

UMA REVISÃO SOBRE O POTENCIAL E A REALIDADE DO USO DE DADOS NÃO ESTRUTURADOS

Rita Maria Cupertino Bastos¹

Nandara Simas Frauches¹

Carla Bernadete Madureira Cruz¹

1. Universidade Federal do Rio de Janeiro - Departamento de Geografia - Rua Lobo Carneiro 6300-6410, Centro de Ciências Matemáticas e da Natureza (CCMN) - Cidade Universitária da Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil (ritamcuper@gmail.com; simasnandara@gmail.com; carlamad@gmail.com)

RESUMO

O Big Data, interpretado como grandes volumes de dados disponíveis nas redes atualmente, é encontrado de maneira crescente no mercado e apresenta grande potencial em pesquisas acadêmicas. Esses dados podem ter sua tipologia dividida em três tipos, organizados pela forma como são encontrados: dados estruturados, semiestruturados e não estruturados. Cerca de 92% dos dados são do tipo não estruturados (CRUZ, 2022), uma vez que esses são encontrados no formato textual, pictórico ou audiovisual, obtidos a partir de diversas fontes diferentes, como em redes sociais. Esses tipos de dados são complexos em seu manejo e interpretação, além necessitar uma extensa capacidade de processamento. Os desafios encontrados para trabalhar e analisar dados não estruturados apontam para uma deficiência de pesquisas acadêmicas, apesar de ter potencialidades cada vez maiores, ou, muitas vezes, acabam não utilizando a nomenclatura da tipologia correta para essa informação, impossibilitando alcançar a capacidade que este tipo de análise poderia gerar. Desta forma, o objetivo da pesquisa se encontra na análise teórica-conceitual sobre Big Data e Dados Não Estruturados e suas potencialidades para o uso em pesquisas geográficas e demais áreas científicas, assim como apresentações de estudos de casos como forma de demonstrar aplicações potenciais na Geografia, além de focar em estudos de casos do *Google Trends*, entendendo a transformação de dados não estruturados já traduzidos em mapeamentos temáticos.

Palavras-chave: Big Data, Dados Não Estruturados, Geotecnologias.

ABSTRACT

Big Data, interpreted as tools that analyze large volumes of data available on networks today, is increasingly found in the market and has great potential in academic research. These data can have their typology divided into three, organized by the way they are found: structured, semi-structured and unstructured data. About 92% of the data are unstructured (CRUZ, 2022), since they are found in textual, pictorial or audiovisual format, obtained from several different sources, such as social networks. These types of data are complex to handle and interpret, in addition to requiring extensive processing capacity. The challenges encountered in working and analyzing unstructured data generate a deficiency of academic research, despite having potential increasingly larger, or, many times, they end up not using the correct nomenclature for this information, making it impossible to reach the capacity that this type of analysis could generate.

Therefore, the objective of the research lies in the theoretical-conceptual analysis of Big Data and Unstructured Data and their potential for use in geographic research and other scientific areas, as well as presentations of case studies as a way of demonstrating potential applications in Geography and the case study from *Google Trends*, in an attempt to understand the transformation of unstructured data into already translated thematic maps.

Keywords: Big Data, Unstructured Data, Geotecnologies..

INTRODUÇÃO

A localização do homem em seu espaço vivido surge junto ao início das civilizações (CASTIGLIONE, 2019) tornando-se necessária, desde tempo imemoriais, a elaboração de mapas e a análise do espaço. Nas últimas décadas, o acesso à informação, aos dados e aos mapas tem sido cada vez mais difundido pelo mundo, muito por conta do que Milton Santos aponta como revolução do meio técnico-científico-informacional (SANTOS, 2002).

Plataformas como Google Earth e Google Maps estão cada vez mais presentes no cotidiano da população, assim como os aplicativos de delivery, à exemplo do iFood e do Uber Eats. Ambos os aplicativos utilizam a geolocalização como um de seus pilares e adotam o uso de diferentes tipos de dados em suas análises e em seu funcionamento interno. Os dados disponíveis atualmente podem ser agrupados em três tipos de formato: os dados estruturados, os dados semiestruturados e os dados não estruturados.

Com o crescente avanço do uso da internet e das redes sociais, há um aumento na produção de dados não estruturados e, cada vez mais, em menor período de tempo. Grande parte dos dados produzidos vindos de softwares computacionais são categorizados neste formato, uma vez que são englobados fotos, vídeos, textos de diversas fontes.

Atualmente, o Big Data e a potencialidade dos dados produzidos neste contexto têm sido explorados com frequência no mercado, porém o mesmo não ocorre no cenário das pesquisas acadêmicas, com especial atenção para a Geografia. Com isso, tornam-se fundamentais estudos que abarquem as dinâmicas teórico-metodológicas referentes ao uso potencial e às aplicabilidades dos dados não estruturados.

Esse trabalho buscou caracterizar e diferenciar tipos de dados com foco nas aplicações e potencialidades dos dados não estruturados em usos de caráter geográfico. Para isso, foi elaborado um trabalho de cunho teórico com apresentação do tema e de estudos de caso. Consiste, dessa forma, em trazer questões conceituais sobre os dados não estruturados e apresentar aplicações e potencialidades de sua utilização em pesquisas geográficas e demais áreas do conhecimento.

METODOLOGIA

Em um primeiro momento, foi realizado um levantamento bibliográfico de cunho teórico-conceitual com o uso de palavras chaves como “dados”, “tipos de dados”, “Big Data”, “dados não estruturados”, “geografia e dados não estruturados” e “geografia e Big Data” no Periódicos da Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES), na plataforma Minerva das bibliotecas da Universidade Federal do Rio de Janeiro (UFRJ), no ResearchGate, no LinkedIn e no Google Acadêmico.

Em seguida, foi realizada uma revisão bibliográfica acerca da origem do uso de dados não estruturados na academia e sobre Big Data, buscando compreender seu contexto de surgimento e suas complexidades. A partir disso, foi feita uma busca e análise de demais pesquisas que tinham como proposta abordar as aplicações dos dados não estruturados, bem como de suas potencialidades e limitações. A princípio, o objetivo geral era realizar essa análise em meio a trabalhos no âmbito da geografia. Contudo, no decorrer da pesquisa foi observado uma baixa quantidade de trabalhos de geógrafos envolvendo dados não estruturados e, por isso, optou-se por uma expansão dessa análise envolvendo outras áreas do conhecimento.

Realizou-se ainda um estudo de caso envolvendo dados não estruturados em pesquisas geográficas. Para a construção do estudo, foi efetuada uma análise acerca da plataforma gratuita Google Trends, obtendo dados não estruturados já traduzidos em formato csv. O dado foi exportado para o Excel onde, após formatação, pode ser incorporado a um sistema de informações geográficas - neste caso, o ArcMap 10.5. Com elementos tabulares semelhantes, foi possível fazer a junção de dados e tornar possível a construção de mapeamentos temáticos passíveis a diversas análises no âmbito científico e acadêmico na Geografia.

RESULTADOS

O CONTEXTO DO BIG DATA

O entendimento do que significa Big Data é fundamental para poder compreender os dados não estruturados. O Big Data tem tradução literal “grandes dados” e pode ser interpretado como grande volume de dados disponíveis em diferentes redes, sejam elas na internet ou em empresas, muitas vezes maior do que a capacidade de que softwares conseguem processar e analisar. (BRUCE et al., 2013; SAES, 2018). Para entender o Big Data, é fundamental entender os principais conceitos no qual esses grandes volumes de dados são atribuídos, sendo eles o volume, variedade e velocidade, também chamado de os “3 Vs”. Em primeiro lugar, temos o volume relacionado diretamente com a grande quantidade crescente de dados que precisam ser processados diariamente. Já a variedade é representada pela grande diversidade

de dados que são possíveis de obter, e que podem ser adquiridos de diferentes fontes e são passíveis de análises distintas. Por último, temos a velocidade se referindo a rapidez eficaz que a grande capacidade de processamento que o Big Data possibilita, por muitas vezes, em tempo real.

Com o passar do tempo, foram adicionados mais conceitos e potencialidades ao explorar a capacidade do Big Data e foram incluídas outras definições, como veracidade e valor, sendo elas relacionadas à exatidão de um dado após o tratamento necessário para sua utilização e ao custo de adquirir e tratar esse enorme volume de dados, respectivamente. Todo este contexto tem apontado para um potencial até pouco tempo pouco explorado, que está relacionado ao conceito de “mercadoria”. (FÉLIX et al., 2018; SAES, 2018). Desta forma, tem crescido bastante os investimentos tecnológicos e científicos para ampliar a exploração destes dados, o que envolve necessidades de armazenamento, mineração e análise. Ressaltando ainda que, mais recentemente, um conjunto particular de dados tem crescido bastante, ao ponto de predominar na atualidade. Este é o caso dos dados geoinformacionais, que são aqueles que carregam localização no espaço e no tempo, além de uma descrição. O Big Data é responsável pelo seu próprio processo de armazenamento e suas análises que são modificados para cada estrutura de dados, uma vez que os dados presentes tendem a se organizar e são reunidos para que consigam interagir entre si e, assim, diferentes tipos de bancos de dados consigam ser utilizados simultaneamente e posteriormente a sua criação. (FÉLIX et al., 2018).

Com a crescente disponibilidade de um volume e variedade de dados tão grandes, salienta-se a importância de estudos sobre Big Data em atendimento a diferentes análises. Apesar de já ter sua relevância reconhecida, observa-se ainda a necessidade de ampliar a literatura em relação ao estudo acadêmico teórico acerca dos mesmos.

DADOS ESTRUTURADOS, SEMI ESTRUTURADOS E NÃO ESTRUTURADOS

Para trabalhar os dados de forma mais adequada possibilitando extrair o máximo de potencial de uso de suas informações, é necessário primeiramente compreender sua tipologia. De acordo com Li et al. (2008), os dados estruturados são aqueles organizados mediante uma estrutura rígida previamente estabelecida para os armazenar, como banco de dados, formulários de cadastro, Cadastro de Pessoa Física (CPF), endereços, entre outros (LI et al. op cit; CRUZ, 2020; REIS, 2020). Por outro lado, os dados não estruturados são aqueles que não seguem nenhuma diretriz de organização (LI et al., 2008; CRUZ, 2020). Os dados não estruturados podem estar

representados em forma de textos, imagens, sons, vídeos etc, tais como em redes sociais, imagens de satélite e sensores (REIS, 2020). Contudo, a questão em relação às imagens de satélite ainda não é consenso entre os pesquisadores, tendo em vista que se trata de um dado matricial que possui uma certa organização. Já os dados semiestruturados consistem em dados cuja organização está entre os dois anteriores, que contam com uma estrutura considerada flexível (BUNEMAN, 1997; CRUZ, 2020). A figura 1, a seguir, exemplifica as três tipologias de organização da estrutura de dados.

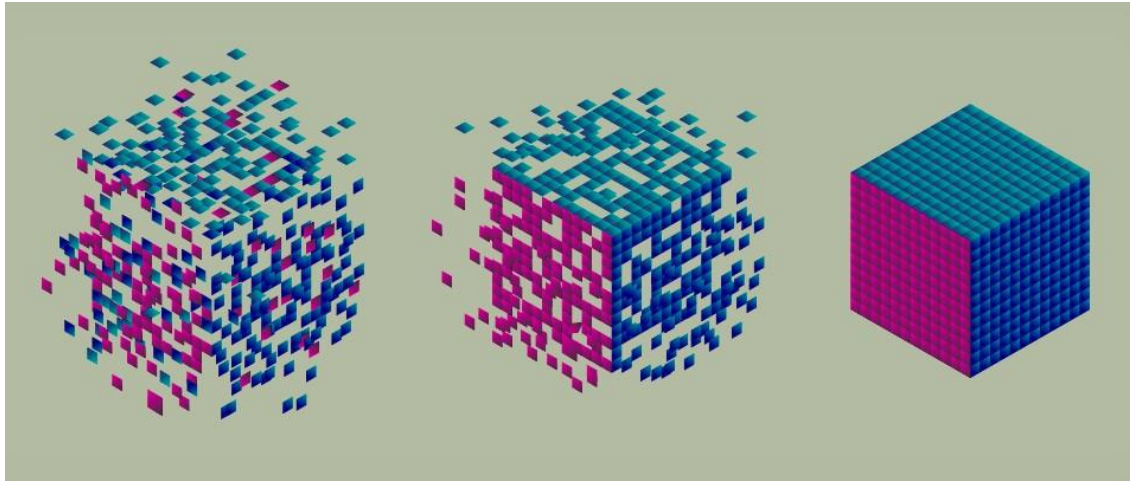


Figura 1: Tipologia da organização de estrutura de dados. Fonte: Google Imagens; Treina Web.

Os dados não estruturados vem crescendo de importância nas análises ao redor do mundo nas últimas décadas, muito por conta da evolução do Big Data e da popularização do uso computacional, principalmente através das redes sociais, atualmente, cerca de 92% das informações geradas por uma organização está no formato não estruturado (CRUZ, 2022). Contudo, dos três tipos de dados listados anteriormente, os dados não estruturados apresentam uma maior complexidade em seu manejo e utilização devido, principalmente, aos 3 Vs que o caracterizam. Este volume tem exigido um tipo de processamento diferenciado, normalmente apoiado pelo uso da nuvem. Outro aspecto negativo é a falta de abordagem no meio acadêmico sobre a diferenciação dos tipos de dados, fazendo com que muitas pesquisas adotem dados não estruturados em suas metodologias de análise mesmo sem saberem que se trata deste tipo de dado, o que gera uma certa invisibilização nas pesquisas que adotam o uso de dados estruturados.

Trabalhar com dados não estruturados traz complexidades que envolvem a necessidade de uma boa mineração (seleção do que é relevante à aplicação) que permita a percepção de padrões espaciais e temporais de fenômenos. O resultado desta mineração é então organizado em uma estrutura tabular que apoia a geração de

gráficos e mapas. Para isso, diferentes métodos de tradução de dados não estruturados se fazem necessários, sendo os mais utilizados o Machine Learning, Deep Learning, Google Trends, nuvem de palavras, gráfico de impacto e análise de conteúdo (SCARINCI, 2010; ARAÚJO E CASTRO E CECÍLIO, 2015; MELO, 2018; SAES, 2018; SANTANA, 2019; IBGE, 2022). Vale ressaltar que a realização dessa tarefa manualmente pode se tornar cansativa e suscetível a erros humanos por conta de seu grande volume. A seguir, nas figuras 2 e 3, apresenta-se exemplos de algumas técnicas de tradução de dados não estruturados mais utilizadas.

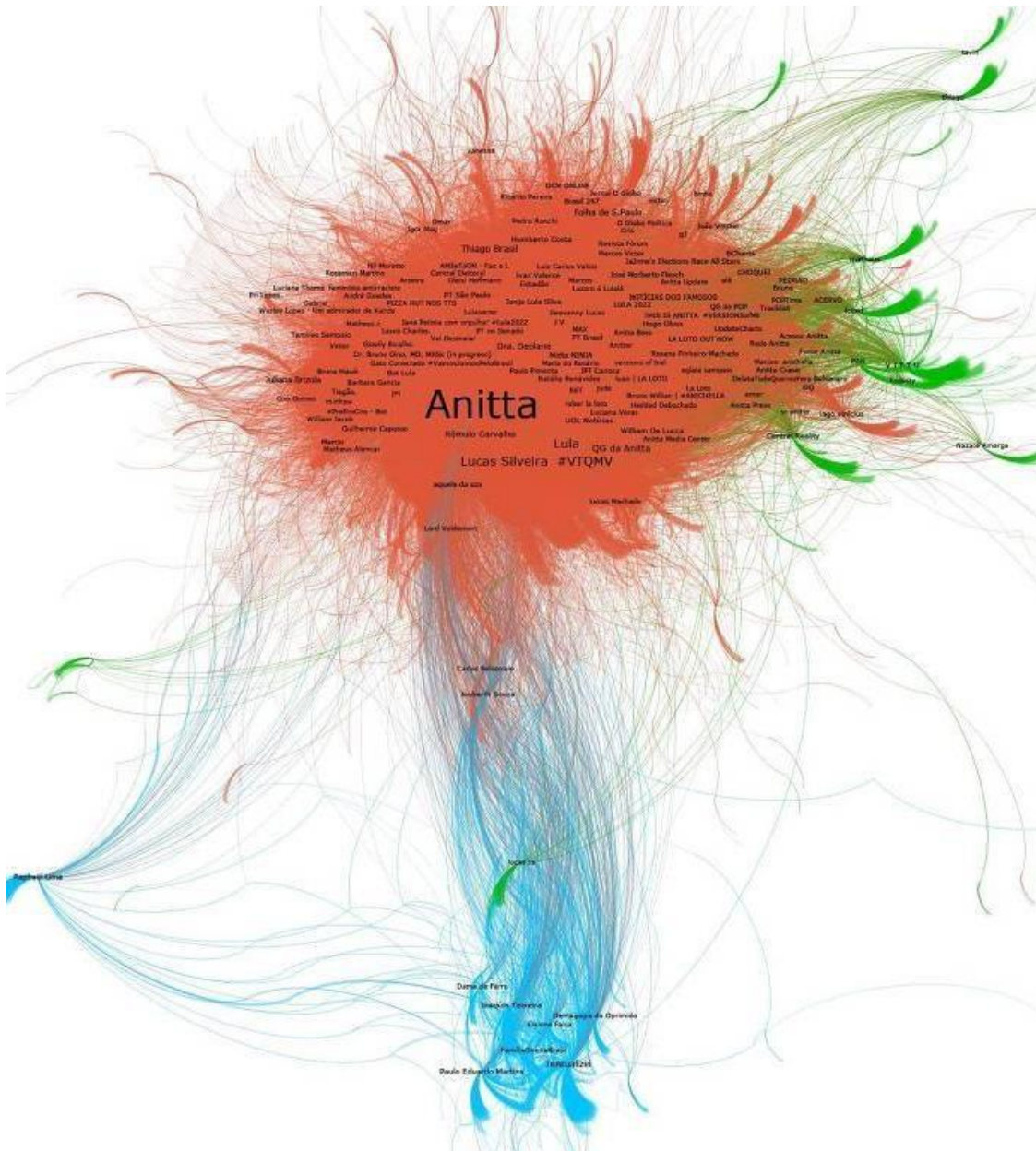


Figura 2: Gráfico de impacto relacionando Lula e Anitta nas redes sociais buscando apontar possíveis impactos do apoio da cantora ao candidato nas eleições de 2022. Fonte: Folha de São Paulo, 2022.



Figura 3: Nuvem de palavras formadas a partir de respostas de alunos à pergunta “O que se espera de um bom monitor?” De Almeida, De Freitas e Santana (2018).

Muitos estudos embasados em dados não estruturados focam em análises que utilizam dados disponíveis nas redes sociais, livros, artigos e mídias gerais, podendo estar representados na forma de vídeos, áudios, textos e imagens, entre outros. Na geografia, sua aplicabilidade gira em torno das representações e conexões espaciais, e temporais, destes dados. Apesar de pouco explorado, observa-se o incremento de interesse no potencial destes dados, principalmente em áreas como geomarketing, geografia política e segurança pública. Todo este contexto demonstra um leque muito amplo de possibilidades para a utilização deste conjunto de dados na ciência geográfica. Esta situação pode ser encontrada em aplicativos de venda e análise de expansão de negócios, como o caso do estudo sobre varejo virtual da Magazine Luiza (FÉLIX, TAVARES E CAVALCANTI, 2018). Outra aplicação interessante é o Fogo Cruzado, que surgiu em 2016 e consiste no monitoramento de tiros e disparos de armas de fogo no Rio de Janeiro, e que recolhe essas informações nas redes sociais, informes policiais e notícias da mídia, organizando, classificando e georreferenciando esses dados.

A seguir, traz-se um estudo de caso recente, a título de exemplo, focado na exploração do Google Trends para a compreensão de padrões espaciais utilizando dados não estruturados.

ESTUDO DE CASO: ANÁLISE DO GEOUPDATES

O GeoUpdatesBR foi um perfil na rede social Instagram criado em 2020 por alunos do departamento de Geografia na Universidade Federal do Rio de Janeiro (UFRJ). O objetivo do perfil era trabalhar com informações atuais e transformá-las em mapas do Brasil (ou em escalas estaduais do território brasileiro), a fim de divulgar dados importantes no período da pandemia de Covid-19. Nesta aplicação, a aquisição de dados foi sempre realizada através do Google Trends, cuja mineração possibilitou a exportação de tabelas com os conteúdos solicitados. Isto permitiu que a geração de mapas temáticos para as análises espaciais fosse efetuada através do Arcmap, tendo-se assim melhores recursos para a geração dos layouts. Os mapas gerados envolveram diferentes conteúdos, trazendo informações interessantes relativas à área da saúde, geomorfologia e socioeconomia.

Inicialmente, utilizou-se dados disponibilizados pelo Ministério da Saúde sobre o número de casos da Covid-19. Com o objetivo de abranger outros temas considerados de interesse atual, expandiu-se a busca para órgãos governamentais que disponibilizam dados estruturados e espacializados, com alto potencial para a elaboração de mapas temáticos. No decorrer do tempo, foi constatado que grande parte dos dados estavam defasados, uma vez que o último censo demográfico, instrumento responsável pela coleta e divulgação de grande parte dos dados socioeconômicos em nível nacional, realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), ocorreu no ano de 2010. Desta forma, buscando-se dados mais atualizados e representativos do momento presente, propôs-se a utilização da enorme base de dados disponibilizada pela Google, alimentada pelas buscas diversas da sociedade, e cuja interface de consulta é o Google Trends. Ressalta-se aqui que esta é uma das grandes potencialidades da análise deste tipo de dado, que possibilita, entre outras facetas, a percepção de dinâmicas temporais aceleradas, sejam elas cíclicas ou eventuais.

O Google Trends, criado em 2006 é, portanto, uma plataforma gratuita da Google responsável que permite disponibilizar ao público assuntos e palavras-chaves utilizados mundialmente com maior frequência nas interfaces de busca da empresa Google. De acordo com StatCounter Global Stats, até junho de 2022, a Google seria responsável por cerca de 91,86% de todas as pesquisas na internet, o que faz com que a empresa receba bilhões de dados não estruturados por dia, dado que os dados recebidos massivamente se encontram em formato textual e sem organização estrutural prévia. A plataforma do Google Trends funciona, desta forma, como uma maneira de disponibilização dos bilhões de dados recebidos nas pesquisas da Google de forma estruturada e já traduzida, gerando gráficos, dados tabulares, mapas em

diferentes recortes espaciais e abrangência, em apoio a análises diversas. Traz ainda os principais contextos relacionados a essas buscas, o que possibilita a compreensão de certos aspectos qualitativos associados aos termos pesquisados. É possível também, fazer a comparação de diversos assuntos/temas, e observar padrões de ocorrência espacial e temporal do interesse da sociedade, podendo-se estabelecer ainda o intervalo espacial (global, nacional, regional) e temporal de interesse.

Lembrando que a intenção do perfil GeoupdatesBR no Instagram era divulgar dados espacializados sobre temas atuais no Brasil, de acordo com a sua relevância no momento da postagem, buscando contribuir para o entendimento de como se comporta a pesquisa e o interesse sobre certos temas no país, sejam de caráter político ou de natureza social, como o apresentado nas figuras 4 e 5. Na plataforma, é possível ver os temas em destaque diariamente, o que pode ser interessante para determinados assuntos. O monitoramento da adesão em termos de interesse (ou likes) de cada tema postado, era considerada na escolha de novos temas, de modo a manter conteúdos de maior interesse ativos.

Os dados foram traduzidos da forma textual original e normalizados (transformados em dados quantitativos) para facilitar a comparação entre dois termos distintos. Essa normalização foi feita considerando os totais obtidos para cada termo, dividindo-se o total de pesquisas realizadas e o intervalo de tempo, pela população da área analisada, utilizando-se, desta forma, uma análise relativa para que seja facilitada a comparação espacial. Os dados foram exportados da plataforma para o Excel, tratados e transferidos para o software ArcMap, onde a partir da junção da tabela de atributos do shapefile do espaço escolhido para o estudo e da tabela de dados normalizados do Google, produzindo mapas temáticos.

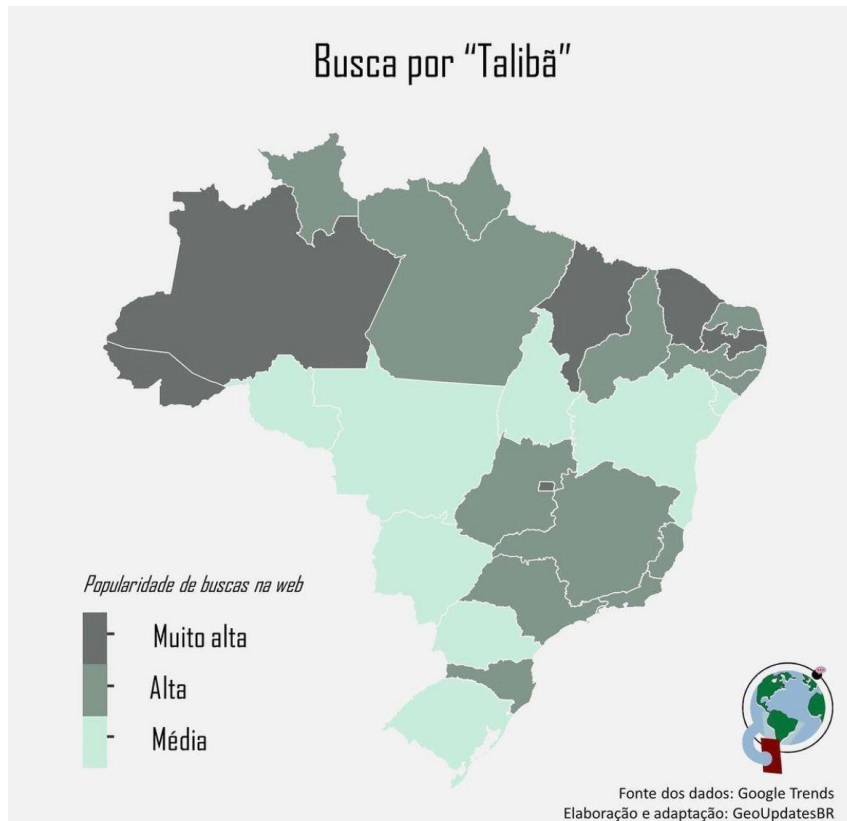


Figura 4. Busca por Talibã após a invasão ao Afeganistão em 2021. Fonte: GeoUpdatesBR. Elaboração: Nandara Simas Frauches.

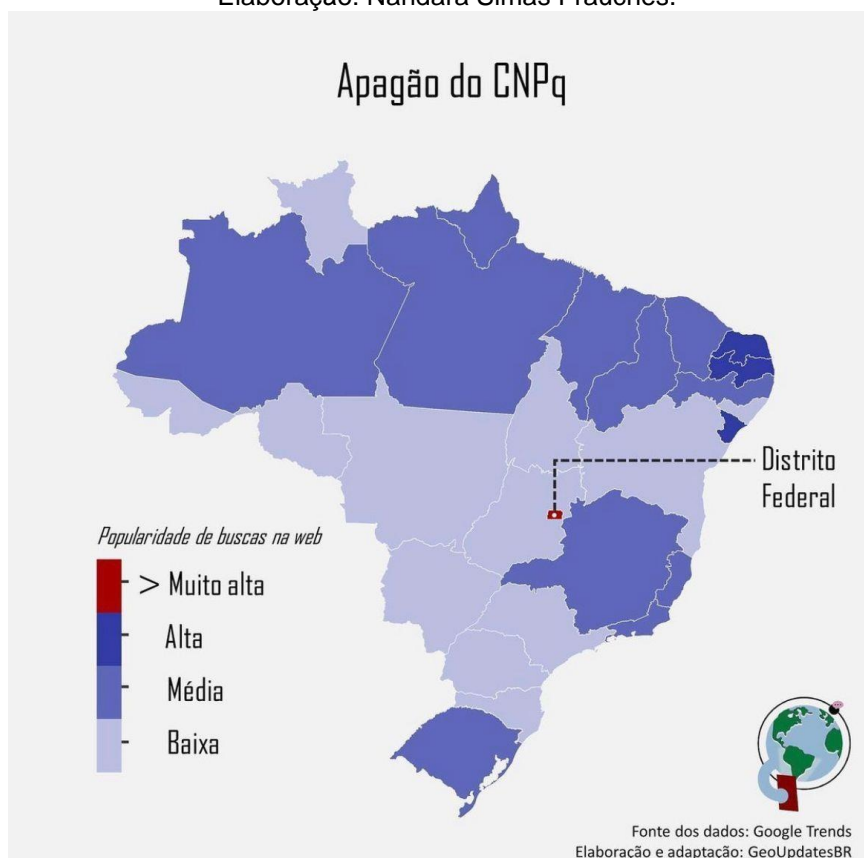


Figura 5. Busca por "Apagão do CNPq" após a plataforma do CNPq estar indisponível por tempo ilimitado no ano de 2021. Fonte: GeoUpdatesBR. Elaboração: Nandara Simas Frauches.

A plataforma possibilita, tanto espacialmente quanto temporalmente, fazer diversos estudos e análises diferentes voltados à percepção do interesse da sociedade em diferentes abrangências e com diferentes escalas de detalhamento. O potencial dessa pesquisa traz novas oportunidades de análises interessantes para a geografia, e que apesar de ainda pouco exploradas, podem gerar um novo ponto de vista a partir da relevância e interesse que diferentes temas são encontrados na web.

CONCLUSÕES

Assim, compreendendo que cerca de 92% dos dados no Big Data são encontrados de forma não estruturada (CRUZ, 2022), reforça-se a importância da compreensão e dos desafios a ultrapassar para que seja possível o seu uso, tanto em termos de armazenamento e disponibilização, como de mineração e processamento. Além disso, a invisibilidade dos dados não estruturados em pesquisas acadêmicas ainda gera uma significativa limitação para uma análise mais aprofundada. Por ser um contexto relativamente novo e dependente de soluções tecnológicas de grande porte, acredita-se que muitas destas lacunas devam ser superadas em um curto período de tempo, principalmente pela importância atual em diversas áreas do conhecimento. Compreender seu significado também envolve a necessidade de se entender as formas de tradução em que esses tipos de dados são submetidos, como é o caso das lógicas computacionais de machine learning e deep learning, ou de representação através de nuvem de palavras, por exemplo.

Apesar das redes sociais e outros meios de divulgação ampla de dados não estruturados estarem presentes nas pesquisas geográficas, ainda é incipiente seu uso na Geografia, apesar do forte apelo para muitas áreas de interesse geográfico. A exploração deste universo de dados, quanto ao potencial latente de percepção de padrões espaciais e temporais, deve ser reconhecida como uma das formas atuais mais efetivas de trazer respostas rápidas e representativas do interesse público em muitas temáticas.

REFERÊNCIAS BIBLIOGRÁFICAS

ARAÚJO E CASTRO, Rodrigo; CECÍLIO, Sumaya Gicarola. Análise de dados não estruturados: Mineração de Textos. 2015. Disponível em <https://www.researchgate.net/publication/313532455_Analise_de_dados_nao_estruturados_Minerao_de_Textos> Acesso em 20 jul 2022.

BRUCE, Weinburg., LENITA, Davis. & PAUL, D. Berger. Perspectives on big data. *Journal of Marketing Analytics*, 2013. v.1(4), p. 187-201.

BUNEMAN, Peter. Semistructured Data. In: Sixteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, 16, 1997. Proceedings [...], p. 117-121, mai. 1995. Disponível em: <https://dl.acm.org/citation.cfm?id=263675>. Acesso em: 02 fev. 2020.

CASTIGLIONE, Luiz Henrique Guimarães. Epistemologia da Geoinformação: uma análise histórico-crítica. Tese de doutorado UFF/IBICT/PPGCI, 2009, 371pg.

CRUZ, Camila Lauria Zenke da. Análise do Classificador Random Forest no Mapeamento da Cobertura e Uso da Terra no Contexto da Mata Atlântica no Estado do Rio de Janeiro. Rio de Janeiro, 2021. Dissertação de Mestrado - Geografia, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, 2021.

CRUZ, Carla Cristina. Uma breve explicação sobre dados estruturados, semi-estruturados e não estruturados. LinkedIn, 2020. Disponível em: <https://pt.linkedin.com/pulse/uma-breve-explica%C3%A7%C3%A3o-sobre-dados-estruturados-e-carla-c-p-cruz> Acesso em: 18 jul 2022.

DE ALMEIDA, Maria Isabel Chicó; DE FREITAS LIMA, Mayara Lopes; SANTANA, Otacílio Antunes. MONITORIA: UMA ANÁLISE NA PRÁTICA EDUCATIVA À LUZ DOS SABERES NECESSÁRIOS DE PAULO FREIRE.

FÉLIX, Bruno Muniz; TAVARES, Elaine; CAVALCANTE, Ney Wagner Freitas. Fatores críticos de sucesso para adoção de Big Data no varejo virtual: estudo de caso do Magazine Luiza. Revista Brasileira de Gestão de Negócios, v. 20, p. 112-126, 2018.

FOGO CRUZADO. Disponível em <https://fogocruzado.org.br/sobre> Acesso em 20 de jul 2022.

FOLHA DE SÃO PAULO. Entenda como Anitta ao declarar voto em Lula pode impactar as eleições de 2022. Disponível em <https://www1.folha.uol.com.br/ilustrada/2022/07/entenda-como-anitta-ao-declarar-voto-em-lula-pode-impactar-as-eleicoes-de-2022.shtml> Acesso em 20 jul 2022.

GEOUPDATESBR. Disponível em <https://www.instagram.com/geoupdatesbr/> Acesso em 19 jul de 2022.

LI, Guoliang. *et al.* EASE: an effective 3-in-1 keyword search method for unstructured, semi-structured and structured data. In: 2008 ACM SIGMOD international conference on Management of data, 2008, Vancouver, Canadá. Proceedings [...], Vancouver: SIGMOD '08, p. 903-914, jun.

MELO, Fabrício Silva. Extração de relações a partir de dados não estruturados baseada em deep learning e supervisão distante. 2018. 71 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Sergipe, São Cristóvão, SE, 2018.

REIS, Cássio. Dados estruturados x não estruturados no contexto de Big Data. LinkedIn, 2020. Disponível em: <https://pt.linkedin.com/pulse/dados-estruturados-x-n%C3%A3o-contexto-de-big-data-cassio-reis?trk=pulse-article> Acesso em: 18 jul 2022.

SAES, Keylla Ramos. Abordagem para integração automática de dados estruturados e não estruturados em um contexto Big Data. 2018. Tese de Doutorado. Universidade de São Paulo.

SANTANA, Brenda Salenave. Extração e aplicação de indicadores no processo de recomendação de recursos urbanos utilizando dados estruturados e não-estruturados. 2019

SANTOS, Milton. A natureza do espaço: técnica e tempo, razão e emoção. Edusp, 2002.

SCARINI, Rui Gureghian, & OLIVEIRA, José Palazzo Moreira de. (2010). Extração de informação como base para descoberta de conhecimento em dados não estruturados. *Cadernos De Informática*, 1(1), 15–20. Recuperado de <https://www.seer.ufrgs.br/index.php/cadernosdeinformatica/article/view/v1n1p15-20>

STATCOUNTER GLOBAL STATS. Search Engine Market Share Worldwide. Disponível em <https://gs.statcounter.com/search-engine-market-share> Acesso em 19 jul 2022.

TREINA WEB. O que são dados estruturados e não estruturados? Disponível em <<https://www.treinaweb.com.br/blog/o-que-sao-dados-estruturados-e-nao-estruturados>> Acesso em 20 jul 2022.

TSAI, Chun-Wei, LAI, Ching-Feng, CHAO, Han-Chieh & ATHANASIOS, Vasilakos. Big data analytics: a survey. Journal of Big Data, 2015.