



EIXO TEMÁTICO:
Organização e Representação da Informação e do Conhecimento

TRANSFORMAÇÃO DIGITAL: UM ENSAIO SOBRE BIG DATA, ANÁLISE DE DADOS E INTELIGÊNCIA ARTIFICIAL

DIGITAL TRANSFORMATION: AN ESSAY ON BIG DATA, DATA ANALYSIS AND ARTIFICIAL INTELLIGENCE

Raíssa Yuri Hamanaka raissa0201@gmail.com
Doutoranda pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina (PPGCI/UEL).

Keli Rodrigues do Amaral Benin keliamaral@hotmail.com
Mestranda pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual de Londrina (PPGCI/UEL).

Francisco Carlos Paletta fcpaletta@usp.br
Doutor em Ciências pela Universidade de São Paulo (USP). Professor do Programa de Pós-Graduação em Ciência da Informação da Universidade de São Paulo (PPGCI/USP) e Universidade Estadual de Londrina (PPGCI/UEL).

Resumo: a era da informação se caracteriza pelas transformações digitais, pelo ambiente de inovações constantes, pelo fenômeno do *Big Data* e pelo enfoque na produção, compartilhamento, curadoria, preservação e uso de dados. O ensaio reflete de forma teórica sobre as relações entre a representação da informação no contexto do *Big Data*, análise de dados e das aplicações de inteligência artificial. A metodologia adotada é exploratório-bibliográfica. Aponta os desafios na gestão de dados em grandes volumes e variados formatos, e como as ferramentas de web semântica e aplicações da inteligência artificial podem contribuir para a estruturação, representação, processamento, visualização e análise dos dados.

Palavras-chave: *Big Data*. Análise de dados. Inteligência artificial. Representação da informação.

Abstract: the information age is characterized by digital transformations, the environment of constant innovation, the Big Data phenomenon and the focus on production, sharing, curation, preservation and use of data. The essay reflects theoretically on the relationships between the representation of information in the context of Big Data, data analysis and artificial intelligence applications. The methodology adopted is exploratory-bibliographic. It points out the challenges in managing data in large volumes and varied formats, and how semantic web tools and artificial intelligence applications can contribute to the structuring, representation, processing, visualization, and analysis of data.

Keywords: Big Data. Data analysis. Artificial intelligence. Information Representation.

1 INTRODUÇÃO

A transformação digital ou indústria 5.0 pode ser caracterizada pelo ambiente de mudanças contínuas e inovações tecnológicas. Aborda temáticas como assistentes virtuais, aprendizado de máquina, *Robotic Process Automation* e trabalhadores digitais (MELLO FILHO; ARAÚJO JÚNIOR, 2021). A era da informação impõe complexidades que exigem novas competências no mercado de trabalho, influenciam os interesses de pesquisas acadêmicas e desafiam os limites éticos de controle e acesso à informação.

O grande volume de dados em variados formatos, que precisa ser acessado e processado rapidamente, denomina-se Big Data (ARAÚJO JÚNIOR; SOUZA, 2016; WAMBA *et al.*, 2015). O *Big Data* oferece inúmeras possibilidades e concomitantemente inúmeros desafios (FRICKÉ, 2015; RIBEIRO; OLIVEIRA; ARAÚJO, 2019). A análise do *Big Data* para agregação de valor aos dados, e consequente, criação de conhecimento para tomada de decisão, depende de seu efetivo armazenamento, estruturação, processamento, visualização e análise.

Nesse contexto, o profissional da informação se depara com o *Big Data*, o uso de técnicas de análise de dados e de aplicações de inteligência artificial, essenciais na organização e processamento do grande volume de dados. É possível indagar sobre como realizar a representação e a gestão de dados nesses cenários. O objetivo da pesquisa é refletir sobre as relações entre a representação da informação no contexto do *Big Data*, análise de dados e das aplicações de inteligência artificial. A metodologia adotada é exploratório-bibliográfica. O ensaio faz reflexões teóricas sobre as temáticas supracitadas.

2 BIG DATA, ANÁLISE DE DADOS E A REPRESENTAÇÃO DA INFORMAÇÃO

Borko (1968) define a Ciência da Informação (CI) como uma disciplina que investiga a produção, coleta, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação. Oliveira (2011) aponta a recuperação da informação como uma das principais preocupações da CI, que engloba como descrever a informação, como especificar a busca e que sistemas, técnicas ou máquinas devem ser utilizados.

Novellino (1996) afirma que a intervenção na comunicação pode ser feita sob

duas perspectivas: da recuperação da informação e da representação da informação. Almeida (2020) define representar como fazer referência a um objeto do mundo real. O autor diferencia os tipos de representação, a pictórica, a linguística, a mental e a lógica, esta necessária ao processamento de informação por computadores. Maiomone, Silveira e Tálamo (2011) distinguem dois tipos de representação: a temática e a descritiva. Portanto, a representação da informação pode ser entendida “[...] como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico” (BRÄSCHER; CAFÉ, 2008, p. 5).

A representação e a recuperação da informação ainda permeiam a CI na era do *Big Data*. O que muda é que o *Big Data* “[...] se aplica a informações que não podem ser processadas e analisadas pelos métodos tradicionais” (CONEGLIAN; GONÇALVEZ; SANTARÉM SEGUNDO, 2017, p. 133).

A análise de dados é utilizada para lidar com grandes volumes de dados. Serrano-Cobos (2014) afirma que a análise de dados tem origem em duas grandes escolas ou disciplinas, a estatística e a informática. Pode ser definida como uma abordagem para gerir o *Big Data*, ou seja, como métodos que auxiliam na análise do *Big Data* e na tradução dos dados em informações úteis para a tomada de decisão (AASHEIM *et al.*, 2015).

Taurion (2013) define quatro etapas para a análise de dados: coleta; limpeza, formatação e validação; integração e agregação e, análise. A CI, tendo a informação como principal objeto de estudo (LE COADIC, 2004), pode facilitar a execução das etapas de análise de dados. Na coleta de dados, o profissional da CI poderá determinar as fontes mais relevantes sobre a temática investigada. A padronização terminológica realizada por meio de teorias e instrumentos de representação da informação poderá facilitar as fases de limpeza, formatação e validação dos dados e de integração e agregação (que permitem a estruturação dos dados). A fase de análise está mais voltada para a vertente tecnológica, exigindo o uso de algoritmos, estatística, fórmulas e mineração de dados, que permitam aos especialistas adicionarem valor aos dados e realizarem previsões.

Os métodos de análise de documentos e da informação, definidos por Le Coadic (2004), como indexação, classificação e *clustering* (agrupamentos de palavras), podem ser utilizados na estruturação do *Big Data*. Reis e Sá (2020) corroboram essa constatação ao afirmarem que a manipulação do *Big Data* exige competências em estatística, matemática, programação e familiaridade com o

negócio, mas conjuntamente com habilidades de classificação, realização de buscas e recuperação da informação, estas próprias ao profissional da CI.

A gestão de dados, entendida como “[...] o desenvolvimento e implementação de políticas, planos e processos que gerenciem esses dados para manter sua integridade, segurança e usabilidade” (SPECHT *et al.*, 2015, p. 145, tradução nossa), também pode ser utilizada na estruturação e preservação do *Big Data*.

3 INTELIGÊNCIA ARTIFICIAL E SUAS APLICAÇÕES NA CIÊNCIA DA INFORMAÇÃO

Quando se pensa em tecnologias com características avançadas, logo vem à mente a Inteligência Artificial (IA), mas o que é a IA e quais as possíveis aplicações na CI?

Segundo Coppin (2010, p. 4) “Inteligência Artificial envolve utilizar métodos baseados no comportamento inteligente de humanos e outros animais para solucionar problemas complexos”. De acordo com Haykin (1994, p. 35) “[...] é uma área de pesquisa que envolve disciplinas como computação, cognição e aprendizagem, e pretende fazer com que as máquinas sejam capazes de ter inteligência”.

A segmentação da IA foi apresentada da seguinte forma: IA Fraca e IA Forte. Na primeira, teríamos computadores que podem simular algumas tarefas realizadas pela mente humana. Na segunda, os computadores seriam autônomos e totalmente capazes de raciocínio e emoções (SEARLE, 1987).

IA é um campo de estudo amplo, que engloba muitas teorias, métodos e tecnologias, assim como os seguintes subcampos:

a) Aprendizado de máquina (*machine learning*): um sistema inteligente capaz de aprender um comportamento e evoluir com a experiência é relevante no “[...] desenvolvimento de softwares práticos para visão de computador, reconhecimento de fala, processamento da linguagem natural, controle de robôs, e outras aplicações” (JORDAN; MITCHELL, 2015, p. 255, tradução nossa).

b) Aprendizado profundo (*deep learning*): nível mais profundo do aprendizado onde se apresenta o desenvolvimento das redes neurais artificiais. Estas possibilitam que a máquina pense e tome decisões baseando-se no conhecimento disponível. Portanto, para dar capacidade cognitiva a uma máquina é necessária uma estrutura complexa que utiliza algoritmos de otimização. Aplicações comuns incluem

reconhecimento de imagem e fala (JORDAN; MITCHELL, 2015).

c) Redes neurais (*neural networks*): devem possuir capacidade de aprendizado e evolução a partir da análise do que foi aprendido, abrindo espaço para um vasto número de aplicações da IA em diferentes setores (FERNEDA, 2006).

d) Sistemas especialistas: são programas elaborados com base em uma série de regras estabelecidas por um especialista. Provê um serviço que auxilia na tomada de decisões, recomenda soluções e aconselha a respeito de problemas (MOGALI, 2014).

e) Computação cognitiva: almeja uma interação natural e humana com máquinas. Consiste em autoaprendizagem, compreensão da fala humana, algoritmo genético, simulações cerebrais, entre outros (MORESI *et al.*, 2017).

f) Visão computacional: tem “[...] a tarefa de identificar, extrair e interpretar automaticamente padrões do ambiente a partir de informações visuais” (SIMÕES, 2006, p. 1).

g) Processamento de Linguagem Natural (PLN): é a capacidade que os computadores têm de analisar, entender e gerar linguagem humana, incluindo a fala. Os diferentes componentes do PLN são, síntese de fala, reconhecimento de fala, tradução automática, abordagens linguísticas, recuperação e extração de informações (MOGALI, 2014).

No trabalho apresentado por Saracevic (1996, p. 7) na *International Conference on Conceptions of Library and Information Science*, a IA é destacada como “[...] uma das áreas chave de interesse para ambas, ciência da computação e CI”, colocando-a como uma linha de pesquisa pertencente às Ciências Cognitivas.

Segundo Martins (2010) há vários usos da IA na CI e uma delas ocorre nos sistemas de recuperação da informação. “A ideia geral é que a partir de uma questão formulada pelo usuário, o sistema seja capaz de lhe apresentar os resultados que sejam compatíveis com a questão apresentada” (MARTINS, 2010, p. 10).

Outros campos em que a IA pode ser empregada na CI segundo Neves (2020), são: na mineração de dados ou textos à procura de padrões consistentes, como regras de associação ou sequências temporais, para detectar relacionamentos sistemáticos entre variáveis; no PLN, por meio da classificação, indexação e estudos cognitivos e de mediação da informação; no apoio ao atendimento e curadoria digital de sistemas especialistas e, na experiência da abordagem documental da visão computacional.

4 CONSIDERAÇÕES FINAIS

Objetivou-se por meio deste ensaio, refletir sobre como os subcampos da CI, recuperação e representação da informação podem ser utilizados no contexto do *Big Data* e análise de dados e, como a inteligência artificial pode ser aplicada no contexto da CI. Foram levantadas possíveis relações entre objetos de transformação digital e os processos de estruturação da informação da CI.

A representação da informação permite adicionar semântica ao grande volume de dados dispersos e não-estruturados do *Big Data*. Desse modo, a recuperação de informações na Web de dados torna-se possível. Na era da informação, a complexidade se desloca da obtenção do acesso aos documentos para a encontrabilidade de documentos que sejam pertinentes ao usuário. Em bases de dados com grande revocação, garantir a precisão e a qualidade na recuperação de documentos relevantes é o maior desafio.

Pensar nas relações entre a representação da informação, *Big Data*, análise de dados e inteligência artificial é pensar nas humanidades digitais. Isto é, nas aproximações entre as ciências sociais e humanas com as transformações digitais, do mundo conectado em rede e com inovações tecnológicas modificando as relações sociais, econômicas e políticas.

REFERÊNCIAS

AASHEIM, C. L.; WILLIAMS, S.; RUTNER, P.; GARDINER, A. Data Analytics vs. Data Science: a study of similarities and differences in undergraduate programs based on course descriptions. **Journal of Information Systems Education**, v. 26, n. 2, Spring 2015.

ALMEIDA, M. B. **Ontologia em Ciência da Informação**: teoria e método. Curitiba: CRV, 2020. 374 p. v. 1. (Coleção Representação do Conhecimento em Ciência da Informação). DOI: <https://doi.org/10.24824/978655578679.8>

ARAÚJO JÚNIOR, R. H.; SOUZA, R. T. B. 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.

BORKO, H. Information Science: what is it? **American Documentation**, p. 3-5, jan. 1968.

BRÄSCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA

- INFORMAÇÃO, 9., 2008, São Paulo. **Anais Eletrônicos** [...]. São Paulo: Universidade de São Paulo, 2008. Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/809>. Acesso em: 20 jun. 2021.
- CONEGLIAN, C. S.; GONÇALVES, P. R. V. A.; SANTARÉM SEGUNDO, J. E. O profissional da informação na era do Big Data. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 22, n. 50, p. 128-143, set./dez. 2017. ISSN 1518-2924. DOI: [10.5007/1518-2924.2017v22n50p128](https://doi.org/10.5007/1518-2924.2017v22n50p128)
- COPPIN, B. **Inteligência artificial**. Rio de Janeiro: LTC, 2010. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=edsmib&AN=edsmib.000007131&lang=pt-br&site=eds-live&scope=site>. Acesso em: 25 jun. 2021.
- FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, Brasília, v. 35, n. 1, 2006. Disponível em: <http://revista.ibict.br/ciinf/article/view/1149>. Acesso em: 27 jun. 2021.
- FRICKÉ, M. Big Data and its epistemology. **Journal of the Association for Information Science and Technology**, v. 66, n. 4, p. 651-661, 2015.
- HAYKIN, S. S. **Neural networks**: a comprehensive foundation. New York: Macmillan College Publishing, 1994.
- JORDAN, M. I.; MITCHELL, T. M. Machine learning: trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255-260, 2015.
- LE COADIC, Y-F. **A Ciência da Informação**. 2. ed. rev. e atual. Brasília: Briquet de Lemos, 2004. 124 p.
- MAIOMONE, G. D.; SILVEIRA, N. C.; TÁLAMO, M. F. G. M. Reflexões acerca das relações entre representação temática e descritiva. **Informação & Sociedade: Estudos**, João Pessoa, v. 21, n. 1, p. 27-35, 2011.
- MARTINS, A. L. Potenciais aplicações da inteligência artificial na Ciência da Informação. **Informação & Informação**, Londrina, v. 15, n. 1, p. 1-16, 2010. DOI: [10.5433/1981-8920.2010v15n1p1](https://doi.org/10.5433/1981-8920.2010v15n1p1)
- MELLO FILHO, L. L.; ARAÚJO JÚNIOR, R. H. Objetos de fronteira: um diálogo entre a Ciência da Informação e a Ciência de dados. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 26, p. 1-22, 2021. DOI: <https://doi.org/10.5007/1518-2924.2021.e77247>
- MOGALI, S. S. **Artificial Intelligence and its applications in libraries**. 2014. Disponível em: https://www.researchgate.net/publication/287878456_Artificial_Intelligence_and_its_applications_in_Libraries. Acesso em: 27 jun. 2021.
- MORESI, E. A. D. *et al.* UCB-bot: um exemplo de aplicação de computação cognitiva. **Revista de Sistemas, Cibernética e Informática**, v. 14, n. 2, 2017. Disponível em: [http://www.iiisci.org/journal/CV\\$/risici/pdfs/CA272IN17.pdf](http://www.iiisci.org/journal/CV$/risici/pdfs/CA272IN17.pdf). Acesso em: 27 jun. 2021.

NEVES, B. C. Quais campos da inteligência artificial o bibliotecário precisa estudar? **Infohome**, 2020. Disponível em:

http://www.ofaj.com.br/colunas_conteudo.php?cod=1228. Acesso em: 18 abr. 2021.

NOVELLINO, M. S. F. Instrumentos e metodologias de representação da informação. **Informação & Informação**, Londrina, v. 1, n. 2, p. 37-45, jul./dez.1996.

Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/1603>.

Acesso em 24 fev. 2021.

OLIVEIRA, M. **Ciência da Informação e Biblioteconomia: novos conteúdos e espaços de atuação**. 2. ed. Belo Horizonte: Ed. UFMG, 2011. 139 p.

REIS, L. C. R.; SÁ, M. I. F. Big Data: um novo campo de atuação para bibliotecários.

Prisma.com, n. 41, p. 231-250, 2020. DOI: <https://doi.org/10.21747/16463153/41a12>

RIBEIRO, N. C.; OLIVEIRA, D. A.; ARAÚJO, R. F. de. Conjecturas da Ciência Aberta na contemporaneidade do Big Data. **Biblos: Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 33, n. 2, p. 163-179, jul./dez. 2019.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SEARLE, J. R. **Mente, Cérebro e Ciência**. Lisboa: Edições 70, 1987.

SERRANO-COBOS, J. Big Data y analítica web: estudiar las corrientes y pescar en uno océano de datos. **El profesional de la información**, v. 23, n. 6, nov./dez. 2014, p. 561-565. DOI: <http://dx.doi.org/10.3145/epi.2014.nov.01>

SIMÕES, A. S. **Aprendizado não-supervisionado em redes neurais pulsadas de base radial**. 2006. 184 f. Tese (Doutorado em Sistemas Digitais) - Escola

Politécnica, Universidade de São Paulo, São Paulo, 2006. Disponível em:

[https://teses.usp.br/teses/disponiveis/3/3141/tde-15092006-](https://teses.usp.br/teses/disponiveis/3/3141/tde-15092006-153353/publico/AlexandreDaSilvaSimo.es.pdf)

[153353/publico/AlexandreDaSilvaSimo.es.pdf](https://teses.usp.br/teses/disponiveis/3/3141/tde-15092006-153353/publico/AlexandreDaSilvaSimo.es.pdf) . Acesso em: 26 jun. 2021.

SPECHT, A. *et al.* Data management challenges in analysis and synthesis in the ecosystem sciences. **Science of the Total Environment**, v. 534, p. 144-158, 2015.

TAURION, C. **Big Data**. Rio de Janeiro: Brasport, 2013.

WAMBA, S. F.; AKTER, S.; EDWARDS, A.; CHOPIN, G.; GNANZOU, D. How 'Big Data' can make big impact: findings from a systematic review and a longitudinal case study. **International Journal Production Economics**, v. 165, p. 234-246, 2015.