

Edgar de Souza

edgarsouza@cetip.com.br

*Mestrando do Programa de
Mestrado Profissional em Gestão e
Tecnologia em sistemas
produtivos do Centro Paula Souza*

Giovana Fadini de Oliveira

gifadini@yahoo.com.br

*Mestrando do Programa de
Mestrado Profissional em Gestão e
Tecnologia em sistemas
produtivos do Centro Paula Souza*

Odir Veiga

odirveiga@ceoconsultoriaempresa.rial.com.br

*Mestrando do Programa de
Mestrado Profissional em Gestão e
Tecnologia em sistemas
produtivos do Centro Paula Souza*

Prof. Dr Marcelo T. Okano

marcelo.okano@fatec.sp.gov.br

*Professor do Programa de
Mestrado Profissional em Gestão e
Tecnologia em sistemas
produtivos*

**SADSJ - South American Development
Society journal – São Paulo, Brasil.**

A UTILIZAÇÃO DA BIBLIOMETRIA PARA A ELABORAÇÃO DE UM PANORAMA DAS PUBLICAÇÕES ACADÊMICAS SOBRE BIG DATA

RESUMO

Com o avanço da tecnologia no desenvolvimento de dispositivos de armazenamento, o volume de dados gerados diariamente são enormes e a preocupação com a manipulação desses dados é cada vez maior. Para se gerenciar essa grande massa de dados encontramos o Big Data, que possui a capacidade de trabalhar com esses dados. Este artigo apresenta a bibliometria do Big Data, para medir os índices de produção e disseminação do conhecimento dessa tecnologia. A base de dados analisada foi a WOS – *Web of Science*, o qual foram selecionados por conveniência 41 artigos entre os anos 2012 a 2016. Esta análise trouxe conhecimento relevante sobre o cenário mundial de pesquisas envolvendo o tema Big Data e seu desenvolvimento.

Palavras-chave: Big Data, bibliometria, Ferramentas para bibliometria

ABSTRACT

With the advancement of technology in the development of storage devices, the volume of data generated daily are huge and concern about the handling of such data is increasing. To manage this massive data found the Big Data, which is able to work with large data sets. This article presents bibliometrics of Big Data to measure the levels of production and dissemination of knowledge of this technology. The analyzed database was WOS - Web of Science, which were selected 41 articles between the years 2012 to 2016. This analysis has relevant knowledge on the world stage of research involving the theme Big Data, as well as details about your development.

Keywords: Big Data, bibliometrics, bibliometrics Tools

INTRODUÇÃO

O constante desenvolvimento da área de Tecnologia da Informação (TI), e os avanços em pesquisas para solucionar problemas em dispositivos de armazenamento de dados com grandes capacidades, permitem que cada vez mais dados sejam armazenados, resultando assim em volumosos bancos de dados.

A simples armazenagem desses dados já traz algum benefício às empresas, pois não é mais necessário procurar informações em grandes arquivos de papel. Mas o que fazer com um enorme volume de dados? Como utilizá-los para trazer mais benefícios às instituições? Basta fazermos uma simples consulta nesses dados?

Em uma pesquisa realizada pela McKinsey, somente 18% das empresas entrevistadas acreditam possuir habilidades necessárias para capturar e utilizar os conhecimentos advindos de grandes massas de dados, visto que o profissional não deve somente entender da tecnologia em si, mas também dos processos de negócio (McKinsey & Company, 2014).

Visto a importância do tema e o seu potencial para expansão e utilização por parte das empresas, este artigo tem o objetivo de elaborar o panorama de publicações na área de *Big Data*, a fim de mostrar a sua importância para a academia, através da produção acadêmica no período de 2012 a 2016, utilizando a bibliometria. Os indicadores fornecidos e encontrados através da bibliometria tem se mostrado relevantes para a análise e avaliação das pesquisas realizadas nas mais diversas áreas do conhecimento em âmbito regional, nacional e internacional (GRÁCIO e OLIVEIRA, 2012).

REFERENCIAL TEÓRICO

Big Data

O termo *Big Data* é o mais recente, dentre as metodologias de análise e manipulação de grandes massas de dados existentes e refere-se a arquiteturas de sistemas capazes de lidar com as novas dimensões dos dados: velocidade, variedade, volume, valor e ganhou bastante importância quando as seguintes tecnologias começaram a ser desenvolvidas (MORGENTHAL, 2012):

- a) Baixo custo dos dispositivos de armazenamento que até então obrigavam as empresas a armazenar apenas a informação essencial.
- b) Virtualização que aumentou a possibilidade de virtualizar quase todo o tipo de hardware necessário.
- c) *Cloud computing* que proporcionou às empresas uma solução de computação elástica, à medida das suas necessidades.
- d) Bases de dados NoSQL que já são desenhadas para lidar com grandes volumes de informação em tempo útil e com suporte para processamento distribuído.

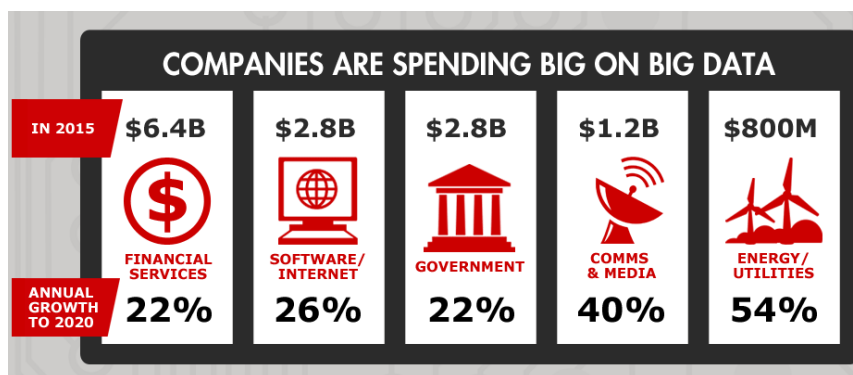
De acordo com Goldchmidt e Passos (2005) dentro de bases de dados volumosas encontramos muito mais do que “coisas” armazenadas, e é inviável a análise desses dados sem o auxílio de alguma ferramenta computacional apropriada.

E para facilitar a leitura, interpretação e manipulação desses dados, é possível identificar várias técnicas que podem ser aplicadas para se obter melhor compreensão

dos mesmos, tais como ferramentas de *Business Intelligence*, Mineração de Dados (*data mining*), estatísticas, algoritmos de similaridades e Visualização de Informação e Big Data.

McAFEE, A.; BRYNJOLFSSON (2012) afirma ainda que soluções de Big Data possuem um potencial maior do que as soluções analíticas tradicionais existentes para trazer benefícios e aumentar a competitividade frente aos concorrentes. Grandes empresas e até governos têm feito o uso de Big Data para ajudar larga gama de consumidores e até mesmo cidades inteiras a planejar suas atividades para alcançar um futuro melhor. (MANYIKA, CHUI, et al., 2011), a figura 1 mostra o quanto as empresas estão investindo em Big Data.

Figura 1: Gastos com Big Data em 2015



Fonte: http://www.bain.com/publications/articles/big_data_the_organizational_challenge.aspx. Acesso em: 30/05/2016

De acordo com Davenport (2012), os dados de um Big Data, podem ser provenientes dos mais diversos locais tais como, voz de um *call center* até dados de genoma de pesquisa biológica e medicinal (DAVENPORT, 2012). Nota-se também que algumas das maiores fontes de dados existentes, são relativamente novas tais como as redes sociais, dentre as quais se destacam o *Facebook* lançado em 2004, *Twitter* em 2006 e o *Youtube*, que aliadas a grande massificação de vendas de *smartphones* e *tablets*, redes wireless e redes de dados 3G e 4G, permitem às pessoas permanecerem conectadas em praticamente todo o tempo, gerando conteúdo, ideias e opiniões.

De acordo com PEARSON, T.; WEGENER, R. (2013), outros diferentes setores de negócio vêm utilizando as ferramentas de Big Data para garantir avanços e melhorias aos negócios, tais setores são Bancos, Ensino, Governo, Saúde e Varejo.

Podemos perceber que gerenciar com análise de dados é uma ideia convencional hoje em dia, porém não é uma prática convencional. Não é surpresa que a Gartner Research identifique levar “a informação certa para a pessoa certa no momento certo” como uma tendência tecnológica estratégica. A Accenture identificou este ponto como uma prioridade máxima para CIOs – *Chief Information Officer*, título este dado ao gerente de TI, superintendente de TI, diretor ou vice presidente de TI (RANSBOTHAM, S.; KIRON, D.; PRENTICE, P. K., 2016).

Neste contexto encontramos o Big Data, que é capaz de trabalhar com grandes massas de dados, com intuito de facilitar o processo de geração do conhecimento, transformando-os dados em informações estratégicas.

Grande parte da literatura define o conceito de Big data com os quatro Vs (Volume, Variedade, Velocidade e Valor), sendo que o volume compreende aos grandes conjuntos de dados ou as grandes massas de dados encontrados em banco de dados gigantescos, a variedade corresponde aos dados que podem vir de diversas bases de dados, a Velocidade no que diz respeito ao processamento desses dados de forma rápida e eficiente e o Valor que esses dados possuem podendo ser transformados em informação caso seja analisado, extraído e interpretado de forma correta.

Bibliometria

O termo Bibliometria foi criado por Paul Otlet em 1934, no Tratado da Documentação, sendo antes conhecida como bibliografia estatística, termo proposto por Hulme em 1923. Todavia, o termo consolidou-se apenas em 1969, após a publicação do artigo de Pritchard, sob o título “Bibliografia estatística ou Bibliometria?” (Vanti, 2002).

Com o advento de novas tecnologias e com maior organização e acesso a bases de dados os estudo bibliométricos se desenvolveram no Brasil a partir da década de 1970 com grande avanço em 1990 (GRÁCIO e OLIVEIRA, 2012).

De acordo com Lopes et al. (2012) a bibliometria é uma técnica quantitativa e estatística que mede índices de produção e disseminação da ciência e do conhecimento. Também acompanha o desenvolvimento de diversas áreas científicas e os padrões de autoria, publicações e o uso de seus resultados.

MÉTODO

Neste trabalho foi realizada uma pesquisa bibliométrica, ou bibliometria, para uma análise quantitativa da comunicação escrita de artigos publicados em periódicos. Foram analisadas citações e concitações. A análise de citação baseia-se na premissa de que autores citam artigos que consideram importantes no desenvolvimento de suas pesquisas.

O levantamento foi realizado na base de dados *Web of Science*, no dia 01 de Julho de 2016. Utilizou-se como palavra-chave ou termo de busca “*Big Data*” AND “*Business Analytics*” no subcampo tópico e no período de 2012 a 2016, a amostra foi definida por conveniência.

Foram utilizadas como ferramentas de análises:

3.1 Bibexcel

BibExcel é uma ferramenta desenvolvida por Olle Persson, Inforsk, Univ Umeå (Suécia). Este software foi concebido para auxiliar o usuário na análise de dados bibliográficos, a ideia é gerar arquivos de dados que pode ser importados para o MS Excel, ou qualquer programa que tenha registros com tabulação de dados, para posterior processamento.

Sua interface gráfica não é nada intuitiva, mas com um pouco de treinamento é possível ao imputar dados de bases de dados científicas, ter a relação e o cruzamento entre esses dados, gerando assim a visualização de periódicos publicados por ano,

divididos em categorias, assuntos, agências de fomento, os autores mais citados, bem como suas referências.

Apesar de sua interface não amigável, sua vantagem é a sua flexibilidade de integração com outras ferramentas como Pajek (Machado Jr. Et al.,2014).

Pajek

O Pajek é uma ferramenta de visualização dos dados que foram processados pela ferramenta Bibexcel.

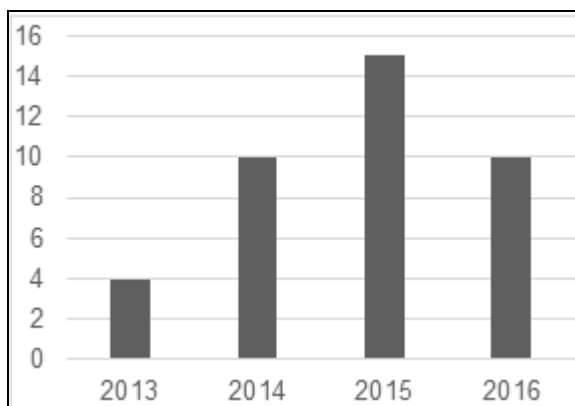
A escolha do Bibexcel como programa de processamento e do Pajek como de visualização foi por apresentarem um melhor custo/benefício quanto a sua facilidade de operação e com seu poder de processamento para o escopo escolhido. Outro quesito importante no processo de seleção foi a compatibilidade entre os softwares Bibexcel e Pajek e os repositórios, pois além de serem livres se mostraram bastante completos de acordo com as necessidades encontradas dentro da pesquisa.

Da base da *Web of Science*, foram obtidos 41 artigos, sendo o universo da pesquisa, e foram analisados os seguintes itens: Tipos de publicação, publicações por ano, principais autores, instituições, países (cidades) e áreas do conhecimento.

RESULTADOS E DISCUSSÃO

Os resultados da pesquisa bibliométrica apresentou um aumento em artigos em periódicos nos últimos 4 anos, passando de 4 artigos em 2013 para 15 em 2015. No ano de 2016 tem-se um valor significativo até o presente momento, pois em 6 meses publicou-se até o presente momento a mesma quantia de publicações realizadas no ano de 2014.

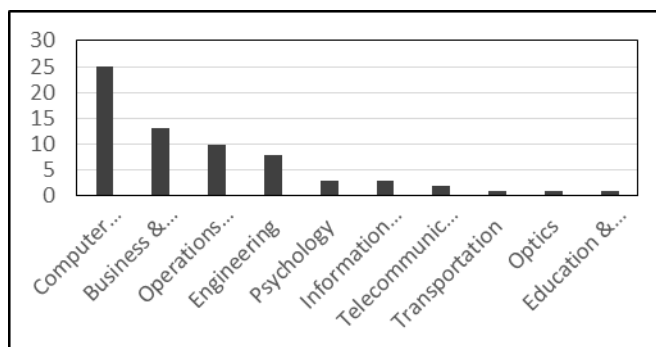
Gráfico 01 – Publicações por ano



Fonte: Elaborado pelos autores

As principais áreas de conhecimento são Ciências da Computação e Economia e Negócios, vide gráfico 02.

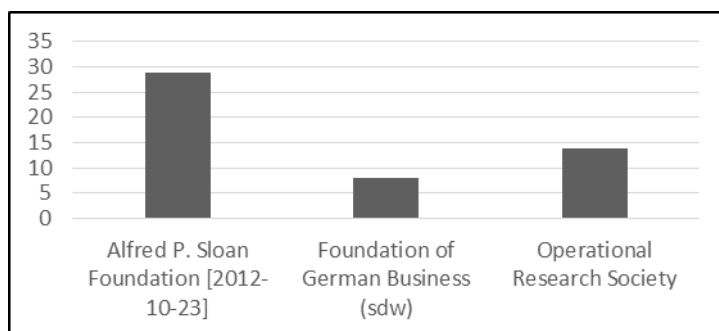
Gráfico 02 – Áreas do conhecimento



Fonte: Elaborado pelos autores

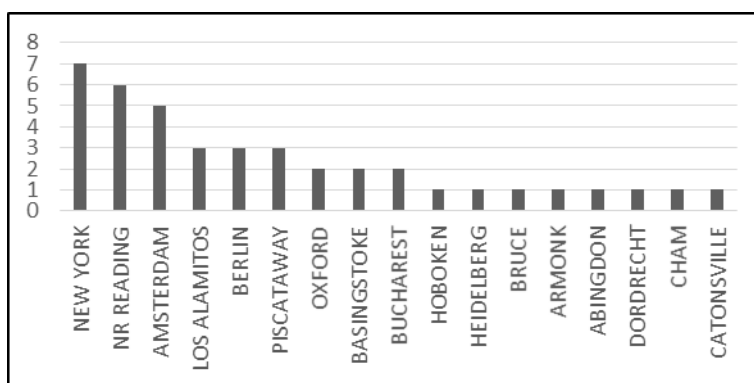
No gráfico 03, encontramos as agências que mais financiam pesquisas na área de Big Data, a que mais se destaca é a *Alfred P. Sloan Foundation*, fundação americana sem fins lucrativos, em segundo lugar tem-se a *Operational Research Society*, que está localizada no Reino Unido e em terceira posição encontra-se a *Foundation German Business*.

Gráfico 03 – Agências de Financiamento



Fonte: Elaborado pelos autores

Gráfico 04 – Publicações por Localidades



Fonte: Elaborado pelos autores

A tabela 01, mostra os principais autores nesse período.

Tabela 01 – Principais autores

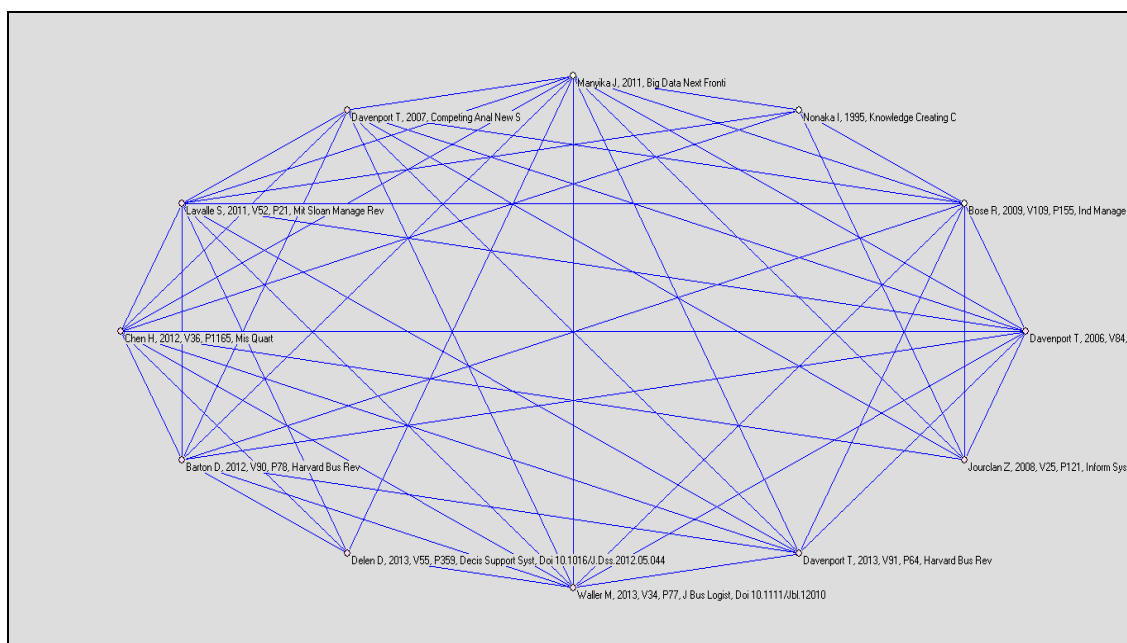
Autor	Publicações
Erickson S	3
Marjanovic O	3
Duan YQ	2
Dinter B	2
Dhar S	2
Rothberg H	2
Sun ZH	2
Ariyachandra T	2
Cao GM	2
Mithas S	2

Mazumdar S	2
Puleo D	1

Fonte: Elaborado pelos autores

Na figura 02, encontramos o sociograma de citações.

Figura 02 – Sociograma de citações



Fonte: Elaborado pelos autores

Desta forma, foram apresentados alguns gráficos que foram gerados a partir da análise da bibliometria.

CONSIDERAÇÕES FINAIS

No mercado corporativo a utilização do Big Data está cada vez mais em ascensão, o que é inquestionável devido ao grande volume de dados gerados diariamente. Para analisar o mesmo desempenho na Academia através de uma panorama de publicações, utilizamos indicadores bibliométricos para análise e avaliação dos artigos na base dados WOS. Foram obtidos 41 artigos no período de 2012 a 2016, visto que por ser um tema relativamente novo, é nítido o crescente

aumento nas pesquisas nos anos pesquisados, mostrando assim a evolução das pesquisas na área de Big Data. Este tema abrange várias áreas do conhecimento, tendo uma maior concentração nas áreas de ciências da computação e Economia de negócios, caracterizando um tema multidisciplinar. Da mesma forma, este tema é pesquisado em diversos países e instituições de ensino e pesquisa pelo mundo, sendo financiado por agências de fomento de grande porte.

Percebe-se a sua importância pelas ocorrências de constantes crescimento encontrado no período pesquisado, caracterizando um aumento nos autores, citações e obras existentes.

REFERÊNCIAS

DAVENPORT, T.; PRUSAK, L. **Conhecimento empresarial: como as organizações gerenciam seu capital intelectual**. Rio de Janeiro: Campus 1998.

GARTNER. **Answering Big Data's 10 Biggest Planning and Implementation Questions**. Disponível em: <<https://www.gartner.com/doc/2957917/answering-big-datas--biggest>>. Acesso em: 30/05/2016.

GARTNER. **Major Myths About Big Data's Impact on Analytics**. Disponível em: <https://www.gartner.com/doc/2846318/major-myths-big-datas-impact>. Acesso em: 30/05/2016.

GOLDSCHMIDT, R.; PASSOS, E. **Data Mining – Um Guia Prático**. Campus 2005.

GRÁCIO, Maria Cláudia Cabrini; OLIVEIRA, Ely Francina Tannuri de. **A inserção e o impacto internacional da pesquisa brasileira em 'estudos métricos': uma análise na base Scopus**. Tendências da Pesquisa Brasileira em Ciência da Informação, p. 1-19, 2012.

LOPES, Sílvia et al. **A Bibliometria e a Avaliação da Produção Científica: indicadores e ferramentas**. In: Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas. 2012.

MACHADO JUNIOR, Celso et al. **As Leis da Bibliometria em Diferentes Bases de Dados Científicos**. Revista de Ciências da Administração, Florianópolis, p. 111-123, 2014.

MANIYKA, J., CHUI, M., BROWN, B., BUGHIN, J., DOBBS, R., ROXBURGH, C., & BYERS A. H. **Big data: The next frontier for innovation, competition, and productivity**. Disponível em: <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>. Acesso em: 30/05/2016

MCKINSEY & COMPANY. Disponível em: <http://www.mckinsey.com/industries/media-and-entertainment/our-insights/global-media-report-2014> >. Acesso em 16/05/2016.

McKinsey Global Institute, May, 2011. Disponível em: http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation. Acesso em: 16/05/2016.

McAFEE, A.; BRYNJOLFSSON, E. **Big Data: The Management Revolution**. Harvard Business Review, October, 2012. p. 1-9. Disponível em: <https://hbr.org/2012/10/big-data-the-management-revolution/ar>>. Acesso em: 30/05/2016

MORGENTHAL, J., 2012. **Evolution in Data Integration From EII to Big Data**. Disponível em: <http://www.infoq.com/articles/DataIntegrationFromEIItoBigData>> Acesso em 14/06/2016.

PEARSON, T.; WEGENER, R. **Bain Insights**. Bain & Company, 2013. Disponível em: http://www.bain.com/publications/articles/big_data_the_organizational_challenge.aspx>. Acesso em: 28/05/2016.

RANSBOTHAM, S.; KIRON, D.; PRENTICE, P. K. **Beyond The Hype: The Hard Work Behind Analytics Success**. MIT Sloan Management Review, 8 Março 2016.

VANTI, Nadia Aurora. **Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do**

conhecimento. *Ciência da Informação*, Brasília, v. 31, n. 2, p. 152-162, maio/ago. 2002.

SETZER, W. **Dado, informação, conhecimento e Competência**. Disponível em: <http://www.ime.usp.br/~vwsetzer/dado-info_Folha.html> - Publicado no jornal do Grupo Folha, *Folha Educação* No. 27, out./nov. 2004, pp. 6 e 7). Acesso em: 30/05/2016.