

EXTROVERSÃO E DESCOBERTA: VISUALIZAÇÃO DE DADOS NO AUXÍLIO A BUSCAS E RECUPERAÇÃO DE INFORMAÇÕES

EXTROVERSION AND DISCOVERY: DATA VISUALIZATION IN AID FOR SEARCHES AND INFORMATION RETRIEVAL

**Eduardo Caetano Barbosa
Nair Yumiko Kobashi**

Resumo: A presente pesquisa tem como questão central “Como a visualização de dados é ou pode ser usada para a recuperação da informação? ”, e busca identificar, a partir da revisão da literatura selecionada em Ciência da Informação e do contexto tecnológico do Big Data, duas potenciais funções da visualização de dados aplicadas em repositórios digitais. A primeira função, de extroversão, aponta o uso da visualização para caracterizar e compreender um, ou um conjunto de repositórios digitais a partir de análises estatísticas, processamento e tratamento de dados. Gráficos, grafos e representações visuais podem trazer à superfície características intrínsecas do acervo de um ou mais repositórios digitais, evidenciar relações entre objetos informacionais e, por conseguinte, entre temas, conceitos e áreas do conhecimento de um ou mais domínios. A segunda, de descoberta, propõe que o usuário possa recuperar a informação usando a visualização dos dados como instrumento na interface de pesquisa em repositórios digitais. O trabalho é, portanto, um estudo exploratório que tem como objetivo refletir sobre a visualização da informação e sua aplicação na organização e recuperação de informações. Está sendo feita uma revisão da bibliografia da área da Ciência da Informação, com recorrência a autores de outras áreas como Ciências da Computação e Ciência dos Dados.

Palavras-chave: Big Data. Classificação automática. Recuperação da Informação. Visualização de Dados.

Abstract: The present research has, as central question, "How data visualization is or could be used for information retrieval? ", and seeks to identify, from the review of the literature selected in Information Science and the technological context of Big Data, two potential functions of the data visualization applied in digital repositories. The first function, of *extraversion*, points to the use of visualization to characterize and understand one or a set of digital repositories from statistical analysis, processing and data processing. Graphs, charts and visual representations can bring to the surface intrinsic characteristics of the collection of one or more digital repositories, to show relationships between information objects and, therefore, between themes, concepts and areas of knowledge of one or more domains. The second, *discovery*, proposes that the user can retrieve the information using the data visualization as an instrument, as a search interface in digital repositories. This work is, therefore, an exploratory study that aims to reflect on the visualization of the information and its application in the organization and information retrieval. A review of the bibliography of the area of Information Science is being made, with recurrence to authors from other areas such as Computer Science and Data Science.

Keywords: Automatic Classification, Big Data, Data Visualization, Information Retrieval

1 Introdução

A presente pesquisa tem como problema central “Como a visualização de dados é ou pode ser usada para a recuperação da informação?”. Procura identificar, a partir das lentes da Ciência da Informação e do contexto tecnológico do Big Data, duas funções da visualização de dados aplicadas a acervos de documentos digitais.

A primeira função, a de **extroversão**, aponta para o uso das metodologias e técnicas de visualização para caracterizar e compreender um, ou um conjunto de repositórios de documentos digitais a partir de análises estatísticas, processamento e tratamento de dados. Gráficos, grafos e representações visuais podem trazer à superfície características intrínsecas do acervo de um ou mais repositórios digitais, evidenciar relações entre objetos informacionais e, por conseguinte, entre temas, conceitos e áreas do conhecimento de um ou mais domínios.

Extroversão é um termo com origem na psicologia, citado pela primeira vez na teoria dos Tipos Psicológicos de Carl Jung (1921). A teoria distingue duas formas de *atitude*, ou *disposição*, atribuídas ao indivíduo – a extroversão e a introversão.

A *extroversão* tem enfoque no objeto exterior, nos fatos e pessoas em que se observa a atenção para a ação, comunicabilidade, sociabilidade e facilidade de expressão. Já a *introversão* tem enfoque no objeto interior, na ação reflexiva, reservada. Os indivíduos assumem as duas atitudes ao longo da vida, mas aquela que se destaca determinará as características dominantes e, assim, caracterizar o indivíduo.

A pesquisa adota o termo *extroversão*, por traduzir qualidades de ação atribuídas à função da visualização da informação ligada à comunicação e à socialização. A VI tem o potencial de expressar, traduzir e evidenciar, de forma clara e fácil, características de uma base de documentos digitais. Não identifiquei o uso do termo no campo da organização e recuperação de informações. Portanto, este termo – extroversão – está aqui sendo utilizado de forma metafórica, para significar explicitação, representação visual de dados.

A segunda, de **descoberta**, propõe que o usuário possa recuperar informações usando a visualização de dados como instrumento na interface de pesquisa em repositórios digitais. A partir da seleção de partes dos gráficos, deve ser possível ao usuário recuperar os documentos que integram o conjunto representado, seja por pontos, áreas ou outras formas de representação que compõe a visualização gráfica a ser explorada.

No campo da Ciência da informação, e mesmo da Computação, utiliza-se a expressão “descoberta de informação” em bancos de dados. As ontologias, por ex. são caracterizadas, entre outras coisas, como instrumentos de descoberta de conhecimentos. Outra expressão que

pode ser usada é “mineração de dados”, comumente aplicados no campo da Ciência de Dados e Computação. Mineração e descoberta também definem, por fim, a capacidade de recuperar informações relevantes.

O objetivo amplo deste trabalho é refletir sobre a visualização da informação aplicado nos processos envolvidos da recuperação da informação, e verificar a viabilidade das funções de Extroversão e Descoberta, como descritas acima.

Já como objetivos específicos, pretende-se (i) estabelecer um paralelo entre Ciência da Informação e Ciência dos Dados, buscando aspectos comuns em suas origens e a complementaridade em suas práticas, (ii) elucidar os conceitos de Big Data, Recuperação de Informação e Visualização da Informação e (iii) identificar o papel do documento e da classificação neste processo.

A metodologia adotada é a de revisão bibliográfica sobre os conceitos levantados tanto no campo da Ciência da Informação, como na Ciência dos dados, relacionando-os.

A seguir, são introduzidos os capítulos que farão parte da dissertação

2 O contexto Big Data

Big Data é abordado como um contexto caracterizado pela capacidade exponencial de produção, armazenamento e processamento de grandes volumes de dados e informações. 4 atributos se destacam dentre diversas definições na literatura sobre o tema: Informação, que é o dado estruturado, que faz sentido em sua aplicação; Tecnologia, a base essencial de hardware e software que viabiliza processos; Métodos, novas abordagens interdisciplinares e domínios específicos de aplicação; e Impacto, relativo às transformações que essas novas práticas causam na organização da sociedade e seus modos de produção. (DE MAURO, A.; GRECO, M.; GRIMALDI, M. 2016)

Para dar conta dos novos desafios neste contexto, é preciso uma abordagem multidisciplinar. As áreas de Computação, Estatística, Engenharia de Sistemas e, mais recentemente, a disciplina de Ciência de Dados, têm sido largamente reconhecidas como partes importantes desse processo. Dito de outra forma, o objetivo macro deste trabalho é explorar os diálogos da Ciência da Informação com as disciplinas acima referidas, no processo de enfrentamento dos problemas ligados à manipulação de documentos digitais no contexto Big Data, para recuperação de informações.

Procura-se abordar a Ciência da Informação em suas origens em paralelo com o estudo e origem da área de documentação, representada, em sua origem, por Paul Otlet e os desdobramentos para uma visão neodocumentalista, representada então por nomes como Revista Brasileira de Biblioteconomia e Documentação, São Paulo, v. 13, n. esp., p. 115-120, set. 2017.

Bernd Frohmann (2013) e que tecem a visão sobre o documento nos dias atuais, como materialização da informação (FROHMANN 2008).

3 Organização e Recuperação de Informações

A Recuperação de Informações é, hoje, objeto de estudo tanto da da Ciência da Informação como da Ciência da Computação (BAEZA-YATES; RIBEIRO-NETO, 1999), e serão exploradas visões das duas áreas, buscando vislumbrar a relação também com a Ciência dos Dados. Dentre os principais autores a serem explorados, estão Mooers (1951), Shera (apud JACOB; SHAW, 1998), e Ingwersen (1992) e Vakkari (1999).

4 Documentos e classificação

Em conjunto com os conceitos de recuperação da Informação, são discutidos também os conceitos de documento e de classificação (em particular a indexação automática) como partes fundamentais no processo de recuperação e visualização das informações. A leitura de autores como Frohmann (2008), Buckland (1991) e Lancaster (2004) direcionam essa parte do trabalho.

5 Visualização da Informação

Por visualização de informações compreende-se o desenvolvimento e aprimoramento de técnicas de representação visual de volumes de dados que não apresentam em si uma natureza gráfica. Refere-se à representação multidimensional de conjuntos de dados volumosos, que apresentam hierarquias muito grandes, e processados através de processos estatísticos. A expressão também é associada à representação dos conjuntos de dados não estruturados ou abstratos (documentos, sites da WWW, menções de redes sociais, entre outros) (CORRÊA, 2013).

A pesquisa deve explorar diversos aspectos ligados à visualização de informações, aspectos cognitivos e os processos estatísticos envolvidos. Apontará, também modelos de visualização que são adequados ao processo de Recuperação da Informação e seu potencial como interface, adicionando recursos de interatividade. Dentre os autores abordados estão Chen (2006) e Tufte (1997), entre outros.

Referências

ABRIL, G. **Teoría general de la información**. Madrid: Catedra, 1997.

Revista Brasileira de Biblioteconomia e Documentação, São Paulo, v. 13, n. esp., p. 115-120, set. 2017.

VI Seminário de Pesquisa em Ciência da Informação do PPGCI 2017
Escola de Comunicações e Artes – Universidade de São Paulo

BURKE, P. **Uma História Social do Conhecimento**: de Gutenberg a Diderot. Rio de Janeiro: Zahar Ed, 2003.

BURKE, P. **Uma História Social do Conhecimento II**: da Enciclopédia à Wikipedia. Rio de Janeiro: Zahar Ed, 2012.

CAPURRO, R. Epistemologia e Ciência da Informação. V Encontro Nacional de Pesquisa em Ciência da Informação, Belo Horizonte, 2003. **Anais...**, Belo Horizonte: 2003 Tradução de Ana Maria Rezende Cabral, Eduardo Wense Dias, Isis Paim, Ligia Maria Moreira Dumont, Marta Pinheiro Aun e Mônica Erichsen Nassif Borges. Disponível em: www.capurro.de/enancib_p.htm

CAPURRO, R. **Knowledge map of information science**: Rafael Capurro's responses to Chaim Zins, 2003. Disponível apenas online em: <http://www.capurro.de/zins.html>

CAPURRO, R. **Desafíos teóricos y prácticos de la ética intercultural de la información**. Conferência inaugural no I Simpósio Brasileiro de Ética da Informação. João Pessoa, 18 de março de 2010. Disponível em: <http://www.capurro.de/paraiba.html>

CAPURRO, R. Entrevista. **Liinc em Revista**, Rio de Janeiro, v.11, n.2, p.322-328, novembro 2015. Disponível em: <http://revista.ibict.br/liinc/index.php/liinc>

CAPURRO, R., HJORLAND, B. O conceito de informação. **Perspectivas em Ciência da Informação**, v.12, n.1, p.148-207, 2007.

CORREA, C. A. **Indexação automática e visualização de informações**: um estudo baseado em lógica paraconsistente, 2011.

DE MAURO, A.; GRECO, M.; GRIMALDI, M. A formal definition of Big Data based on its essential features, **Library Review**, Vol. 65 ,Issue 3 pp. 122 - 135, 2016.

FACELI, K.; LORENA, A.C.; GAMA, J.; CARVALHO, A.C.P.L.F. **Inteligência Artificial**: uma abordagem de aprendizado de máquina, Rio de Janeiro, LTC, 2015

FROHMANN, Bernd. Revisiting “what is a document?”. **Journal of Documentation**, v. 65, n. 2, p. 291-303, 2008.

FROHMANN, B. O caráter social, material e público da informação. In: FUJITA, M.S.L.; MARTELETO, R. M.; LARA, M. L. G. (org). **A dimensão epistemológica da Ciência da Informação e suas interfaces técnicas, políticas e institucionais nos processos de produção, acesso e disseminação da informação**. São Paulo: Cultura Acadêmica/Marília: FUNDEPE, 2008. p. 13-34.

GARCIA GUTIÉRREZ, A. Cientificamente favelados: uma visão crítica do conhecimento a partir da epistemografia. **Transinformação**, Campinas, v.18, n.2, p.103-112, maio/ago.2006. Disponível em: <http://revistas.puc-campinas.edu.br/transinfo/viewarticle.php?id=151>

GILILAND-SWETLAND. A.J. **Introduction to metadata**: pathways to digital information. Getty Research Institute, Los Angeles, p. 1-19, 2000.

INGWERSEN, P., **Information Retrieval Interaction**. Michigan: Taylor Graham, 1992.

VI Seminário de Pesquisa em Ciência da Informação do PPGCI 2017
Escola de Comunicações e Artes – Universidade de São Paulo

INGWERSEN, P.; JÄRVELIN K. **The Turn**: integration of information seeking and retrieval in context. Springer, 2005.

LUND, N.W. Document theory. **Annual Review of Information Science and Technology**, v.43, n.1, p.1-55, 2009.

MAYER-SCHÖNBERGER, V; CUKIER, K. **Big Data**: a revolution that will transform how we live, work, and think. London: John Murray. 2013

MAZZOCHI, F. Images of thought and their relation to classification: the tree and the net. **Knowledge Organization**, v.40, n.6, p.366-374.

GARHARDT, T. E; SILVEIRA, D. T. (Orgs). **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

MÉNDEZ RODRÍGUEZ, Eva. **Metadatos y recuperación de información**: estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón: Trea, 2002.

OLSON, H. Classification and universality: application and construct. **Semiotica**, v.139, n.1/4, p.377-391, 2002.

SANTOS, B. S. **Um discurso sobre as ciências**. Porto: Afrontamento, 1987. Acessível em http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40141988000200007

SAYÃO, L. F. Bibliotecas digitais e suas utopias. **Ponto de Acesso**, Salvador, v. 2, n. 2, p. 2-36, ago./set. 2008.

SAYÃO, L. F., SALES, L.F. Curadoria Digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Sociedade**: Estudos, João Pessoa, v.22, n.3, p. 179-191, set./dez. 2012.

SMIT, J. A informação na Ciência da Informação. **InCID**: R. Ci. Inf. e Doc., Ribeirão Preto, v. 3, n. 2, p. 84-101, jul./dez. 2012.

SOWA, J. **Knowledge representation: logical, philosophical, and computational foundations**. Pacific Grove: Brooks Cole Publishing Co, 2000.

WU, X., ZHU, X., WU, G. Q., & DING, W. Data mining with big data. **IEEE Transactions on Knowledge and Data Engineering**, v. 26, n. 1, p. 97-107, 2014.

Sobre o autor e a autora:

Eduardo Caetano Barbosa
Mestrando no PPGCI/ECA/USP

Nair Yumilo Kobashi
Docente no PPGCI/ECA/USP