APRENDER FÍSICA, BIOLOGIA, QUÍMICA E MATEMÁTICA COM BIG DATA

Learning Physics, Biology, and Mathematics with Big Data

Renato P. dos Santos

Resumo

Vivemos num universo digital que, em 2010, atravessou a marca de um *zettabyte* de dados. Essa enorme quantidade de dados, processada em computadores extremamente velozes, com técnicas otimizadas, permite encontrar *insights* em novos e emergentes tipos de dados e conteúdos para responder a perguntas que foram anteriormente consideradas fora de nosso alcance. Essa é a ideia de Big Data. Tal como aconteceu com os PCs, a internet e a comunicação móvel, nós, como indivíduos, seremos capazes de fazer mais com nossos próprios dados do que as grandes empresas tão logo nos sejam disponibilizadas ferramentas acessíveis tais como várias que estamos vendo surgir. A empresa Google disponibiliza ao público a ferramenta de análise *Google Correlate* que, para um termo de busca ou uma série de dados temporais ou regionais, fornece uma lista das consultas no Google cujas frequências seguem padrões que melhor se correlacionam com os dados, segundo o coeficiente de determinação R^2. Naturalmente, "correlação não implica causação"; no entanto, acreditamos haver um potencial nessas ferramentas de Big Data para encontrar correlações inesperadas, e até inusitadas, que poderão servir de pistas para fenômenos interessantes do ponto de vista pedagógico e até científico. Neste trabalho, apresentamos alguns exemplos de possibilidades de aplicação de uma proposta didática para a utilização de Big Data no ensino de Física, Biologia, Química e Matemática, tendo, como mediadores, o computador e as ferramentas públicas e gratuitas de Big Data tais como o *Google Correlate*.

Palavras-chave: Big Data. Ensino de Física. Ensino de Biologia. Ensino de Química. Educação Matemática.

Abstract

We live in a digital world that, in 2010, crossed the mark of a zettabyte data. This huge amount of data processed on computers extremely fast with optimized techniques, allows one to find insights in new and emerging types of data and content, to answer questions that were previously considered beyond reach. This is the idea of Big Data. As it happened with the PC, the Internet and mobile communication, we, as individuals, will be able to do more with our own data than the large enterprises, as soon as affordable tools are available, such as the several ones that are seen coming out. Google now offers the *Google Correlate* analysis public tool that, from a search term or a series of temporal or regional data, provides a list of queries on Google whose frequencies follow patterns that best correlate with the data, according to the determination coefficient R^2. Of course, "correlation does not imply causation"; however, we believe that there is potential for these big data tools to find unexpected, even unusual, correlations that may serve as clues to interesting phenomena, from the pedagogical and even scientific point of view. In this paper, a few examples of application possibilities are presented of a didactic proposal for the use of Big Data in the Teaching of Physics, Biology, Chemistry and Mathematics, taking as mediators the computer and the public and free tools such as *Google Correlate*.

Keywords: Big Data. Physics Teaching. Biology Education. Chemistry Education. Mathematics Education.

Introdução

O universo digital em que vivemos atravessou a marca de um *zettabyte* (aproximadamente 10^{21}) de dados (ZIKOPOULOS et al., 2013, p.9), correspondente a postagens e curtida nas redes sociais, em imagens e vídeos de telefones celulares enviados para o YouTube, filmes digitais de alta definição, movimentações bancárias, imagens de câmaras de segurança, colisões subatômicas registradas pelo LHC do CERN, chamadas telefônicas, mensagens SMS, etc. (GANTZ; REINSEL, 2012). Vivemos na onda do Big Data.

Entre as muitas definições de Big Data encontradas, preferimos a seguinte, por a julgarmos mais esclarecedora para os propósitos deste trabalho:

> Big Data é mais do que simplesmente uma questão de tamanho, é uma oportunidade de encontrar *insights* em novos e emergentes tipos de dados e conteúdos, para tornar seu negócio mais ágil e para responder a perguntas que foram anteriormente consideradas fora de seu alcance. (IBM, s.d.)

Segundo Mattmann (2013), para resolver os desafios do Big Data, é necessária uma nova raça denominada 'cientistas de dados'. Para Kate Mueller (DUMBILL et al., 2013), dizer que só os profissionais da ciência da computação se tornam bons especialistas de Big Data ou bons analistas de dados é um erro. Mattmann (2013) insiste em que os cientistas naturais também se devem familiarizar com Big Data.

Acreditamos, como Searls (2013), que, tal como aconteceu com os PCs, a internet e a comunicação móvel, nós, como indivíduos, seremos capazes de fazer mais com nossos próprios dados do que as grandes empresas, sem necessidade de aprender *Hadoop*[1], *MapReduce*[2] e outras tantas, tão logo nos sejam disponibilizados meios para isso. De fato, estamos vendo surgir ferramentas de Big Data, muito poderosas, mas acessíveis, tais como *BigSheets*[3], *Tableau*[4], *Karmasphere*[5], *Revolution Analytics*[6] e *HDInsight*[7].

O principal objetivo deste projeto é investigar a viabilidade do uso Big Data no Ensino de Ciências, tendo, como mediadores, o computador e as ferramentas públicas e gratuitas do Big Data, tais como o *Microsoft Power Map*[8] (anteriormente conhecido como *GeoFlow*), o *Google Trends*[9], o *Google Correlate*[10] e outras que devem vir a surgir em breve.

Nossa proposta (dos SANTOS, 2014a, 2014b) tem embasamento no construcionismo de Papert (1985), o qual ressalta a importância de ferramentas, mídias e contextos no desenvolvimento humano e em como seus diálogos com artefatos promovem a autoaprendizagem e facilitam a construção de novos conhecimentos (ACKERMANN, 2001).

Entre propostas anteriores do uso de ferramentas de Big Data no ensino, podemos citar Baram-Tsabari e Segev (2009a, 2009b, 2013), Segev e Baram-Tsabari (2012), que propõem utilizar as ferramentas *Google Trends*, *Google Zeitgeist* (encerrado em 2007 e substituído por *Hot Trends*, um recurso dinâmico em *Google Trends*) e *Google Insights for Search* (incorporado ao *Google Trends* em 2012) para a pesquisa e a discussão sobre a compreensão da ciência pelo público e sobre a distinção entre ciência e pseudociência em sala de aula. Bülbül (2009) e Yin et al. (2013) propõem determinar e discutir tendências em Física e em Educação através de pesquisas de palavras chave através de Google, *Google Scholar* (conhecido no Brasil como *Google Acadêmico*) e *Google Trends*.

[1] *Hadoop* é uma plataforma de *software* em Java de computação distribuída voltada para *clusters* e processamento de grandes massas de dados.
[2] *MapReduce* é um modelo de programação para o processamento de grandes conjuntos de dados, usado para fazer a computação distribuída em *clusters* de computadores.
[3] http://www-01.ibm.com/software/ebusiness/jstart/bigsheets/
[4] http://www.tableausoftware.com/
[5] http://karmasphere.com/what-we-do
[6] http://www.revolutionanalytics.com/
[7] http://www.windowsazure.com/pt-br/home/features/hdinsight/
[8] http://www.microsoft.com/en-us/download/details.aspx?id=38395
[9] http://www.google.com/trends/
[10] http://www.google.com/trends/correlate

Nossa proposta é distinta e baseia-se na ferramenta *Google Correlate*, mais moderna, discutida a seguir.

Como se sabe, o motor de busca Google não apenas realiza alegadas cem bilhões de pesquisas mensais de termos na *web* (SULLIVAN, 2012) como as armazena todas, identificadas por hora e local de origem em seus gigantescos *data centers* ao redor do mundo. Essas informações são posteriormente utilizadas pelos programas de publicidade geridos pelo Google, tais como *DoubleClick*, *Google Analytics*, *Google AdWords* e *Google AdSense*, de onde provêm mais de 90% da renda da empresa Google (GOOGLE INC., 2013).

Desde maio de 2011, a empresa Google disponibiliza ao público a ferramenta de análise *Google Correlate*. Nela, introduz-se um termo de busca ou uma série de dados temporais ou regionais e se obtém, dentre milhões de candidatos, através de um processo automático, uma lista das consultas no Google cujas frequências seguem padrões que melhor se correlacionam com os dados, segundo o coeficiente de correlação de Pearson R^2 (MOHEBBI et al., 2011).

Google Correlate recebe uma entrada de dados que pode ser:

- um termo de busca individual;
- uma série temporal de dados carregada pelo utilizador;
- uma série espacial de dados carregada pelo utilizador;
- um esboço de um gráfico feito com o recurso '*Search by Drawing*' (busca por desenho).

Vários trabalhos acadêmicos baseados no *Google Correlate* podem ser encontrados na literatura científica em diversos campos de conhecimento, tais como Saúde Pública, Economia, Sociologia e Meteorologia.

Naturalmente, há que ter consciência de que as causas subjacentes aos comportamentos de busca podem não ser conhecidas, que os utilizadores do Google não representam uma amostra aleatória da população e que essa correspondência pode não se manter no futuro, devido a mudanças no comportamento de busca dos usuários (MOHEBBI et al., 2011).

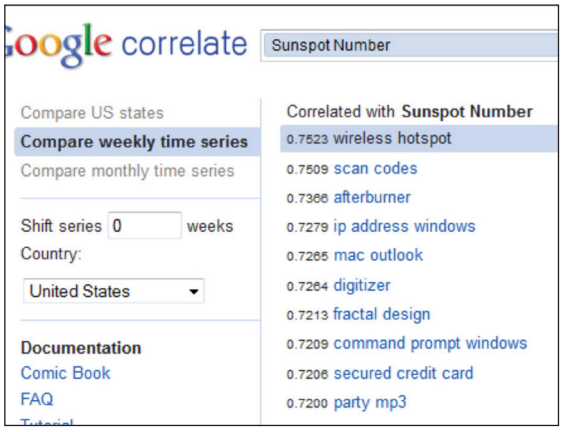
Vale também a pena lembrar o alerta dos estatísticos "correlação não implica causação" (FIELD, 2003, p.10), o que significa que estabelecer uma correlação não implica estabelecer uma relação causal, até porque não sabemos o que causou o quê. No entanto, acreditamos que "correlação não é causalidade, mas com certeza é uma pista" (TUFTE, 2006, p.5) e que há um potencial nessas ferramentas de Big Data para encontrar correlações inesperadas, e até inusitadas, que poderão, no entanto, servir de pistas para fenômenos interessantes do ponto de vista pedagógico e até científico.

Na sequência, apresentaremos alguns exemplos, demonstrando as possibilidades de utilização de Big Data para o ensino de Química, Biologia, Física e Matemática.

Física

Como primeiro exemplo, utilizamos (dos SANTOS, 2014a, 2014b) a variação semanal da atividade solar, medida pela variação do número de manchas solares[11] de 5 jan. 2003 a 31 mar. 2013. Observa-se, da Figura 1, uma boa correlação para vários termos, sendo que o que melhor correlacionou (R^2=0, 7523) foi '*wireless hotspot'* que significa locais em que a tecnologia *wi-fi* está disponível.

Figura 1 – Termos de busca no Google com frequências correlacionadas positivamente à variação semanal do número de manchas solares de 5 jan. 2003 a 31 mar. 2013.



Fonte: *Google Correlate* (http://www.google.com/trends/correlate).

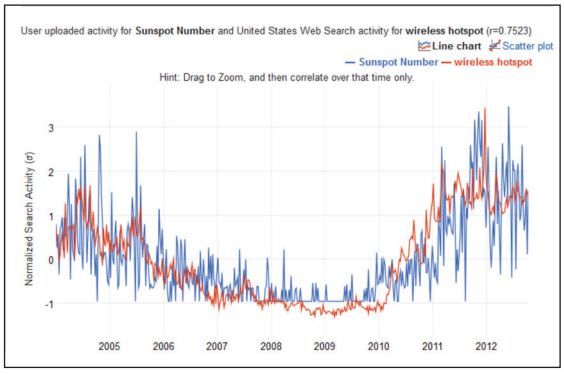
[11] Dados obtidos em http://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-indices/sunspot-numbers/international/listings/

No gráfico produzido, pelo *Google Correlate*, para o termo '*wireless hotspot*' (Figura 2), essa correlação fica bastante aparente.

Inicialmente, pode-se não ver relação causal entre as buscas por esses locais no Google e as variações do número de manchas solares ou com a atividade solar. Nossa proposta (dos SANTOS, 2014a, 2014b) de um possível mecanismo causal seria o de que máximos nessa atividade prejudicam o alcance dos *hotspots* e, por isso, usuários acostumados a utilizar determinados *hotspots* se veriam obrigados a procurar novos *hotspots* para se conectarem.

Pela nossa proposta (dos SANTOS, 2014a, 2014b), este é um momento frutífero de aprendizado para o estudante de Ciências: observado um fenômeno novo (a correlação), buscar uma explicação científica (causação) para ele. Para confirmar ou não a hipótese aventada acima, seria necessário que os estudantes aprofundassem suas pesquisas em várias outras fontes, o que seria extremamente produtivo em termos de aprendizado de Ciências.

Figura 2 – Comparação entre a frequência de pesquisa do termo 'wireless hotspot' no Google e a variação semanal do número de manchas solares de 5 jan. 2003 a 31 mar. 2013.

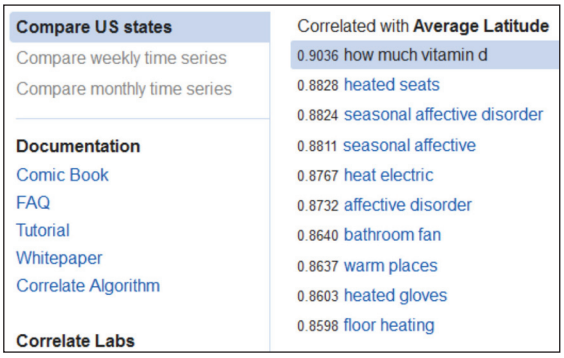


Fonte: *Google Correlate* (http://www.google.com/trends/correlate).

Biologia

Utilizando, agora, uma série de dados espaciais de latitudes médias dos estados norte-americanos (infelizmente, o *Google Correlate* ainda só tem bases de dados espaciais para os EUA), obtêm-se os seguintes termos de busca melhor correlacionados (Figura 3).

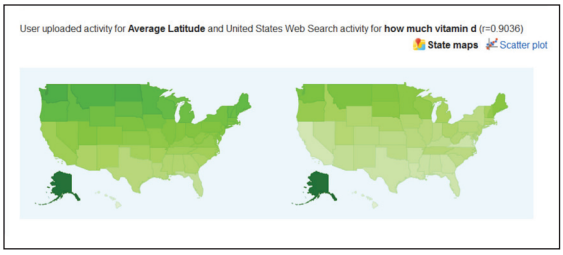
Figura 3 – Termos de busca no Google com frequências correlacionadas positivamente às latitudes médias dos estados norte-americanos.



Fonte: *Google Correlate* (http://www.google.com/trends/correlate).

Além de termos relativos a sistemas e meios de aquecimento e de termos referentes à chamada síndrome de inverno, ou Transtorno Afetivo Sazonal (em inglês, *Seasonal Affective Disorder*), o termo de busca com maior correlação é a pergunta "quanta vitamina D" [é recomendada...]. É bastante razoável que, quanto maior a latitude do Estado, com menor incidência de luz solar, maior a preocupação de seus habitantes com sua ingestão de vitamina D (Figura 4), já que esta é sintetizada na pele por ação da radiação solar.

Figura 4 – Comparação entre a frequência de pesquisa do termo 'quanta vitamina d' no Google e a latitude do Estado.

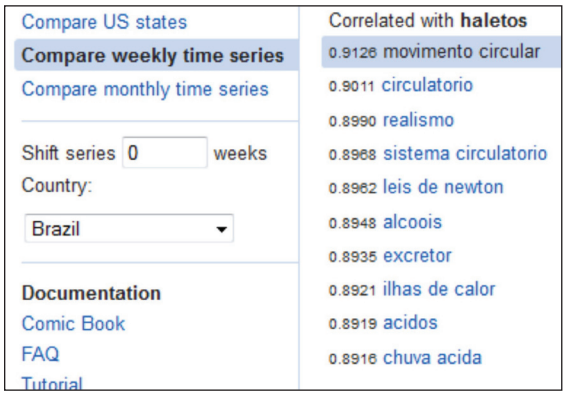


Fonte: *Google Correlate* (http://www.google.com/trends/correlate).

Química

Agora, em vez de introduzir uma série de dados, vamos utilizar um termo de busca individual. Entre os vários conceitos relevantes para a Química, a Figura 5 abaixo mostra os termos de busca que mais se correlacionam com a palavra 'haletos'.

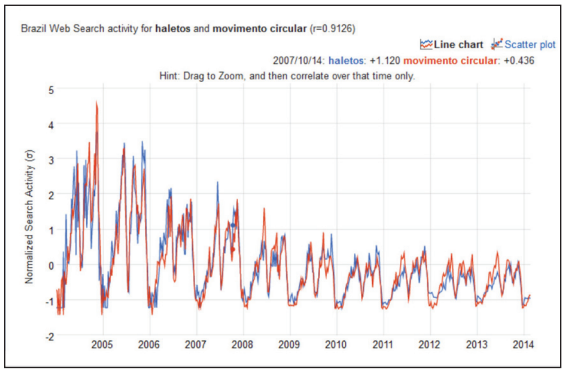
Figura 5 – Termos de busca no Google com frequências correlacionadas positivamente ao termo de busca 'haletos'.



Fonte: *Google Correlate* (http://www.google.com/trends/correlate).

Da Figura 6 abaixo, observa-se, porém, certa periodicidade nas frequências de busca, com picos em meados do mês de junho e no final do mês de novembro, o que corresponde aos períodos usuais de avaliação final semestral nas escolas brasileiras. De fato, tanto 'haletos' como os termos de busca correlacionados na Figura 5 são tópicos do ensino médio, o que corrobora nossa suposição. Aqui, verifica-se, portanto, não uma correlação causal, mas temporal, decorrente da causa externa comum a todas essas séries temporais: as avaliações escolares. Observamos esse mesmo padrão de periodicidades nas correlações para muitos termos de busca correspondentes a conceitos básicos em Física, Química, Biologia e Matemática.

Figura 6 – Comparação entre as frequências de pesquisa dos termos de busca 'haletos' e 'movimento circular' no Google.



Fonte: *Google Correlate* (http://www.google.com/trends/correlate).

Matemática

Por último, mas não menos importante, vejamos um exemplo envolvendo a Matemática. Para ela, faremos uso da outra forma de entrada de dados no *Google Correlate*, qual seja um esboço de um gráfico feito com o recurso '*Search by Drawing*'.

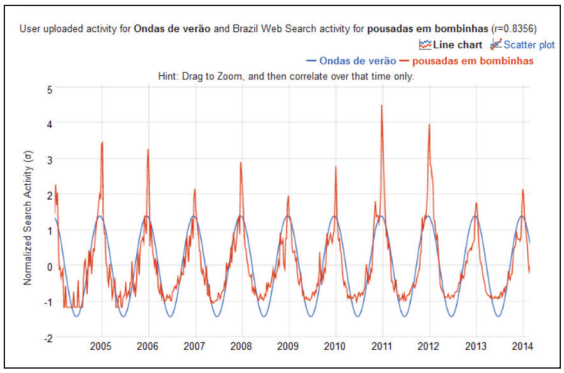
De um gráfico aproximadamente senoidal (Figura 8), com máximos nos picos das estações de verão ("ondas de verão"), resultam os seguintes termos de busca mais bem correlacionados (Figura 7).

Figura 7 – Termos de busca no Google com frequências correlacionadas positivamente ao esboço senoidal das ondas de verão.

Fonte: *Google Correlate* (http://www.google.com/trends/correlate).

De fato, faz sentido todos os termos acima serem mais buscados na época de verão, pois se correlacionam bem a atividades das férias de verão.

Figura 8 – Comparação entre as frequências de pesquisa dos termos de busca 'pousada em Bombinhas' no Google e o esboço senoidal das ondas de verão.



Fonte: *Google Correlate* (http://www.google.com/trends/correlate).

Temos, aqui, um recurso poderoso para o ensino e a aprendizagem de Matemática, tanto em termos de modelagem quanto em termos de estudo de funções, verificando como uma função matemática tem aplicação na modelagem de comportamentos humanos do mundo real.

Conclusão

Acreditamos que esta é a primeira proposta de utilização de Big Data no Ensino de Ciências (dos SANTOS, 2014a, 2014b), com um viés que não é meramente o de uma capacitação em infraestruturas computacionais ou de treinamento em análise preditiva, mas uma preparação de nossos estudantes, futuros profissionais de Ciências, tanto em termos técnicos como em éticos, para os desafios científicos propostos pelo Big Data ao mundo real, no qual vão exercer suas profissões, além de uma melhor compreensão, embasada na prática do Big Data, sobre a construção do conhecimento físico, especialmente numa melhor compreensão das noções de fenômeno, observação, medida, leis físicas, teoria, causalidade, entre outras.

Dentro do projeto de pesquisa que apoia esta proposta, também está sendo construída uma base de dados[12] de correlações obtidas com o *Google Correlate*, a qual pode servir de subsídio a outros professores que venham a se interessar por aplicar esta proposta em suas salas de aula.

Referências

ACKERMANN, E. K. Piaget's Constructivism, Papert's Constructionism: What's the difference? *Future of learning group publication*, v.5, n.3, p.438, 2001.

BARAM-TSABARI, A.; SEGEV, E. Exploring new web-based tools to identify public interest in science. *Public Understanding of Science*, v.20, n.1, p.130-143, 2009b.

_____. Just Google it! Exploring New Web-based Tools for Identifying Public Interest in Science and Pseudoscience. In: ESHET-ALKALAI, Y.; CASPI, A.; EDEN, S.; GERI, N.; YAIR, Y. (Eds.). CHAIS CONFERENCE ON INSTRUCTIONAL TECHNOLOGIES RESEARCH 2009: LEARNING IN THE TECHNOLOGICAL ERA. *Proceedings....* Raanana: The Open University of Israel, 2009a. p.20-28.

_____. The half-life of a "teachable moment": The case of Nobel laureates. *Public understanding of science*, p.83-89, 2013.

BÜLBÜL, M. Ş. *Google Centered Search Method in Pursuit of Trends and Definitions in Physics and Education*. Disponível em: <www.fizikli.com/piwi/fizikli6.pdf>. Acesso em: 7 fev. 2014.

Dos SANTOS, R. P. Big Data as a Mediator in Science Teaching: A Proposal. *JETERAPS – Journal of Emerging Trends in Educational Research and Policy Studies*, v.5, n.2, 2014b.

_____. Big Data como um mediador no Ensino de Ciências: um estudo de caso. XV EPEF – Encontro de Pesquisa em Ensino de Física, 27 a 31 de outubro de 2014, Maresias, SP. *Anais....* São Paulo: SBF – Sociedade Brasileira de Física, 2014a.

FIELD, H. Causation in a Physical World. In: LOUX, M. J.; ZIMMERMAN, D. W. (Eds.). *Oxford Handbook of Metaphysics*, 2003. Oxford: Oxford University Press.

GANTZ, J.; REINSEL, D. *The Digital Universe in 2020*: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. Framingham, MA, 2012.

GOOGLE INC. *Google Inc. Announces First Quarter 2013 Results*. Mountain View, CA, 2013.

[12] http://www.searchcorrelations.com

IBM. *What is big data?* s.d. Disponível em: <http://www-01.ibm.com/software/data/bigdata/>. Acesso em: 10 maio 2013.

MATTMANN, C. A. Computing: A vision for data science. *Nature*, v.493, n.7433, p.473-475, 2013.

MOHEBBI, M. H.; VANDERKAM, D.; KODYSH, J. et al. *Google Correlate Whitepaper*. 2011.

PAPERT, S. A. *Logo: computadores e educação*. São Paulo: Brasiliense, 1985.

SEARLS, D. *People will do more with Big Data than big companies can* [Blog post]. Disponível em: <http://blogs.law.harvard.edu/doc/2013/05/01/people-will-do-more-with-big-data-than-big-companies-can/>. Acesso em: 7 maio 2013.

SEGEV, E.; BARAM-TSABARI, A. Seeking science information online: Data mining Google to better understand the roles of the media and the education system. *Public Understanding of Science*, v.21, n.7, p.813-829, 2012.

SULLIVAN, D. *Google*: 100 Billion Searches Per Month, Search To Integrate Gmail, Launching Enhanced Search App For iOS. Disponível em: <http://searchengineland.com/google-search-press-129925>. Acesso em: 8 maio 2013.

TUFTE, E. R. *The Cognitive Style of PowerPoint: Pitching Out Corrupts Within*. Cheshire, CT: Graphics Press, 2006.

YIN, C.; SUNG, H.-Y.; HWANG, G.-J. et al. Learning by Searching: A Learning Environment that Provides Searching and Analysis Facilities for Supporting Trend Analysis Activities. *Journal of Educational Technology & Society*, v.16, n.3, p.286-300, 2013.

ZIKOPOULOS, P. C.; DEROOS, D.; PARASURAMAN, K.; et al. *Harness the Power of Big Data*: The IBM Big Data Platform. New York: McGraw-Hill, 2013.

Renato P. dos Santos – Docente e pesquisador do Programa de Pós-Graduação em Ensino de Matemática e Ciências da ULBRA Canoas.