

Resenha Crítica: Metodologia do Artigo “Exploração do acervo da RAE-Revista de Administração de Empresas (1961-2016) à luz da bibliometria, text mining, rede social e geoanálise”

Introdução

O artigo de Favaretto e Francisco (2017) propõe uma abordagem metodológica inovadora ao integrar técnicas quantitativas e qualitativas para analisar o acervo da RAE. Esta resenha avalia criticamente a metodologia, com foco nos processos de coleta e tratamento dos dados, tecnologias empregadas, técnicas aplicadas e limitações, conforme solicitado.

Coleta e Tratamento dos Dados

A coleta dos 2.381 documentos foi realizada via protocolo OAI-PMH, garantindo interoperabilidade e padronização dos metadados (FGV, 2017a). Scripts em Perl e R automatizaram a extração, filtragem e padronização dos dados, solucionando desafios como caracteres não padronizados e duplicidade de registros por meio de códigos únicos. A separação de idiomas foi feita por análise de títulos, palavras-chave e resumos. A extração dos textos completos dos PDFs, essencial para as análises textuais, utilizou ferramentas robustas para minimizar perdas de informação.

Apesar do sucesso na automação, persistiram desafios na padronização dos metadados, especialmente em períodos antigos, e na identificação de instituições de coautores. Recomenda-se, para estudos futuros, o uso de técnicas avançadas de machine learning, como modelos de reconhecimento de entidades nomeadas (NER) e algoritmos de desambiguação de nomes (por exemplo, spaCy ou BERT), além da integração com bases externas (como ORCID e Cross-Ref) para enriquecer os metadados. A replicabilidade pode ser ampliada com a disponibilização dos scripts e dados tratados em repositórios públicos, como o GitHub, permitindo que outros pesquisadores repliquem e aprimorem o estudo.

Tecnologias e Linguagens Utilizadas

O uso combinado de Perl e R permitiu automação eficiente e flexibilidade na manipulação dos dados (CPAN, 2017; R Core Team, 2017). O Perl foi fundamental para a extração e limpeza inicial, enquanto o R viabilizou análises estatísticas e text mining, além de facilitar a integração com bibliotecas especializadas. O emprego dessas linguagens evidencia maturidade técnica e alinhamento com as melhores práticas em ciência de dados.

Técnicas Utilizadas

Bibliometria

Foram aplicadas métricas como frequência de publicações, coautoria e centralidade de autores, permitindo mapear a produção científica e identificar padrões de colaboração (Hood & Wilson, 2001; Zupic & Čater, 2015). Além da contagem de publicações por autor, os autores aplicaram métricas de centralidade e intermediação, como grau e betweenness, para identificar os pesquisadores mais influentes e os principais termos do acervo, conforme apresentado nas Tabelas 1, 2 e 5 do artigo. A análise bibliométrica foi potencializada pela padronização dos dados, possibilitando comparações temporais e temáticas.

Text Mining

A análise de conteúdo utilizou recursos do R para identificar palavras-chave, frequências e gerar nuvens de palavras, além de aplicar a Lei de Zipf para verificar a distribuição dos termos (Feinerer et al., 2008; Zipf, 1949). O estudo utilizou a Lei de Zipf para ranqueamento das palavras mais frequentes e gerou nuvens de palavras a partir dos campos título, resumo e palavras-chave, visualizadas nas Figuras 2, 3 e 4, o que permitiu identificar os temas centrais do periódico ao longo do tempo. O uso de stopwords específicas do português e a segmentação por tipo de documento agregaram valor à análise.

Análise de Redes Sociais (SNA)

O mapeamento das redes de coautoria foi realizado por meio de SNA, identificando componentes principais, centralidade e intermediação dos autores (Haneman & Riddle, 2005). A construção das redes de coautoria foi realizada pelo método t-mode, com análise de centralidade dos autores e visualização dos componentes principais, conforme ilustrado no Gráfico 5, evidenciando a colaboração entre pesquisadores. A visualização das redes permitiu compreender a dinâmica colaborativa e a formação de comunidades científicas ao longo do tempo.

Geoanálise

A dimensão geográfica foi incorporada por meio de softwares como ArcGIS, permitindo visualizar a distribuição espacial das colaborações e temas (Câmara et al., 2009; ESRI, 2016). A dimensão geográfica foi incorporada por meio do uso do ArcGIS e da extensão Spatial Analyst, permitindo mapear a distribuição espacial das palavras-chave e autores, como apresentado no Gráfico 7. Essa abordagem inovadora ampliou o entendimento sobre a internacionalização e a abrangência dos estudos publicados.

Pontos Fortes da Metodologia

Destaca-se a integração inovadora de múltiplas técnicas, a automação e padronização dos dados, e a robustez das análises. A principal inovação reside na integração de técnicas de bibliometria, text mining, análise de redes sociais e geoanálise, proporcionando uma visão abrangente e replicável, com potencial de aplicação em outros periódicos científicos. O uso de diferentes abordagens permitiu uma visão abrangente e multifacetada do acervo, contribuindo significativamente para a área de Sistemas de Informação. A replicabilidade e a transparência metodológica são pontos altos, embora possam ser ainda mais fortalecidos.

Limitações e Possíveis Melhorias

Entre os principais desafios, destacam-se a padronização de nomes de instituições e autores, além da ausência de metadados completos em edições antigas, o que dificultou a automatização total do processo. Para superar essas limitações, sugere-se a implementação de pipelines automatizados de limpeza de dados utilizando bibliotecas como OpenRefine, além da integração de APIs de bases externas (ORCID, CrossRef) para validação e enriquecimento dos metadados. O uso de identificadores persistentes (como ORCID) para autores pode ser incorporado diretamente no processo de extração. A abertura dos scripts e dados tratados em plataformas como GitHub é fundamental para potencializar a reprodutibilidade e o impacto científico, permitindo que a comunidade contribua com melhorias contínuas.

Conclusão

A metodologia analisada representa um avanço relevante ao integrar técnicas de bibliometria, text mining, SNA e geoanálise, com automação e rigor. O trabalho se destaca pela originalidade ao propor uma abordagem multidisciplinar e replicável, servindo de referência para pesquisas na área de Sistemas de Informação. Recomenda-se, para estudos futuros, o aprofundamento na padronização dos dados, a implementação de pipelines automatizados e a abertura dos scripts em repositórios públicos, potencializando a reprodutibilidade e o impacto científico. A integração de métodos, aliada à transparência e inovação, torna este estudo exemplar e digno de nota máxima.

Referências

Abbasi, A., Hossain, L., & Leydesdorff, L. (2012). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics*, 6(3), 403-412. doi:10.1016/j.joi.2012.01.002

Câmara, G., Vinhas, L., Davis, C., Fonseca, F., & Camargo, T. (2009). Geographical information engineering in the 21st century. In G. Navratil (Ed.), *Research Trends in Geographic Information Science* (pp. 23-41). Berlin, Germany: Springer.

CPAN. (2017). The Comprehensive Perl Archive Network. Recuperado de <http://www.cpan.org>

ESRI. (2016). ArcGIS Desktop: Release 10.5. Redlands, USA: ESRI. Recuperado de <http://desktop.arcgis.com/arcmap>

Favaretto, J. E. R., & Francisco, E. R. (2017). Exploração do acervo da RAE-Revista de Administração de Empresas (1961-2016) à luz da bibliometria, text mining, rede social e geoanálise. *RAE-Revista de Administração de Empresas*, 57(4), 365-390. doi:10.1590/S0034-759020170404

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, 25(5), 1-54. doi:10.18637/jss.v025.i05

FGV. (2017a). Repositório FGV de periódicos e revistas. Biblioteca digital da RAE – Revista de Administração de Empresas. Recuperado de <http://bibliotecadigital.fgv.br/ojs/index.php/rae/issue/archive>

Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California.

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, 52(2), 291-314. doi:10.1023/A:1017919924342

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Recuperado de <https://www.r-project.org>

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429-472. doi:10.1177/1094428114562629

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.