em 2004, como “*read-*

*write web*¹⁵, tendo como principal atrativo é a possibilidade de interação, ou seja, os usuários tornam-se colaboradores, pois são capazes de relacionarem-se com o ambiente digital e por meio dele (AGHAENI, NEMATBAKHS; FARSANI, 2012). O aumento dessa colaboração faz surgir novas aplicações na *web*, como ocorre nas redes sociais, por exemplo, a partir do gerenciamento de uma “[...] multidão de quantitativo mundial com interesses em comum para interagir socialmente” (CHOUDHURY, 2014, p. 8097, tradução nossa). Por outro lado, significa que há “[...] mais interação com menos controle [...]” (AGHAENI, NEMATBAKHS; FARSANI, 2012, p. 3, tradução nossa).

A possibilidade de criação e atualização de conteúdos na *web 2.0*, também chamada de *web social*, aumenta os casos de “[...] vandalismo, confusões e erros de informação ou de uso das ferramentas (como apagamento incidental [*sic*] de dados) [...]” (PRIMO, 2007, p. 6), pois se trata de um meio flexível e dinâmico (ALVIM, 2007). Nesta fase da WWW, já não é necessário ser autoridade de um assunto para se escrever sobre tal. Amadores tornam-se autores e, assim, os conteúdos tendem a perder a qualidade, ao tratar de assuntos de forma superficial, buscarem ser velozes e desafiam a linha limite que separa a opinião pessoal do conhecimento humano (PARDO KUKLINSKI, 2007). Apesar dos supracitados desafios, a *web 2.0* assume a seguinte essência:

A filosofia da Web 2.0 prima pela facilidade na publicação e rapidez no armazenamento de textos e ficheiros, ou seja, tem como principal objectivo tornar a web um ambiente social e acessível a todos os utilizadores, um espaço onde cada um selecciona e controla a informação de acordo com as suas necessidades e interesses. (COUTINHO; BOTTENTUIT JUNIOR, 2007, p. 200)

Ademais, o carácter participativo da segunda versão da *web* inaugura-se em uma fase interativa de uso e de desenvolvimento de serviços, em que a Internet “[...] implica em *web design* flexível, reuso criativo, atualizações, conteúdos colaborativos quanto à criação e à modificação [...]” (CHOUDHURY, 2014, p. 8097, tradução nossa). Além disso, a *web* se transforma em uma plataforma que provê novas formas de comunicação e interação entre usuários, ofertando aplicações que facilitam a

¹⁷ *Web* para leitura e escrita.

transferência e a produção em colaboração e elevam os negócios a outro patamar, ou seja, misturam *software* e gestão empresarial (CHOUDHURY, 2014).

Por mais que a *web* 2.0 tenha trazido uma revolução na forma de produzir informação e de se comunicar, ela também trouxe limitações relacionadas a ética, em outras palavras, a confiabilidade e a validade das informações, pois ocorrem constantes mudanças e atualizações nos conteúdos e nos serviços da *web* (CHOUDHURY, 2014).

A expansão e popularidade da *web*, ao combinar velocidade, alto volume de informação e participação ativa dos usuários resultou em euforia. Contudo, o resultado foi o surgimento de novas preocupações relacionadas à veracidade das informações disponibilizadas. Pestana (2001) elenca que o aumento de páginas pessoais, informações desatualizadas, dificuldades de acesso e falta de exatidão são apenas alguns dos problemas a serem enfrentados, e que justificam a necessidade de mecanismos de avaliação de fontes.

Critérios de avaliação, de fontes na Internet, servem como uma forma de identificar inconstâncias na qualidade das informações, ao partir-se do pressuposto de que, normalmente, “[...] não existem avaliações prévias do que é disponibilizado. O acúmulo de informações sem relevância aponta para a necessidade de filtros que permitam a recuperação de informações de qualidade e com maior revocação” (TOMÁEL *et al.*, 2001, p. 3). E somado ao excesso de informações irrelevantes, o tom de informalidade, empregado nesse ambiente digital, aumenta o número de fontes incertas, ainda que facilite no aspecto da comunicação (Ibid).

O destaque na participação ativa do usuário na criação de conteúdos, permitida pela *web* 2.0, está presente em ambientes como *blogs* e fóruns. No caso da comunicação entre usuários, ela é possibilitada pelas redes sociais, por exemplo, por meio de aplicações como *Facebook*¹⁶ e *Twitter*¹⁷. Outros ambientes, como o *Dropbox*¹⁸ e o *Google Drive*¹⁹, permitem o armazenamento em nuvem e compartilhamento de arquivos, enquanto que o *Netflix*²⁰ e o *Youtube*²¹ atuam no ramo do fluxo de mídia (DA SILVA; DIAS; CASTRO, 2016).

¹⁶ Site: <https://www.facebook.com/>.

¹⁷ Site: <https://twitter.com/>.

¹⁸ Site: <https://www.dropbox.com/>.

¹⁹ Site: https://www.google.com/intl/pt-BR_ALL/drive/.

²⁰ Site: <https://www.netflix.com/browse>.

²¹ Site: <https://www.youtube.com/?gl=BR>.

A participação na criação, atualização e compartilhamento de informações é reforçada, através de uma Internet que funciona como plataforma, enquanto uma característica da *web 2.0*. Neste sentido, o usuário não se restringe ao *desktop*, o que “[...] acaba com a dependência dos média físicos de armazenamento de dados, pois, através das ferramentas disponibilizadas, o utilizador pode manter tudo online de forma pública ou privada [...]” (BOTTENTUIT JUNIOR; COUTINHO, 2008, p. 5). O *download* deixa de ser uma ação obrigatória, ainda que o usuário tenha a opção de “baixar” o conteúdo/arquivo na própria máquina ou dispositivo móvel, uma vez que o acesso aos documentos ocorre via Internet.

A *web 2.0* introduziu novas maneiras de publicação e circulação de informações ao permitir que os usuários se tornassem criadores. Espaços como os *blogs* permitiram o fortalecimento “[...] das bordas da rede” (PRIMO, 2007, p. 4), ou seja, a partir da reunião de pequenos nichos, com interesses segmentados, ganha-se “[...] peso na rede a partir de sua interconexão com outros sub-sistemas [*sic*]” (Ibid, p. 4). Assim, os *blogs* ultrapassam o *status* de ferramenta de publicações individuais para tornarem-se instrumentos de disseminação da informação, como exemplificado na figura cinco (5), um *blog* voltado à tecnologia e negócios, apresentando dois *posts* sobre *big data*.

Figura 5: Exemplo de *blog*.

The image shows a screenshot of a blog interface. At the top left, it says "Posts da categoria _". To the right is a search bar with the text "Pesquisar ...". Below the search bar is a sidebar titled "Categorias" with a list of categories: Atualidades, Big Data, E-Commerce, Internet, Negócios, and Tecnologia. The main content area displays two blog posts. The first post has a header image of a laptop on a desk and a red "BIG DATA" tag. The title is "Mitos do Big Data Parte 4: as Legislações de Privacidade Vão Matar o Big Data" and it is dated "Por BigData Corp em 15/04/2019". The second post has a header image of hands holding a glowing sphere and a red "BIG DATA" tag. The title is "Dados Ignorados, Oportunidades Perdidas e Desperdício: O Caso de Flint, nos EUA" and it is dated "Por BigData Corp em 08/04/2019".

Fonte: Bigdatacorp, 2019.

Prosseguindo na temática de criação de conteúdo e participação ativa dos usuários, os fóruns surgem como ambientes de potencialização e de motivação de troca de dúvidas e experiências entre usuários (LISBÔA; COUTINHO, 2011). Por exemplo, a figura seis (6) ilustra um fórum de tecnologia e que possui vários assuntos, como telefonia e *games*. A partir das perguntas dispostas pelos usuários, é informada a quantidade de respostas e *tags* relacionadas as questões.

Figura 6: Exemplo de fórum.

The screenshot shows a forum interface with the following elements:

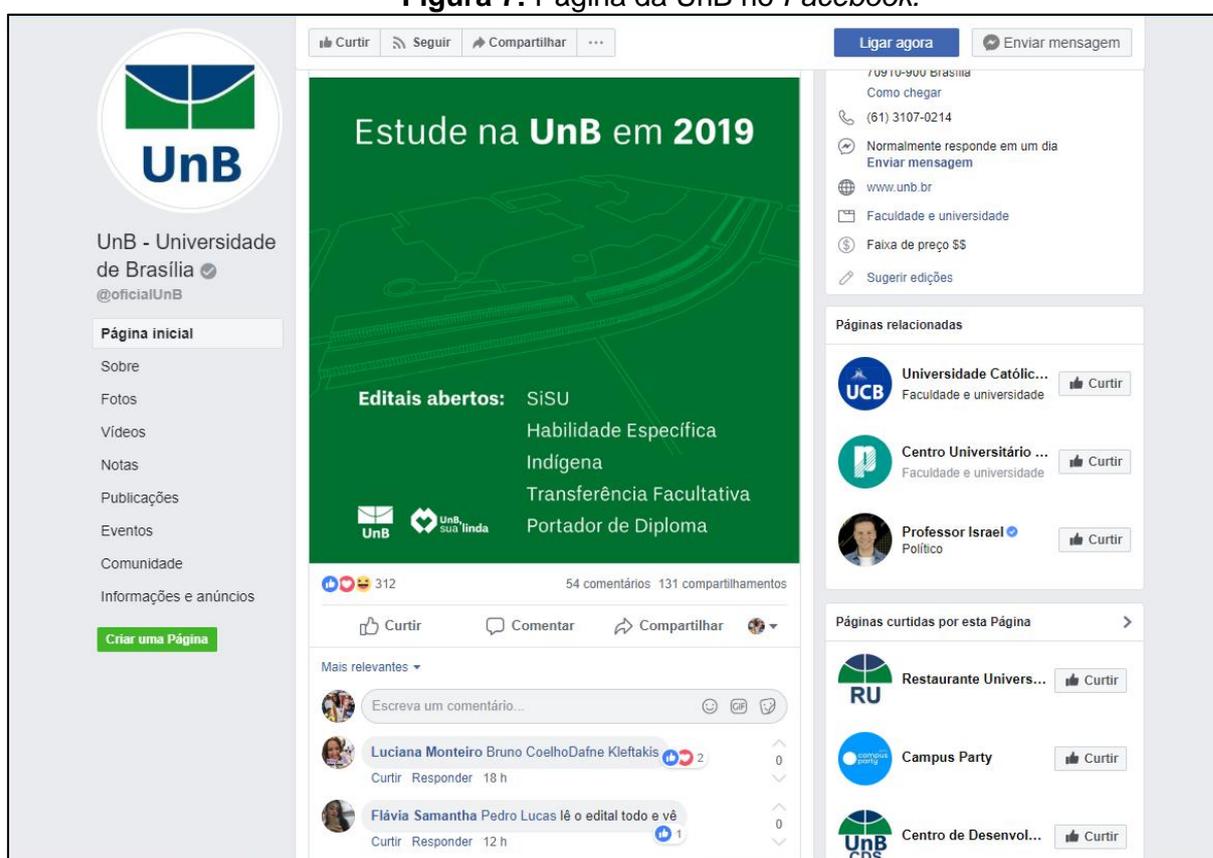
- Header:** "TODAS AS PERGUNTAS" with tabs for "Ativo", "Novas", and "Mais Votadas".
- Question 1:** "Como descobrir o modelo do celular?". Statistics: 0 votes, 1 response, 0 visits. Posted 1 hour ago by PaoloV57 (7.1k). Tags: smartphone, modelo, motorola. Social sharing icons for Facebook, Twitter, and Google+ are present.
- Question 2:** "Alguém conhece DHgate?". Statistics: 0 votes, 6 responses, 21.0k visits. Posted 1 hour ago by nande007 (1). Tags: celular, site, dhgate. Social sharing icons for Facebook, Twitter, and Google+ are present.
- Question 3:** "Esse computador roda jogos de boa?". Statistics: 0 votes, 1 response, 0 visits. Posted 1 hour ago by marcmira (29.9k). Tags: pcgamers, placadevideo, computador. Social sharing icons for Facebook, Twitter, and Google+ are present.
- Summary Box:** Located on the right, it displays "60943 PERGUNTAS" and "157199 RESPOSTAS".
- Tags Section:** Titled "TAGS RELACIONADAS", it lists: pc (4,988), smartphone (4,843), celular (4,796), and notebook (3,719).

Fonte: Techtudo, 2019.

Outro ponto a se destacar é a possibilidade de interação entre pessoas pela Internet a partir de ambientes que viabilizam uma comunicação dinâmica, especificamente pelas chamadas redes sociais, as quais “[...] funcionam com o primado fundamental da *interação social* [...]” (RECUERO, 2005, p. 11, destaque do autor), como o *Facebook* e o *Twitter*. O primeiro tem o foco voltado à disponibilização de informações pessoais enquanto “[...] um ambiente com recursos para a interação

entre os usuários, sendo basicamente os *feeds*, aplicativos, *fanpages*²², marcação de eventos, organização de álbuns e criação de grupos virtuais” (DA SILVA; RUFINO, 2016, p. 23, destaque dos autores). Na figura sete (7), como exemplo, apresenta-se a página da UnB na rede social, intencionada em disponibilizar informações gerais sobre a instituição, com a possibilidade de interação entre os participantes por meio de “curtidas”, compartilhamentos e comentários, além de funcionar como uma espécie de agregador e centralizador de informações para estudantes, professores, pesquisadores e funcionários da Instituição, bem como para as comunidades externas que estão interessadas na Universidade.

Figura 7: Página da UnB no Facebook.



Fonte: Facebook, 2019.

Diferentemente do *Facebook*, em que a experiência é focada no uso pessoal, o *Twitter* tem a experiência voltada ao conteúdo (HUGHES *et al*, 2012). Em textos limitados a 280 caracteres, o ambiente tem foco em disseminar mensagens rápidas e

²² Compreende-se *fanpage* como “[...] um tipo de página direcionada a empresas e instituições devido aos seus recursos oferecidos como dados estatísticos e a não limitação no número de amigos” (DA SILVA; RUFINO, 2016, p. 23).

objetivas, mesclando com imagens, *gifs*²³ e vídeos, como demonstra o *tweet* da página do perfil da UnB na figura oito (8): “Amanhã começa a XXII Jornada Odontológica da UnB! Participe”.

Figura 8: Página da UnB no *Twitter*.



Fonte: *Twitter UnB*, 2019.

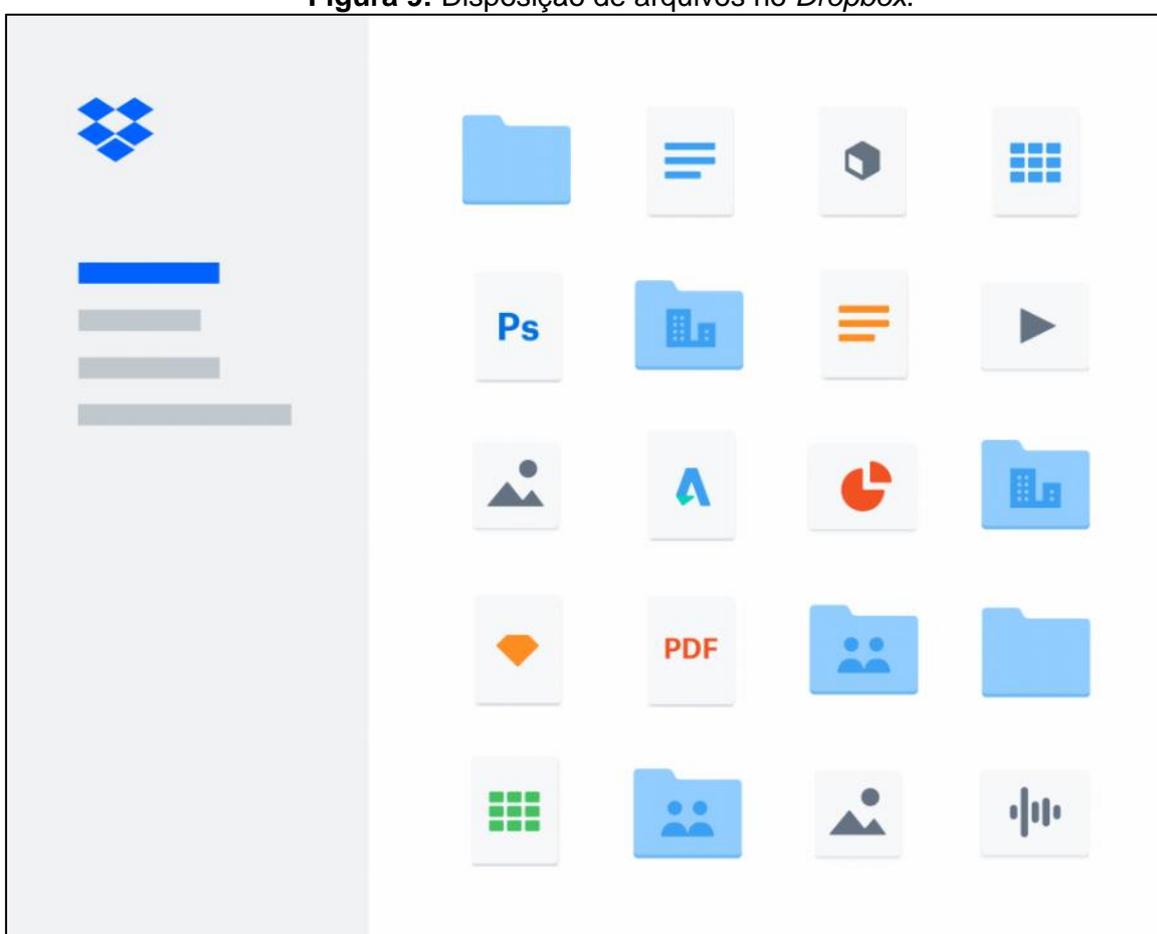
Quanto aos ambientes de armazenamento, existem os que possibilitam guardar, em nuvem (*cloud*), arquivos de formatos variados, recuperá-los nos mais diversos dispositivos, a partir da Internet, entre computadores e *mobiles*, editá-los e compartilhá-los com outros usuários. Por meio da computação em nuvem (*cloud computing*), compreendido como um serviço que revolucionou o armazenamento de arquivos, ao “[...] permitir, além da guarda de arquivos, a possibilidade de realizar tarefas simples na rede como assistir vídeos, postar fotos ou mesmo enviar mensagens em redes sociais” (LAGO, 2014, p. 57), o usuário possui maior liberdade para interações na *web*, sem comprometer o espaço interno de armazenamento do

²³ *Graphics Interchange Format*.

seu equipamento, pois não há obrigatoriedade em realizar *download* de aplicativos e arquivos, por exemplo (Ibid). Nesse contexto, tem-se como exemplo o *Dropbox*.

O *Dropbox* é um serviço de armazenamento em nuvem de “[...] arquivos salvos pelos usuários [...] normalmente não-editáveis, como imagens JPEGs²⁴, documentos PDFs²⁵ e programas executáveis” (DRAGO; VIEIRA; SILVA, [2013], p. 110), como ilustra a figura nove (9), ao apresentar uma variedade de documentos e elementos suportados pelo ambiente. O *Dropbox* caracteriza-se também por permitir sincronização de “[...] vários dispositivos através de uma mesma conta” (Ibid, p. 111), bem como sincronização seletiva de arquivos.

Figura 9: Disposição de arquivos no *Dropbox*.



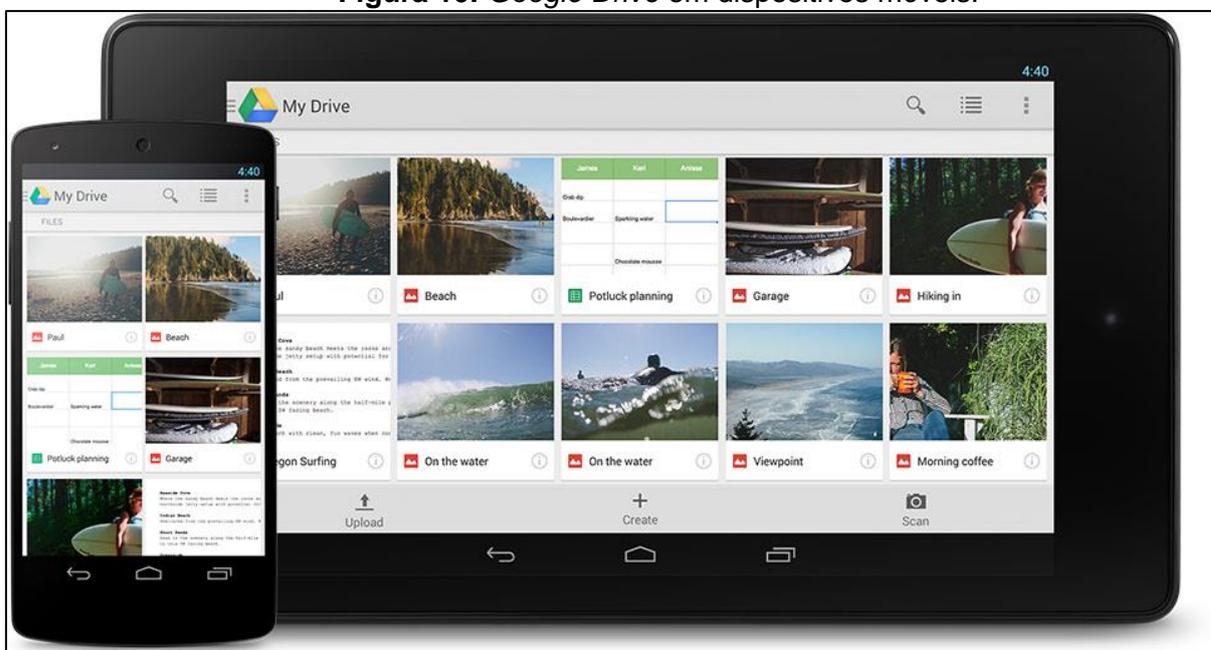
Fonte: *Dropbox*, 2019.

²⁴ Joint Photographics Experts Group.

²⁵ Portable Document Format.

Na mesma linha do *Dropbox*, o *Google Drive* atua como armazenador de arquivos em nuvem, entretanto diferencia-se ao potencializar a geração de conteúdo, por meio da disposição de ferramentas que possibilitam a criação, edição e o compartilhamento de documentos na própria *web*, permitindo ainda a elaboração conjunta dos arquivos no formato de texto pelo *Google Docs*²⁶, de planilhas pelo *Google Sheets*²⁷, de apresentações *powerpoint* pelo *Google Slides*²⁸ e de formulários pelo *Google Forms*²⁹. Referendando, a figura dez (10) ilustra o *Google Drive*, acessado, por exemplo tanto por *smartphone*, como por *tablet*. Assim como o *Dropbox*, o *Google Drive* sincroniza uma mesma conta de usuário em diversos dispositivos, permitindo acessar e realizando criações e edições de documentos sem a necessidade do *download*, uma vez que o aplicativo funciona no prisma da Internet plataforma. Mesmo assim, o usuário pode “baixar” a aplicação para o dispositivo móvel, como para o *desktop*, neste caso, variando em funcionalidades/recursos. Quanto aos tipos de arquivos, o *Google Drive* guarda “[...] fotos, histórias, projetos, desenhos, gravações, vídeos e muito mais” (GOOGLE DRIVE, [201-?]).

Figura 10: *Google Drive* em dispositivos móveis.



Fonte: *Google Drive*, [201-?].

²⁶ Site: https://www.google.com/intl/pt_br/docs/about/.

²⁷ Site: https://www.google.com/intl/pt_br/sheets/about/.

²⁸ Site: https://www.google.com/intl/pt_br/slides/about/.

²⁹ Site: https://www.google.com/intl/pt_br/forms/about/.

Outra característica da segunda geração da *web* refere-se à popularização de ambientes de fluxo de mídia, denominados de *streaming*, o qual corresponde à “[...] disponibilização de sons e/ou vídeos diretamente pela internet [...]” (SZELBRACIKOWSKI, 2017, p. 1), ausentes da necessidade de realizar *download* e serem acessíveis por dispositivos variados, como o *Netflix* (figura 11).

Figura 11: *Netflix* disponível em dispositivos diversos.

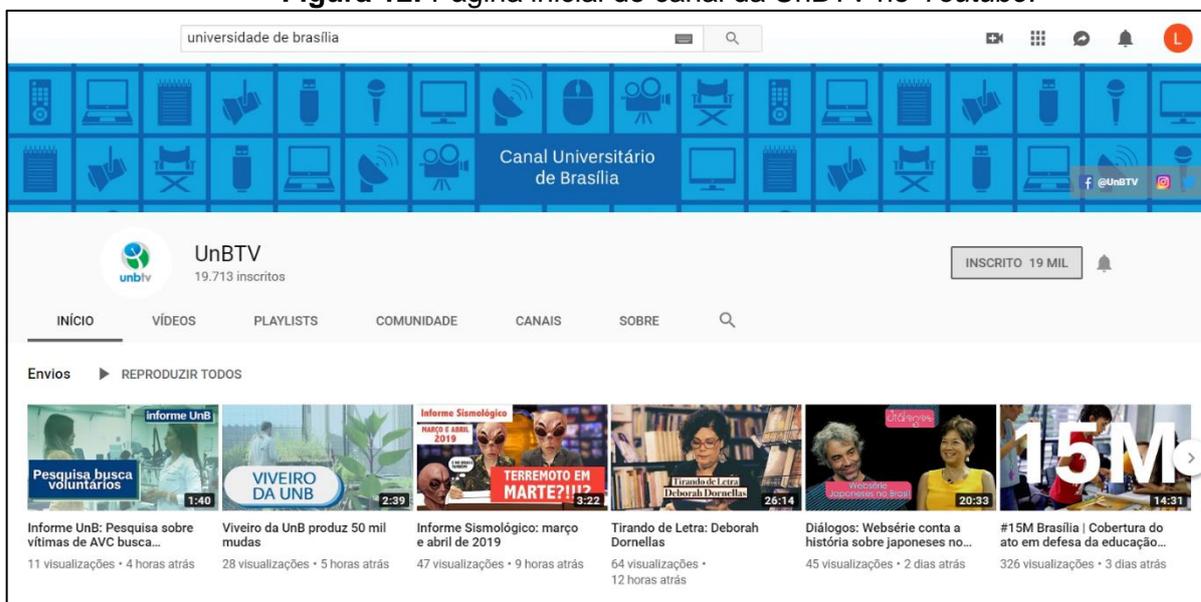


Fonte: *Techradar*, 2019.

Os ambientes de fluxo de mídia facilitam a visualização e compartilhamento de mídias. No caso do *Youtube*, defendendo “[...] que todos devam ter acesso livre e fácil às informações e que o vídeo tem grande influência na educação, na construção do entendimento e na transmissão de informações sobre acontecimentos no mundo [...]” (YOUTUBE, 2019), aceita formatos do tipo filmes, séries, videoclipes, documentários e *vlogs*. Além de possibilitar *uploads* de conteúdos dos usuários, o *Youtube* promove a interação entre o usuário e o proprietário do canal, como entre os próprios usuários, por meio de comentários. Para tanto, faz-se necessário realizar uma breve inscrição no canal desejado. Outro ponto que merece destaque, contextualizado anteriormente, é o fato, assim como promulga a *web 2.0* na oferta de ambientes plataforma, é a possibilidade de assisti-lo diretamente pela Internet, sem ocupar espaço de armazenamento nos dispositivos. Contudo, o usuário pode “baixar” o aplicativo correspondente na própria máquina/*mobile*.

Como exemplo, na figura doze (12) é possível a disponibilização de vídeos referentes à UnB a partir do espaço UnBTV, que tem como nome “canal universitário de Brasília”.

Figura 12: Página inicial do canal da UnBTV no Youtube.



Fonte: UnBTV, 2019.

Por fim, o grande fluxo de informação propiciado pela *web 2.0*, causado pelo alto volume de criação e transformação de informação, coloca em uso o discernimento humano, que nem sempre obtém êxito para filtrar o conteúdo exposto no ambiente digital para seu consumo, bem como realizar as escolhas mais adequadas diante da variedade de informações/opções. Neste sentido, em um cenário com múltiplas possibilidades e opções de escolha, a *web 3.0* mostra-se como uma tentativa de, por meio da tecnologia semântica, direcionar o usuário dentre à ampla gama de resultados, considerando os perfis dos usuários, especificamente sobre os seus interesses, gostos, preferências.

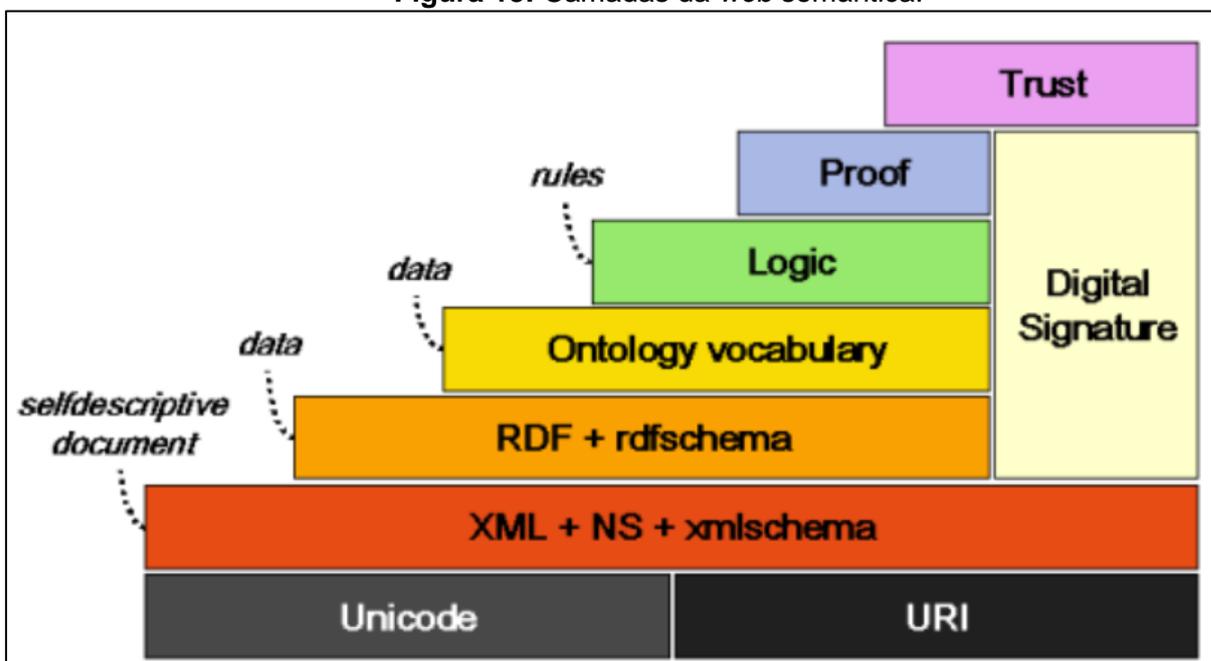
2.2.3 Web 3.0

A terceira geração da *web* caracteriza-se como *read-write-execute* (*web* para leitura, escrita e execuções), fundamentada na intenção de “[...] definir dados estruturados e conectá-los de forma a possibilitar descobertas mais efetivas, automação, integração, e reuso por meio de aplicações diversas” (CHOUDHURY, 2014, p. 8097, tradução nossa). A ideia básica é, através de conjuntos de dados diversos, analisá-los, integrá-los e “linká-los” para obter novas informações (AGHAEI; NEMATBAKHSH; FARSANI, 2012).

Também conhecida como *web semântica*, a geração 3.0 tem o objetivo de tornar a *web* legível por máquinas e não apenas para humanos (AGHAEI; NEMATBAKHSI; FARSANI, 2012). A “educação” de máquinas consiste em programá-las para pensarem e atuarem como uma mente humana, como uma representação do cérebro. A *web semântica* permite que máquinas “[...] entendam e respondam demandas humanas complexas baseadas em seus significados [...]” (CHOUDHURY, 2014, p. 8098, tradução nossa), isto é, a ação de entender requer contexto e envolve um processo cognitivo de atribuir significado a algo.

A terceira geração da *web* significa a adoção de um conjunto de tecnologias disponíveis e complementares, o que, de certa forma, oferece maior transparência para o usuário. Em outras palavras, por meio da interconexão entre documentos e dados relacionados, da ampliação do alcance das informações publicadas *online*, do processamento automático das informações e, conseqüentemente, da maior integração das informações, a *web semântica* contribui para ampliação da transparência (GARCÍA-MARCO, 2010). Para tanto, a supracitada transferência é possível devido às camadas de tecnologias e padrões *web* que formam a *web semântica*, ilustradas na figura treze (13).

Figura 13: Camadas da *web semântica*.



Fonte: World Wide Web Consortium (W3C), 2001.

As camadas Unicode e URI estabelecem conjuntos de caracteres internacionais e provêm meios de identificar os objetos. O Unicode consiste em um padrão universal “[...] que define, em um só lugar, todos os caracteres necessários para escrever na maioria dos idiomas em uso em computadores” (WORLD WIDE WEB CONSORTIUM, 2018, p. 1, tradução nossa). Quanto a URI, permite a “[...] identificação única de recursos, bem como os relacionamentos entre eles” (LIMA; CARVALHO, 2004, p. 5). Em seguida, o *Extensible Markup Language* (XML), juntamente com os *namespace* (NS)³⁰ e *schema*³¹, garante a integração das “[...] definições da *web* semântica com outros padrões baseados em XML” (KOIVUNEN; MILLER, 2002, p. 34, tradução nossa). Na camada *Resource Description Framework* (RDF) e esquema RDF ocorre o estabelecimento das declarações sobre os objetos por meio de vocabulários específicos e utilizando URI. É justamente nessa camada que se pode “[...] atribuir tipos para recursos e *links*” (Ibid, p. 34, tradução nossa), utilizando-se do RDF para fazer afirmações/inferências [...] sobre coisas particulares (pessoas, *Web sites*) [...] (LIMA; CARVALHO, 2004, p. 6) e estabelecer propriedades e valores. Desta forma, mapeia-se as informações de um objeto específico e, por meio do URI, estabelece-se relacionamentos e *links* entre objetos (Ibid). A camada *Ontology Vocabulary* (vocabulário de ontologia) dedica-se à evolução de vocabulários, bem como estabelecer relações entre conceitos. Sobre a camada *Logic* (lógica), “[...] possibilita a escrita de regras [...]” (LIMA; CARVALHO, 2004, p. 4). No caso das camadas *Proof* (prova) e *Trust* (validação), são responsáveis por executar e avaliar as regras definidas, estabelecendo-se em quais aplicações confiar segundo as provas realizadas. Por fim, a *Digital Signature* (assinatura digital) é capaz de reconhecer alterações em documentos (KOIVUNEN; MILLER, 2002) com vistas à “[...] garantir a procedência de um documento ou definição, o que é vital para decidir se uma informação é confiável ou não” (RAMALHO; VIDOTTI; FUJITA, 2005, p. 6).

A *web* 3.0 suprime, tanto os tradicionais conceitos de *website*, como a dinâmica de posse de dados para compartilhamento, por meio de serviços especializados (CHOUDHURY, 2014). Os dados, nessa fase da *web*, são compartilhados e os

³⁰ XML *namespace* é um método adotado “[...] para evitar conflitos relacionados aos nomes dos elementos” (W3SCHOOLS, [201-?]b, p. 1, tradução nossa), considerando que elemento refere-se ao que esteja entre *tags* (<*element*></*element*>), como “[...] texto, atributos, outros elementos ou um mix” (Ibid, [201-?]a, p. 1, tradução nossa).

³¹ A W3C define que o padrão XML *schema*, também referido como XML *schema definition* (XSD), é responsável por “[...] descrever a estrutura de um documento XML” (W3SCHOOLS, [201-?]c, p. 1, tradução nossa).

serviços possuem perspectivas variadas, ou seja, diversas aplicações em dispositivos e múltiplos formatos. O essencial é que os serviços tenham “[...] foco no contexto e na personalização, e ambos serão alcançados pela pesquisa vertical³²” (TORRES, 2007, p. 1, tradução nossa). Por exemplo, enquanto que no *website* (tradicional) de hipertexto, o ambiente é construído por documentos, “[...] onde os *links* são as âncoras de relacionamento em documentos de hipertexto escritos em HTML [...]” (BERNERS-LEE, 2009, p. 1), na *web* de dados, ainda que seja construída por documentos, assim como os *websites* de hipertexto, os *links* são feitos “[...] entre objetos arbitrários descritos por RDF” (Ibid, p. 1, tradução nossa), utilizando URI que são capazes de identificar “[...] qualquer tipo de objeto ou conceito” (Ibid, p. 1, tradução nossa), ou seja, torna os documentos recuperáveis com maior facilidade.

A terceira geração, também chamada de *web* de dados, intenciona-se em integrar e conectar formatos diversos para permitir os “[...] usuários de achar, compartilhar e combinar informação mais facilmente” (CHOUDHURY, 2014, p. 8098, tradução nossa). Para isso, “[...] requer que as fontes de informação relevante estejam semanticamente estruturadas” (Ibid, p. 8098, tradução nossa) e um dos caminhos tem sido o esquema RDF, o qual tem a função de definir o relacionamento entre objetos.

O RDF “[...] permite a tecnologia de expressar o significado de termos e de conceitos em formato que os computadores podem processar” (Ibid, p. 8099, tradução nossa). Em suma, a partir do RDF é possível organizar a *web* por *links* entre elementos digitais, configurando o modelo de dados ligados (*linked data*), regido por quatro princípios básicos essenciais para conectar a *web* semântica. Conforme Berners-Lee (2009, p. 1), são eles:

1. Use URIs como nome para objetos;
2. Use HTTP URIs para que as pessoas possam procurar por esses nomes;
3. Quando alguém visitar uma URI, provê informação útil, usando os padrões (RDF, SPARQL³³);
4. Inclua *links* para outros URIs, assim eles podem descobrir mais coisas. (BERNERS-LEE, 2009, p. 1)

³² Pesquisa vertical tem o foco na qualidade dos resultados do processo de busca por ser mais abrangente, “[...] envolvendo imagens, vídeos, notícias [...]” (POZZEBOM, 2011, p. 1). A ideia é que a pesquisa não forneça apenas o resultado, “[...] mas sim, informações da própria página de resultados” (Ibid, p. 1).

³³ Do inglês *SPARQL Protocol and RDF Query Language*, trata-se de uma linguagem de consulta “[...] oficialmente introduzida em 2008 para recuperar um conjunto de dados RDF e fornecer *endpoints* para consultar fontes distribuídas” (IZQUIERDO, 2017, f. 11, tradução nossa).

Como consequência da possibilidade de relacionar elementos e reuni-los em conjuntos, novas áreas se desenvolveram e vem contribuindo com a terceira versão da *web*, como é o caso da Inteligência Artificial (IA), assim como novas ferramentas surgiram/surgem com o objetivo de facilitar o processo de “linkar” dados e gerar novas informações, como, por exemplo, as tecnologias *machine learning* e *big data*. Enquanto a *machine learning* significa “[...] permitir um sistema a aprender a partir do passado ou do presente e use esse conhecimento para realizar previsões ou decisões relativas a eventos futuros desconhecidos” (LANDSET *et al.*, 2015, p. 2), a tecnologia *big data* “[...] é um corpo de informação que é volumosa e variada. (MARTIN, 2018, p. 5, tradução nossa), além de ser gerado rapidamente. Neste sentido, a figura catorze (14) representa um exemplo da *web 3.0* a partir do serviço *online WolframAlpha*³⁴, que responde a perguntas dos usuários para recuperar informações acerca dos termos buscados.

Figura 14: Resultado de busca no *Wolfram|Alpha*.

The image shows a screenshot of the Wolfram|Alpha search results for a comparison between Brazil and Argentina. The results are organized into two main sections: 'Geographic properties' and 'Demographics'. Each section contains a table with three columns: the property name, the value for Brazil, and the value for Argentina. The 'Geographic properties' section includes total area, land area, and continent. The 'Demographics' section includes population, population density, population growth, life expectancy, and median age. The interface also features buttons for 'More', 'Show non-metric', 'Satellite image', 'Show rates', and 'Units'.

Geographic properties:		
	Brazil	Argentina
total area	8.515 × 10 ⁶ km ² (square kilometers) (world rank: 5 th)	2.767 × 10 ⁶ km ² (square kilometers) (world rank: 8 th)
land area	8.459 million km ² (square kilometers) (world rank: 5 th)	2.737 million km ² (square kilometers) (world rank: 8 th)
continent	South America	South America

Demographics:		
	Brazil	Argentina
population	209 million people (world rank: 5 th) (2017)	44.3 million people (world rank: 31 st) (2017)
population density	25 people/km ² (world rank: 192 nd) (2017)	16.2 people/km ² (world rank: 212 th) (2017)
population growth	0.784 %/yr (world rank: 150 th) (2017)	0.961 %/yr (world rank: 136 th) (2017)
life expectancy	75.5 years (world rank: 100 th) (2016)	76.6 years (world rank: 83 rd) (2016)
median age	31.3 years (world rank: 78 th) (2015)	30.8 years (world rank: 82 nd) (2015)

Fonte: *WolframAlpha*, 2018.

³⁴ Site: <https://www.wolframalpha.com/>.

No uso da expressão “Brazil x Argentina” (figura 14) é possível perceber que o sistema oferece dados comparativos entre os países, organizados por propriedades geográficas e demografias, ao invés de apresentar uma lista de resultados de páginas *web* variadas como ocorre tradicionalmente nos buscadores da Internet.

Outro exemplo é o serviço *Siri*³⁵, fornecido pela *Apple*³⁶ em seus dispositivos móveis, no qual, utilizando a voz, é possível requisitar funções a serem realizadas e também fazer perguntas para o robô (IA). Como ilustrado na figura quinze (15), o usuário, fez uma pergunta em voz relacionada ao tempo e como resultado obteve a previsão do tempo da sua localidade e na data que o questionamento foi realizado. Ou seja, ao dizer “Siri vai chover”, obteve-se como resposta a previsão do dia (segunda-feira), em Curitiba, constando-se que a localização e a data são levadas em consideração para responder aos usuários.

Figura 15: Serviço *Siri* sobre a previsão de chuva.



Fonte: Tecmundo, 2017.

³⁵ Site: <https://www.apple.com/br/siri/>.

³⁶ Site: <https://www.apple.com/br/>.

Enquanto a *web* semântica demonstra o potencial de recuperar e reunir dados úteis em um vasto conjunto de informações disponíveis, ferramentas como *big data* possuem um objetivo mais ambicioso, ou seja, gerar conhecimento a partir de uma extensa e variada gama de informações, disponibilizadas em alta velocidade, contribuindo diretamente na geração de sabedoria pelo usuário e, conseqüentemente, embasá-lo na tomada de decisões.

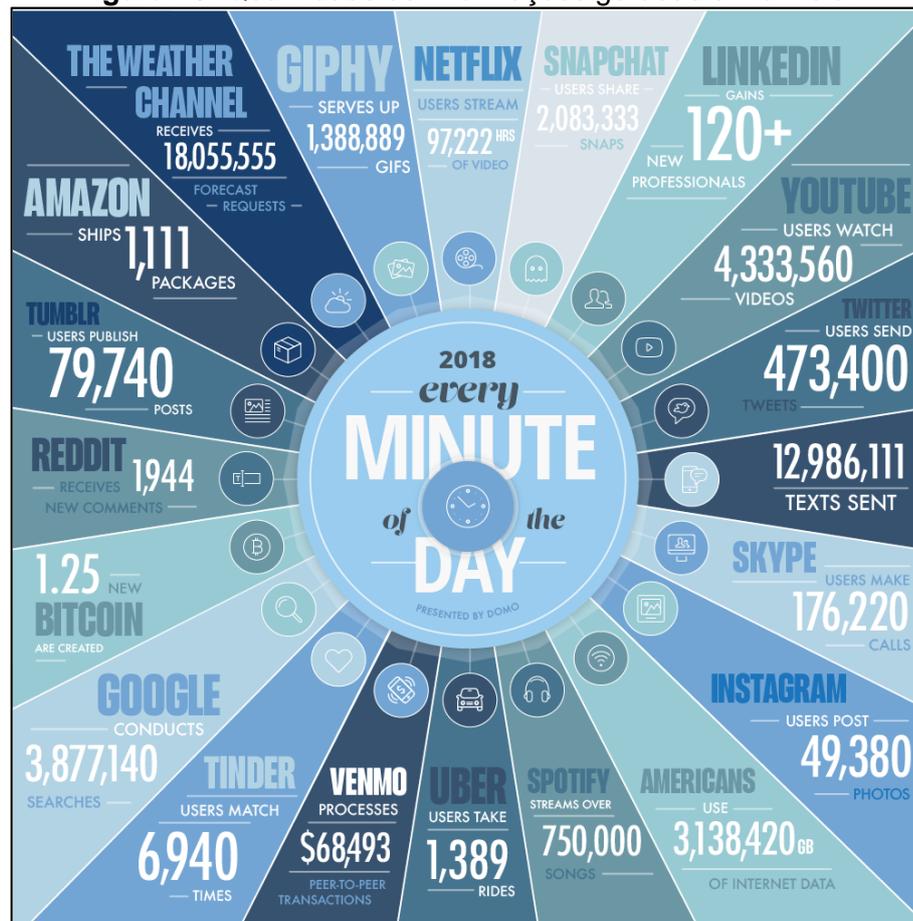
2.3 *BIG DATA*: conceitos e possibilidades de aplicação

A tecnologia *big data* configura-se como uma ferramenta imergida no prisma da *web* 3.0 e compreendida como um recurso “[...] aplicado à informação que não pode ser processada ou analisada por métodos usando processos ou ferramentas tradicionais” (ZIKOPOULOS *et al.*, 2012, p. 3, tradução nossa). Em outras palavras, a contemporaneidade, permeada por tecnologias envolvidas em processos de diferentes complexidades, da comunicação via aparelhos móveis até o uso de sensores em estradas, presencia uma demasiada geração de dados, de diferentes formatos, em grande quantidade e em pouco espaço de tempo, resultando em um cenário fomentador à aplicação da tecnologia *big data*.

Considerando as possibilidades de adoção, a *big data* caracteriza-se por três elementos básicos: “[...] volume, variedade e velocidade” (ZIKOPOULOS *et al.*, 2012, p. 5, tradução nossa), também conhecidos como “3V”. O **volume** refere-se à quantidade de dados gerada diariamente, de forma exorbitante (figura 16). Essa dimensão de produção de dados, exponencialmente quantitativa, vem se apresentando desde os anos 2000, prospectando vislumbres de armazenamento até 2020. Nesta perspectiva, os autores asseveram que:

No ano 2000, 800,000 *petabytes* (PB) de dados estavam armazenados no mundo. [...] Nós esperamos que esse número alcance 35 *zettabytes* (ZB) até 2020. Sozinho, o Twitter gera mais de 7 *terabytes* (TB) de dados todos os dias, Facebook 10 TB, e algumas empresas geram *terabytes* em dados a cada hora, todos os dias do ano. (ZIKOPOULOS *et al.*, 2012, p. 5-6)

Figura 16: Quantidade de informações geradas em um dia.



Fonte: Domo, 2018.

Além do quantitativo, a tecnologia *big data* exhibe-se como uma solução de análise de grandes volumes de dados considerando o elemento **variedade**, alude a todos os tipos de formatos, uma vez que, nos dias atuais, a explosão das tecnologias gera dados complexos e em multiplicidade (figura 17). Desta forma, a tecnologia *big data* atinge a seguinte cobertura de dados:

[...] não inclui apenas dados relacionais tradicionais, mas também dados brutos, semiestruturados, e não estruturados provindos das páginas *web*, arquivos de *login* da *web* [...], buscas indexadas, fóruns de mídia social, *e-mail*, documentos, [...] e assim por diante. (ZIKOPOULOS et al., 2012, p. 7)

Figura 17: Exemplos de variedade de formatos na *big data*.



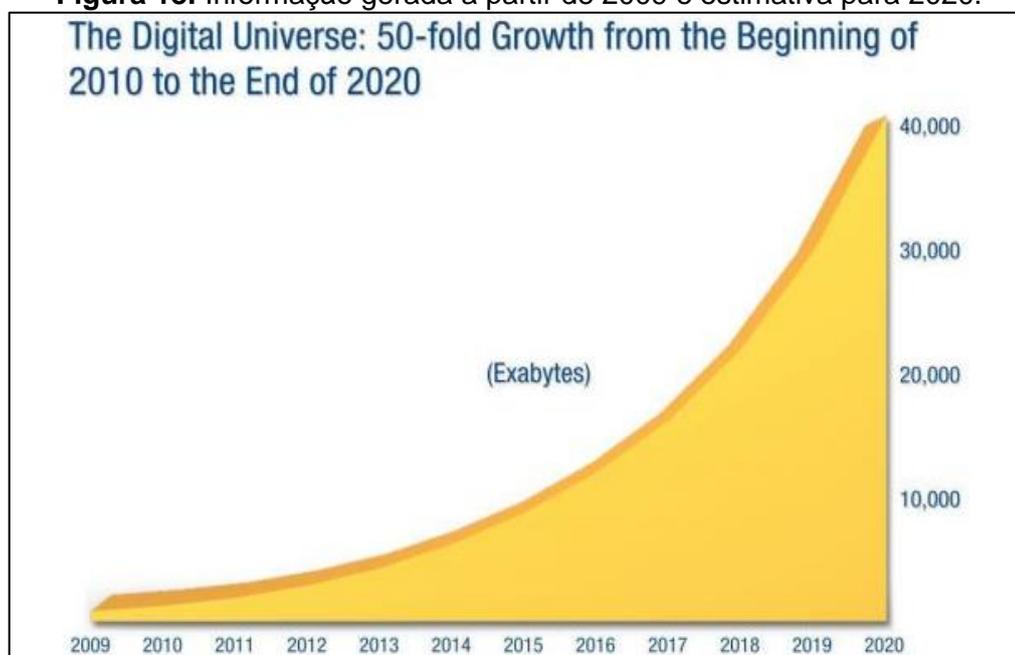
Fonte: Proativa, 2015.

A existência de sistemas que analisam dados de formatos variados é aparentemente uma realidade impraticável, de difícil exercício. Por outro lado, trabalhar com dados estruturados indica a possibilidade de recuperação devido à organização que possuem, neste caso, baseada nas relações e nas classificações específicas, ou seja, “[...] em grupos possuindo as mesmas descrições” (NASCIMENTO; COUTO, 2017, p. 5). Assim, os dados estruturados caracterizam-se como elementos que podem ser armazenados em um sistema de gerenciamento de base de dados (SGBD). Por outro lado, ao tratar de dados desestruturados, as questões de guarda e recuperação tornam-se complexas, uma vez que se tratam, por exemplo, de “[...] textos, documentos, gravações de serviços ao consumidor, fotos e imagens e vídeos [...]” (Ibid, p. 5) que envolvem informações aleatórias e estruturas não definidas. É nesse contexto que a *big data* se torna uma solução viável a partir de potenciais análises baseadas em dados estruturados e não estruturados para apoiar tomadas de decisão (ZIKOPOULOS *et al.*, 2012).

Por último, o elemento **velocidade** considera a rapidez com que são gerados os dados, segundo certo volume e dimensão de variedade (figura 18). A referida característica “[...] refere-se a aspectos da dinâmica de crescimento e processamento dos dados, em que o processo de tomada de decisão está muito interligado ao momento e a rapidez que tal decisão é tomada” (CONEGLIAN *et al.*, 2018, p. 139).

Outra forma de se enxergar a velocidade em *big data* seria considerando o dado como um objeto em movimento, ou seja, “[...] a velocidade com que os dados estão fluindo” (ZIKOPOULOS *et al.*, 2012, p. 8). Sendo assim, a rapidez nas ações de gerenciamento e análise de dados é posicionada como um componente de vantagem competitiva, como a extração de uma informação relevante para a tomada de decisão em tempo real, por exemplo. O objetivo torna-se a gerenciar e analisar dados em tempo real, traduzido como *streams computing*, enquanto um processamento de “[...] dados que são gerados continuamente ou ao vivo à medida que estes dados se tornem disponíveis [...]” (CHEDE, 2009, p. 1).

Figura 18: Informação gerada a partir de 2009 e estimativa para 2020.



Fonte: International Data Corporation (IDC), 2012.³⁷

Entre as características da tecnologia *big data*, ainda podem ser mencionadas as propriedades de **complexidade**, **veracidade** e **valor**. A complexidade diz respeito ao estabelecimento de relações e *links* entre os dados e os relacionamentos e hierarquias que provém de fontes variadas, uma vez que é de grande facilidade perder o controle sobre os dados. A propriedade veracidade refere-se à validade e inconsistências dos conjuntos de dados. Por sua vez, o valor da informação está relacionado com o conhecimento que ela pode gerar, principalmente em

³⁷ Relatório patrocinado pela EMC Corporation. Site: <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>

organizações, em que a geração de conhecimento reflete diretamente nas estratégias adotadas (KATAL; WAZID; GOUDAR, 2013).

A partir das características em “V” que formalizam a tecnologia *big data*, infere-se a sua pertinência “[...] quando todos, ou a maioria, dos dados precisam ser analisados *versus* uma amostra de dados [...]” (ZIKOPOULOS *et al.*, 2012, p. 16, tradução nossa) e a sua adoção “[...] para análises interativas e exploratórias quando as medidas da análise dos dados não foram pré-determinadas pelas organizações” (Ibid, p. 16, tradução nossa). A partir dos resultados obtidos pelas análises é tomar decisões, tanto para traçar metas, objetivos, ações, como para atingi-los, conseqüentemente (ARAÚJO JÚNIOR; SOUZA, 2018).

A tecnologia *big data* tem a capacidade de atuar em segmentos variados, incentivando o aumento da produtividade e competitividade das organizações. Por exemplo, o uso de aplicações *web* viabilizam previsões de mercado (corporativo) com base em informações provindas de fontes variadas, como a análise de redes sociais e o histórico de buscas dos usuários (CHEN; ZANG, 2014). Outro exemplo refere-se a empresa *WalMart*³⁸ ao implementar a *big data* como uma ferramenta para “[...] rastrear toda venda gravada em seus terminais de venda” (Ibid, p. 3, tradução nossa) e, a partir da análise do grande volume de dados disponibilizado, cerca de 267 milhões de transações, foi capaz de “[...] aumentar a eficiência deles nas estratégias dos preços e nas campanhas de publicidade” (Ibid, p. 3, tradução nossa).

Na era informacional, além do volume em ascensão de informação, as organizações têm que lidar com a velocidade na qual essa informação é gerada. Assim, a *big data* apresenta-se como uma funcionalidade útil para empresas por meio de plataformas como a *FICO Falcon*³⁹. Trata-se de um sistema que detecta fraudes em cartões de crédito e “[...] gerencia mais de 2.1 bilhões de contas válidas ao redor do mundo” (CHEN; ZANFG, 2014, p. 3, tradução nossa). Para atuar, a velocidade torna-se parâmetro essencial para a *FICO Falcon*, uma vez que necessita receber informações em tempo real para identificar eventuais anomalias.

Não somente empresas do setor privado utilizam dos benefícios oferecidos pela tecnologia *big data*, como também órgãos da administração pública. Ao lidar com centenas de pessoas, de idades diferentes e com necessidades variadas, a população anseia que a “máquina” governamental “[...] seja mais efetiva na administração

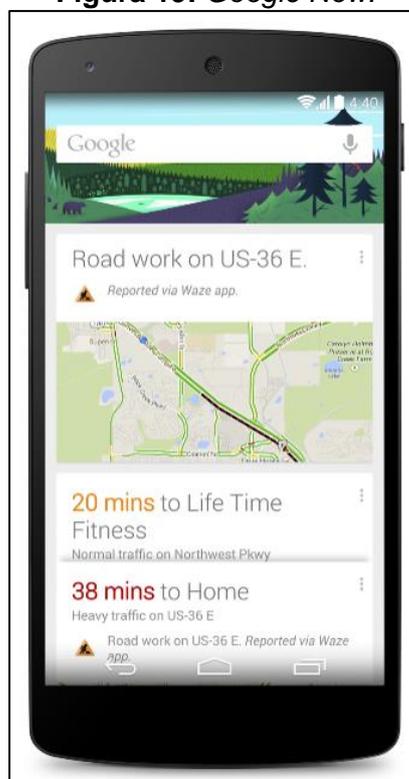
³⁸ Site: <https://www.walmart.com.br/>.

³⁹ Site: <https://www.fico.com>.

pública” (CHEN; ZANG, 2014, p. 3, tradução nossa). Desta forma, para alcançar o potencial de descobrir soluções alternativas para diminuir déficits orçamentários, o uso da *big data* destaca a propriedade de valor da informação para gerar uma tomada de decisão e aumentar a produtividade (CHEN; ZANG, 2014), como, por exemplo, a *Gendarmerie Nationale*⁴⁰, grupo policial francês comprometido com a segurança pública do Estado. Utilizando da velocidade em tempo real, veracidade e volume de informações dos sistemas aos quais possuem acesso, a *Gendarmerie Nationale* aumentou a economia de tempo e de recursos nas tomadas de decisões (SAP SE, 2017).

Presente em variados segmentos, entre sociais, corporativas e/ou governamentais, a tecnologia *big data* encontra-se também em aplicações de uso cotidiano. A partir dos dados coletados dos usuários, o aplicativo *Google Now*, por exemplo, é capaz de prever o melhor trajeto e o tempo entre a casa e o trabalho (figura 19), ou indicar eventos de entretenimento com base na localização e interesses pessoais do usuário (MASS, 2016).

Figura 19: *Google Now*.

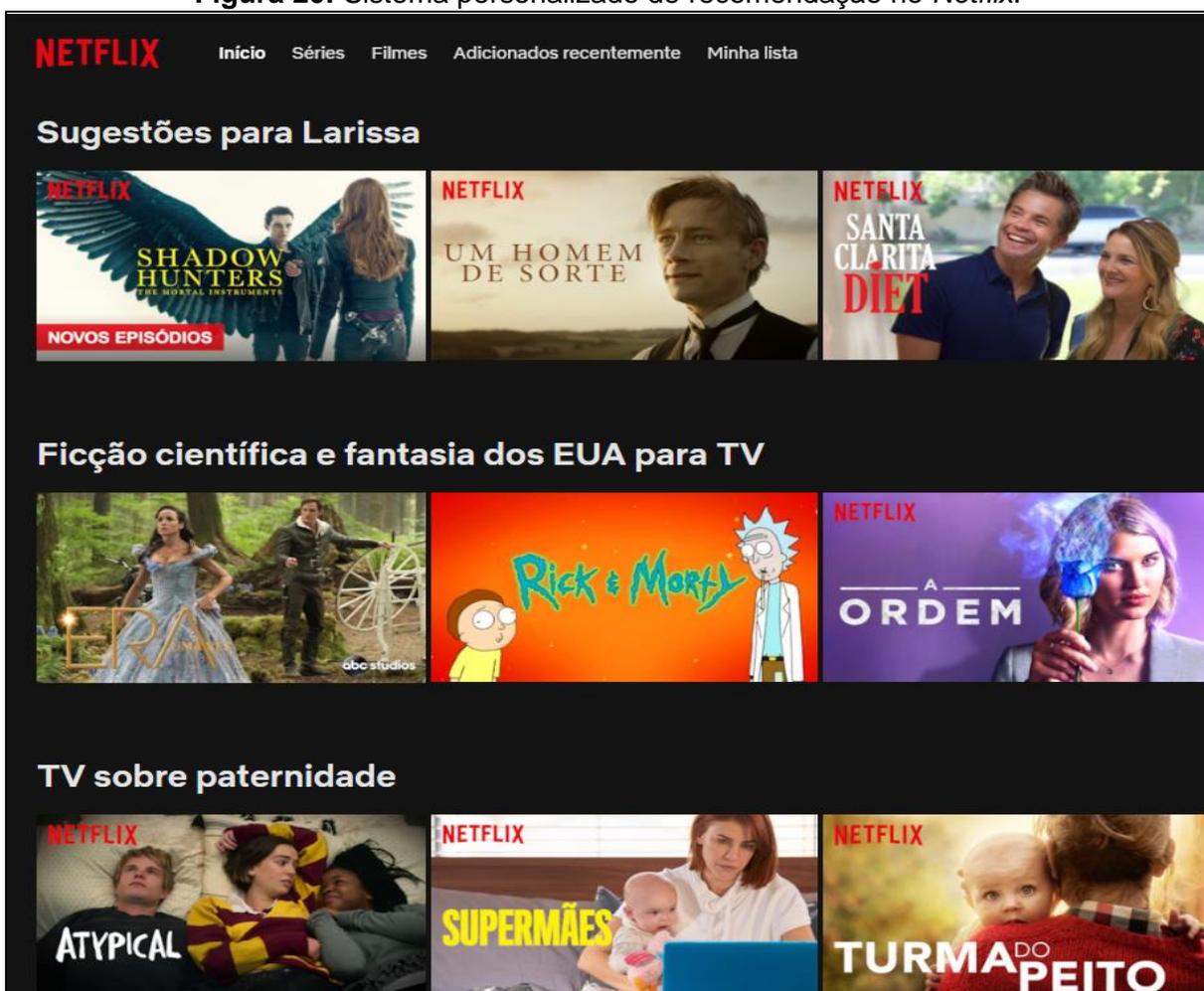


Fonte: Chavez, 2014.

⁴⁰ Site: <https://www.gendarmerie.interieur.gouv.fr/>.

Outro exemplo remete-se, novamente, ao *Netflix*, considerado uma referência em *streaming* de filmes e séries e que utiliza de algoritmos para coletar dados sobre os seus clientes. Após tratar os dados, o sistema fornece uma experiência única, ao oferecer sugestões personalizadas (RICCIULLI, 2016), como é possível visualizar na figura 20 a partir de “sugestões para o usuário”, “ficção científica e fantasia dos EUA para TV”, “TV sobre paternidade” etc. Contudo, abre-se um parêntese sobre a carência de profissionais da informação, como os bibliotecários, envolvidos com o *design* de interface e com a organização da informação, ao elencar possíveis limitações quanto a inexistência de buscas avançadas, como por assunto ou por palavras-chave indexadas⁴¹, por exemplo.

Figura 20: Sistema personalizado de recomendação no *Netflix*.



Fonte: Netflix, 2019.

⁴¹ Indexação, “[...] também chamada de tratamento da informação, seja manual ou automática, [...] ocorre a partir das fases de análise conceitual [...] e de tradução [...]” (DA SILVA, 2018, p. 27). A análise conceitual “[...] implica decidir do que trata um documento – isto é, qual o seu assunto” (LANCASTER, 2004, p. 90). A tradução “[...] envolve a conversão da análise conceitual de um documento num determinado conjunto de termos de indexação” (LANCASTER, 2004, p. 18).

O mesmo princípio de coleta e tratamento de dados, com fins de recomendações personalizadas, aplica-se ao *website* da *Amazon* (figura 21). Com base no que o usuário comprou, salvo na “lista de desejos”, por exemplo, a *Amazon* é capaz de realizar sugestões personalizadas, ou seja, “[...] recomendações [que] são geradas baseadas nas preferências do usuário alvo e dos outros usuários da comunidade” (BORGES; OLIVEIRA, [2010], p. 8). Ainda, o sistema recomenda um produto bastando o cliente tê-lo visualizado pelo menos uma vez. A partir das recomendações estabelecidas, a ordem de apresentação dos produtos sugeridos levará em consideração as avaliações e os itens mais acessados e comprados por outros clientes, em relação aos produtos conhecidos pelo próprio usuário.

Figura 21: Sistema personalizado de recomendação na *Amazon*.



Fonte: Amazon, 2019.

O emprego da tecnologia *big data* mostra potencial de uso também no segmento de pesquisas científicas, como “[...] o grande telescópio de levantamento sinóptico (*the Large Synoptic Survey Telescope – LSST*) [que] irá gravar 30 trilhões de bytes de dados em imagens em um único dia” (ZANG; CHEN, 2014, p. 4, tradução nossa) (figura 22). A análise desses dados resultará em informações acerca da origem do universo, ao validar o “amontoado” de informações capturadas pelo telescópio.

Figura 22: *The Large Synoptic Survey Telescope.*



Fonte: National Science Foundation, 2012.

Em suma, a tecnologia *big data* constitui a combinação de características, entre **volume**, **variedade**, **velocidade**, **complexidade**, **veracidade** e **valor**, as quais subsidiam o estabelecimento de diretrizes acerca do uso, aplicação e resultados possíveis a serem gerados. A ferramenta agrega valor aos dados, oferta informações tratadas e compreensíveis, contribui na possibilidade de constituir conhecimento e potencializa e tomada de decisão a partir da sabedoria construída, e, assim, estipular e avaliar estratégias e metas nas organizações.

3 METODOLOGIA

Os fenômenos presentes no cotidiano dos seres humanos são possíveis de serem demonstrados e explicados graças às ciências. A ciência é constituída por objetivo ou finalidade, função e objeto, enquanto “[...] uma sistematização de conhecimento, um conjunto de proposições logicamente correlacionadas sobre o comportamento de certos fenômenos [...]” (MARCONI; LAKATOS, 2010, p. 62).

No prisma das ciências, pesquisas são realizadas para alcançar a comprovação, ou não, de hipóteses e, conseqüentemente, conseguir explicar fenômenos da natureza investigados. Para tanto, adota-se a pesquisa científica, compreendida como uma “[...] investigação disciplinada das relações entre os fenômenos naturais [...]” (KERLINGER, 1973, p. 321), ou ainda como “[...] a atividade científica pela qual descobrimos a realidade” (DEMO, 1987, p. 23).

Entre as partes que constituem uma pesquisa científica está a metodologia, ou seja, etapa que requer a definição do caminho pelo qual os resultados serão alcançados de acordo com um “[...] conjunto de métodos ou caminhos que são percorridos na busca do conhecimento” (ANDRADE, 2010, p. 117). Neste caso, direcionando ao presente trabalho, os caminhos adotados almejam alcançar o objetivo de apresentar um cenário literário de uso da tecnologia *big data* em bibliotecas.

3.1 CARACTERIZAÇÃO DA PESQUISA

Enquanto uma pesquisa científica, o presente TCC (quadro 1) caracteriza-se segundo o método científico, a natureza, os objetivos propostos, o procedimento técnico de investigação e a abordagem de coleta de dados. Segundo Lakatos e Marconi (2017b), essas características referem-se ao percurso metodológico a ser seguido pelo pesquisador no andamento do estudo.

Quadro 1: Caracterização da pesquisa.

Método	Indutivo
Natureza	Básica
Objetivo	Exploratória
Procedimento técnico	Pesquisa bibliográfica
Abordagem	Qualitativa

Fonte: Da autora, 2019; com base em Silva e Menezes, 2005.

Intencionado em apresentar um panorama macro, o método utilizado na pesquisa é o indutivo, pois “[...] a generalização deriva de observações de casos da realidade concreta” (SILVA; MENEZES, 2001, p. 26). Em outras palavras, o método adotado “[...] possibilita o desenvolvimento de enunciados gerais sobre as observações acumuladas de casos específicos ou proposições que possam ter validades universais” (OLIVEIRA, 2001, p. 119).

No que tange à natureza, a pesquisa caracteriza-se como básica, pois “objetiva gerar conhecimentos novos, úteis para o avanço da Ciência, sem aplicação prática prevista” (GERHARDT; SILVEIRA, 2009, p. 34), e assim contribuir na solução de demandas concretas que, no caso da presente investigação, é apresentar um levantamento de exemplos que elenquem aplicações da *big data* em bibliotecas, de forma a aprimorar e melhorar produtos e serviços oferecidos aos seus usuários.

Quanto ao estudo exploratório, tem a finalidade de acumular mais informações sobre determinado assunto, torná-lo familiar, uma vez que costumam ser temas com bibliografia escassa ou em desenvolvimento (RAMOS, 2009). Neste caso, de acordo com os objetivos propostos, a pesquisa exploratória contribui no maior aprofundamento sobre a tecnologia *big data*, com ênfase em bibliotecas, esclarecendo e apontando possibilidades de aplicações.

Sustentando a pesquisa exploratória, o procedimento técnico é a pesquisa bibliográfica, intencionado em abranger o maior número possível de publicações relacionadas aos assuntos abordados no TCC, especialmente sobre a tecnologia *big data* continuando justificando sobre a importância de fazer um levantamento bibliográfico sobre um assunto que ainda exige de publicações e experiências em bibliotecas, especialmente no Brasil. Neste sentido, a pesquisa bibliográfica objetiva “[...] conduzir o leitor a determinado assunto e a produção, coleção, armazenamento, reprodução, utilização e comunicação das informações coletadas para o desempenho da pesquisa” (FACHIN, 2001, p. 125), o que possibilita “[...] descobrir novo tipo de enfoque [...]” (ANDRADE, 2010, p. 112) para o objeto de estudo.

Sobre à abordagem, a coleta de dados caracteriza-se como qualitativa, pois “[...] preocupa-se em analisar e interpretar aspectos mais profundos, descrevendo a complexidade do comportamento humano. Fornece análise mais detalhada sobre investigações, hábitos, atitudes, tendências de comportamento etc.” (LAKATOS; MARCONI, 2017b, p. 299), resultando em um panorama geral do assunto, nesse caso, a tecnologia *big data* em bibliotecas.

3.2 ETAPAS DA PESQUISA

Primeiramente, foram pesquisados conceitos (básicos) na expectativa de auxiliar na definição e compreensão do tema da pesquisa e, conseqüentemente, na contextualização dos objetivos propostos. Sendo assim, as pesquisas se deram sobre pirâmide informacional, explosão da informação, Internet, *web* e *big data*. Em seguida, estabelecida uma compreensão sobre o tema da pesquisa, que é a tecnologia *big data* nas bibliotecas, foi possível realizar o levantamento bibliográfico a partir de discussões sobre a evolução da Internet até se chegar à particularidades e ambientes de cada fase da *web*, em situações cotidianas e em ambientes específicos, caminho que subsidiou ao alcance do assunto principal da pesquisa, ou seja, a *big data*. A partir desse levantamento teórico foi possível identificar aplicações e usos da tecnologia *big data* em bibliotecas, formalizando a coleta de dados. Na sequência, ocorreram o tratamento, a análise e a interpretação dos dados coletados a fim de responder aos problemas da pesquisa. Em suma, compreendido ao período de agosto de 2018 até março de 2019, o percurso metodológico dividiu-se nas seguintes etapas:

1. Delimitação do tema;
2. Pesquisa bibliográfica a partir de consultas realizadas no buscador *Google*⁴², na Biblioteca Digital de Dissertações e Teses (BDTD)⁴³, na Base de Dados Referencial de Artigos e Periódicos em Ciência da Informação (BRAPCI)⁴⁴, em livros e em *blogs* especializados nas áreas de Tecnologia e Informação;
3. Identificação dos ambientes de aplicação da *big data* para a coleta de dados;
4. Coleta de dados nos ambientes selecionados, utilizando da combinação das palavras-chave “*big data*” e “biblioteca” como forma de filtrar os resultados obtidos.
5. Tratamento e análise dos dados coletados;
6. Confecção de um quadro sugestão de uso da *big data* em bibliotecas.

⁴² Site: <https://www.google.com.br/>.

⁴³ Site: <http://bdt.d.ibict.br/vufind/>. Biblioteca promovida pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

⁴⁴ Site: <http://www.brapci.inf.br/>.

4 RESULTADOS DA PESQUISA

Considerando os exemplos elencados no cotidiano, em setores diversos, entre pessoal, corporativo e governamental, infere-se a *big data* como uma tecnologia de ampla implementação. Neste sentido, visando ao alcance do objetivo de apresentar um cenário literário de uso da referida tecnologia em bibliotecas, por meio do percurso metodológico adotado para a coleta e seleção de informações, identificou-se a aplicação da *big data* em bibliotecas, o que gerou a detecção de três ambientes, especificamente sendo a *Harvard University Library*⁴⁵, a *Washington University Library*⁴⁶ e a *Seattle Public Library*⁴⁷.

De acordo com o levantamento teórico realizado na pesquisa, três são as características básicas que formalizam a tecnologia *big data*, a partir da representação chamada de “3V”, ou seja, volume, variedade e velocidade, e complementadas pelos predicados complexidade, veracidade e valor, assim estratificadas no quadro dois (2).

Quadro 2: Características básicas da *big data*.

ESSENCIAIS			COMPLEMENTARES		
VOLUME	VARIEDADE	VELOCIDADE	COMPLEXIDADE	VERACIDADE	VALOR
Refere-se a grande quantidade de dados e informações gerada diariamente por instrumentos tecnológicos diversos.	Relaciona-se aos diferentes tipos e formatos de informações que são gerados e coletados.	Associa-se ao volume e a variedade, uma vez que a grande quantidade de dados, de variados tamanhos e categorias, é gerada em um curto espaço de tempo.	Baseia-se na estruturação de relações, <i>links</i> e hierarquias entre dados e relacionamentos, provindos de fontes diversas.	Verifica-se a validade e fragilidades dos conjuntos de dados.	Constitui-se no conhecimento que pode ser gerado a partir de informações e ser usado como estratégia pelas Organizações.

Fonte: Da autora, 2019.

De posse das características mínimas da tecnologia em estudo, identificaram-se cenários de uso da *big data* no cotidiano. Por exemplo, o sistema de recomendação personalizado para os usuários da *Amazon*, a partir da coleta de informações de dados, entre interesses pessoais, produtos visualizados e compras realizadas na plataforma, é capaz de apresentar sugestões, tanto com base no histórico do cliente, quanto nas pesquisas e compras semelhantes de outros usuários. Neste sentido, ao abrir a página de descrição do produto “*notebook Dell*”, por exemplo, há

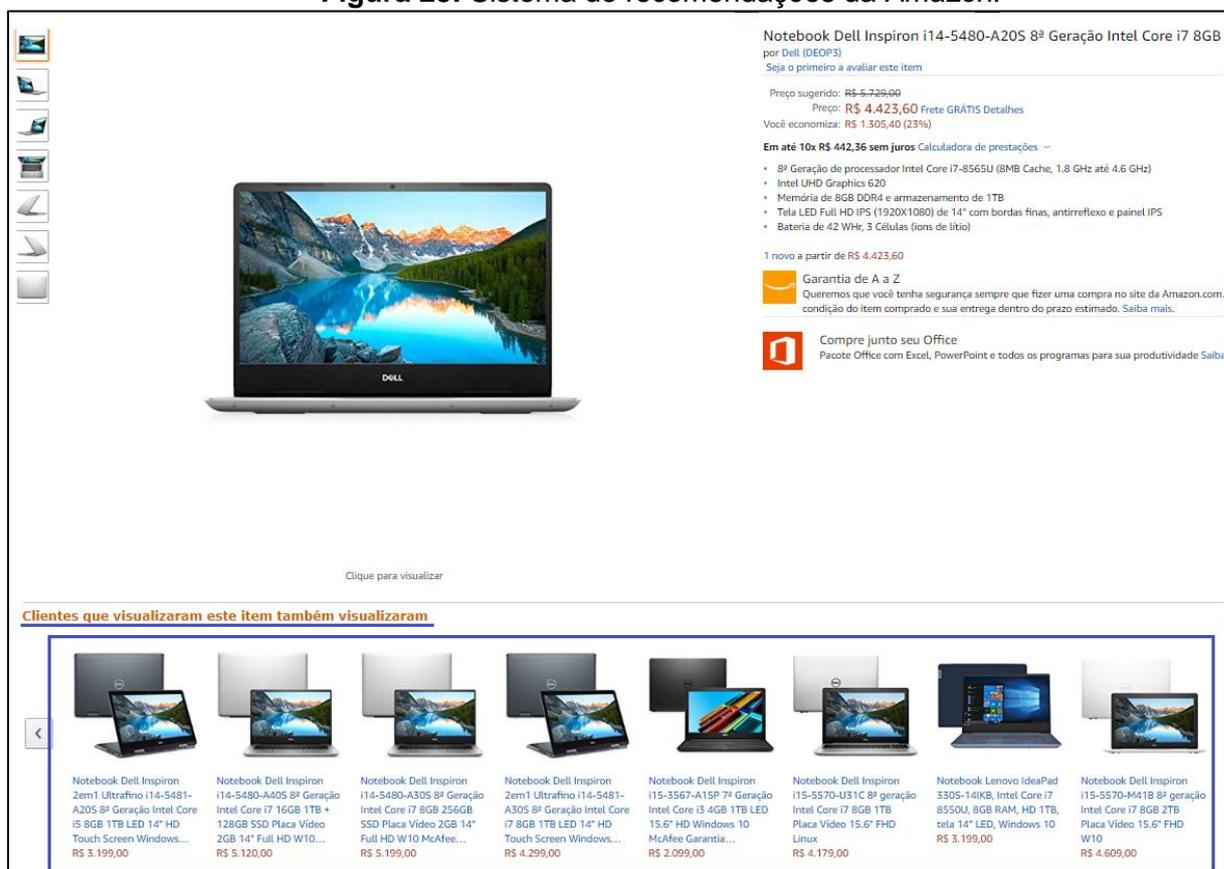
⁴⁵ Site: <https://library.harvard.edu/>.

⁴⁶ Site: <https://library.wustl.edu/>.

⁴⁷ Site: <https://www.spl.org/>.

recomendação de produtos semelhantes de acordo com a seguinte informação: “clientes que visualizaram este item também visualizaram” (figura 23) outros computadores semelhantes.

Figura 23: Sistema de recomendações da Amazon.



Notebook Dell Inspiron i14-5480-A20S 8ª Geração Intel Core i7 8GB por Dell (DEOP3)
Seja o primeiro a avaliar este item

Preço sugerido: R\$ 5.799,00
Preço: **R\$ 4.423,60** Frete GRÁTIS Detalhes
Você economiza: R\$ 1.305,40 (23%)

Em até 10x R\$ 442,36 sem juros Calculadora de prestações

- 8ª Geração de processador Intel Core i7-8565U (8MB Cache, 1,8 GHz até 4,6 GHz)
- Intel UHD Graphics 620
- Memória de 8GB DDR4 e armazenamento de 1TB
- Tela LED Full HD IPS (1920x1080) de 14" com bordas finas, antirreflexo e painel IPS
- Bateria de 42 Whr; 3 Células (ions de lítio)

1 novo a partir de R\$ 4.423,60

Garantia de A a Z
Queremos que você tenha segurança sempre que fizer uma compra no site da Amazon.com, condição do item comprado e sua entrega dentro do prazo estimado. Saiba mais.

Compre junto seu Office
Pacote Office com Excel, PowerPoint e todos os programas para sua produtividade Saiba mais.

Clique para visualizar

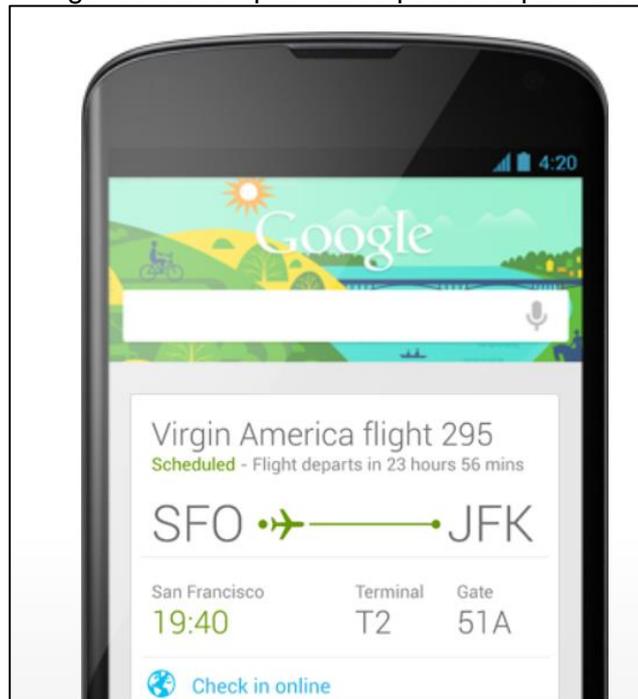
Clientes que visualizaram este item também visualizaram

 Notebook Dell Inspiron Zem1 Ultrafino i14-5481-A20S 8ª Geração Intel Core i5 8GB 1TB LED 14" HD Touch Screen Windows... R\$ 3.199,00	 Notebook Dell Inspiron i14-5480-A40S 8ª Geração Intel Core i7 16GB 1TB + 128GB SSD Placa Vídeo 2GB 14" Full HD W10... R\$ 5.120,00	 Notebook Dell Inspiron i14-5480-A30S 8ª Geração Intel Core i7 8GB 256GB SSD Placa Vídeo 2GB 14" Full HD W10 McAfee... R\$ 5.199,00	 Notebook Dell Inspiron Zem1 Ultrafino i14-5481-A30S 8ª Geração Intel Core i7 8GB 1TB LED 14" HD Touch Screen Windows... R\$ 4.299,00	 Notebook Dell Inspiron i15-3567-A1SP 7ª Geração Intel Core i3 4GB 1TB LED 15,6" HD Windows 10 McAfee Garantia... R\$ 2.099,00	 Notebook Dell Inspiron i15-5570-U31C 8ª geração Intel Core i7 8GB 1TB Placa Vídeo 15,6" FHD Linux R\$ 4.179,00	 Notebook Lenovo IdeaPad 330S-14IKB, Intel Core i7 8550U, 8GB RAM, HD 1TB, tela 14" LED, Windows 10 R\$ 3.199,00	 Notebook Dell Inspiron i15-5570-M418 8ª geração Intel Core i7 8GB 2TB Placa Vídeo 15,6" FHD W10 R\$ 4.609,00
--	---	---	--	--	---	---	---

Fonte: Amazon, 2019.

Citando novamente o *Google Now*, foi outro recurso digital identificado e que faz uso da *big data*. Neste caso, a referida aplicação de uso cotidiano, a partir do histórico de pesquisas realizadas pelo usuário enquanto “logado” em sua conta (*Google*), além de considerar a localização e o tempo, consegue apresentar sugestões conforme os interesses pessoais. Com isso, o sistema, por exemplo, “aprende” sobre a rotina do usuário e, assim, recomendar melhores rotas do trabalho até a sua casa, mostrar notícias e eventos, informar horários de filmes e a distância dos cinemas conforme a sua localização real, apresentar lembretes sobre compromissos pessoais, como viagens e, conseqüentemente, o melhor trajeto até o aeroporto e informações acerca do voo, como a ilustração da figura 24, entre São Francisco (SFO) e Nova Iorque (JFK), nos EUA.

Figura 24: Agenda de compromissos pessoais pelo *Google Now*.



Fonte: Google, [201-].

Outro exemplo, é o *Netflix*, novamente citado. Destaca-se não apenas como *streaming* de filmes e séries, mas por proporcionar sugestões personalizadas aos clientes, embasado nas avaliações atribuídas aos seus conteúdos a partir de interesses individuais, séries e filmes assistidos e consequentes gêneros mais escolhidos (figura 25).

Figura 25: Recomendação do *Netflix*.



Fonte: Netflix, 2019.

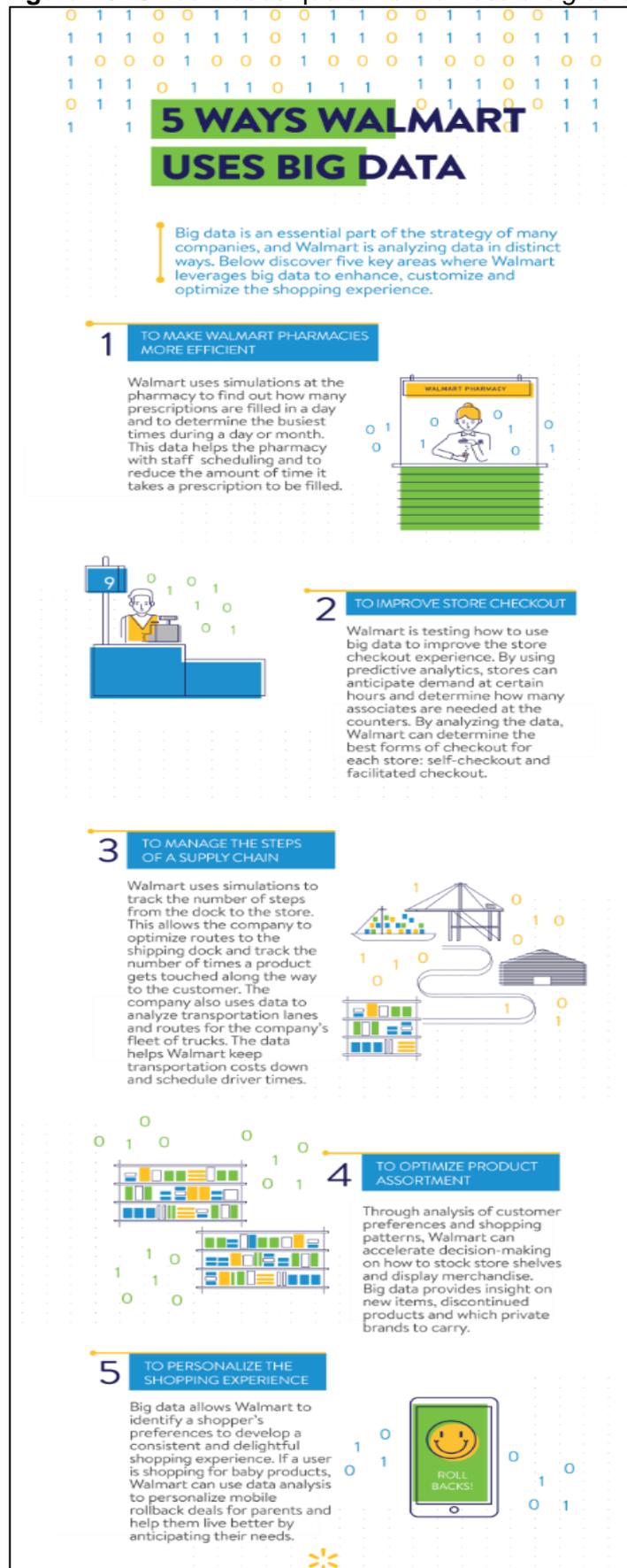
Desta forma, a figura 25 demonstra, numericamente, a relevância de sugerir a série *Stranger Things* diante do resultado de 95%, representado pela comparação entre o que foi assistido pelo indivíduo e a pontuação aferida ao conteúdo pelos usuários que também consumiram o mesmo produto. Portanto, quanto maior a avaliação do conteúdo, maior será a chance de recomendação.

No ramo corporativo, empresas como o *WalMart*, desfrutam de dados de tráfego e seleção de produtos dos clientes para gerar estratégias e campanhas, buscando conhecer as suas necessidades/costumes e ajuda-los na tomada de decisões. Atuando dessa forma, o *WalMart* almeja otimizar a experiência de compra dos seus usuários. Como exemplo (figura 26), o *WalMart* (2017) elenca as áreas de aplicação da *big data* e como a tecnologia ocorre, respeitando a correspondente sequência:

1. Inicia-se pela análise de demandas dos clientes na farmácia (*to make WalMart pharmacies more efficient*);
2. Passa para a etapa de pagamento e entrega dos produtos (*“to improve store checkout”*), o que permite definir a carga-horária e a quantidade de funcionários necessárias para otimizar os serviços;
3. Ocorre o monitoramento do transporte de produtos (*“to manage the steps of a supply chain”*), desde a fornecedora até a loja, traçando melhores rotas e horários;

Acelera-se a tomada de decisão dos clientes com base nas preferências de produtos e marcas, no momento de selecionar os produtos das prateleiras (*“to optimize product assortment”*), afora possibilitar personalização na experiência de compras *online* (*“to personalize the shopping experience”*) por meio do acompanhamento e análise dos produtos pesquisados.

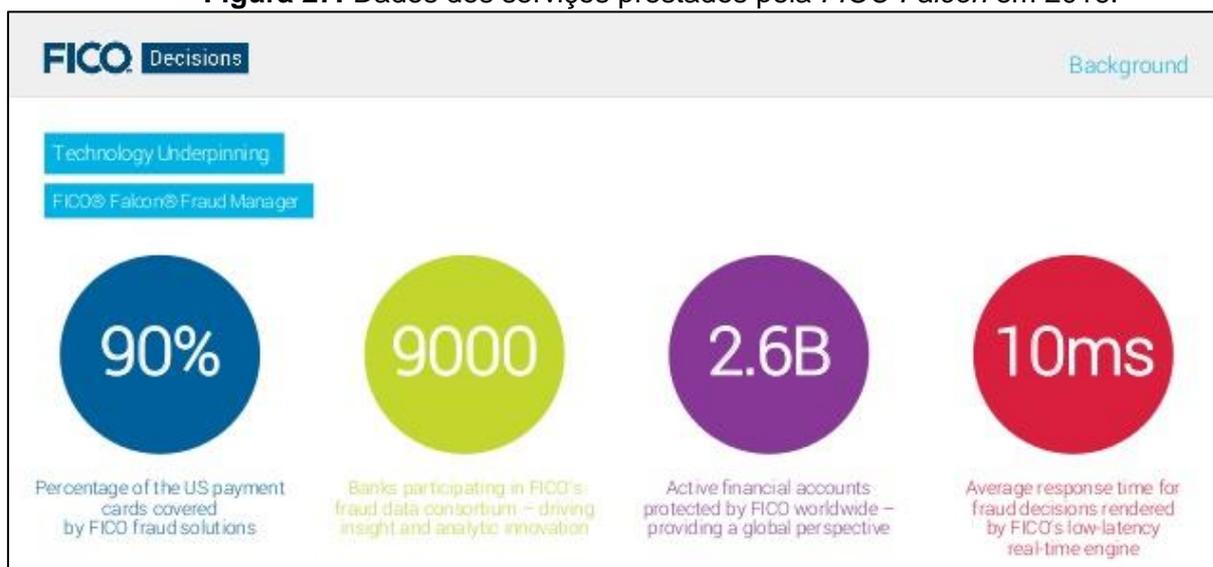
Figura 26: Cinco modos que o WalMart usa a big data.



Fonte: WalMart, 2017.

Empresas como a *FICO Falcon*, novamente citada, utilizam de algoritmos para detecção de fraudes em cartões de crédito e em alta velocidade. A organização dispõe de larga quantidade de dados/informações, mas o ponto-chave para o sucesso de suas operações é a análise desse montante em tempo real. Por exemplo (figura 27), conforme informações organizacionais de 2016, dentre os 9000 bancos afiliados, dos quais concedem informações para o banco de dados (BD) da empresa, e os 2.6 bilhões de contas financeiras que estão aos cuidados da *FICO Falcon* em todo o mundo, o tempo de resposta para decisões voltadas para possíveis fraudes custa, em média, 10 milissegundos. O baixo tempo para tomadas de decisões, somado às informações privilegiadas que são disponibilizadas pelos bancos, garantiram à organização o alcance de 90% de sucesso na cobertura de soluções de fraudes relacionadas aos cartões de crédito na região dos EUA.

Figura 27: Dados dos serviços prestados pela *FICO Falcon* em 2016.

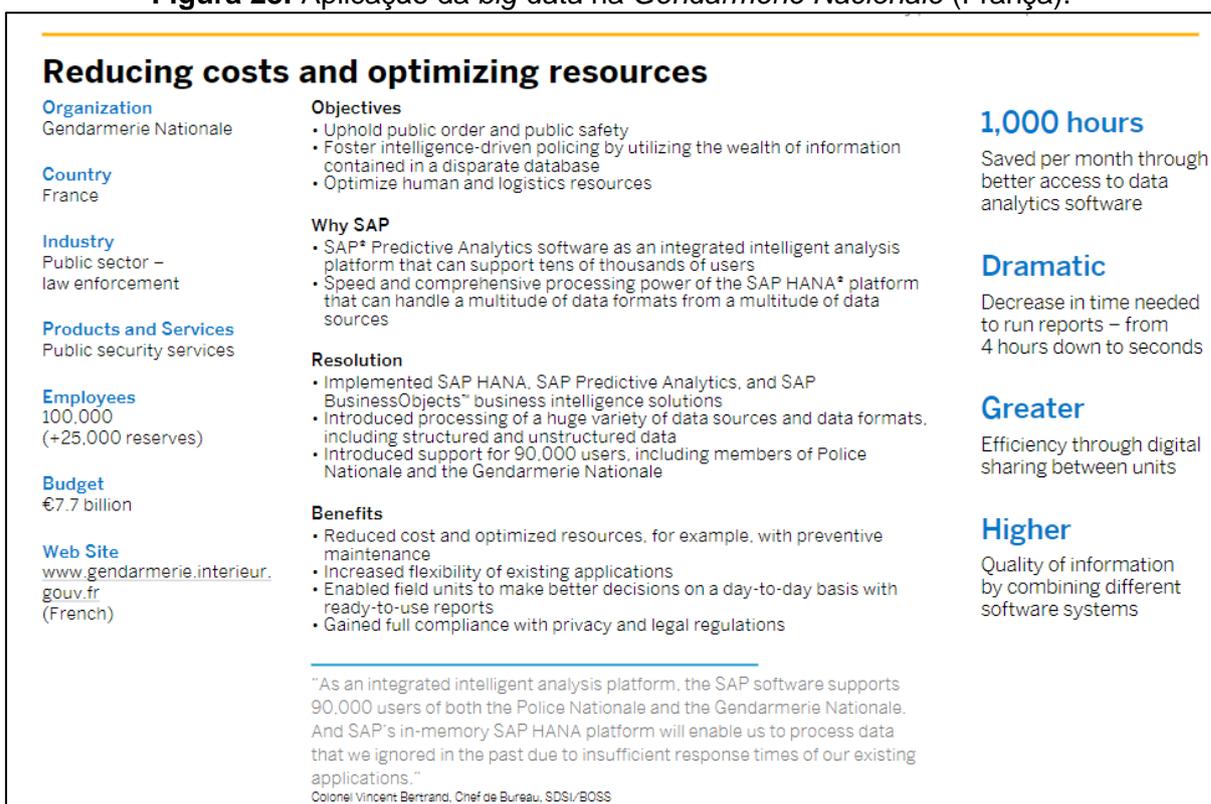


Fonte: *FICO Falcon*, 2016.

No segmento governamental, o uso de *big data* torna possível obter soluções, embasar tomadas de decisões e aumento da produtividade, posto que dispõem de grande quantidade de informações, em variados formatos, sobre os cidadãos, as instituições e as organizações. Para Chen e Zang (2014), a *big data* tem a capacidade de influenciar na definição de políticas públicas e orçamentos, por exemplo, tornando a administração pública mais eficiente e produtiva, ao passo que resulta no crescimento da nação.

A adoção da *big data* pode ser exemplificada pela *Gendarmerie Nationale* (figura 28), força militar francesa encarregada da segurança do Estado Francês, objetivando defender a ordem e a segurança pública (*uphold public order and public safety*); promover o policiamento orientado pela inteligência, usando o valor de informações contidas em BD díspar (*foster intelligence-driven policing by utilizing the wealth of information contained in a disparate database*); e otimizar recursos humanos e logísticos (*optimize human and logistics resources*). A partir da tecnologia, a Organização alcançou benefícios relacionados à economia de tempo, tanto para o acesso a análises de dados, quanto elaborar relatórios; maior eficiência ao permitir uma melhor tomada de decisão no dia a dia; e aumento da qualidade da informação ao combinar diferentes sistemas de informação.

Figura 28: Aplicação da *big data* na *Gendarmerie Nationale* (França).

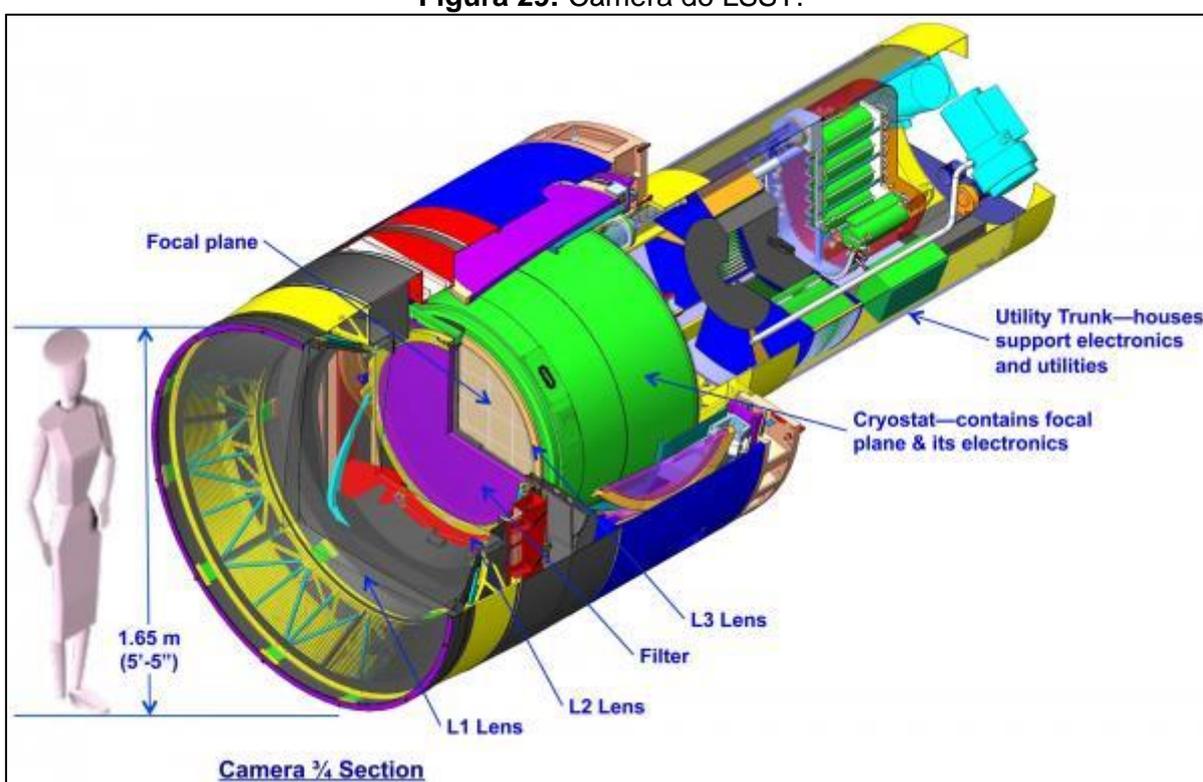


Fonte: SAP SE, 2017.

Quanto a aplicação da tecnologia *big data* no setor de pesquisas científicas, rememora-se o LSST enquanto um projeto colaborativo entre 23 países, focados em "[...] construir novos modos de colaboração interdisciplinar nas áreas de astronomia, física, ciência da computação, matemática e ciência da informação" (LARGE SYNOPTIC SURVEY TELESCOPE, [20-?]a, p. 1, tradução nossa). Por meio da

câmera que faz parte de sua constituição, ilustrada na figura 29, o telescópio tem a capacidade de “[...] entregar um conjunto de 500 petabytes de imagens e dados [...]” (Ibid, p. 1, tradução nossa). Assim, o LSST obterá, como resultado, a catalogação do sistema solar e o mapeamento de galáxias, e, a partir da análise de um amontoado de dados capturados, ao longo de 10 anos, conduzirá pesquisas para obter respostas sobre questionamentos que hoje não são possíveis de serem respondidos, como, por exemplo, “o que é matéria escura, como está distribuída, e como suas propriedades afetam na formação de estrelas, galáxias e estruturas maiores?” (Id., [20-?]b, p. 1, tradução nossa).

Figura 29: Câmera do LSST.



Fonte: LARGE SYNOPTIC SURVEY TELESCOPE, 2011.

Somadas as características básicas e exemplos de aplicação da tecnologia *big data* no contexto geral, bem como explanado na fundamentação teórica, a possibilidade de aplicação, em nichos diversos, torna o campo da Educação um segmento passível de adoção dessa tecnologia, especificamente nas bibliotecas, enquanto unidades onde o fluxo de informação é intenso, tanto na produção, quanto na disseminação, o que faz com que o tradicional conceito de biblioteca seja repensado. Em outras palavras, com o progresso tecnológico, o conceito de biblioteca

foi revisado e/ou ampliado, fomentando o surgimento de espaços extensivos ao ambiente físico como as bibliotecas digitais, por exemplo.

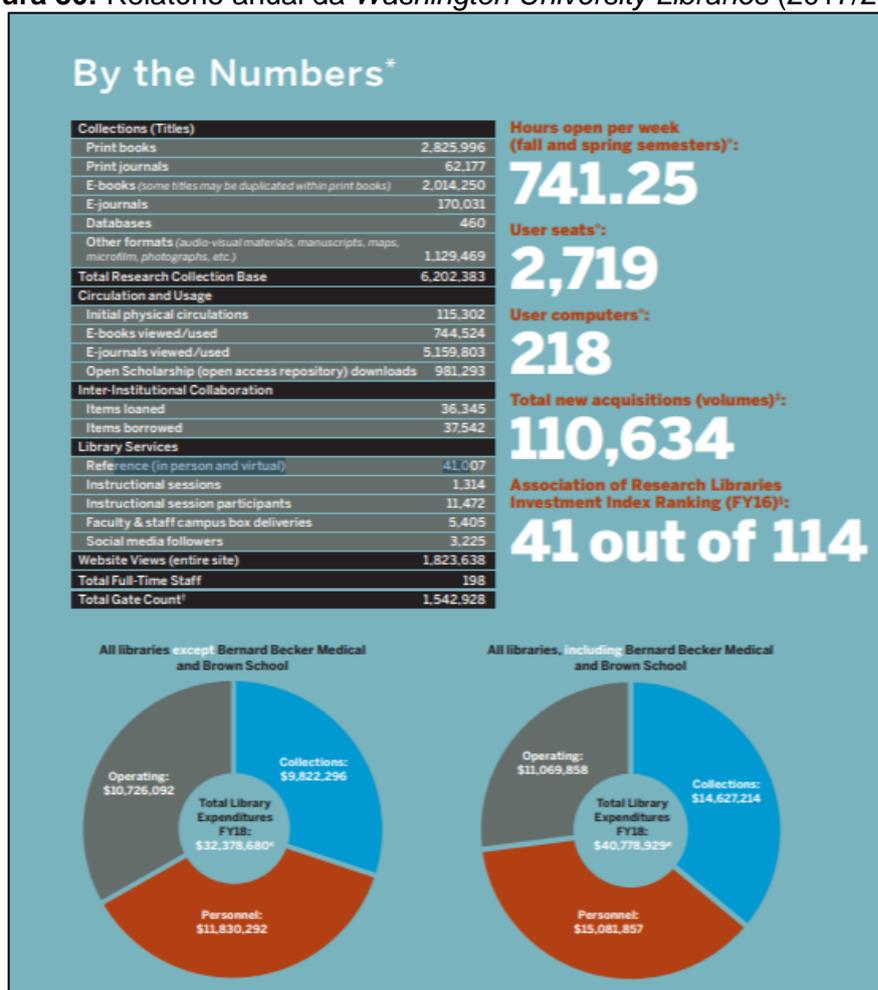
A biblioteca, comumente relacionada a espaços físicos e que objetiva guardar e organizar informações gravadas em suportes físicos para disponibilizar a sociedade (SANTOS; ASSUNÇÃO, 2012), passa a romper paradigmas a partir da sua versão digital, especialmente nos seguintes aspectos:

[...] tempo (na localização, obtenção e conservação de documentos); com a questão da distância (entre pontos de produção ou disponibilização de informação e a localização dos utilizadores); e com a questão dos custos (ao pretender assegurar a partilha social dos custos no acesso à informação) (FURTADO, p. 3, 1998)

Por mais que rupturas abruptas obrigaram a biblioteca a adaptar-se e surgir em novos formatos, a missão essencial não foi alterada, ou seja, de “[...] oferecer um conjunto organizado de informações afins, armazenadas com o objetivo de consulta ou utilização posterior” (LUCAS, 2004, p. 16). Portanto, considerando a biblioteca como organização, é necessário, como qualquer outro negócio, conhecer seus concorrentes, bem como o ambiente em que está inserida, como defende Amaral (1990). Ao considerar bibliotecas como empresas, defende-se “[...] informação como um bem a ser comercializado como outro qualquer, o que não impede que os usuários sejam bem atendidos e recebam as informações que necessitam” (Ibid, p. 314). Dessa forma, estabelece que como organização, bibliotecas necessitam de gerenciamento e, por conseguinte, de tomadas de decisão, tornando-se assim apta para a implantação de *big data* em seus processos.

A *big data* contribui para a implementação de ferramentas em bibliotecas voltadas, principalmente, para a visualização de dados e, a partir disso, gerar conhecimentos para embasar as tomadas de decisão. Na *Washington University*, por exemplo, todos os dados coletados foram reunidos pelas bibliotecas setoriais da universidade. A partir do conjunto formado, nasceu o projeto *Making the numbers speak* que consiste em representar os dados graficamente e interpretá-los (CHEN *et al.*, 2015), conforme ilustra a figura 30. Os resultados guiaram à direção estratégica das bibliotecas que se preocupavam em “[...] entender melhor os estudantes e membros [...] [atendidos e] “[...] tomar melhores decisões em relação às bibliotecas” (Ibid, p. 2, tradução nossa).

Figura 30: Relatório anual da *Washington University Libraries* (2017/2018).



Fonte: *Washington University Libraries*, 2018, p. 13.

A partir da reunião de relatórios semestrais e anuais (figura 30) geram-se informações, e combinadas com o método de *Balanced Scorecard*⁴⁸, as bibliotecas da *Washington University* embasam os seus planejamentos estratégicos. Os resultados, de visualizar graficamente os volumes de dados coletados, refletem nos comportamentos organizacionais. A partir dos dados referentes aos projetos desenvolvidos em 2015-2016, por exemplo, foi lançado o projeto *Geospatial Information Systems* em 2017 (CHEN *et al.*, 2016), ilustrado na figura 31.

⁴⁸ O método *Balanced Scorecard* (BSC) surgiu na década de 90, introduzido pelos professores Robert S. Kaplan e David P. Norton, “[...] como um modelo de avaliação e performance empresarial” (SILVA; DIAS, 2011, p. 1). A evolução da ferramenta possibilitou que o BSC atualmente alinhe “[...] os resultados de desempenho da empresa, com as metas traçadas, buscando identificar as falhas” (Ibid, p. 1).

Figura 31: Projeto *Geospatial Information Systems* (2017).

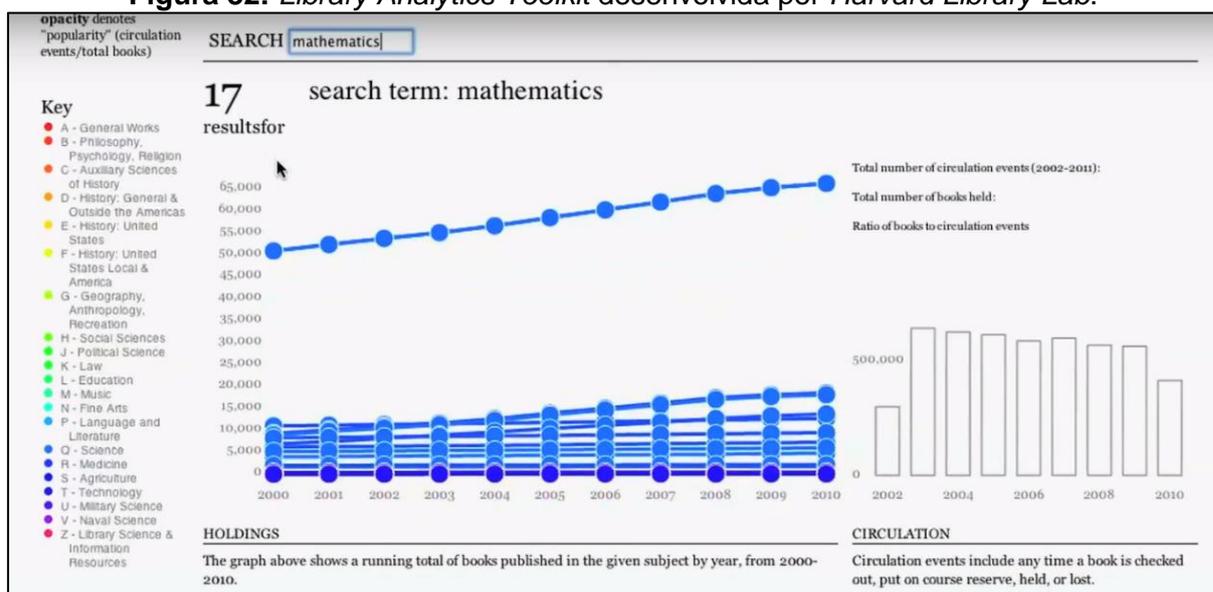


Fonte: *Washington University Libraries*, 2017, p. 16.

O projeto *Geospatial Information Systems*, de colaboração entre as bibliotecas da *Washington University* e o departamento administrativo de *Facilities Planning and Management*, resultou em um mapa interativo do campus, no qual, visitantes, alunos e profissionais podem ter uma visão geral, via *desktop* e dispositivos *mobiles*, para localizar estacionamentos, bibliotecas setoriais, prédios acadêmicos, lanchonetes e ainda obter informações sobre os estabelecimentos (WASHINGTON UNIVERSITY LIBRARIES, 2017).

Outro exemplo se refere aos laboratórios de inovação da biblioteca da *Harvard University* ao proporem a expansão das funcionalidades da página *Check out the checkouts*, pois o *site* apresenta as obras devolvidas em formato de *ranking*. Essa funcionalidade, chamada de *Library Analytics Toolkit* (figura 32), permite, além de ser configurada por cada biblioteca setorial da Universidade individualmente, que os dados sejam estudados por meio de um painel visual (HARVARD LIBRARY LAB, [2011]), ou seja, conforme um “[...] *dashboard* de análise interativo que providenciaria informação detalhada e em tempo real sobre os eventos de uso da biblioteca no sistema de *Harvard*” (Ibid, p. 2, tradução nossa).

Figura 32: *Library Analytics Toolkit* desenvolvida por *Harvard Library Lab*.



Fonte: *Harvard Library Lab*, [2014].

A figura 32 exemplifica o *dashboard* interativo desenvolvido pela *Harvard Library Lab*, em que é possível visualizar, em *timeline*, informações de circulação referentes às obras relacionadas ao termo pesquisado pelo usuário. Por meio dessas informações, é possível identificar comportamentos e interesses dos usuários e, assim decidir sobre a constituição do acervo da *Harvard University*. Na figura 32, o gráfico de elementos azuis (*holdings*) apresenta o número total de obras sobre determinado assunto, no caso matemática, que foram publicadas por ano, no intervalo de 2002 a 2010. Quanto ao gráfico em barras (*circulation*), corresponde a quantitativos, por ano, de eventos de circulação, ou seja, inclui quando o livro foi emprestado, reservado, tirado da estante ou perdido.

Em 2004, o projeto *Making visible the invisible* propôs reunir os dados coletados pela *Seattle Public Library* e apresentá-los aos seus usuários por meio de painéis (LEGRADY, 2005). A proposta centra-se na visualização de mudanças no fluxo de informação ao longo do tempo, através do processamento de dados. Conseqüentemente, esses dados são apresentados eletronicamente, refletindo “[...] em tempo real a natureza dinâmica da biblioteca como recurso de informação [...]” (Ibid, p. 5, tradução nossa). Nos painéis, de hora em hora, são apresentadas informações sobre os livros e as mídias devolvidos (figura 33), em números e em ordem cronológica, bem como a frequência de palavras-chave indexadas, conectando-as em relacionamentos de igualdade e hierarquia com outros itens do acervo (Ibid).

Figura 33: Painel da *Seattle Public Library* sobre empréstimos de obras.



Fonte: *Blog Future of the Book*, 2005.

A *Seattle Public Library* explorou o uso da *big data* não apenas para o tratamento das suas informações geradas, como também ser um meio de interagir com os usuários da Instituição, consolidando-se como uma espécie de campanha institucional ao proporcionar a divulgação dos serviços e produtos proporcionados, exemplificados pela circulação das obras do acervo, bem como a transparência dos resultados alcançados, neste caso, a partir de relatórios em tempo real. Desta forma, fazendo uma analogia aos aeroportos, as informações obtidas pela *Seattle Public Library* são apresentadas em painéis semelhantes aos quadros de informações dos voos, entre partidas e decolagens, os quais atualizam-se de “tempos em tempos” e veiculam mensagens aos transeuntes.

Por fim, a partir dos exemplos apresentados de aplicação da *big data*, especialmente com foco em bibliotecas, foi possível sugerir um conjunto de ações para o uso da referida tecnologia, elencado no quadro três (3). De forma ampla, as sugestões vão ao encontro da iniciativa da *Seattle Public Library*, ao viabilizar (recomendar), tanto a aproximação do usuário com os serviços disponibilizados pela biblioteca, quanto a divulgação da oferta de produtos e serviços, além de métricas alcançadas pela unidade de informação, como a quantidade de obras emprestadas em um certo momento, materiais mais solicitados etc., neste caso, por meio de painéis que possibilitam a visualização dessas informações, como um exemplo prático das características velocidade, veracidade, volume e valor. Em consequente, a *Harvard University Library* e a *Washington University Library* inspiram o setor da administração, especificamente no planejamento estratégico das bibliotecas e consequentes tomadas de decisão. A *Harvard University Library*, por meio da identificação de comportamentos organizacionais, do gerenciamento e da elaboração de relatórios dos

eventos da biblioteca, combina-se as características da variedade, valor e complexidade. Quanto a *Washington University Library*, deslumbram-se possibilidades de colaborações entre setores diversos com o intuito de analisar e apresentar o alto volume de informações de maneira personalizada, mais precisa, além de considerar a devida complexidade para criar relações/conexões entre as informações e, conseqüentemente, apoiar as cooperações setoriais e respectivas tomadas de decisão.

Deve-se destacar que ambas as Universidades, *Harvard Library* e *Washington Library*, compartilham do estímulo de uso do atributo volume com fins de realizar estudos de usuários. A identificação dos interesses, necessidades e comportamentos dos usuários, pelo setor de referência, por exemplo, serviria de informações ao setor de desenvolvimento de coleção para direcionar a tomada de decisões na constituição do acervo, utilizando-se das características de complexidade, que conectaria informações diversas sobre as obras, quanto aos formatos variados em que estão disponíveis, e do volume ao dispor de grandes quantidades de informações acerca do acervo, dos usuários, das possibilidades de busca e recuperação, por exemplo.

Quadro 3: Sugestão de aplicação da *big data* em bibliotecas.

CARACTERÍSTICA BÁSICA	SETOR DA BIBLIOTECA	AÇÃO
Variedade/Valor/Complexidade	Administração	Identificar comportamentos organizacionais para embasar e direcionar o planejamento estratégico da biblioteca, a partir dos resultados apresentados.
Velocidade/Valor	Administração	Gerenciar e relatar eventos de uso da biblioteca em tempo real, possibilitando tomadas de decisão com maior rapidez.
Volume/Complexidade	Administração	Permitir visão geral da situação da biblioteca, resultando no direcionamento às tomadas de decisão e visualização de novas oportunidades, como, por exemplo, possibilidades de novos projetos e colaborações entre setores.
Velocidade/Volume/Veracidade	Circulação	Aproximar e interagir com os usuários, através da exibição de informações em tempo real no momento do empréstimo e devolução de materiais, por exemplo. E apresentar de maneira transparente o fluxo (dos processos) de funcionamento da biblioteca e os resultados alcançados, tanto pela organização como um todo, como por setores específicos.
Complexidade/Volume/Variedade	Desenvolvimento de Coleção	Conectar informações acerca das obras, da relevância dos assuntos e do interesse sobre o comportamento dos usuários, aos quais subsidiam as decisões sobre a constituição do acervo.
Valor	Referência	Divulgar produtos e serviços ofertados pela biblioteca, além de apresentar os resultados alcançados, pela organização como um todo e/ou por seus setores, aos seus clientes, ou seja, aos usuários da biblioteca.
Volume	Referência	Realizar estudos de usuários para identificar interesses, necessidades e comportamentos, como forma de orientar, ajustar e desenvolver melhorias voltadas para o usuário nos processos e setores da biblioteca.

Fonte: Da autora, 2019.

Em suma, elencadas algumas aplicações que utilizam a tecnologia *big data*, acentuando suas características básicas, elencando exemplos de aplicação no contexto geral, pontuando exemplos de uso da *big data* em bibliotecas e exibindo um quadro sugestivo de aplicação da referida tecnologia no âmbito das bibliotecas, infere-se que a tecnologia *big data* é uma ferramenta de considerável utilidade e versatilidade, do poder público até empresas de *marketing*, do *e-commerce* até as bibliotecas. Dada a sua capacidade de adoção em ambientes diversificados, implica-se que o uso da *big data* se relaciona, diretamente, a problemas ligados ao alto volume de dados gerados e disseminados diariamente, como ocorre nos mercados financeiros, de bolsas de valores, ao lidarem com um fluxo crescente de transações entre empresas e países, valores de moedas distintas e em tempo real; e, conseqüentemente, a geração de dados e disponibilização de informações nas bibliotecas não seria diferente, obviamente considerando as devidas proporções. Desta forma, a partir da *big data* infere-se que o seu uso agregaria, tanto nos serviços ofertados, quanto nos procedimentos internos realizados, qualidade no atendimento aos seus usuários e agilidade na tomada de decisão pelas unidades de informação.

5 CONSIDERAÇÕES FINAIS

O presente trabalho considerou a pirâmide informacional (dado, informação, conhecimento e sabedoria) como o fundamento de que os processos cognitivos geram conhecimentos, que por sua vez produzem informações, que assumirão valor, ou não, e possibilitarão, ou não, a tomada de decisões estratégicas.

Os processos de decisão, em tempos atuais, estão cada vez mais atrelados à geração de conhecimento, formalizando uma espécie de relacionamento de produção entre homens e máquinas. Na terceira fase da *web*, por exemplo, a semântica possibilita o relacionamento por *links* entre objetos, criando conexões entre dados e, a partir dessas relações, geram-se informações personalizadas. Ao oferecer maior transparência, por meio da ampliação do alcance, da interconexão entre documentos e informações, e dos padrões e camadas que formam a *web*, a dinâmica dos ambientes e interações são alterados para um cenário caracterizado pela maior facilidade na recuperação de documentos e integração de formatos diversos. Em outras palavras, a *web* 3.0 objetiva tornar a *web* um ambiente legível para as máquinas, justificando a nomeação de *web* semântica. Neste cenário, marcado principalmente pela adoção do RDF e das URI, ambos assumem a responsabilidade de organizar a *web*, apoiado na conexão de elementos por *links*, criando relacionamentos entre eles.

A partir do desenvolvimento da *web* 3.0, em paralelo a evolução de áreas como a IA e ferramentas como o *machine learning* e a própria *big data*, objetivando facilitar o processo de integração de dados e a geração de novas informações, observa-se o alcance ao topo da pirâmide informacional, ou seja, a sabedoria contribuindo no processo de tomada de decisões.

No aspecto de aplicações da tecnologia *big data* foi possível identificar exemplos, nos mais diversos segmentos da sociedade, tais como: corporativos, com aplicação voltada à análise do fluxo e tráfego de clientes, experiência de compra personalizada e identificação de riscos em tempo real para tomadas de decisões em menor tempo; governamental, com iniciativas e possibilidades de alteração na estrutura política e geração de eficiência e otimização de recursos da administração pública; cotidiano social, em que destacam-se sistemas de recomendações personalizadas com base, principalmente, nos interesses pessoais e históricos dos

usuários; e, por último, no científico, relacionado à geração de grande quantidade de informações para embasar evidências científicas, validar hipóteses e formular teorias.

Ainda nas possibilidades do uso da *big data*, em setores diversos da sociedade, foi possível reunir sugestões de adoção da ferramenta nos serviços realizados e ofertados pelas bibliotecas, por meio de exemplos elencados, sendo: nos setores de gestão e de tomada de decisão das bibliotecas, ao possibilitar a identificação de “fragilidades” informacionais da organização, alinhar metas com recursos disponíveis e, tanto embasar, quanto traçar estratégias e objetivos para a biblioteca, como a promoção de cooperação entre setores; no mapeamento dos perfis e interesses dos usuários, o que permitiria captar tendências ao monitorar e acompanhar o uso do acervo, constituindo uma forma de estudo do usuário, e, a partir dos resultados, embasar decisões de novas aquisições ao acervo, bem como possibilitar um meio de integrar a biblioteca ao cotidiano do campus e dos usuários; por fim, na apresentação dos serviços e produtos ofertados pela biblioteca, como forma de engajar os usuários, assim como divulgar os resultados alcançados, tanto pela unidade de informação como um todo, quanto pelos setores, o que pode, conseqüentemente, estabelecer uma espécie de comunicação transparente entre usuários e organização.

Conclui-se, enquanto um reflexo da produção de dados no cotidiano social, que as características volume, velocidade e variedade (3V), encaradas como básicas da tecnologia *big data* e complementadas pelas propriedades complexidade, veracidade e valor, podem auxiliar na tomada de decisão em setores das bibliotecas, como na administração, na circulação, no desenvolvimento de coleções e na referência, em um cenário de produção e compartilhamento de informações promovidos pelas TIC, como a Internet, e em sentido exponencial.

Para estudos futuros, além do interesse pessoal em aprofundar-se sobre o assunto investigado em nível de pós-graduação, espera-se que a presente pesquisa seja um objeto de incentivo e direcionamento para temas voltados à tecnologia *big data* nas bibliotecas, para futuros graduandos, pesquisadores e entusiastas pelo tema. Pensar desta forma é considerar a *big data* em bibliotecas como uma ampliação no leque de possibilidades de pesquisas e práticas nos campos da Biblioteconomia e da CI, ainda tímido em relação ao que vem ocorrendo no *e-commerce*, quanto a análise de dados, por exemplo.

REFERÊNCIAS

AGHAEI, Sareh; NEMATBAKHSH, Mohammad Ali; FARSANI, Hadi Khosravi. Evolution of the World Wide Web: from web 1.0 to web 4.0. **International Journal of Web & Semantic Technology**, v. 3, n. 1, jan. 2012. Disponível em: <https://pdfs.semanticscholar.org/8cb3/93c3229e8f288febfa4dac12a0f6298efb93.pdf>. Acesso em: 25 set. 2018.

ALBRECHT, Karl. Um modelo de inteligência organizacional. **HSM Management**, v. 44, maio/jun. 2004. Disponível em: http://www.quemel.blog.br/curso/m02/a04/m02a04t03_inteligencia_organizacional.pdf. Acesso em: 11 out. 2018.

ALLEN, G. Donald. Hierarchy of knowledge – from data to wisdom. **International Journal of Current Research in Multidisciplinary**, v. 2, n. 1, p. 15-27, 2017. Disponível em: http://www.ijcrm.com/publish_article/edition_8/IJCRM_2123.pdf. Acesso em: 1 nov. 2018.

ALVIM, Luísa. Blogues e bibliotecas: construir redes na Web 2.0. **Cadernos Bad**, n. 1, p. 38-74, 2007. Disponível em: <https://www.bad.pt/publicacoes/index.php/cadernos/article/view/781/780>. Acesso em: 10 jan. 2019.

AMARAL, Sueli Angélica do. Marketing e gerência de biblioteca. **Revista de Biblioteconomia**, Brasília, v. 18, n. 2, p. 311-317, jul./dez. 1990. Disponível em: http://www.brapci.inf.br/_repositorio/2011/07/pdf_550341fd3b_0017716.pdf. Acesso em: 4 nov. 2018.

AMAZON. **Amazon**. 2019. Página do usuário logado. Disponível em: <https://www.amazon.com.br/>. Acesso em: 28 mar. 2019.

_____. **Notebook Dell inspirtation**. 2019. Apresenta informações do produto. Disponível em: https://www.amazon.com.br/Dell-Inspiron-i14-5480-A20S-14-Dom%C3%ADcilio/dp/B07Q2LPGBX/ref=sr_1_3?keywords=notebook+dell&qid=1558666358&s=gateway&sr=8-3. Acesso em: 25 abr. 2019.

ANDRADE, Maria Margarida de. **Introdução à metodologia do trabalho científico**. 10. ed. São Paulo: Atlas, 2010. 158 p.

ANGELONI, Maria Terezinha. Elementos intervenientes na tomada de decisão. **Ciência da Informação**, Brasília, v. 32, n. 1, p. 17-22, jan./abr. 2003. Disponível em: <http://www.scielo.br/pdf/ci/v32n1/15969.pdf>. Acesso em: 14 out. 2018.

AQUINO, Mirian de Albuquerque. O novo status da informação e do conhecimento na cultura digital. **Informação & Sociedade: Estudos**, João Pessoa, v. 18, n. 1, p. 79-100, jan./abr. 2008. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/ies/article/view/1555/1640>. Acesso em: 13 out. 2018.

ARAÚJO JÚNIOR, Rogério Henrique de; SOUZA, Renato Tarciso Barbosa de. Estudo do ecossistema de Big Data para conciliação das demandas de acesso, por meio da representação e organização da informação. **Ciência da Informação**, Brasília, v. 45, n. 3, p. 187-198, set./dez. 2016. Disponível em: <http://www.brapci.inf.br/index.php/res/v/18140>. Acesso em: 15 out. 2018.

ARAÚJO, Eliany Alvarenga; OLIVEIRA, Marlene de. A produção de conhecimento e a origem das bibliotecas. OLIVEIRA, Marlene de (Coord.). In: **Ciência da informação e biblioteconomia: novos conteúdos e espaços de atuação**. Belo Horizonte: UFMG, 2005. p. 29-44.

BACHELARD, Gaston. **A formação do espírito científico**. 3. ed. Rio de Janeiro: Contraponto, 2002. 316 p.

BANCO INTERAMERICANO DE DESAROLLO. **Evaluación**: una herramienta de gestión para mejorar el desempeño de los proyectos. New York: BID, 1997. 72 p. Disponível em: http://www.cedet.edu.ar/Archivos/Bibliotecas_Archivos/72BID-OVE%20-%20Marco%20L%C3%B3gico.pdf. Acesso em: 6 nov. 2018.

BARRETO, Aldo de Albuquerque. A condição da informação. **São Paulo em Perspectiva**, São Paulo, v. 16, n. 3, p. 67-74, 2002. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-88392002000300010. Acesso em: 20 out. 2018.

BEIJO, Helena Soterio; VALOTO, Ana Maria. O bibliotecário como mediador da informação, conhecimento e uso da tecnologia da informação em biblioteca universitária. **Ciências Aplicadas em Revista**, v. 14, n. 27, jul./dez. 2014, p. 220-233. Disponível em: <http://e-revista.unioeste.br/index.php/csaemrevista/article/view/15225>>. Acesso em: 18 julho 2019.

BERNERS-LEE, Timothy John. **Linked data**. jun. 2009. Disponível em: <https://www.w3.org/DesignIssues/LinkedData>. Acesso em: 14 jan. 2019.

BIGDATACORP. **Bigdatacorp**. 2019. Página do “Blog” de posts da categoria big data. Disponível em: <https://www.bigdatacorp.com.br/blog/index.php/category/big-data/>. Acesso em: 18 maio 2019.

BORDENAVE, Juan E. Diaz. **Além dos meios e mensagens**: introdução à comunicação como processo, tecnologia, sistema e ciência. 2. ed. Petrópolis, RJ: Vozes, 1984. 110 p.

BORGES, Deise Miranda; OLIVEIRA, Fernando Luiz de. **Análise e comparação dos sistemas de recomendação de produtos existentes em três empresas de comércio eletrônico (CE)**: Saraiva, Submarino e Amazon. [2010]. Disponível em: <http://sistemas-humano-computacionais.wdfiles.com/local--files/capitulo%3Aredes-sociais/Mir2010.pdf>. Acesso em: 20 maio 2019.

BOTTENTUIT JUNIOR, João Batista; COUTINHO, Clara Maria Pereira. As ferramentas da web 2.0 no apoio à tutoria na formação em e-learning. In: COLÓQUIO DA AFIRSE, 16, 2008, Lisboa, Portugal. **Anais...** Lisboa: AFIRSE, 2008. Disponível em: <http://hdl.handle.net/1822/7767>. Acesso em: 24 maio 2019.

BRIGGS, Asa; BURKE, Peter. **Uma história social da mídia**: de Gutenberg à Internet. 3. ed. Rio de Janeiro: Zahar, 2004. 432 p.

CANDIOTTO, Cesar; BASTOS, Cleverson Leite; CANDIOTTO, Kleber B. B. **Fundamentos da pesquisa científica**: teoria e prática. Petrópolis, RJ: Vozes, 2011. 166 p.

CAÑEDO ANDALIA, Rúben. Aproximaciones para una historia de Internet. **ACIMED**, Ciudad de La Habana, v. 12, n. 1, feb. 2004. Disponível em: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352004000100005#cargo. Acesso em: 29 nov. 2018.

CARVALHO, Gracilene Maria de; LIMA, Gracirlei Maria de Carvalho. Mediação da informação e os repositórios institucionais: otimização do acesso à informação. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 27., 2017, Fortaleza. **Anais...** São Paulo: FEBAB, 2017. Disponível em: <https://portal.febab.org.br/anais/article/view/1715/1716>. Acesso em: 18 jul. 2019.

CASTELLS, Manuel. **A galáxia da internet**: reflexões sobre a internet, os negócios e a sociedade. Rio de Janeiro: Zahar, 2003. 243 p.

CHARAUDEAU, Patrick. **Discurso das mídias**. 2. ed. São Paulo: Contexto, 2018. 285 p.

CHAVEZ, Chris. **Google Now adds real time incident reports to help users avoid traffic before their commute**. 2014. Disponível em: <https://phandroid.com/2014/03/24/google-now-adds-real-time-incidents-reports-to-help-users-avoid-traffic-before-their-commute/>. Acesso em: 10 abr. 2019.

CHEDE, Cezar. **Um novo paradigma**: stream computing e o InfoSphere Streams da IBM. 2009. Disponível em: https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/um_novo_paradigma_stream_computing_e_o_infosphere_streams_da_ibm2?lang=en. Acesso em: 14 jan. 2019.

CHEN, Chun-I Philip; ZANG, Chun-Yang. Data-intensive applications, challenges, techniques and technologies: a survey on big data. **Information Sciences**, v. 275, n. 10, p. 314-347, aug. 2014. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0020025514000346>. Acesso em: 17 jan. 2019.

CHEN, Hsin-liang et al. Library assessment and data analytics in the big data era: practice and policies. **Proc. Assoc. Info. Sci. Tech.**, v. 52, n. 1, p. 1-4, 2016. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/pras.2015.14505201002>. Acesso em: 21 jan. 2019.

CHOUSHURY, Nupur. World Wide Web and its journey from Web 1.0 to Web 4.0. **International Journal of Computer Science and Information Technologies**, v. 5, n. 6, 2014. Disponível em: <http://ijcsit.com/docs/Volume%205/vol5issue06/ijcsit20140506265.pdf>. Acesso em: 20 out. 2018.

CONEGLIAN, Caio Saraiva et al. O papel da web semântica nos processos do big data. **Encontros Bibli**, v. 23, n. 53, p. 137-146, set./dez. 2018. Disponível em: <http://www.brapci.inf.br/index.php/res/download/52521>. Acesso em: 15 jan. 2019.

CONEGLIAN, Caio Saraiva; GONÇALVEZ, Paula Regina Ventura Amorim; SANTARÉM SEGUNDO, José Eduardo. O profissional da informação na era do Big Data. **Encontros Bibli**, Santa Catarina, v. 22, n. 50, p. 128-143, set./dez. 2017. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n50p128>. Acesso em: 6 ago. 2018.

COUTINHO, Clara Pereira; BOTTENTUIT JUNIOR, João Batista. Blog e wiki: os futuros professores e as ferramentas da web 2.0. In: SIMPÓSIO INTERNACIONAL DE INFORMÁTICA EDUCATIVA, 9, 2007, Porto, Portugal. **Anais...** Porto: ESSE-IPP, 2007. p. 199-204. Disponível em: <https://repositorium.sdum.uminho.pt/bitstream/1822/7358/1/Com%20SIIIE.pdf>. Acesso em: 11 jan. 2019.

DA SILVA, Márcio Bezerra; DIAS, Luana Gomes; CASTRO, Hallison Phelipe Lopes de. A presença de recursos da web 2.0 nos sistemas de automação de bibliotecas: uma sugestão de interface. In: ASOCIACIÓN DE EDUCACIÓN E INVESTIGACIÓN EN CIENCIA DE LA INFORMACIÓN DE IBEROAMÉRICA Y EL CARIBE, 10, 2016, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2016. Disponível em: https://www.researchgate.net/publication/328411050_A_presenca_de_recursos_da_web_20_nos_sistemas_de_automacao_de_bibliotecas_uma_sugestao_de_interface. Acesso em: 5 dez. 2018.

_____; RUFINO, Fernanda Maciel. A web 2.0 na informatização de bibliotecas: um estudo propositivo. **PontodeAcesso**, Salvador, v. 10, n. 2, p. 17-38, ago. 2016. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/14447/11534>. Acesso em: 29 maio 2019.

_____. **Estudo teórico-analítico sobre o uso de facetas na organização da informação e na estruturação de ambientes digitais**. 2018. Tese (Doutorado em Ciência da Informação) – Instituto de Ciência da Informação, Universidade Federal da Bahia, Salvador, 2018. Disponível em: <https://repositorio.ufba.br/ri/handle/ri/28130>. Acesso em: 20 mar. 2019.

DAVENPORT, Thomas. H.; PRUSAK, Laurence. **Working knowledge**: how organizations manage what they know. Boston, MA: Harvard Business School Press, 2000.

DEMO, Pedro. **Metodologia científica em ciências sociais**. São Paulo: Atlas, 1981.

DOMO. **Data never sleeps 6.0**. 2018. Disponível em: <https://www.domo.com/learn/data-never-sleeps-6>. Acesso em: 18 abr. 2019.

DRAGO, Idilio; VIEIRA, Alex Borges; SILVA, Ana Paula Couto da. Caracterização dos arquivos armazenados no Dropbox. In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS, 31, 2013, Brasília. **Anais...** SBC, 2013, p. 109-114. Disponível em: <http://sbrc2013.unb.br/files/anais/wp2p/artigos/artigo-11.pdf>. Acesso em: 20 maio 2019.

DROPBOX. **Dropbox basic**. 2019. Apresentação do serviço. Disponível em: https://www.dropbox.com/pt_BR/basic. Acesso em: maio 2019

DRUCKER, Peter Ferdinand. **As fronteiras da administração**: onde as decisões do amanhã estão sendo determinadas hoje. São Paulo: Pioneira, 1989.
Estados Unidos: IEEE, 2013. Disponível em: <https://ieeexplore.ieee.org/document/6612229>. Acesso em: 15 jan. 2019.

FACEBOOK. **UnB – Universidade de Brasília**. 2019. Página do perfil da universidade. Disponível em: https://www.facebook.com/pg/oficialUnB/posts/?ref=page_internal. Acesso em: abr. 2019.

FACHIN, Odília. **Fundamentos de metodologia**. 3. ed. São Paulo: Saraiva, 2001. 216 p.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, n. 3, p. 37-54, 1996. Disponível em: <https://www.aaai.org/ojs/index.php/aimagazine/article/viewFile/1230/1131>. Acesso em: 8 out. 2018.

FICO FALCON. **FICO Falcon fraud manager** - customer success stories. 2016. Disponível em: <https://www.slideshare.net/FICO/fico-falcon-fraud-manager-customer-success-stories>. Acesso em: 1 maio 2019.

FIGUEIREDO, Nice Menezes de. Evolução e avaliação do serviço de referência. **Revista de Biblioteconomia de Brasília**, Brasília, v. 2, n. 2, p. 175-198, jul./dez. 1974. Disponível em: <http://www.brapci.inf.br/index.php/res/v/74833>. Acesso em: 17 out. 2018.

_____. **Paradigmas modernos da Ciência da Informação**. São Paulo: Polis, 1999.

FURGERI, Sérgio. **Representação de informação e conhecimento**: estudo das diferentes abordagens entre a Ciência da Informação e a Ciência da Computação. 2006. 161 p. Dissertação (Mestrado em Ciência da Informação) - Pontifícia Universidade Católica de Campinas, Campinas, 2006. Disponível em: http://bdtd.ibict.br/vufind/Record/PCAM_ac6b10abd6bab652844d870f0eb22d01. Acesso em: 30 out. 2018.

FURTADO, José Afonso. Bibliotecas na era digital. **Revista de Biblioteconomia de Brasília**, Brasília, v. 22, n. 1, p. 3-17, 1998. Disponível em: <http://www.brapci.inf.br/index.php/res/download/90368>. Acesso em: 17 jan. 2019.

FUTURE OF THE BOOK. **Making visible the invisible**: George Legrady installation at Seattle Central Library. 2005. Disponível em: http://futureofthebook.org/blog/2005/09/16/making_visible_the_invisible_g/. Acesso em: 19 abr. 2019.

GARCÍA-MARCO, Francisco-Javier. El reto de la transparencia y la web 3.0: um nuevo contexto para las publicaciones oficiales. In: ROVER, Aires José; GALINDO, Fernando (Eds.). **O governo eletrônico e suas múltiplas facetas**. Zaragoza: Pressas Universitarias de Zaragoza, 2010. p. 241-264. Disponível em: http://www.egov.ufsc.br/portal/sites/default/files/lefis_10.pdf. Acesso em: 10 jan. 2019.

GEISLER, Eliezer. **The metrics of Science and Technology**. Connecticut: Quorum, 2000. 400 p.

GENDARMERIE NATIONALE. **Gendarmerie nationale**: Ministère de L'Interieur. [201-?]. Disponível em: <https://www.gendarmerie.interieur.gouv.fr/>. Acesso em: 25 maio 2019.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (Orgs). **Métodos de Pesquisa**. Porto Alegre: UFRGS, 2009. 120 p. (Série Educação a Distância).

GIORDAN, Marcelo. Ferramentas de busca na Web. **Química Nova na Escola**, n. 7, p. 15-16, maio 1998. Disponível em: <http://qnesc.sbq.org.br/online/qnesc07/eqm.pdf>. Acesso em: 20 maio 2019.

GOOGLE. **Google Now**. [201-]. Disponível em: <https://www.google.com/intl/pt-BR/landing/now/#gnow-cards-manage>. Acesso em: 18 maio 2019.

GOOGLE DRIVE. **Google Drive**. [201-?]. Apresenta informações acerca do serviço. Disponível em: https://www.google.com/intl/pt-BR_ALL/drive/. Acesso em: 7 abr. 2019.

HARVARD LIBRARY LAB. **Library Analytics Toolkit**: Library Lab grant proposal. [2011]. Disponível em: <https://osc.hul.harvard.edu/liblab/projects/library-analytics-toolkit>. Acesso em: 20 jan. 2019.

_____. **Library Analytics Toolkit**. [2014]. Disponível em: <https://osc.hul.harvard.edu/liblab/projects/library-analytics-toolkit>. Acesso em: 20 abr. 2019.

HUGHES, David John et al. A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. **Computers in Human Behavior**, v. 28, n. 2, 2012, p. 561-569. Disponível em: <https://doi.org/10.1016/j.chb.2011.11.001>. Acesso em: 22 maio 2019.

IKEMATU, Ricardo Shoiti. Gestão de metadados: sua evolução na Tecnologia da Informação. **DataGramZero**, v. 2, n. 6, dez. 2001. Disponível em: <http://www.brapci.inf.br/index.php/article/view/0000001249/9137060e5be654b068648fda011ee021>. Acesso em: 16 out. 2018.

INTERNATIONAL DATA CORPORATION. **The digital universe in 2020**: big data, bigger digital shadows, and biggest growth in the far East. Massachusetts: EMC, dec. 2012. Disponível em: <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>. Acesso em: 17 maio 2019.

IZQUIERDO, Yenier Torres. **Keyword search over federated RDF graphs by exploring their schemas**. 2017. 66 p. Dissertação (Mestrado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, Rio de Janeiro, 2017. Disponível em: <https://www.maxwell.vrac.puc-rio.br/30739/30739.PDF>. Acesso em: 20 abr. 2019.

JAPIASSU, Hilton. **A pedagogia da incerteza e outros estudos**. Rio de Janeiro: Imago, 1983. 171 p.

KATAL, Avita; WAZID, Mohammad; GOUDAR, R. H. Big data: issues, challenges, tools and good practices. In: INTERNATIONAL CONFERENCE ON CONTEMPORARY COMPUTING, 6., 2013, Noida, India. **Anais eletrônicos ... IEEE**, 2013, p. 404-409. Disponível em: https://www.researchgate.net/publication/261454906_Big_data_Issues_challenges_tools_and_Good_practices. Acesso em: 17 dez. 2018.

KEEN, Andrew. **O culto do amador**: como blogs, MySpace, YouTube e a pirataria digital estão destruindo nossa economia, cultura e valores. Rio de Janeiro: Zahar, 2009. 207 p.

KERLINGER, Fred Nichols. **Foundations of behavioral research**. New York: Holt, Rinehart & Winston, 1973.

KLAPWIJK, Wouter. **The library (big) data scientist**. 2016. 35 slides. Material apresentado durante o evento IFLA/ALA webinar. Disponível em: <https://npsig.files.wordpress.com/2016/04/bd-sig-wouter-klapwijk.pdf>. Acesso em: 19 jan. 2019.

KÖCHE, José Carlos. **Fundamentos de metodologia científica: teoria da ciência e prática da pesquisa**. Petrópolis, RJ: Vozes, 2001. 180 p.

KOIVUNEN, Marja-Riita; MILLER, Eric. W3C semantic web activity. In: HYVÖNEN, Eero (Ed.). SEMANTIC WEB KICK-OFF IN FINLAND, 2001, Helsinki, Finland. **Anais...** Helsinki: HIIT, 2002. chap. 2, p. 27-43. Disponível em: <https://www.cs.helsinki.fi/u/eahyvone/stes/semanticweb/kick-off/proceedings.pdf>. Acesso em: 12 jan. 2019.

LAGO, Gessyca da Silva. **Computação em nuvem: estudo sobre o uso e implementação na visão dos bibliotecários da BCE/UnB**. 2014. 105 f. Monografia (Graduação em Biblioteconomia) - Universidade de Brasília, Faculdade de Ciência da Informação, Brasília, 2014. Disponível em: <http://bdm.unb.br/handle/10483/9062>. Acesso em: 20 maio 2019.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos de metodologia científica**. 8. ed. São Paulo: Atlas, 2017a. 337 p.

_____. **Metodologia científica**. 7. ed. São Paulo: Atlas, 2017b. 360 p.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. Tradução de Antonio Agenor Briquet de Lemos. Brasília, DF: Briquet de Lemos/Livros, 2004.

LANDSET, Sara *et al.* A survey of open source tools for machine learning with big data in the Hadoop ecosystem. **Journal of Big Data**, v. 2, n. 24, 2015. Disponível em: <https://journalofbigdata.springeropen.com/track/pdf/10.1186/s40537-015-0032-1>. Acesso em: 12 jan. 2019

LE COADIC, Yves-François. **A ciência da informação**. 2. Ed. Brasília: Brinquet de Lemos, 2004. 124 p.

LIMA, Júnior César de; CARVALHO, Cedric Luiz de. **Uma visão da web semântica**. Goiás: UFG, 2004. 15 p. Disponível em: http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-04.pdf. Acesso em: 12 jan. 2019.

LARGE SYNOPTIC SURVEY TELESCOPE. **About LSST**. [20-?]a. Disponível em: <https://www.lsst.org/about>. Acesso em: 31 maio 2019.

_____. **Design of the LSST câmera**. 2011. Disponível em: <https://www.lsst.org/gallery/design-lsst-camera>. Acesso em: 31 maio 2019.

_____. **LSST Science goals**. [20-?]b. Disponível em: <https://www.lsst.org/science>. Acesso em: 31 maio 2019.

LUCAS, Clarinda Rodrigues. O conceito de biblioteca nas bibliotecas digitais. **Informação & Sociedade**, v. 14, n. 2, p. 15-32, jul./dez. 2004. Disponível em: <http://www.periodicos.ufpb.br/ojs/index.php/ies/article/view/58/1530>. Acesso 17 jan. 2019.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de metodologia científica**. 7. ed. São Paulo: Atlas, 2010. 297 p.

MARTELETO, Regina Maria. Informação, rede e redes sociais – fundamentos e transversalidades. **Informação & Informação**, Londrina, v. 12, n. esp., 2007. Disponível em: http://www.brapci.inf.br/_repositorio/2010/07/pdf_691c714087_0011336.pdf. Acesso em: 15 out. 2018.

MARTIN, Brett. **Big data and machine learning**. Chicago: Norwood House Press, [2018]. 47 p.

MARTINS, António. Da origem do conhecimento à inovação organizacional. **Economia Global e Gestão**, Portugal, v. 5, n. 1-2, p. 9-27, 2000.

MASS, Bruno. **Cinco exemplos práticos que provam que já vivemos no mundo do big data**. 2016. Disponível em: <https://computerworld.com.br/2016/09/26/cinco-exemplos-praticos-que-provam-que-ja-vivemos-no-mundo-do-big-data>. Acesso em: 27 fev. 2019.

MIRANDA, Guilhermina Lobato. Limites e possibilidades das TIC na educação. **Sísifo: Revista de Ciências da Educação**, v. 3, maio/ago. 2007. Disponível em: <http://ticsproeja.pbworks.com/f/limites+e+possibilidades.pdf>. Acesso em: 3 out. 2018.

MIRANDA, Roberto Campos da Rocha. O uso da informação na formulação de ações estratégicas pelas empresas. **Ciência da Informação**, Brasília, v. 28, n. 3, p. 286-292, set./dez. 1999. Disponível em: http://www.scielo.br/scielo.php?pid=s0100-19651999000300006&script=sci_abstract&tlng=pt. Acesso em: 13 out. 2018.

MOREIRA, Fábio Mosso; SANT'ANA, Ricardo César Gonçalves; JORENTE, Maria José Vicentini. A complexidade da disponibilização e acesso a dados governamentais na Web. **Perspectivas em Ciência da Informação**, v. 21, n. 3, p. 70-88, jul./set. 2016. Disponível em: <http://dx.doi.org/10.1590/1981-5344/2540>. Acesso em: 17 out. 2018.

MORIN, Edgar. **A cabeça bem feita: repensar a reforma, reformar o pensamento**. Rio de Janeiro: Bertrand do Brasil, 2000.

MUELLER, Susana Pinheiro Machado. Bibliotecas e sociedade: evolução da interpretação de função e papéis da biblioteca. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 13, n. 1, mar. 1984. p. 7-50.

NATIONAL SCIENCE FOUNDATION. **Large Synoptic Survey Telescope headed to the final design stage**. 2012. Disponível em: <https://scitechdaily.com/large-synoptic-survey-telescope-headed-to-the-final-design-stage/>. Acesso em: 10 abr. 2019.

NETFLIX. **Netflix**. 2019. Página inicial do usuário logado. Disponível em: <https://www.netflix.com/browse>. Acesso em: 15 abr. 2019.

NONAKA, Ikujiro; TAKEUCHI, Hirotaka. **Criação do conhecimento na empresa: como as empresas geram a dinâmica da inovação**. Rio de Janeiro: Campus, 1997.

NONATO, Marcos Daniel de Arruda. **World Wide Web**: www. Tefé, AM: DanG3, 2015. 26 p. Disponível em: <https://livrosdigitais.org.br/baixar-livro/1447A68L1BQ3T>. Acesso em: 4 nov. 2018.

OLIVEIRA, Silvio Luiz de. **Tratado de metodologia científica: projetos de pesquisas, TGI, TCC, monografias, dissertações e teses**. São Paulo: Pioneira, 2001. 320p.

PARDO KUNKLINSKI, Hugo. Nociónes básicas alrededor de la web 2.0. In: COBO ROMANÍ, Cristobal; PARDO KUNKLINSKI, Hugo (Org.). **Planeta web 2.0**: inteligencia colectiva o medios *fast food*. Cataluña, España: Universitat de Vic, 2007. cap. 1, p. 27-42.

PESTANA, Olívia. Elementos para uma avaliação de fontes de informação na Internet. **Páginas a&b**, n. 6, 2001, p. 41-50. Disponível em: <http://ojs.letras.up.pt/index.php/paginasueb/article/view/132>. Acesso em: 7 jan. 2019.

POZZEBOM, Rafaela. **Diferença entre busca vertical e busca horizontal**. 2011. Disponível em: https://www.oficinadanet.com.br/artigo/otimizacao_seo/diferenca_entre_busca_vertical_e_busca_horizontal. Acesso em: 20 maio 2019.

PINTO, Maria Martha D'Angelo. Escrita, memória e educação: dos hieróglifos aos textos eletrônicos. In: CONGRESSO GALAICO PORTUGUÊS DE PSICOPEDAGOGIA, 7, 2005. **Anais...** Braga, Portugal: Universidade do Minho, 2005. p. 1887-1896.

PRIMO, Alex. O aspecto relacional das interações na Web 2.0. **E-Compós**, Brasília, v. 9, ago. 2007. Disponível em: <http://www.e-compos.org.br/e-compos/article/view/153/154>. Acesso em: 7 ago. 2018.

PROATIVA. **Big data e o dilema dos 4Vs**. 2015. Disponível em: <http://proativasoft.com.br/noticias/o-big-data-e-o-dilema-dos-4vs/>. Acesso em: 20 mar. 2019.

RAMALHO, Rogério Aparecido Sá; VIDOTTI, Silvana Aparecida Borsetti Gregorio; FUJITA, Mariângela Spotti Lopes. Web semântica: aspectos interdisciplinares para a organização e a recuperação de informações. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 6., 2005, Florianópolis, SC. **Anais...** Florianópolis: ENANCIB, 2005. Disponível em: http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/369/GT2_Ramalho_Vidotti_Fujita.pdf?sequence=1. Acesso em: 19 maio 2019.

RECUERO, Raquel da Cunha. Redes sociais na Internet: considerações iniciais. **E-Compós**, Brasília, v. 2, abr. 2005. Disponível em: <http://e-compos.org.br/e-compos/article/view/28/2>. Acesso em: 21 maio 2019.

RICCIULLI, Stefania Ludescher Souza. Data is the new black: o papel da Netflix na web. In: CHIACHIRI, Roberto; PERSICHETTI, Simonetta (Orgs.). **Imagem e inserção social II**. São Paulo: UNI, 2016. p. 51-70. Disponível em: <https://casperlibero.edu.br/wp-content/uploads/2014/07/INSER%C3%87%C3%83O-SOCIAL-II.pdf>. Acesso em: 2 mar. 2019.

RODRIGUES, Agostinha Maria. **O uso do tesouro na arquitetura da informação em Websites**. 2014. 156 f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2014.

ROSA, Anntônio Machuco. A filosofia do código-fonte aberto: de Richard Stallman a LINUX. **Interact**, v. 10, p. 1-12, fev. 2004. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/74557/2/73539.pdf>. Acesso em: 29 nov. 2018.

ROZA, Rodrigo Hipólito. Revolução informacional e os avanços tecnológicos da informática e das telecomunicações. **Ciência da Informação em Revista**, Maceió, v. 4, n. 3, p. 3-11, set./dez. 2017. Disponível em: <http://www.seer.ufal.br/index.php/cir/article/view/3482>. Acesso em: 8 ago. 2018.

ROZADOS, Helen Beatriz Frota. **Indicadores como ferramenta para gestão de serviços de informação tecnológica**. 2004. 239 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004. Disponível em: <https://lume.ufrgs.br/handle/10183/5668>. Acesso em: 16 out. 2018.

SANTOS, Carla Marques do; ASSUNÇÃO, Suelene Santana. Biblioteca digital: uma evolução da biblioteca convencional. In: ENCONTRO REGIONAL DE ESTUDANTES DE BIBLIOTECONOMIA, DOCUMENTAÇÃO, CIÊNCIA E GESTÃO DA INFORMAÇÃO, 15., 2012, Juazeiro do Norte. **Anais ...** Juazeiro do Norte: Universidade Federal do Cariri, 2012. Disponível em: <http://www.brapci.inf.br/index.php/res/download/98806>. Acesso em: 15 jan. 2019.

SANTOS, Plácida Leopoldina Ventura Amorim da Costa; SANT'ANA, Ricardo César Gonçalves. Transferência da informação: análise para valoração de unidades de conhecimento. **DataGramZero**, v. 3, n. 2, abr. 2002. Disponível em: <http://www.brapci.inf.br/index.php/article/view/0000001259>. Acesso em: 15 out. 2018.

SAP SE. **GN**: processing data quickly for effective action with SAP HANA. 2017. Disponível em: <https://www.sap.com/documents/2017/07/d8b227c3-c67c-0010-82c7-ed71af511fa.html>. Acesso em: 22 maio 2019.

SETZER, Valdemar. Dado, informação, conhecimento e competência.

DataGramaZero, v. 0, n. 0, dez. 99. Disponível em:

<http://www.brapci.inf.br/index.php/article/view/0000009680/a006e804d12f9c436a7357686a4c923b>. Acesso em: 15 out. 2018.

SEVERINO, Antônio Joaquim. **Metodologia do trabalho científico**. 22. Ed. São Paulo: Cortez, 2002. 335 p.

SHANNON, Claude; WEAVER, Warren. **Teoria matemática da comunicação**. 11. ed. São Paulo: Difel, 1975.

SILVA, Edna Lúcia da; MENEZES, Estera Muszkat. **Metodologia da pesquisa e elaboração de dissertação**. 3. ed. Florianópolis: Laboratório de Ensino a Distância da UFSC, 2001. 121 p.

SILVA, Elvis Magno da; DIAS, Louise Maria Aninceto. O Balanced Scorecard como ferramenta estratégica para a tomada de decisões na empresa Microboard. In: SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA, 8, 2011, Resende, Rio de Janeiro. **Anais...** Rio de Janeiro: AEDB, 2011, Disponível em: <http://adm.aedb.br/seget/artigos11/811465.pdf>. Acesso em: 30 abr. 2019.

SILVA, Terezinha Elizabeth da. Bibliotecas: metáforas da memória. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. 21, jan./jun. 2006. p. 85-94. Disponível em:

<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2006v11n21p85>. Acesso em 16 out. 2018.

SMITH, Thomas. Pesquisa científica: um episódio no desenvolvimento da tecnologia da aviação. In: KITTS, David B. (Org.). **Iniciação à história da ciência**. São Paulo: Cultrix, 1966. p.149-160.

SQUEFF, Flávia de Holanda Schmidt; NEGRI, Fernanda de. Ciência e tecnologia de impacto: uma análise do caso DASPA. In: RAUEN, André Tortato (Org.). **Políticas de inovação pelo lado da demanda no Brasil**. Brasília: Ipea, 2017. cap. 10. Disponível em:

http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=30404. Acesso em: 30 out. 2018.

SZELBRACIKOWSKI, Daniel Corrêa. **Novas tecnologias e a necessidade de reforma tributária**. Consultor Jurídico, jan. 2017. Disponível em:

http://www.dsa.com.br/dsa/media/ckeditor/uploads/2016/03/19/1083111_1.pdf. Acesso em: 20 maio 2019.

TAKAHASHI, Tadao. **Sociedade da informação no Brasil**: livro verde. Brasília: Ministério da Ciência e Tecnologia, 2000. 203 p.

TECHTUDO. **Fórum Techtudo**. 2019. Apresenta as perguntas do fórum. Disponível em: <https://forum.techtudo.com.br/perguntas/>. Acesso em: 17 maio 2019.

_____. **Relembre sites que marcaram os anos 1990 e 2000**. 2015. Disponível em: <https://www.techtudo.com.br/listas/noticia/2015/11/relembre-sites-que-marcaram-os-anos-90-e-2000.html>. Acesso em: 16 maio 2019.

TECHRADAR. **Netflix review**. 2019. Disponível em: <https://www.techradar.com/reviews/audio-visual/av-accessories/netflix-1065801/review>. Acesso em: 20 maio 2019.

TECMUNDO. **17 maneiras como a Siri pode te ajudar no dia a dia**. 2017. Disponível em: <https://www.tecmundo.com.br/siri/118398-17-maneiras-siri-te-ajudar-dia-dia.htm>. Acesso em: 27 abr. 2019.

TOMÁÉL, Maria Inês *et al.* Avaliação de fontes de informação na internet: critérios de qualidade. **Informação & Sociedade**, v. 11, n. 2, p. 13-35, 2001. Disponível em: <http://www.brapci.inf.br/index.php/article/view/0000001061/a9f7ed402ee5bd1ff45ead513a74e0cb/>. Acesso em: 6 jan. 2018.

_____; ALCARÁ, Adriana Rosecler; DI CHIARA, Ivone Guerreiro. Das redes sociais à inovação. **Ciência da Informação**, Brasília, v. 34, n. 2, p. 93-104, maio/ago. 2005. Disponível em: <http://www.brapci.inf.br/index.php/res/v/21277>. Acesso em: 13 out. 2018.

TORRES, Marcos Daniel Marado. **What is web 3.0?** ago. 2007. Disponível em: <http://mindboosternoori.blogspot.com/2007/08/what-is-web-30.html>. Acesso em: 13 jan. 2019.

TWITTER. **UnB**. 2019. Página do perfil da universidade. Disponível em: https://twitter.com/unb_oficial. Acesso: 9 maio 2019.

UNBTV. **UnBTV**. 2019. Página do canal da universidade. Disponível em: <https://www.youtube.com/channel/UCUUU5zBnxg55vaXymuFJDPw>. Acesso em: 14 maio 2019.

URDANETA, Iraset Páez. **Gestión de la inteligência. Aprendizaje tecnológico y modernización del trabajo informacional**: retos y oportunidades. Caracas, Venezuela: Universidad Simon Bolivar, 1992.

VYGOTSKY, Lev Semyonovich. **A formação social da mente**. São Paulo: Martins Fontes, 1998. 224 p.

W3SCHOOLS. **XML Elements**. [201-?]a. Disponível em: https://www.w3schools.com/xml/xml_elements.asp. Acesso em: 18 maio 2019.

_____. **XML Tutorial**. [201-?]b. Disponível em: https://www.w3schools.com/xml/xml_namespaces.asp. Acesso em: 27 abr. 2019.

_____. **XSD Schema Tutorial**. [201-?]c. Disponível em: https://www.w3schools.com/xml/schema_intro.asp. Acesso em: 27 abr. 2019.

WALMART. **5 ways WalMart uses big data**. 2017. Disponível em: <https://blog.walmart.com/innovation/20170807/5-ways-walmart-uses-big-data-to-help-customers>. Acesso em: 1 maio 2019.

WASHINGTON UNIVERSITY LIBRARIES. **Off the shelf**. Washington, fall 2017. 26 p. Disponível em: https://library.wustl.edu/wp-content/uploads/2017/12/120517_OTIS_F17.pdf. Acesso em: 15 abr. 2019.

_____. **Annual report 2017/2018**. Washington, 2018. 14 p. Disponível em: https://library.wustl.edu/wp-content/uploads/2018/12/Libraries_2017_18_AR_FINAL.pdf. Acesso em: 15 abr. 2019.

WAYBACK MACHINE. **Hotel pousada dos Pirineus**. 2014. Página “Apartamentos” em 2002. Disponível em: <https://web.archive.org/web/20020728234143/http://www.pousadapirineus.com.br/800/hotel.html>. Acesso em: 16 maio 2019

_____. **Universidade de Brasília**. 2014. Página inicial em 1997 Disponível em: <https://web.archive.org/web/19970524095255/https://www.unb.br/>. Acesso em: 15 maio 2019.

WOLFRAMALPHA. **Brazil vs Argentina**. 2018. Apresenta resultado da pesquisa. Disponível em: <https://www.wolframalpha.com/input/?i=brazil+vs+argentina>. Acesso: 5 abr. 2019.

WORLD WIDE WEB CONSORTIUM. **Character encodings**: essential concepts. 2018. Disponível em: <https://www.w3.org/International/articles/definitions-characters/>. Acesso em: 28 abr. 2019.

_____. **W3C Semantic Web activity**. 2001. Disponível em: <https://www.w3.org/2001/12/semweb-fin/w3csw>. Acesso em: 2 abr. 2019.

YOUTUBE. **Youtube**. 2019. Página “Sobre”. Disponível em: <https://www.youtube.com/intl/pt-BR/yt/about/>. Acesso em: 25 maio 2019.

ZIKOPOULOS, Paul *et al.* **Understanding big data**: analytics for enterprise class hadoop and streaming data. New York: McGraw Hill, 2015. 141 p.