

PERIÓDICO TRANSINFORMAÇÃO: CARACTERIZAÇÃO DE TERMOS PARA ANÁLISE DE CO-OCORRÊNCIA DE PALAVRAS – ACP (2007-2021)

CAMILA DE AZEVEDO GIBBON¹; RODRIGO AQUINO DE CARVALHO²

¹*Universidade Federal do Rio Grande - FURG – camila_gibbon@hotmail.com*

²*Universidade Federal do Rio Grande - FURG – racfurg@gmail.com*

1. INTRODUÇÃO

A análise de co-ocorrência de palavras é uma técnica que estuda as características de uma determinada área do conhecimento científico por intermédio da análise de termos e/ou expressões utilizados em partes de documentos como o título, as palavras-chave, os resumos etc. (GALVEZ, 2018). A ocorrência conjunta desses termos é, posteriormente, investigada a fim de concluir quais são as associações temáticas mais estabelecidas em um determinado contexto. A pesquisa aqui descrita representa a primeira parte da análise de co-palavras do periódico científico da área da Ciência da Informação, TransInformação¹, entre os anos de 2007 a 2021.

Desse modo, os objetivos do estudo são: caracterizar os termos e expressões que foram retirados dos documentos quanto aos seus metadados de origem, o idioma que possuem e quantificar as frequências/ocorrências.

O interesse em desenvolver esta pesquisa justifica-se pela relevância de estudar, por meio de métricas, as características que a produção científica em Ciência da Informação vêm assumindo ao longo dos últimos anos.

2. METODOLOGIA

Caracteriza-se como uma pesquisa exploratória e cientométrica. Ademais, possui uma abordagem quantitativa, ou seja, trabalha com resultados possíveis de serem transformados em informações estatísticas. (PRODANOV; FREITAS, 2013).

A definição da revista, da tipologia documental e dos anos a serem abordados ocorreu entre os meses de outubro a novembro de 2021. Para mais, a extração dos termos e expressões tanto das palavras-chave quanto do metadado de título ocorreu entre os meses de dezembro de 2021 a março de 2022. Por fim, esses dados foram inseridos em planilhas online a fim de serem organizados e analisados. A primeira conferência da qualidade dos dados ocorreu no mês de junho de 2022 e os resultados encontrados podem ser vistos a seguir. Vale ressaltar ainda que todos os dados foram extraídos, planilhados e revisados manualmente.

A Tabela 1 apresenta a distribuição dos artigos por ano de publicação e o Quadro 1 os tipos de documentos elegíveis e não para a análise proposta.

¹ ISSN: 0103-3786. Disponível em: <https://www.scielo.br/j/tinf/>.

TABELA 1 - Distribuição dos artigos utilizados na pesquisa por ano e bloco de três anos.

Ano	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Totais
Nº de artigos	21	22	18	18	18	18	21	28	25	28	29	27	27	26	14	340
%	6,2	6,5	5,3	5,3	5,3	5,3	6,2	8,2	7,4	8,2	8,5	7,9	7,9	7,6	4,1	100
Três anos		61			54			74			84			67		340
%		17,9			15,9			21,8			24,7			19,7		100

Fonte: dados da pesquisa, 2022.

QUADRO 1 - Tipologias documentais.

Utilizadas na pesquisa	Não utilizadas na pesquisa
Artigos de revisão/revisão de literatura; relatos de pesquisa; comunicações (dossiês e números especiais).	Expedientes; editoriais; entrevistas; ensaios; pontos de vista/comentários; relatos de experiência; resenhas; resumos de dissertações; traduções; suplementos em vídeo.

Fonte: autores.

Vale salientar que os termos apresentados ainda não passaram por um processo de limpeza, que teve um protocolo definido no mês de julho de 2022 e está ocorrendo atualmente. Assim, a caracterização dos termos ocorre a partir de como foram extraídos de cada um dos 340 documentos.

3. RESULTADOS E DISCUSSÃO

Os termos e/ou expressões foram extraídos dos metadados palavras-chave, como estava no artigo, e dos títulos, caso ocorresse a presença de um termo não indicado nas palavras. Vale ressaltar que o periódico em questão publica em três idiomas: português, inglês e espanhol. Sempre que possível, foram coletados os referidos dados no idioma português, no entanto, em textos que não possuíam as informações em português nem como a língua do segundo resumo, coletou-se nos demais idiomas.

A Figura 1 explana as ocorrências dos termos extraídos nos dois metadados.

FIGURA 1 - Quantidade de ocorrências por metadado extraído



Fonte: autores

Por meio da Figura 1 é possível observar que o total de ocorrências dos termos extraídos dos documentos foi de 2025, sendo que 72,1% dessas foram oriundas do metadado de palavras-chave e 27,9% do título. Além disso, desse montante total, 1649 são ocorrências em português, 365 em espanhol e 11 em inglês.

Com o auxílio da Tabela 2, analisa-se as médias, modas e a medianas das ocorrências dos termos/expressões retirados das palavras-chave, dos títulos e na soma dos dois metadados. Bem como apresenta os valores de menor e maior ocorrência apresentados em um documento e o seu desvio padrão.

TABELA 2 - dados estatísticos

	Nº total de termos	Termos do Título	Termos da PC
Total	2025	565	1460
Média	5,69	1,66	4,29
Mediana	6,00	1,00	4,00
Moda	5,00	1,00	4,00
Mínimo	3,00	0,00	2,00
Máximo	12,00	8,00	9,00
Desvio padrão	1,69	1,39	0,99

Fonte: dados de pesquisa, 2022

O total de termos extraídos nessa primeira análise de dados foi de 1415. Desses, 1187 são termos que ocorreram apenas uma vez e 228 possuem quantidade diferentes de repetições.

TABELA 3 - 42 termos/expressões mais utilizados e suas ocorrências

	Termos/expressões				Termos/expressões				
	PC	T	Total		PC	T	Total		
1	Ciência da Informação	25	06	31	22	España	02	05	07
2	Produção científica	17	09	26	23	Colaboração científica	06	-	06
3	Brasil	07	17	24	24	Classificação	05	01	06
4	Informação	15	05	20	25	Indicadores bibliométricos	05	01	06
5	Comunicação científica	16	02	18	26	Análise bibliométrica	03	03	06
6	Bibliometria	13	-	13	27	Web of Science	02	04	06
7	Gestão do conhecimento	13	-	13	28	Análise de redes sociais	05	-	05
8	Web semântica	13	-	13	29	Bases de dados	05	-	05
9	Acesso aberto	09	01	10	30	Documento	05	-	05
10	Representação da informação	08	02	10	31	Indexação	05	-	05
11	Internet	08	01	09	32	Interdisciplinaridade	05	-	05
12	Bibliotecas universitárias	07	01	08	33	Ontologia	05	-	05
13	Metadados	07	01	08	34	Organização da informação	05	-	05
14	Recuperação da informação	07	01	08	35	Periódico científico	05	-	05
15	Conhecimento	05	03	08	36	Políticas públicas	05	-	05
16	Estudo de caso	-	08	08	37	Redes sociais	05	-	05
17	Biblioteconomia	07	-	07	38	Letramento informacional	04	01	05

18	Gestão da informação	07	-	07	39	Memória	04	01	05
19	Epistemologia	06	01	07	40	Representação descritiva	04	01	05
20	Organização do conhecimento	06	01	07	41	Web 2.0	04	01	05
21	América Latina	05	02	07	42	Scopus	02	03	05

Legenda: PC – palavras-chave; T – título.

Fonte: dados de pesquisa, 2022.

Com o auxílio da Tabela 3, percebe-se que os 10 termos/expressões com maior destaque são: Ciência da Informação, com 31 ocorrências, produção científica com 26, Brasil com 24, informação com 20, comunicação científica com 18, bibliometria, gestão do conhecimento e web semântica com 13 ocorrências, bem como, acesso aberto e representação da informação com 10 cada.

4. CONCLUSÕES

O estudo aqui apresentado qualifica-se como sendo a base inicial para uma pesquisa que visa descobrir quais assuntos e temáticas estão recebendo maior destaque dentro da produção científica brasileira da área da Ciência da Informação na contemporaneidade. Por isso, esta primeira fase necessitou de uma construção sólida, com diversas revisões e constatações da qualidade e veracidade dos dados obtidos.

É importante destacar que, futuramente, serão realizadas limpezas nos termos extraídos do periódico a fim de padronizar o idioma dos mesmos, bem como evitar redundâncias nas variações gráficas e morfossintáticas no índice da pesquisa. Somente após essas limpezas é que as co-ocorrência das palavras serão abordadas e analisadas.

5. REFERÊNCIAS BIBLIOGRÁFICAS

GALVEZ, C. Análisis de co-palabras aplicado a los artículos muy citados en Biblioteconomía y Ciencias de la Información (2007-2017). **Transformação**, v. 30, n. 3, p. 277-286, 2018. Disponível em: <http://dx.doi.org/10.1590/2318-08892018000300001>. Acesso em: 10 ago. 2022.

PRODANOV, C. C.; FREITAS, E. C. de. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Feevale. 2013. Disponível em: https://aedmoodle.ufpa.br/pluginfile.php/291348/mod_resource/content/3/2.1-E-bo ok-Metodologia-do-Trabalho-Cientifico-2.pdf. Acesso em: 10 ago. 2022.