



Transinformação

ISSN: 2318-0889

Pontifícia Universidade Católica de Campinas

FURLAN, Patricia Kuzmenko; LAURINDO, Fernando José Barbin
Agrupamentos epistemológicos de artigos publicados sobre *big data analytics*
Transinformação, vol. 29, núm. 1, 2017, Janeiro-Março, pp. 91-100
Pontifícia Universidade Católica de Campinas

DOI: 10.1590/2318-08892017000100009

Disponível em: <http://www.redalyc.org/articulo.oa?id=384357140009>

- Como citar este artigo
- Número completo
- Mais informações do artigo
- Site da revista em redalyc.org

UABM
redalyc.org

Sistema de Informação Científica Redalyc
Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal
Sem fins lucrativos acadêmica projeto, desenvolvido no âmbito da iniciativa
acesso aberto

Agrupamentos epistemológicos de artigos publicados sobre *big data analytics*

Epistemological grouping of published articles on big data analytics

Patricia Kuzmenko FURLAN¹

Fernando José Barbin LAURINDO¹

Resumo

A era do *big data* já é realidade para empresas e indivíduos, e a literatura acadêmica sobre o tema tem crescido rapidamente nos últimos anos. Neste artigo, pretendeu-se identificar quais são os principais nichos e vertentes de publicação sobre o *big data analytics*. A opção metodológica foi realizar pesquisa bibliométrica na base de dados ISI *Web of Science*, utilizando-se aquele termo para focar as práticas de gestão de *big data*. Foi possível identificar cinco grupos distintos dentre os artigos encontrados: evolução do *big data*; gestão, negócios e estratégia; comportamento humano e aspectos socioculturais; mineração dos dados (*data mining*) e geração de conhecimento; e *Internet das Coisas*. Concluiu-se que o tema é emergente e pouco consolidado, apresentando grande variação nos termos empregados, o que influencia nas buscas bibliográficas. Como resultado complementar da pesquisa, foram identificadas as principais palavras-chave empregadas nas publicações sobre *big data analytics*, o que contribui para as pesquisas bibliográficas de estudos futuros.

Palavras-chave: Análise de dados. Bibliometria. *Big data*.

Abstract

The era of *big data* is yet a reality for businesses and individuals. In recent year, the academic literature exploring this field has grown rapidly. This article aimed to identify the main fields and features of the published papers about *big data analytics*. The methodological approach considered was a bibliometric research at the ISI *Web of Science* platform, whose focus was given to the *big data* management issues. It was possible to identify five distinct groups within the published papers: evolution of *big data*; management, business and strategy; human behavior and the social and cultural aspects; data mining and knowledge generation; *Internet of Things*. It was possible to conclude that *big data* corresponds to an emerging theme, which is not yet consolidated. There is a wide variation in the terms used, which influences the bibliographic searches. Therefore, as a complimentary contribution of this research, the main keywords used in such articles were identified, which contributes for bibliometric research of future studies.

Keywords: Data analysis. Bibliometric. *Big data*.

Introdução

O termo *big data* se refere ao alto volume de dados virtuais, que são complexos, diversos, heterogêneos e que provêm de múltiplas e autônomas fontes, com

controles distribuídos e descentralizados (MCAFEE & BRYNJOLFSSON, 2012; WU *et al.*, 2014; ZHANG *et al.*, 2014b). Esse movimento já é realidade para empresas e indivíduos, e apresenta tendência emergente graças à expansão das redes computacionais, do armazenamento

¹ Universidade de São Paulo, Escola Politécnica, Departamento de Engenharia de Produção. Av. Prof. Almeida Prado, Travessa 2, 128, Cidade Universitária, 05508-070, São Paulo, SP, Brasil. Correspondência para/Correspondence to: P.K. FURLAN. E-mail: <patricia@kuzmenko.com.br>.

Recebido em 22/7/2015, reapresentado em 23/2/2016 e aprovado para publicação em 29/3/2016.

de dados e da capacidade de coleta de dados (CHEN *et al.*, 2014; ZHANG *et al.*, 2014b).

Segundo Sun *et al.* (2014), muitos esforços têm sido feitos no campo da gestão de dados e no processamento de informações. Assim, a coleta e o armazenamento de dados têm crescido rapidamente, sendo que a capacidade dos *softwares* comuns está aquém da necessária para capturar, gerenciar e processar tais dados num período de tempo conveniente.

McAfee e Brynjolfsson (2012) caracterizam o movimento do *big data* com 3V's: Velocidade, Volume e Variedade. Eles consideram que os 3V's aumentaram significativamente na era do *big data*, o que muda o ambiente competitivo dos negócios. Chen *et al.* (2014) salientam que o movimento do *big data* traz oportunidades para se descobrirem novos valores, bem como contribui para ganhar uma maior compreensão de valores escondidos e vivenciar novos desafios.

A era do *big data* compreende diferentes facetas e processos, que envolvem procedimentos de geração, seleção e manipulação de dados. Distintos campos técnicos e do conhecimento são afetados por esse novo ambiente, o que gera o desenvolvimento de estudos e aplicações em distintas áreas científicas, como computação, elétrica, ciências sociais, dentre outras. Ademais, estudos em estratégia e gestão têm explorado o tema como um vetor revolucionário dos negócios e do modo como as empresas adquirem vantagens competitivas.

Chen *et al.* (2012) identificam que os temas de *business intelligence* (inteligência de negócios) e *big data analytics* possuem sinergia. O *business intelligence* começou a ser explorado no início dos anos 1990 e, após crescimento vertiginoso do número de publicações, elas se estabilizaram, ao passo que as publicações a respeito de *big data* e *business analytics* obtiveram rápido crescimento nos últimos anos, incluindo-se publicações na *Harvard Business Review* e no *Management Information Systems (MIS) Quarterly*, de alta relevância.

Desse modo, dada a importância do tema e a abrangência de assuntos correlatos, este artigo pretendeu desenvolver um estudo bibliométrico, de modo a identificar os principais artigos e temas desenvolvidos sobre *big data analytics*.

Procedimentos metodológicos

McAfee e Brynjolfsson (2012) afirmam haver correlação entre os termos "*big data*" e "*analytics*", com movimento atual (*big data*) muito maior do que foi no passado (apenas *analytics*), pois envolve maior volume, velocidade e variedade nas informações. Desse modo, na presente pesquisa, almejou-se o estudo das principais referências bibliométricas que combinam ambos os termos para identificar os principais trabalhos e temas explorados.

Inicialmente, conduziu-se uma busca na plataforma ISI *Web of Science*, em junho de 2015, com as seguintes palavras-chave: "*big data*" OR "*big data analytics*" OR "*big data analysis*", resultando em 5 174 publicações. Destas, 1 673 consistiam em artigos, considerando que a amostra se restringiu a esse formato de publicação. Posteriormente, conduziu-se uma filtragem por temas, com o uso de ferramentas disponibilizadas pelo ISI *Web of Science*, obtendo-se uma amostragem de 1 132 artigos. Foram excluídos temas relativos à biologia, saúde, medicina, física, literatura, filosofia, dentre outros assuntos não correlatos ao mundo dos negócios. Foram, portanto, considerados temas referentes a manufatura, matemática, computação, negócios, economia e ciências sociais. Desse modo, obteve-se a seguinte restrição:

Refinado por tipos de documento (*Article*) e áreas de pesquisa (*Computer science or engineering or business economics or mathematics or information science library science or transportation or mathematical methods in social sciences or energy fuels or international relations or urban studies or government law or social sciences other topics or public administration or operations research management science or food science technology or cultural studies or social issues*).

Observando a Figura 1, pode-se notar que a busca realizada no ISI *Web of Science* trouxe principalmente artigos dos últimos quatro anos, cujas curvas denotam rápido crescimento no número de publicações.

Os resultados foram, então, extraídos para o *Excel* para análise dos dados, quando se notou, ainda, a necessidade de exclusão de alguns artigos com temas sobre medicina, biologia e saúde. Esse procedimento foi

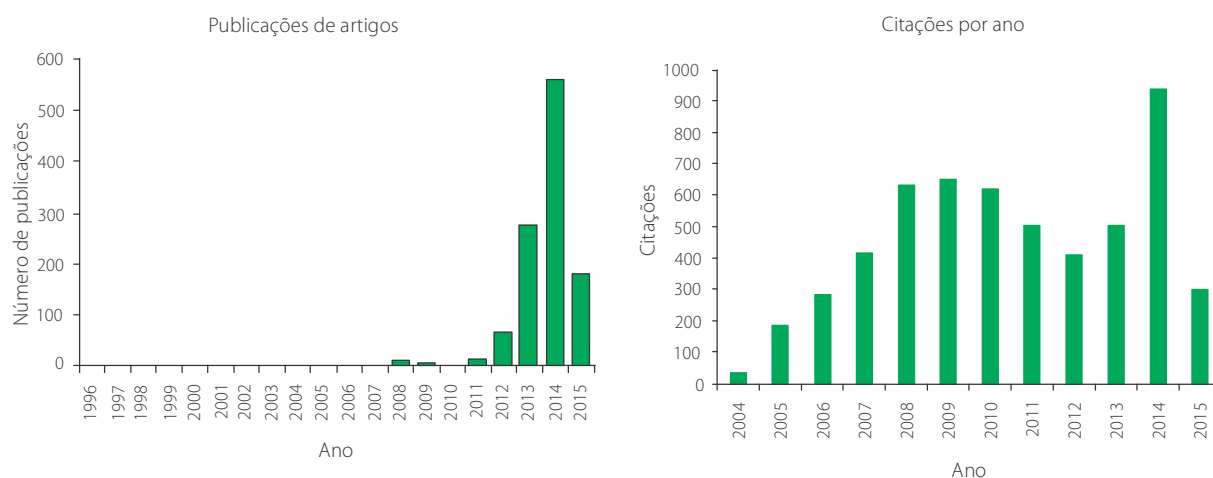


Figura 1. À esquerda, publicação de artigos por ano; à direita, citações por ano.

Fonte: ISI Web of Science.

realizado pela leitura do título do artigo e pela observação do periódico que abrigou a publicação. Foram mantidos na amostra artigos com temas relacionados a negócios (inteligência de negócios, privacidade, nichos específicos como o *m-business*, logística, cadeia de suprimentos), estratégia, tecnologia (incluindo plataformas digitais sociais, como o *Twitter*), computação (desenvolvimento computacional para mineração e análise de dados) e técnicas específicas para manipulação de dados.

Resultados e Discussão

Com auxílio do *software VosViewer* (Center for Science and Technology Studies, Luden, Holanda), foi possível realizar um levantamento das principais palavras utilizadas nos títulos dos artigos da amostra, cujo mapeamento pode ser visto na Figura 2. Notou-se forte ênfase na etapa de coleta e mineração dos dados, com palavras referentes a mineração, medida, relacionamento, teoria e criação. Foi identificado outro nicho consistente, englobando aspectos relativos à análise dos dados no ambiente computacional, com as seguintes palavras-chave: análise de dados, inteligência de negócios, computação na nuvem, ambiente computacional, ontologia, vantagem competitiva. Por fim, notou-se uma série de palavras-chave em títulos que aparecem de forma mais dispersa, com referência ao futuro, ao século 21, à

era *big data*, o que enaltece a importância do tema para discussões futuras na academia e para uso futuro da análise de *big data* pelas corporações.

As cores dos *clusters* são aleatórias e servem como forma de ilustrar os vários grupos de palavras, não possuindo outro significado. Contudo, o local onde aparecem no mapa não é aleatório, pois ele indica a proximidade que os termos possuem entre si. O tamanho de cada *cluster* também possui significado específico, indicando a quantidade de vezes que tal palavra aparece no grupo de artigos considerados. A relevância da palavra é diretamente proporcional ao diâmetro de seu *cluster*.

Dada a amostra considerável de artigos, o presente estudo bibliométrico almejou identificar os principais e, então, prosseguir com análises de conteúdo que agregassem valor no mapeamento dos nichos de publicação. Para refinar as buscas e definir um montante de artigos com maior relevância, optou-se pelo uso da ferramenta da plataforma ISI Web of Science para identificar aqueles mais citados. Nesse contexto, foram definidos 15 artigos como os principais, os quais se encontram listados no Quadro 1.

Pela análise de conteúdo, contudo, notou-se que o artigo 5 da Tabela 1, de Razaviyayn *et al.* (2013), não menciona a palavra *big data* no corpo do texto. O artigo foi então descartado de classificação, segundo o nicho epistemológico que aborda a *big data analytics*.

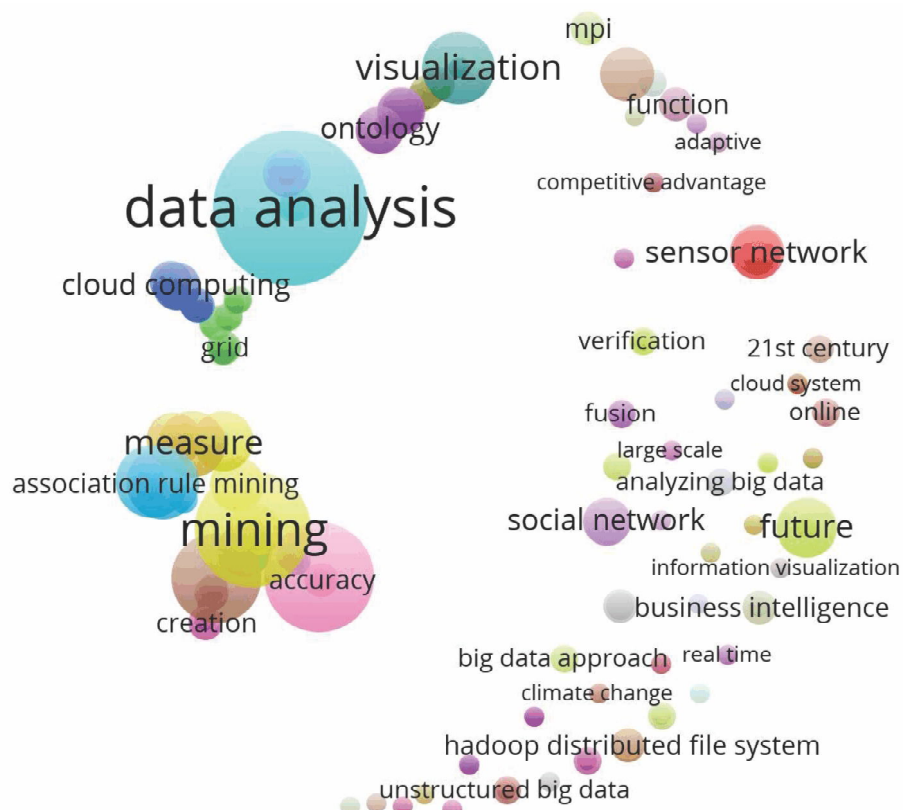


Figura 2. Mapeamento das palavras-chave contidas no título das publicações.
Fonte: Elaboração própria.

Observa-se que os autores dos artigos com mais citações têm menor quantidade de trabalhos publicados. Esse fato foi evidenciado pelo mapa de coautoria gerado pelo *VosViewer*, utilizando-se a amostragem completa dos artigos identificados pelo *ISI Web of Knowledge* (e não apenas os mais citados). Na Figura 3, pode-se perceber que somente três autores dentre aqueles com artigos mais citados (identificados na Tabela 1) são evidenciados no mapa por meio de *clusters*. Os demais autores ilustrados apresentam mais de uma publicação. O mapa não avalia a relevância dos artigos, e sim a quantidade de publicações por autores, ao serem classificados em *clusters* de coautoria.

O número do *cluster* corresponde ao número do artigo listado no Quadro 1 em que o autor foi contemplado.

O *cluster 7* corresponde ao de maior magnitude, como pode ser visto na Figura 3. Isso significa que os

autores que compõem tal *cluster* (principalmente Chen) são bastante ativos em *big data analytics* e são os que mais publicam a respeito do tema. As publicações consideradas nesse *cluster* abordam assuntos da ciência da computação relativos ao processamento e armazenamento de *big data*, cujos *periódicos* mais comuns alvo de publicação correspondem ao *IEEE Transactions on Parallel and Distributed Systems* (fator de impacto 2.173), da *IEEE Computer Society*, e o *Journal of Computer and System Sciences* (fator de impacto 1.091).

Com vistas à diversidade de assuntos abordados pelos artigos mais relevantes em *big data analytics*, foi possível propor uma classificação segundo o tema que exploram. Essa classificação foi realizada a partir da análise de conteúdo de cada artigo, sendo possível definir os seguintes grupos:

– Grupo A. Evolução do *big data*: foi contemplado apenas um artigo neste grupo, combinando aspectos

Quadro 1. Artigos mais citados segundo ferramenta da ISI Web of Science.

	Título	Autores	Periódico	Ano	Citação
1	<i>Business intelligence and analytics: From big data to big impact</i>	Chen, Hsinchun; Chiang, Roger H. L.; Storey, Veda C.	<i>MIS Quarterly</i>	2012	68
2	<i>Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon</i>	Boyd, Danah; Crawford, Kate	<i>Information Communication & Society</i>	2012	56
3	<i>Private traits and attributes are predictable from digital records of human behavior</i>	Kosinski, Michal; Stillwell, David; Graepel, Thore	<i>Proceedings of the National Academy of Sciences of the United States of America</i>	2013	45
4	<i>Big data: The management revolution</i>	McAfee, Andrew; Brynjolfsson, Erik	<i>Harvard Business Review</i>	2012	43
5	<i>A unified convergence analysis of block successive minimization methods for nonsmooth optimization</i>	Razaviyayn, Meisam; Hong, Mingyi; Luo, Zhi-Quan	<i>SIAM Journal on Optimization</i>	2013	19
6	<i>Data mining with big data</i>	Wu, Xindong; Zhu, Xingquan; Wu, Gong-Qing; Ding, Wei	<i>IEEE Transactions on Knowledge and Data Engineering</i>	2014	16
7	<i>A scalable two-phase top-down specialization approach for data anonymization using MapReduce on cloud</i>	Zhang, Xuyun; Yang, Laurence T.; Liu, Chang; Chen, Jinjun	<i>IEEE Transactions on Parallel and Distributed Systems</i>	2014	12
8	<i>Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management</i>	Waller, Matthew A.; Fawcett, Stanley E.	<i>Journal of Business Logistics</i>	2013	12
9	<i>A comparison of parallel large-scale knowledge acquisition using rough set theory on different MapReduce runtime systems</i>	Zhang, Junbo; Wong, Jian-Syuan; Li, Tianrui; Pan, Yi	<i>International Journal of Approximate Reasoning</i>	2014	9
10	<i>Big data: A survey</i>	Chen, Min; Mao, Shiwen; Liu, Yunhao	<i>Mobile Networks & Applications</i>	2014	8
11	<i>Internet of things in industries: A survey</i>	Xu, Li Da; He, Wu; Li, Shancang	<i>IEEE Transactions on Industrial Informatics</i>	2014	7
12	<i>Trends in big data analytics</i>	Kambatla, Karthik; Kollias, Giorgos; Kumar, Vipin; Grama, Ananth	<i>Journal of Parallel and Distributed Computing</i>	2014	6
13	<i>Semantic-based QoS management in cloud systems: Current status and future challenges</i>	Kourtesis, Dimitrios; Alvarez-Rodriguez, Jose Maria; Paraskakis, Iraklis	<i>Future Generation Computer Systems – The International Journal of Grid Computing and E-science</i>	2014	6
14	<i>Big data: New tricks for econometrics</i>	Varian, Hal R.	<i>Journal of Economic Perspectives</i>	2014	6
15	<i>Constructing the web of events from raw data in the web of things</i>	Sun, Yunchuan; Yan, Hongli; Lu, Cheng; Bie, Rongfang; Zhou, Zhangbing	<i>Mobile Information Systems</i>	2014	5

Fonte: ISI Web of Science.

técnicos e de negócios e fazendo uma varredura do tema. Dado que o estudo não evidenciou os nichos de publicação e de desenvolvimento do tema na comunidade científica, abriu-se uma lacuna de conhecimento, para cuja superação o presente texto visa contribuir;

– Grupo B: Gestão, negócios e estratégia: foram incluídos neste grupo artigos que exploram temas de

negócios, gestão e estratégia, abrangendo publicações de periódicos consagrados como a *Harvard Business Review* e a *MIS Quarterly*. São artigos bastante relevantes para a compreensão do movimento do *big data* sob o ponto de vista da gestão;

– Grupo C: Comportamento humano e aspectos socioculturais: o grupo abrange pesquisas que estudam



– Grupo D: Mineração dos dados e geração de conhecimento: os artigos pertencentes a este grupo foram desenvolvidos pela comunidade científica da computação e exploram questões ligadas ao desenvolvimento de técnicas para o *data mining*, para garantir a proteção dos dados estruturados e não estruturados em *Cloud Computing* e gerar conteúdo e/ou conhecimento a partir do processamento dos dados. Corresponde ao maior grupo e está em linha com o que foi visualizado na Figura 3, gerada pelo *software VosViewer*:

nas cadeias de suprimentos e nas redes sociais, de modo a viabilizarem informações no espaço digital, assim permitindo o aumento de *big data*.

A seguir, são apresentados com mais detalhes os agrupamentos epistemológicos identificados e listados acima. A identificação dos artigos em cada grupo remete ao Quadro 1.

O grupo é compreendido por apenas um artigo, o Artigo 10, de Chen *et al.* (2014). Os autores realizam um estudo bastante amplo sobre *big data*, identificando sua evolução e o estado-da-arte da computação para *big data*. Tratam dos aspectos tecnológicos do *big data*, de suas aplicações e dos desafios dessa era. Devido à abrangência de assuntos, não se limitando apenas à gestão ou à computação, o artigo foi classificado neste grupo, à parte dos demais, sendo de alta relevância para a compreensão do tema.

Grupo B: Gestão, negócios e estratégia (Artigos 1, 4, 8)

Os negócios contam, cada vez mais, com ferramentas digitais que aumentam significativamente a quantidade de informações disponibilizadas. Contudo, como posto por McAfee e Brynjolfsson (2012), a dificuldade não está mais na geração de informações, e sim na etapa de análise de conteúdo para orientar ações futuras de forma diferenciada.

McAfee e Brynjolfsson (2012), no artigo 4, identificam duas eras no ambiente de negócios: uma referente ao "*analytics*" e outra referente ao "*big data*", em que a primeira fomentou técnicas rigorosas para a tomada de decisão, de modo que a segunda pôde tornar mais simplificado e poderoso o ambiente de negócios. Assim, os autores classificam ambos os termos como dois estágios de desenvolvimento de dados virtuais, o que correspondeu a um valioso achado neste estudo bibliométrico.

Por outro lado, Chen *et al.* (2012), no artigo 1, utilizam as expressões "*Business Intelligence and Analytics* (BI&A)" e "*big data analytics*", apresentando uma outra forma de incluir a *analytics*, agora também junto à inteligência de negócios (em BI&A). Os autores mencionam que a BI&A, em conjunto com as práticas de *big data analytics*, tornou-se importante para a academia e para a comunidade dos negócios a partir das últimas duas décadas. Chen *et al.* (2012, p.1166, tradução nossa) definem BI&A como "*as técnicas, tecnologias, sistemas, práticas, metodologias e aplicações para analisar dados de negócios críticos para ajudar as organizações a melhor compreender seus negócios e mercado e a tomar melhores decisões*". Os autores exploram o *big data analytics* como um campo relacionado ao BI&A, o que concede novas direções e oportunidades à pesquisa em inteligência de negócios e *analytics*. Os autores também evidenciam outras palavras-chave de pesquisas emergentes que utilizam *analytics* e contribuem para a BI&A: *text analytics*; *web analytics*; *network analytics*; *mobile analytics*. Para eles, as oportunidades emergentes de pesquisas sobre *analytics* estão classificadas nessas quatro áreas técnicas somadas pela área de (*big*) *data analytics* (Chen *et al.*, 2012).

Waller e Fawcett (2013), no artigo 8, exploram a intersecção do ambiente de *big data* com as cadeias de

suprimentos e visualizam grandes transformações na gestão das cadeias. No artigo, introduzem outro termo para se referenciar à análise dos dados: ciência dos dados, análise preditiva e *big data* (em inglês, *Data science*, *Predictive analytics* and *Big Data*, DPB).

Grupo C: Comportamento humano e aspectos socioculturais (Artigos 2 e 3)

Neste grupo, incluem-se trabalhos que estudam a influência do *big data* no comportamento humano e avaliam os impactos sociais e culturais do movimento.

Para Boyd e Crawford (2012), o termo *big data* corresponde a uma era que já se iniciou. Os autores salientam, também, que profissionais como cientistas da computação, físicos, economistas, matemáticos e cientistas políticos são alguns daqueles que almejam o maior acesso massivo de dados produzidos por pessoas, coisas ou por suas interações. Para os autores, não há distinção entre a era *big data* e a era *analytics*, como identificado por McAfee e Brynjolfsson (2012). Ambos os termos são utilizados como palavras-chave separadamente. Contudo, Boyd e Crawford (2012) ressaltam que o termo *big data* é fraco, pois a quantidade de dados processados é relativa à capacidade dos computadores.

Kosinski *et al.* (2013) retratam um estudo particular referente ao *Facebook*, em que são estudados comportamentos humanos a partir de uma função disponibilizada aos usuários. Nesse caso, *big data* aparece no contexto humano, em que são estudadas interações humanas e o perfil dos indivíduos, de modo a aprimorar as ações de *marketing*.

Grupo D: Mineração dos dados e geração de conhecimento (Artigos 6, 7, 9, 12, 13, 14)

Os assuntos aqui explorados se referem aos aspectos técnicos e computacionais para o processamento e mineração de dados, para o estabelecimento de interfaces entre plataformas na nuvem ou bases de dados, e para a geração de conhecimento a partir dos dados. É o grupo com maior número de publicações, dentre as quais o artigo de Kambatla *et al.* (2014), artigo

12, destaca-se por oferecer uma visão geral das principais tendências tecnológicas em *big data analytics*.

A mineração dos dados e as descobertas de conhecimento a partir dos conteúdos digitais se tornaram o novo desafio, devido ao aumento na disponibilização de dados, os quais são também gerados em alta velocidade (Zhang *et al.*, 2014a). Os artigos como o de Wu *et al.* (2014), artigo 6, e o de Zhang *et al.* (2014a), artigo 9, focam as questões técnicas relativas aos *data mining*. Adicionalmente, o estudo de Varian (2014), artigo 14, focaliza o ambiente de transações econômicas para geração de conteúdo. Ele estuda ferramentas para a manipulação dos dados econômicos, defendendo que há necessidade de outras ferramentas para análise dos dados, além da estatística convencional e das técnicas de econometria.

É comum artigos desse grupo explorarem questões relativas ao *MapReduce*, às tecnologias semânticas e ao *Cloud Computing*. Segundo Zhang *et al.* (2014a), o *MapReduce* corresponde a um sistema, a uma infraestrutura ou a um *framework* que realiza filtros e buscas, resumindo-as com auxílio de uma outra variável. Assim, a grande quantidade de dados pode ser sistematicamente processada e analisada, de modo a ser transformada num montante menor de dados e com mais relevância. Outra importante questão desse grupo corresponde à preocupação com a privacidade e a segurança dos dados, que são a questão central da computação na nuvem, como explorado por Zhang *et al.* (2014b) no artigo 7.

Kourtesis *et al.* (2014), no artigo 13, mencionam a necessidade de se formularem indicadores relevantes com base em *big data* para a tomada de decisão em *Cloud Computing*, onde é ambição da computação alcançar o nível em que a tomada de decisão é feita pela própria máquina, excluindo-se, portanto, a interferência humana. Contudo, como o próprio artigo cita, essa situação é inalcançável, pois é necessária a mente humana para observar os dados e tomar decisões considerando também outros conteúdos. Kourtesis *et al.* (2014), Zhang *et al.* (2014a) nos Artigos 9 e 13, apresentam técnicas distintas para a seleção de dados dentre diferentes fontes: o *MapReduce* e o *Quality of Service (QoS) Management* utilizando *semantics*.

Os artigos desse grupo tentam suprir as lacunas técnicas e operacionais para aumentar a extração de conteúdos a partir das informações disponibilizadas, com o objetivo de gerar conhecimento para moldar ações futuras de empresas e indivíduos. Zhang *et al.* (2014b) ressaltam que esse é o principal desafio das aplicações do *big data*.

Todos os artigos aqui observados utilizam como palavras-chave os termos técnicos associados às tecnologias e mencionam, também, o termo *big data*.

Grupo E: *Internet das Coisas* (Artigos 11 e 15)

A “*internet das Coisas*” (em inglês, *Internet of Things*, IoT) ou “*Web das Coisas*” (em inglês, *Web of Things*, WoT), segundo Sun *et al.* (2014), almeja alavancar o cotidiano das pessoas por meio da geração de conhecimento de eventos que estão para acontecer, servindo como interface entre o mundo físico (no qual se enquadram objetos materiais e pessoas) e o virtual. Em relação aos eventos na *web*, Sun *et al.* (2014) identificaram uma tendência na comunidade acadêmica de afirmar que os eventos correspondem à abstração fundamental no estudo do mundo dinâmico.

Segundo Xu *et al.* (2014), as aplicações da *Internet das Coisas* na indústria têm avançado nos últimos anos, incluindo tecnologias como o *radio-frequency identification*, *wireless*, aparelhos móveis e sensores, para o monitoramento integrado das “coisas” físicas com o virtual. Diferentemente dos outros grupos, este focaliza a integração dos fluxos físico e virtual de matérias ou produtos das manufaturas, e trata de informações específicas dessas “coisas”. Trata-se, portanto, de um grupo peculiar que não necessariamente manipula dados relativos ao *big data*. Xu *et al.* (2014), no artigo 11, fazem uma varredura sobre o estado-de-arte da *Internet das coisas* na manufatura, identificando a literatura acadêmica atual, as tecnologias envolvidas e as principais aplicações industriais. Já Sun *et al.* (2014) se preocupam com a “*Web of Events*”, eventos na *web*, retratando a importância de tratar eventos na *Internet* e seus elos, para absorver mais conhecimento dos dados.

Este grupo está voltado para a interação do mundo físico com o virtual, estando presente em diferentes

atividades, como a manufatura, a cadeia de suprimentos e as atividades das pessoas, com o apoio de tecnologias que transmitem a informação física para o ambiente digital. O ambiente de *big data* é aquele que a *Internet* das coisas quer alcançar, pois ele almeja gerar, no mundo virtual, o máximo possível de dados do mundo físico.

Conclusão

O termo *big data*, em específico *big data analytics*, objeto de estudo deste artigo, tem sido utilizado por grupos distintos na comunidade científica, em temas correlatos a: mineração de dados, análise de dados e geração de conhecimento; gestão, estratégia e negócios; influência da tecnologia no comportamento humano e nas alterações sociais; *Internet* das Coisas, que basicamente envolve grupos de estudo das cadeias globais de suprimentos (*global supply chains*) e logística. Ademais, periódicos de alta relevância têm se preocupado em publicar acerca do tema, o que demonstra sua importância na atualidade.

Por meio da análise das palavras dos títulos, pôde-se observar a existência de dois núcleos bem definidos e um outro mais disperso. No que tange aos núcleos definidos, inclui-se um relativo aos procedimentos de mineração de dados (*data mining*) e outro relativo à análise dos dados, onde “inteligência de negócios” aparece como termo correlato. Constatou-se ainda a menção dispersa de palavras relativas ao futuro, ao século atual, donde se conclui que *big data analytics* corresponde a um movimento que tende a crescer e a se fortalecer nos próximos anos ou décadas.

No grupo B, relativo às publicações sobre negócios, gestão e estratégia, apareceram formas diferentes de se referenciar às palavras-chave empregadas nesta pesquisa bibliográfica, como: *analytics* e *big data*, correspondendo a dois momentos distintos no ambiente de negócios; BI&A e *big data analytics*, indicando sintonia entre a área de inteligência de negócios e a de analítica; Ciência dos dados, análise preditiva e *big data*, expressão para se referir à análise de dados. Nos demais grupos, inclusive nos de ciência humana ou computacional, não houve diferenciação dos termos *big data* e *analytics*, como identificado no grupo relativo aos negócios, estratégia e gestão.

A variação das formas para se referir ao *big data analytics* chama atenção, pois estas devem ser empregadas em conjunto na pesquisa bibliográfica, principalmente quando se almeja a busca por literatura sobre negócios, gestão e estratégia. Notou-se, portanto, falta de padronização dos termos para explorar a questão, o que é comum em pesquisas emergentes contemplando novos temas.

Finalmente, espera-se que o presente artigo contribua para o avanço na compreensão do tema *big data* e que novos estudos sejam fomentados para a consolidação do tema em suas diversas facetas, contemplando a computação, as ciências humanas e computacionais, os negócios, a gestão e a estratégia.

Colaboradores

P.K. FURLAN e F.J.B. LAURINDO contribuíram na concepção e desenho do estudo, análise de dados e redação final.

Referências

- BOYD, D.; CRAWFORD, K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication & Society*, v. 15, n. 5, p. 662-679, 2012.
- CHEN, H.; CHIANG, R. H. L.; STOREY, V. C. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, v. 36, n. 4, p. 1165-1188, 2012.
- CHEN, M.; MAO, S.; LIU, Y. Big data: A survey. *Mobile Networks & Applications*, v. 19, n. 2, p. 171-209, 2014.

- KAMBATLA, K. et al. Trends in big data analytics. *Journal of Parallel and Distributed Computing*, v. 74, n. 7, p. 2561-2573, 2014.

- KOSINSKI, M.; STILLWELL, D.; GRAEPEL, T. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, v. 110, n. 15, p. 5802-5805, 2013.

- KOURTESIS, D.; ALVAREZ-RODRIGUEZ, J. M.; PARASKAKIS, I. Semantic-based QoS management in cloud systems: Current status and future challenges. *Future Generation Computer*

Systems: The International Journal of Grid Computing and E-science, v. 3, p. 307-323, 2014.

McAFEE, A.; BRYNJOLFSSON, E. Big data: The management revolution. *Harvard Business Review*, v. 90, n. 10, p. 60, 2012.

RAZAVIYAYN, M.; HONG, M.; LUO, Z. Q. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization*, v. 23, n. 2, p. 1126-1153, 2013.

SUN, Y. *et al.* Constructing the web of events from raw data in the Web of Things. *Mobile Information Systems*, v. 10, n. 1, p. 105-125, 2014.

VARIAN, H. R. Big data: New tricks for econometrics. *Journal of Economic Perspectives*, v. 28, n. 2, p. 3-27, 2014.

XU, L. D.; HE, W.; LI, S. Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics*, v. 10, n.4, p. 2233-2243, 2014.

ZHANG, J. *et al.* A comparison of parallel large-scale knowledge acquisition using rough set theory on different MapReduce runtime systems. *International Journal of Approximate Reasoning*, v. 55, n. 3, p. 896-907, 2014b.

ZHANG, X. *et al.* A Scalable Two-Phase Top-Down specialization approach for data anonymization using MapReduce on cloud. *IEEE Transactions on Parallel and Distributed Systems*, v. 25, n. 2, p. 363-373, 2014b.

WALLER, M. A.; FAWCETT, S. E. Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, v. 34, n. 2, p. 77-84, 2013.

WU, X. *et al.* Data Mining with Big data. *IEEE Transactions on Knowledge and Data Engineering*, v. 26, n. 1, p. 97-107, 2014.