Visualização de dados e de informação em bibliotecas: estudo exploratório preliminar para pesquisa sobre Inteligência Artificial

Antonio Victor Wolf Tadini[a], *Ricardo César Gonçalves Sant'Ana*[b] *e José Eduardo Santarem Segundo*[c]

1. Introdução

A contemporaneidade tem se caracterizado pela valorização da informação e dos dados, contexto em que têm sido tomados como ativos informacionais, insumos estratégicos para as mais variadas finalidades: inteligência competitiva; desenvolvimento de ciência e tecnologia; gestão pública e transparência governamental; diagnósticos em saúde; monitoramento na agricultura e na indústria; apenas para citar alguns exemplos.

As bibliotecas, como espaços privilegiados para a circulação de conhecimento, são em essência ambientes de informação, profícuos para a utilização de dados. A visualização de dados, por exemplo, tem muito a contribuir na dinamização e potencialização desses ambientes, uma vez que "ajuda o cérebro

a Mestrando em Ciência da Informação pela Universidade Estadual Paulista – UNESP. E-mail: antonio.vw.tadini@unesp.br. ORCID: https://orcid.org/0000-0002-7234-4617.

b Doutor em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Professor na Universidade Estadual Paulista – UNESP. E-mail: ricardo.santana@unesp.br. ORCID: https://orcid.org/0000-0003-1387-4519.

c Doutor em Ciência da Informação pela Universidade Estadual Paulista – UNESP. Professor na Universidade de São Paulo – USP. E-mail: santarem@usp.br. ORCID: https://orcid.org/0000-0003-3360-7872.

humano a processar dados de modo mais veloz e mais eficaz do que informações baseadas em texto" (CHEN, 2017c, p. 24).

> Os dados são inúteis se não formos capazes de visualizar o que estamos procurando […] e, para isso, é necessário ter as melhores ferramentas de visualização. Com o tempo é verdade que os modelos de inteligência artificial e *machine learning* serão aqueles que processam essas enormes quantidades de dados (já é feito, de fato, em alguns campos), mas sempre haverá um fator humano decisivo (COUTO, 2019, p. 220).

Sobre a diferença entre a visualização de dados e a visualização da informação, pode-se afirmar que, apesar de serem conceitos distintos, tendo campo próprio cada um, com interesses próprios, têm muitos pontos em comum. Kim, Zhu e Chen (2016, p. 123), em estudo que compara as tendências desses domínios entre 2000 e 2014, concluem que

> Na visualização de dados, tópicos emergentes e novos progressos abarcam a redução da dimensionalidade e aplicações de técnicas visuais […]. A pesquisa de visualização de informação está examinando aspectos cognitivos e teóricos. Em conclusão, a visualização das informações e a visualização dos dados se evolveram uma à outra. Ao mesmo tempo, ambos os campos estão se desenvolvendo distintamente com seus próprios interesses científicos.

Esses aspectos cognitivos são referidos por Saracevic (1996) para descrever as relações interdisciplinares da Ciência da Informação com a Biblioteconomia e também com as Ciências

Cognitivas, dentro das quais se encontra o campo da Inteligência Artificial.

Segundo Santaella (2019, p. 392; p. 402), a inteligência artificial, a despeito de remontar a meados do século XX, encontra-se em uma fase de "acelerado aumento de complexidade" nos resultados de suas pesquisas, no que considera uma "explosão até agora bem-sucedida".

No contexto das bibliotecas não é diferente. A *American Library Association* (ALA), assim como publicou um número inteiro do seu periódico *Library Technology Reports* dedicado à visualização da informação, em 2017, publicou mais recentemente, em 2019, um sobre inteligência artificial em bibliotecas. Evidentemente, ambos os temas constituem tendências na área.

Sendo assim, este trabalho tem como objetivo geral explorar, em nível teórico, possibilidades de aplicação de tecnologias de visualização de dados e de informação no âmbito das bibliotecas, como um levantamento preliminar que forneça subsídios para pesquisa futura sobre inteligência artificial em bibliotecas.

2. Procedimentos metodológicos

A pesquisa se caracteriza como qualitativa (STRAUSS; CORBIN, 2008) e, quanto ao tipo, como exploratória conforme seu objetivo geral, bem como bibliográfica quanto aos procedimentos técnicos empregados (GIL, 2002). Para coleta de dados, procedeu-se a um levantamento bibliográfico nas bases de dados orientado pela questão de pesquisa: quais são as

possibilidades de aplicação de tecnologias de visualização de dados e de informação em bibliotecas, considerando eventuais contribuições de tecnologias de inteligência artificial para essas atividades? O Quadro 1 apresenta a estratégia de busca adotada para recuperação dos textos.

Foram selecionados estudos considerando a ordem de relevância oferecida pelas bases de dados, e de acordo com o exame do título, das palavras-chave e do resumo. A leitura do texto completo foi utilizada subsidiariamente para efeito de seleção. Em seguida, os textos selecionados foram analisados para realização da revisão de literatura.

Quadro 1 – Expressões de busca nas bases de dados

Base de dados	**Expressão de busca**
Base de Dados em Ciência da Informação (BRAPCI)	Todos os campos: "visualização da informação"
	Todos os campos: "visualização de dados"
	Campo Título: "inteligência artificial"
Web of Science (WoS)	*Todos os campos*: ("*information visualization*" OR "*data visualization*") AND ("*artificial intelligence*") AND ("*information science*" OR "*library science*" OR *librarianship*)
Library and Information Science Abstracts (LISA)	Todos os campos, exceto texto completo: ("*information visualization*" OR "*data visualization*") AND ("*artificial intelligence*")

Fonte: Autores.

São três os principais tópicos abordados, conforme os quais se estrutura o trabalho: (i) visualização de dados e de informação em bibliotecas; (ii) inteligência artificial em bibliotecas; e (iii)

resultados preliminares sobre possibilidades com inteligência artificial e visualização de dados e de informação em bibliotecas.

3. Visualização de dados e informação em bibliotecas

Sobre o escopo da visualização de dados, Yang (2017, p. 852) explica que se trata de "uma disciplina abrangente que atravessa a computação gráfica, a interação homem-computador, a estatística, a psicologia".

Técnicas para visualização de dados e de informação consistem na sua transformação em representações imagéticas (tipos de gráficos, linhas do tempo, infográficos, etc.) com o objetivo de beneficiar a compreensão humana, uma vez que, como pontua Couto (2019), "a capacidade humana de reconhecimento de padrões visuais é inata".

> A visualização ajuda a cognição devido às formas como ela aumenta a memória e o processamento dos recursos disponíveis através de propriedades pré-atentivas, Leis de Gestalt e muitas outras propriedades perceptivas (CHEN, 2017b, p. 9).

Dentro disso, existem dois sentidos para o conceito, que podem ser relacionados à função da visualização: (i) a visualização de dados "estuda principalmente como aproximar a percepção humana sobre os dados da percepção natural da visualização humana" (YANG, 2017, p. 852); e (ii) a visualização de dados "também estuda a interação visual da expressão de dados para aumentar a consciência humana, exibindo as informações

implícitas dos dados. Visa explorar a lei dos dados" (YANG, 2017, p. 852).

Nesse segundo sentido, a visualização se serve à tarefa de apoiar a exploração de dados. É uma demanda crescente tendo em vista a consolidação do paradigma da ciência orientada por dados, em que pesquisas exploratórias partem dos dados dispensando a colocação de uma hipótese de pesquisa.

Uma experiência bastante interessante nesse sentido é a relatada por Hayes, Yi e Villaveces (2009), e referida também por Chen (2017c). A Biblioteca de Ciências da Saúde da Universidade da Carolina do Norte, em parceria com o Instituto Renaissance de Computação, implementou um centro de visualização com um painel para interação em tempo real, com função exploratória, isto é, de apontar para novas questões de pesquisa e direcionamentos. Os pesquisadores passaram a poder consultar e colaborar com especialistas em visualização para desenvolver aplicações customizadas para análise de dados.

A visualização pode então ensejar raciocínios, conferindo mais valor aos dados e informações, uma vez que, "ao simplificar a busca de informações e melhorar o reconhecimento de padrões, a visualização permite operações de inferência perceptiva" (CHEN, 2017b, p. 9).

Essa possibilidade se torna especialmente interessante quando se está lidando com grandes conjuntos de dados ou volumes de informação textual, por exemplo.

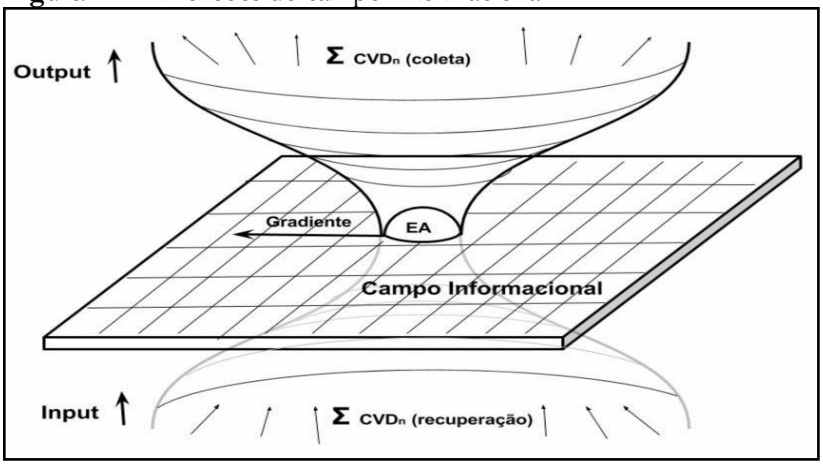
> […] a capacidade para processar e interpretar mais facilmente os dados através do uso da visualização da informação é significativa na retransmissão de mensagens e temas e proporciona uma forma mais fácil de se compreender para grandes quantidades de informação (CHEN, 2017c, p. 25).

É como se as técnicas para beneficiamento da visualização encurtassem um percurso de "camadas de abstração que emergem dos processos de interpretação quando da transdução informacional dos conteúdos registrados nos suportes digitais altamente estruturados até o conteúdo próprio para resposta aos usuários" (SANT'ANA, 2019, p. 117).

Conforme Sant'Ana (2019), a dinâmica dos dados pode ser representada na forma de um ciclo, ao qual se refere como "ciclo de vida dos dados". O ciclo é composto por quatro fases: (i) coleta; (ii) armazenamento; (iii) recuperação e (iv) descarte. Dentro desse dispositivo de análise, a visualização de dados se insere como um elemento integrante da fase de recuperação, e a fase de recuperação do ciclo de vida dos dados se caracteriza como uma dimensão *input*, como se pode visualizar na Figura 1.

As intervenções que beneficiam a visualização de dados podem funcionar como catalisadores do processamento da informação realizado tanto pelos usuários das bibliotecas, quanto pelos bibliotecários. Ambos podem estar na posição de alvo da dimensão *input* do campo informacional, como mentes interpretadoras, que realizam transdução informacional.

Figura 1 – Dimensões do campo informacional



Fonte: Sant'Ana (2019).

Com os usuários na posição de alvo da dimensão *input* pode-se citar o uso de diversos tipos de dados – referentes à coleção, à circulação de itens, a serviços prestados, acesso de recursos digitais, etc. – para comunicação com os usuários, por meio de ferramentas de visualização em seu website (TERRIL, 2018).

A comunicação pode se dar também de modo interativo, estimulando-se o *storytelling* e o engajamento dos usuários, como descrito por Chen (2017c), apontando para um caminho que as bibliotecas poderiam seguir tomando como exemplo uma ferramenta de visualização da 'CNN.com' destinada à interação de e com familiares de combatentes falecidos no Iraque e no Afeganistão, conforme Figura 2. Couto (2019, p. 219) ressalta que

é "recomendável adicionar interação em nossas visualizações de dados".

Figura 2 – Ferramenta de visualização do projeto "*Home and Away*" da CNN.com



Fonte: Chen (2017c).

Ainda com o usuário na posição alvo da dimensão *input*, outro caso muito interessante é a experiência de Legrady e Hoetzlein (2019) na Biblioteca Pública de Seattle – referido também por Chen (2017c). O projeto chama "*Making Visible the Invisible*". Nesse projeto, seis grandes telas de LCD foram instaladas atrás do balcão principal de informações dos bibliotecários.

Neste caso, as representações cumprem a função de expressar os cenários de circulação dos itens da biblioteca, sendo geradas em tempo real a partir da coleta de dados de saída dos itens. São considerados sua classificação na CDD, seu código na coleção e seus termos de indexação. No início, pode até parecer algo trivial, mas com um trabalho estético primoroso, a prática

demonstrou que, para a dinamização e potencialização da biblioteca, é muito interessante que indicadores relacionados àquilo que a comunidade de clientes considera informação útil em um determinado momento esteja acessível visualmente.

Figura 3 – Projeto "*Making Visible the Invisible*" na Biblioteca Pública de Seattle



Fonte: Legrady e Hoetzlein (2019). © George Legrady.

Como mencionado, o bibliotecário também pode estar na posição de alvo da posição *input*, interpretando a partir das representações de visualização de dados e informação. Podem ser utilizadas, por exemplo, para avaliação de dados da biblioteca, para melhoria dos processos de gestão e tomada de decisão. Murphy (2013) relata a experiência com o programa de avaliação das bibliotecas da Universidade de Ohio, que utiliza dados sobre a contagem diária dos portões e sobre os serviços de pesquisa oferecidos, além de uma métrica da *Association for Research Libraries* denominada ARL Ranking.

4. Inteligência Artificial em bibliotecas

Griffey (2019a, p. 9), que é bibliotecário e consultor em tecnologia, defende que "a IA e os sistemas de aprendizado de máquina [*machine learning*] têm o potencial de alterar funções básicas dentro das bibliotecas" e que, em se tratando de tecnologias emergentes, os bibliotecários devem compreendê-las, experimentá-las, e construir algumas de suas próprias ferramentas.

Esforços nesse sentido são relatados na literatura, e tais tecnologias têm sido utilizadas para diversas finalidades. São alguns exemplos: recomendação na recuperação da informação (ASEMI; ASEMI, 2018; YELTON, 2019); serviço de referência (ASEMI; ASEMI, 2018; KIM, 2019; LANCASTER, 1994); representação descritiva (ASEMI; ASEMI, 2018; KIM, 2019; LANCASTER, 1994); indexação automática de documentos (GRIFFEY, 2019b; KIM, 2019; SCHREUR, 2020); classificação automática de documentos (SCHREUR, 2020); entre outras tarefas.

Desse modo, o próximo tópico apresenta resultados preliminares para subsidiar pesquisas futuras que apontem em que medida podem ser relacionadas tecnologias de inteligência artificial com tarefas de visualização de dados e de informação em bibliotecas.

5. Possibilidades com Inteligência Artificial e visualização de dados e de informação em bibliotecas

Uma confluência possível entre a visualização e o *machine learning* – que é um subconjunto da inteligência artificial – é a descrita em Harvey e Porter (2016), em estudo sobre um aspecto da interação humano-computador chamado de estratégias de amostragem orientadas para o usuário. A pesquisa investiga, em vez de situações nas quais o computador seleciona quais dados devem ser rotulados a cada iteração, situações em que quem faz isso é o usuário.

> Tanto a análise visual quanto o *machine learning* tentam aproveitar os pontos fortes complementares dos seres humanos e das máquinas para resolver tarefas complexas de exploração de dados. Estes campos se sobrepõem mais significativamente quando o treinamento está envolvido: a visualização ou a ferramenta de *machine learning* melhoram com o tempo, explorando as observações da interação humano-computador (HARVEY; PORTER, 2016, p. 64).

Em artigo sobre redes neurais – que é um tipo de *machine learning*, e, portanto, de inteligência artificial –, Yelton (2019) relata que tem como objetivo a visualização de dados em interfaces que funcionem com base nesse tipo de tecnologia.

> A rede neural codifica informação sobre ligações entre textos, ligações que, no entanto, não são facilmente exploráveis apenas em uma interface de texto. Imagine, em vez disso, um mapa onde círculos maiores, mais ou menos espaçados, mostrem os *clusters* de significado

> presentes no corpus. "Dando um zoom" nos *clusters*, seria possível ver as ligações individuais entre textos que os compõem (YELTON, 2019, p. 14).

Sobre as redes neurais em sua relação com a tradição dos metadados em bibliotecas, a autora mostra algumas disparidades: (i) redes neurais não resultam em categorias; (ii) redes neurais podem resultar em *clusters*, mas esses *clusters* não têm (e às vezes não podem ter) cabeçalhos de assunto (*subject headings*); e (iii) redes neurais podem operar em espaços onde metadados tradicionais são indisponíveis ou inadequados (YELTON, 2019).

Jiang e Carter (2018, p. 16) sugerem que "a visualização de dados em bibliotecas no futuro pode dispor de inteligência artificial para fins de navegação", de modo a ampliar suas funções, na medida em que,

> com o auxílio de inteligência e a busca de bancos de dados, visualizações de dados devem ser capazes de conduzir os usuários para os resultados adaptados por máquinas que correspondam com precisão ao que ele/ela propõe (JIANG; CARTER, 2018, p. 17).

Os autores também apresentam o seguinte cenário: "a análise de dados levantados para avaliação da biblioteca geralmente encontra respostas ausentes ou incompletas" e utilizar esses dados desbalanceados ou apagar os registros incompletos pode gerar imprecisões e, assim, viciar a análise. Como solução eles mostram que, com uma técnica de inteligência artificial chamada imputação,

> é possível preencher as células ausentes no conjunto de dados e depois alimentar a visualização com um conjunto de dados "completados" artificialmente. O pré-processamento deste tipo não está disponível na maioria das aplicações de visualização de dados (JIANG; CARTER, 2018, p. 16-17).

Note-se que, por essa perspectiva, os algoritmos de inteligência artificial atuam no pré-processamento dos dados, de modo que, em um primeiro momento se encontram na posição de alvo da dimensão *input* do campo informacional, e em seguida realizam uma tarefa de tratamento sobre os dados. O objetivo final é gerar uma visualização, como facilitadora da compreensão por um humano como alvo da dimensão *input*.

A visualização aqui contribui na medida em que, para tratar os dados, o algoritmo opera estruturando e codificando os conteúdos, sobrepondo camadas que o colocam em uma relação de antagonismo com a compreensão humana.

> As diversas interfaces envolvidas neste processo transformativo, ao mesmo tempo em que viabilizam os seus usos e interpretação, necessariamente erguem barreiras à percepção dos usuários sobre as transduções exigidas para que os conteúdos possam ser tratados pelos dispositivos digitais (SANT'ANA, 2019, p. 122).

Sendo assim, um caminho que preliminarmente começa a se delinear é uma composição em que as tecnologias de inteligência artificial e de visualização promovam processos transformativos complementares, com o objetivo de estabelecer um processo

maior que beneficia a interpretação e a exploração dos recursos por humanos.

Essa complementaridade pode ser observada na descrição que Gonzalez Aguilar et al. (2017) apresentam para o processo de visualização analítica, como um método analítico que combina técnicas de análise automatizada com visualizações interativas para melhor compreensão sobre grandes conjuntos de dados complexos, de modo a favorecer o raciocínio e a tomada de decisões.

6. Considerações finais

Este trabalho se propôs a explorar, em nível teórico, possibilidades para desenvolvimento e implementação de atividades de visualização de dados e informação no âmbito das bibliotecas. A pesquisa bibliográfica realizada permite afirmar que as tecnologias e práticas de visualização de dados e de informação constituem oportunidades e desafios contemporâneos para as bibliotecas, no que se refere à melhoria tanto de seus processos, como dos produtos e serviços oferecidos.

O trabalho lançou também alguns indícios para pesquisas futuras sobre como tecnologias de inteligência artificial podem eventualmente contribuir e se combinar a atividades de visualização em bibliotecas – seja na "clusterização" de documentos, no beneficiamento da navegação do usuário, ou no pré-processamento de dados, por exemplo, conforme apresentado.

Referências

ASEMI, A.; ASEMI, A. Artificial intelligence (AI) application in library systems in Iran: a taxonomy study. **Library Philosophy and Practice**, [*s.l.*], e1840, 2018.

CHEN, H. M. An overview of information visualization. **Library Technology Reports**, Chicago, v. 53, n. 3, p. 5-7, 2017.

CHEN, H. M. Information visualization: principles, techniques, and software. **Library Technology Reports**, Chicago, v. 53, n. 3, p. 8-16, 2017.

CHEN, H. M. Real-world uses for information visualization in libraries. **Library Technology Reports**, Chicago, v. 53, n. 3, p. 21-25, 2017.

COUTO, F. C. S. Visualização de dados: passado, presente e futuro. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 205-223, 2019.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GONZALEZ AGUILAR, A. *et al.* **Visualização de dados, informação e conhecimento**. Florianópolis: Ed. UFSC, 2017.

GRIFFEY, J. Introduction. **Library Technology Reports**, Chicago, v. 55, n. 1, p. 5-9, 2019.

GRIFFEY, J. Conclusion. **Library Technology Reports**, Chicago, v. 55, n. 1, p. 26-28, 2019.

HARVEY, N.; PORTER, R. User-driven sampling strategies in images exploitation. **Information Visualization**, [*s.l.*], v. 15, n. 1, p. 64-74, 2016.

HAYES, B.; YI, H.; VILLAVECES, A. Information visualization services in a library?: a public health case study. **Bulletin of the American Society for Information Science and Technology**, [*s.l.*], v. 35, n. 5, p. 13-18, 2009.

JIANG, Z.; CARTER, R. Visualizing library data interactively: two demonstrations using R language. **Library Hi Tech News**, [*s.l.*], v. 35, n. 5, p. 14-17, 2018.

KIM, B. AI and creating the first multidisciplinar AI lab. **Library Technology Reports**, Chicago, v. 55, n. 1, p. 16-20, 2019.

KIM, M. C.; ZHU, Y.; CHEN, C. How are they different?: a quantitative domain comparison of information visualization and data visualization (2000-2014). **Scientometrics**, [*s.l.*], v. 107, n. 1, p. 123-165, 2016.

LANCASTER, F. W. Ameaça ou oportunidade?: o futuro dos serviços de biblioteca à luz das inovações tecnológicas. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 23, n. 1, p. 7-27, 1994.

LEGRADY, G.; HOETZLEIN, R. Making visible the invisible: a data-driven media artwork, in continuous operation for 15 years. *In:* SIGGRAPH, 2019, Los Angeles. **Short Art Papers**. [*S. l.*]: ACM, 2019. e14. p. 1-6.

MURPHY, S. A. How data visualization supports academic library assessment: three examples from the Ohio State University libraries using Tableau. **College and Research Libraries**, [*s. l.*], v. 76, n. 9, p. 482-486, 2015.

SANT'ANA, R. C. G. Transdução Informacional: impactos do controle sobre os dados. *In:* MARTÍNEZ-ÁVILA, D.; SOUZA, E. A.; GONZALEZ, M. E. Q. (Orgs.). **Informação, conhecimento, ação autônoma e big data**: continuidade ou revolução? São Paulo: Cultura Acadêmica: FiloCzar, 2019. p. 117-128.

SANTAELLA, L. A concepção ampliada da mente segundo C. S. Peirce. **Cognitio**, São Paulo, v. 20, n. 2, p. 392-403, 2019.

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

SCHREUR, P. E. Artificial intelligence and complicit bias. **Technicalities**, [*s. l.*], v. 40, n. 5, p. 1, 4-6, 2020.

STRAUSS, A. L.; CORBIN, J. M. **Pesquisa qualitativa**: técnicas e procedimentos para o desenvolvimento de teoria fundamentada. 2. ed. Porto Alegre: Artmed, 2008.

TERRIL, L. J. Telling their story with data: what academic research libraries share on their websites. **Journal of Web Librarianship**, [*s. l.*], v. 12, n. 4, p. 232-245, 2018.

YANG, M. Research progress on data analysis in big data technology. **Advances in Computer Science Research**, [*s. l.*], v. 74, p. 851-854, 2017.

YELTON, A. HAMLET: neural-net-powered prototypes for library discovery. **Library Technology Reports**, Chicago, v. 55, n. 1, p. 10-14, 2019.