

XVI CIAEM 

Conferencia Interamericana de Educación Matemática
Conferência Interamericana de Educação Matemática
Inter-American Conference of Mathematics Education

 UNIVERSIDAD DE LIMA

Lima - Perú
30 julio - 4 agosto 2023



xvi.ciaem-iacme.org

Análise de Conteúdo Automatizada: evidências nas pesquisas em Educação Matemática

Leonardo Martins

Universidade Anhanguera de São Paulo

professor@leomartins.net

Karla Jocelya Nonato

Universidade Federal de Mato Grosso do Sul

karlanonato@yahoo.com.br

Nielce Meneguelo Lobo da Costa

Universidade Anhanguera de São Paulo

nielce.lobogmail.com

Resumo

Este estudo visa identificar evidências do uso de *software* de apoio na análise de conteúdo nas pesquisas em Educação Matemática, nas quais a Análise de Conteúdo Automatizada foi utilizada como metodologia de pesquisa, em particular com os *softwares* Iramuteq, Nvivo e/ou CHIC. Para tanto, os procedimentos metodológicos, dos quais foram constituídos por busca de conjuntos de termos relacionados ao estudo, via programa Harzing's Publish or Perish, na base do Google Scholar. Os dados foram analisados com apoio do *Software* Iramuteq e os resultados apresentados em Diagrama de Zipf e em Nuvem de Palavras. Constatamos, o uso automatizado para análises dos dados da pesquisa, por parte de pesquisadores sem, contudo, haver a menção de que se tratava de uma Análise de Conteúdo Automatizada.

Palavras-chave: Recursos Tecnológicos; Software para análise de dados; Iramuteq; Nvivo; CHIC.

Resumen

Este estudio tiene como objetivo identificar evidencias del uso de software de apoyo en el análisis de contenido en la investigación en Educación Matemática, en el que se utilizó el Análisis de Contenido Automatizado como metodología de investigación, en particular con el software Iramuteq, Nvivo y/o CHIC. Para ello, los

procedimientos metodológicos, que se constituyeron por la búsqueda de conjuntos de términos relacionados con el estudio, a través del programa Harzing's Publish or Perish, en la base de datos de Google Scholar. Los datos se analizaron utilizando el software Iramuteq y los resultados se presentaron en Zipf Diagram y Word Cloud. Encontramos el uso automatizado para el análisis de datos de investigación por parte de investigadores sin mencionar, sin embargo, que se trataba de un Análisis de contenido automatizado.

Palabras llave: Recursos Tecnológicos; Software de soporte para el analizar datos; Iramuteq; Nvivo; CHIC.

Introdução

Atualmente há muito investimento para desenvolver a Inteligência Artificial (IA), que por um lado vai facilitar e auxiliar o homem, por outro lado, causa temor quanto ao seu alcance. A IA reproduz a inteligência humana, via sistemas ou máquinas, realiza tarefas e pode aprimorar seus procedimentos a partir de informações coletadas quando o device interage com o humano (Lee, 2019). A IA pode ser aplicada de várias formas, como por exemplo, nos smartphones, na compilação de dados, nos mecanismos de segurança.

O temor em relação a IA é que ela venha nos substituir, tomando decisões por nós, ganhando vida própria, inclusive quanto à manipulação e uso de dados pessoais, que podem até ser compartilhados sem autorização prévia. Esse é um risco que pesquisadores não podem correr, pois a autonomia do cientista é fundamental para o desenvolvimento de suas pesquisas. É ele quem toma as decisões definindo referencial teórico e procedimentos metodológicos a partir das características do objeto estudado e do objetivo. A seleção, organização e interpretação dos dados é exclusividade do pesquisador, entretanto, ele pode se valer de recursos tecnológicos e da IA para tais processos.

A IA pode auxiliar no auxílio de pesquisas, incluindo nas sobre o ensino de Matemática (Jess, 2004) e na área de Educação Matemática, por exemplo, a Análise de Conteúdo (AC) (Bardin, 1977) é um referencial metodológico utilizado por inúmeros cientistas para análise dos dados. Com o avanço das tecnologias, foi desenvolvida a Análise de Conteúdo Automatizada (ACA) (Grimmer & Stewart, 2013), que é a AC realizada com o auxílio de recursos tecnológicos (*software*). Neste contexto, a proposta desta pesquisa foi identificar evidências do uso de *software* de apoio na análise de conteúdo nas pesquisas em Educação Matemática.

Análise de Conteúdo Automatizada

A AC surgiu na 1ª Guerra Mundial, enquanto a IA, ao contrário do lembrado em nosso imaginário, despontou na década de 1940 e é mais velha que a ACA. A ACA é um procedimento de síntese qualitativa e quantitativa para dados manifestados em forma de texto. Tal recurso, ao minar o texto, utiliza aprendizagem de máquina (um tipo de Inteligência Artificial) e análise de texto. O aprendizado de máquina que a ACA utiliza “se concentra no reconhecimento de padrões e em fazer previsões a partir de dados, para identificar e definir conceitos/tópicos [...] dentro de um corpo” (Nunez-Mir *et al*, 2016, p. 1263).

Vivemos um momento em que há documentos digitais disponíveis em grande quantidade, a ACA agiliza o processo de tabulação dos dados, sem contudo excluir a necessidade de uma leitura minuciosa do pesquisador, ou seja, do controle das variáveis pelo pesquisador.

Desta forma, a ACA é a AC executada com o auxílio de recursos tecnológicos e mesmo com *software* para auxiliar na ACA, a seleção e organização dos dados, a interpretação e as conclusões são realizadas pelo pesquisador (Grimmer & Stewart, 2013). Os *software* que realizam a inferência dos dados, proporcionam ao cientista, além do ganho de tempo, a visualização de relações entre os dados, pois evidencia Classes Hierárquicas e outras relações nem sempre simples de se perceber manualmente. Mais uma vez, frisamos que os *software* não substituem o pesquisador, pois é ele quem deve preparar o corpus, levantar as categorias (ou reconhecê-las a posteriori), interpretar os dados, estabelecer as relações entre categorias e chegar às conclusões.

Existem pequenas diferenças entre a ACA e a AC. A ACA examina a palavra, enquanto a AC considera os sentidos das permutas comunicativas nas unidades de análise. Os tratamentos dos dados realizado com *software* são feitos por análises matemáticas em relação ao verbete (Cúrcio, 2006). Assim a ACA e a AC são compatíveis, entretanto não são idênticas.

É primordial ao pesquisador, que opta por utilizar a ACA, se apropriar dos resultados veiculados pelos software e, a partir deles, constatar e reafirmar quais palavras são relevantes para atingir o objetivo de pesquisa. Para tanto deve seguir os procedimentos metodológicos de modo a manter a confiabilidade.

Metodologia

A IA pode e deve ser utilizada a nosso favor. Para realizarmos as buscas, utilizamos o *software Harzing's Publish or Perish*, dotado de IA, que funciona a partir de comandos humanos. Com a finalidade de atingir o objetivo da pesquisa, empregamos no método a pesquisa bibliográfica do tipo Revisão Sistemática de Literatura (Kitchenham, 2004). Sem delimitação temporal, de nacionalidade e de idioma dos textos, a consulta e identificação bibliográfica se deu no banco de dados do Google Scholar, no mês de julho de 2022, via *software* "Harzing's Publish or Perish" versão 8.2.3944.8118.

Foram utilizados os operadores booleanos AND e OR e realizadas combinações de termos de busca, sendo: I) "Análise de conteúdo automatizada" AND "Educação Matemática"; II) "Análise de Conteúdo" AND "Educação Matemática" AND "Iramuteq"; III) "Análise de Conteúdo" AND "Educação Matemática" AND "Nvivo"; IV) "Análise de conteúdo" AND "Educação Matemática" AND "CHIC"; e V) "Iramuteq" OR "Nvivo" OR "CHIC" AND "Análise de Conteúdo" AND "Educação Matemática".

Foi construído um *corpus* textual com as palavras-chave das pesquisas identificadas nas cinco consultas realizadas, no qual foi submetido a processamento no IRaMuTeQ na função estatística textual descritiva e, dentre os resultados disponibilizados, optamos pelo Diagrama de Zipf – que expõe a quantidade de vezes que uma palavra-chave foi utilizada pelos pesquisados e

a Nuvem de Palavras – que é um método de análise lexical simples, pelo qual se organiza graficamente as palavras por tamanho, em função da frequência com que aparecem nos textos.

Análise dos dados

A primeira busca realizada por ele (termos I) resultou em 3 pesquisas, que após a leitura fluante foram descartadas, pois eram de outras áreas. A segunda consulta (termos II) retornou 101 pesquisas; a terceira (termos III) apresentou 141 artigos; enquanto a quarta consulta (termos IV) 47 estudos.

O *software* Iramuteq classificou o *corpus* textual em 910 ocorrências (segmentos de textos) e em 477 formas, encontrando 362 hápax (palavras com única frequência) em um texto. O *corpus* textual foi submetido à função de estatística textual descritiva no Iramuteq. Nesse processamento, o Diagrama de Zipf forneceu o comportamento de frequência das palavras por meio de uma ilustração gráfica (Figura 1). O Diagrama de Zipf apresenta a relação entre a frequência das formas (y) e a quantidade de formas (x).

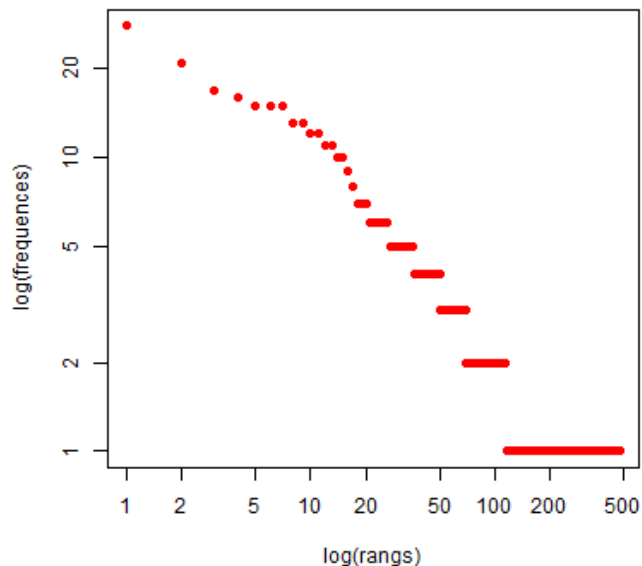


Figura 1. Diagrama de Zipf de frequência de palavras
Fonte. Autores (2022).

Observamos o seguinte resultado na Figura 1: um termo com frequência superior a 20; outro termo com aproximadamente 20 repetições; e cinco termos entre 15 e 20 usos pelos autores. Aproximadamente 350 verbetes foram utilizados uma vez como palavra-chave.

Na nuvem de palavras (Figura 1) criada, os verbetes com fonte maior são as mais utilizadas como palavras-chaves nos textos encontrados nas pesquisas, aparecendo com maior frequência no *corpus* textual e no Diagrama de Zipf (Figura 1).

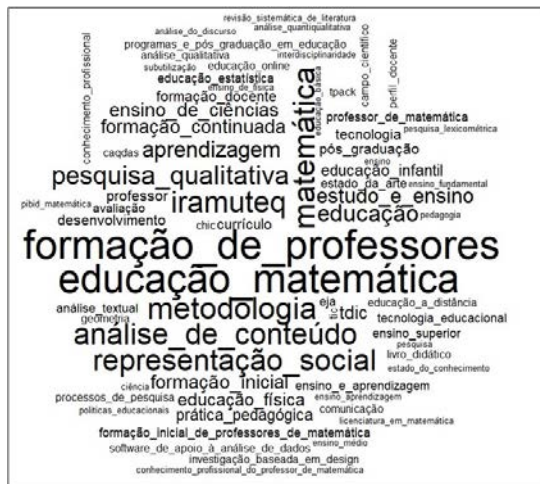


Figura 2. Nuvem de Palavras
Fonte. Autores (2022).

Na Nuvem de Palavras (Figura 2) a palavra-chave que mais se destacou foi “formação de professores”, e, segundo o Diagrama de Zipf (Figura 1), ela foi utilizada mais de 20 vezes. “Educação Matemática”, que se destaca na Nuvem de Palavras, foi mencionada como palavra-chave, aproximadamente, 20 vezes. Já “Matemática”, “Metodologia”, “Iramuteq”, “Análise de Conteúdo” e “Representação Social” estão entre 15 e 20 citações.

Em contrapartida, de acordo com o hápax, temos 362 verbetes únicos, como por exemplo, “Abordagem Investigativa”, “Ciência e Tecnologia”, “Modelagem Matemática”, “Pesquisa Quantitativa” e “Metodologia Científica”. O hápax representa justamente as palavras presentes uma vez no *corpus* textual.

Para as buscas, utilizamos os nomes dos *software* de apoio à Análise de Conteúdo Automatiza, o termo “Iramuteq” foi o mais relevante na Nuvem de Palavras, seguido pelo termo “CHIC”, como visibilidade menor, a exceção é o termo “Nvivo”, que foi inexpressivo. O termo, “Análise de Conteúdo”, dentre todos, foi o de maior destaque, pois está no grupo que foi citado entre 15 e 20 vezes.

Considerações Finais

Na ACA, os *software* de apoio utilizados, realizam uma série de combinações de análises quantitativas e qualitativas do *corpus* textual durante o tratamento dos dados. Para tal, utilizam desde a teoria dos grafos até a lematização, entre outros. Mas, a máquina, dotada ou não de IA é dependente dos comandos humanos, os resultados, com indicativo percentual de confiabilidade, dependem da organização dos dados e interpretação realizada pelo pesquisador.

Quanto as pesquisas que utilizaram *software* como Iramuteq, NVivo e CHIC, para análise de dados, mesmo não sendo citada a realização de ACA, a partir do momento em que a metodologia utilizada foi a AC, combinada ou não com outra, e o *software* utilizado gerou gráficos, representações, inferências e índices em geral, ela se caracteriza no grupo das análises de conteúdo automatizadas.

Ao buscarmos evidências do uso de *software* de apoio na análise de conteúdo no desenvolvimento de pesquisa em Educação Matemática, mostramos que a ACA é utilizada como metodologia. Mesmo com a a inexpressividade do termo “Análise de Conteúdo Automatizado” propriamente dito, a expressividade dos termos vinculados ao tema representadas no Diagrama de Zipf e na Nuvem de palavras apontam para o emprego da ACA nas pesquisas em Educação Matemática.

Apoio

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Referências e bibliografia

- Barrientos-Fernández, A., Pericacho-Gómez, F.-J. y Sánchez-Cabrero, R. (2020). Competencias sociales y emocionales del profesorado de Educación Infantil y su relación con la gestión del clima de aula. *Estudios sobre Educación*, 38, 59–78. <https://doi.org/10.15581/004.38.59-78>
- Bourguignon, R., Garaudel, P., & Porcher, S. (2020). Global framework agreements and trade unions as monitoring agents in transnational corporations. *Journal of Business Ethics*, 165(3), 517–533. <https://doi.org/10.1007/s10551-019-04115-w>
- Cúrcio, V. R. (2006). Estudos estatísticos de textos literário. *Revista Texto Digital*, (2)
- Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, (pp. 1–31).
- Jess, G. M. (2004). Dissertação. <https://www.acervodigital.ufpr.br/bitstream/handle/1884/32969/R%20-%20D%20-%20GIL%20MARCOS%20JESS.pdf?sequence=1&isAllowed=y>
- Lee, K. F. (2019). *Inteligência artificial*. Globo Livros.
- Nunez-Mir et al. Automated content analysis: addressing the big literature challenge in ecology and evolution. *Methods in Ecology and Evolution* (pp. 1262-1272). Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004), (pp. 1-26).
- Tur Marí, J. A. y Pons Biescas, A. (2005). La alimentación en el mundo Púnico. En J. Salas-Salvadó, P. García Lorda y J. M. Sánchez Ripollés (Eds.), *La alimentación y la nutrición a través de la historia* (pp. 82–112). Glosa.