Analysis of the functions of keywords assigned by authors in scientific publications in events and journals

Mariangela Spotti Lopes Fujita

ABSTRACT
Introduction: Authors, with their prior knowledge of the subject of their article, can be considered domain expert indexers when they assign keywords from their domain expert language that represent meaningful content in submission systems of journals, events, and repositories. Objective: The study on the indexing of the scientific author as an indexer aimed at investigating the functions assigned to keywords in scientific publications for events and journals in two areas of knowledge, Information Science and Education. Methodology: Exploratory research was carried out with papers from the ENANCIB (Information Science) and ANPED (Education) events and articles from Qualis A1 and A2 journals in the SciELO Information Science and Education collections. Method of keyword function annotation and category determination were applied, according to the context of the abstract and title. Results: The most frequently used keywords are represented by the functions “Research topic” and “Research object” and the least frequent ones represented by the functions “Data”, “Research area”, “Other” and “Research method”. Conclusion: The study concluded that the two main functions are “Research object” and “Research topic”.

KEYWORDS
Keywords. Indexing. Knowledge representation

Análise das funções de palavras-chave atribuídas por autores em publicações científicas de eventos e periódicos

RESUMO
Introdução: O autor, com seu conhecimento prévio sobre o assunto de seu artigo, pode ser considerado um indexador especialista de domínio, quando atribui palavras-chave de sua linguagem de domínio especializado que representam o conteúdo significativo, em sistemas de submissão de periódicos, eventos e repositórios. Objetivo: O estudo sobre a indexação do autor científico como indexador teve como objetivo investigar as funções atribuídas às palavras-chave, em publicações científicas de eventos e periódicos de duas áreas de conhecimento, Ciência da Informação e Educação. Metodologia: Realizou-se pesquisa exploratória
Com trabalhos dos eventos ENANCIB (Ciência da Informação) e Reunião Nacional da ANPED (Educação) e com artigos de periódicos Qualis A1 e A2 da coleção SciELO de Ciência da Informação e de Educação. Foi aplicado método de anotação de funções das palavras-chave e determinação da categoria, conforme contexto do resumo e do título. **Resultados:** As palavras-chave mais frequentes são representadas pelas funções “Tópico de pesquisa” e “Objeto de pesquisa”, enquanto as menos frequentes, representadas pelas funções “Dados”, “Área de pesquisa”, “Outros” e “Método de pesquisa”. **Conclusão:** Conclui-se que as duas principais funções são “Objeto de pesquisa” e “Tópico de pesquisa”.

**PALAVRAS-CHAVE**
Palavras-chave. Representação do conhecimento.

**CRediT**
- **Acknowledgments:** Not applicable
- **Funding:** The scientific research was funded by the Brazilian government’s development agency, the National Council for Scientific and Technological Development (CNPq), with PQ and PIBIC scholarships.
- **Conflicts of interest:** The author certifies that she has no commercial or associative interest that represents a conflict of interest in relation to the manuscript.
- **Ethical approval:** Not applicable.
- **Availability of data and material:** Not applicable.
- **Authors’ contributions:** Conceptualization; Research Methodology; Project administration; Supervision; Visualization; Writing – original draft; Writing – proofreading & editing: FUJIT, M. S. L.
- **Image:** Author’s collection.
- **Translation:** Nakano Aulas e Traduções

**JITA:** IC. Index languages, processes and schemes.
**ODS:** 4. Quality education

*Article submitted to the similarity system*

**Submitted:** 06/04/2024 – **Accepted:** 20/05/2024 – **Published:** 16/07/2024

**Editor:** Gildenir Carolina Santos
1 INTRODUCTION

Information, to be properly retrieved, must be organized and represented in the most faithful way possible in order to comply with the informational content of the document and the user who needs it. This process takes place through the creation of informational products, such as abstracts, classification numbers, keywords and descriptors. The concepts, chosen and synthesized in keywords, are also configured as representations of the document content, even if these are assigned by other indexers, the authors, who, in this case, become expert indexers (Lardera; Hjørland, 2020).

Authors, with their prior knowledge of the subject matter of an original article, can be taken as domain expert indexers, generally without access to controlled vocabularies or indexing guidelines. In this way, when they assign keywords which characterize or “encapsulate” the knowledge of their original texts in journal submission systems, events and repositories, they become indexers that use keywords from their specialized domain language.

Furthermore, the search for keywords is also fundamental in retrieving information, and it is through natural language that the same thing is described in different ways. If, on the one hand, natural language has the advantage of specialized domain language and vocabulary enrichment, on the other, choosing the terms might lead to some problems such as identical terms with different meanings; queries with high recall, generating lists with large number of documents; users, having little knowledge of the subject they are researching and not knowing the specific terms in the desired area of knowledge, struggle to formulate queries that generates relevant documents (Wives; Oliveira, 2010).

Currently, the systems for submitting theses and dissertations and original articles used by journals, events and repositories, in general, employ a metadata filling system to identify titles, authors, subjects by keywords, etc. In the case of assigning subject metadata, the task is generally carried out by the authors themselves, who often do not consult a controlled vocabulary. This vocabulary is not always recommended for authors to consult at the time of assignment. Consequently, the vocabulary used is that of the author, who is considered an expert on the topic of the original content.

This situation can be seen from two perspectives: the authors' natural language is rich in new terms from their domain of knowledge; the authors' natural language has no vocabulary control. Both are valid and, in this regard, we understand that there will always be opportunities to update controlled vocabularies using the author's natural language, especially considering that they are experts in their knowledge domain, and each keyword has relevance to information access and retrieval contained in their written text. This hybrid situation of information systems using natural language and controlled language, today constitutes a reality that cannot be denied (Fujita, 2020). This solution, although conflicting, can bring mutual benefits if there is an indexing policy inclusive of the different actors that perform the indexing.

However, indexing policies can only guide authors, based on the observation and identification of the procedures and criteria they adopt when assigning keywords, in the context of the knowledge domain in which they are experts. In this sense, the objective of the investigation is to identify features and functions of keywords to improve guidance for authors within indexing policies.

Considering the need to develop indexing policies that provide authors with guidance on the indexing process in information systems, such as in self-archiving systems for scientific production in institutional repositories, scientific journals, events and the filling in curriculums. This research proposes to examine the indexing of scientific authors as domain expert indexers, aiming at investigating the functions adopted by
authors when assigning keywords to scientific publications for events and journals in two areas of knowledge, Information Science and Education.

2 KEYWORD ANALYSIS AND THEIR FUNCTIONS

The study on keywords began with the uniterm and intensified with automatic indexing at the end of the 1950s (Lancaster, 2002; Chu, 2003; Gomes, 1996; Paes, 2004), that is, long before the popularization of the web. The coexistence of natural language with controlled language is older than we know now. Research on keywords assigned by authors or by automatic extraction has evaluated the influence of keywords on information representation and retrieval with both quantitative approaches (Gonçalves, 2008; Wolverton; Hoover; Fowler, 2011; Névéol; Dogan; Zhiyong, 2010) and qualitative approaches (Peset, 2020; Li, 2018; Lu et al., 2019).

Névéol, Dogan and Zhiyong (2010) conducted a comparative study of article representations independently provided by authors and indexers, evaluating the extent to which indexers represent topics considered important by the authors. They agree that there are significant differences between keyword assignment and subject indexing, because “[...] authors are asked to choose a small number of keywords, without reference to a controlled vocabulary whereas indexers are trained to select indexing terms according to a specific protocol” (Névéol; Dogan; Lu, 2010, p. 537).

The comparison of keywords with words that telegraph what happens in the text is understood by Peset et al. (2020) (na referencia está et al.) as a form of telegraphic description. In their investigation, they observed two types of keywords: those that authors choose based on terms already used by others in their field, which indicates the influence of previous discoveries on their research; conversely, authors contribute to a new topic and use words from other areas or simply create new terms.

The study by Wolverton, Hoover and Fowler (2011) on the analysis of the subject of theses and dissertations with 280 university libraries, specifically also on keywords, revealed that, out of the 82 institutions that assigned keywords, only eight specified the number of one to five assigned keywords on average. They were also asked to describe the instructions provided to non-indexing contributors regarding term selection. According to the authors, the obtained responses were varied: selection instructions through ProQuest, OhioLink, Dissertation Abstracts and DSpace repositories; institutional printed instructions and online applications; instructions provided by graduate courses through submission forms; and instruction provided by academic departments and library staff.

Specific study by Gonçalves (2008) on the number of keywords assigned by Brazilian authors in journal articles in the areas of Social Sciences confirms that in no area was there a number of keywords lower than three, and the average was 4 to 5 keywords. Although keywords lack standardization, Gonçalves (2008) identified some with relevant functions and characteristics, such as type of research (exploratory, theoretical, etc.) and context of the study. Due to the very characteristics of Social Sciences, proper names perform the function of keywords, as exemplified in the citation of proper names, such as “Gramsci” or “Kant” (Gonçalves, 2008).

Considering the different functions that keywords assume, Lu et al. (2019) observed how the selected keywords function semantically in scientific publications. To this end, they manually processed 693 articles from the Journal of Informetrics, with a total of 3,311 keywords. They developed a manual annotation scheme for keyword functions, such as "research topic", "research method", "research object", "research area", "..."
"data" and "other", based on content analysis of the texts. The results demonstrated that the diversity of keyword functions decreases, but irregularity increases with the number of keywords assigned by the author. The conclusions indicated that research must consider the different types of keywords selected by authors.

In an investigation on the application of bibliometric analysis in automatic keyword extraction, Li (2018) used the differentiation of functions for keywords combined with the classification of topical terms from texts, concluding that the differentiation of functions proposed for keywords partially improves the selection of high-frequency words, especially for text topic queries. A case study implemented by Zhang et al. (2016) addressed the connections and differences between author keywords and Web of Science vocabulary keywords, finding that the latter seem to emphasize research methods and techniques rather than author keywords, while author keywords seem to contain more information on research topics (Zhang et al., 2016).

Studies carried out around guidelines for authors to assign keywords, based on analysis of journals in the areas of Psychology (Terra; Agustín Lacruz; Fujita, 2022), Communication and Information Science (Fujita; Agustín Lacruz; Terra, 2018), as well as theses and dissertations (Terra et al., 2021), confirm that there are no specific recommendations regarding the use of keyword functions, but they highlight the minimum and maximum number, which can range from one to 10 keywords.

These investigations also revealed that 21.8% of Psychology journals, and 22% of Communication and Information Science journals mention the possibility of using controlled vocabulary without saying which one, and only a few cited MeSH. Regarding the study on keywords in theses and dissertations, the study by Terra et al. (2021) found that in the metadata records of the digital repository of São Paulo University (USP), submitted between 2001 and 2019, a total of 223,867 keywords, with an average of 4.62 keywords in Portuguese and 4.59 in English per record. Keyword assignment has gained particular relevance in the current open scientific communication ecosystem in digital environments, therefore, the importance of investigating the representation functions of keywords, through the practice of indexing authors as experts in the domain, who have “[...] comprehensive knowledge about the representational function, as well as the structural aspects of summaries and keywords” (Costa; Moura, 2013, p.45).

3 METHODOLOGY

This research is characterized as qualitative, and its development was conducted through an exploratory search in a sample of complete papers presented in WG-2 – Knowledge Organization and Representation – of ENANCIB, the main graduate scientific event and research in the area of Information Science, and WG-22 – Environmental Education – of the ANPED National Meeting, the main graduate scientific and research event in the field of Education, as well as articles published in journals Qualis A1 and A2 (Quadriennium 2017-2020) in the areas of Information Science and Education belonging to the SciELO Periódicos collection.

In this investigation, the analysis will have two stages: analysis of the guidelines for authors for assigning keywords by ENANCIB, by the ANPED National Meeting and by the journals, in addition to analysis of the keyword functions in the papers published by ENANCIB and by the ANPED National Meeting and the articles published in journals.

To analyze the guidelines and keyword functions, the study includes the latest published proceedings of ENANCIB held in 2022 (24 papers) and the ANPED National Meeting held in 2021 (30 papers), and the articles from the issues published by the
journals in the area of Information Science, Em Questão (2023 - 25 articles), Transinformação (2022 - 30 articles) and Perspectivas em Ciência da Informação (2022 - 7 articles) and, from the area of Education, the journals Educação e Realidade (2022 - 56 articles) and Educação & Sociedade (2022 - 65 articles).

The analysis of the term functions of the keywords assigned by the sample authors was conducted based on the methodology used by Lu et al. (2019), specifically to the method of function annotation and the use of categories for characterizing functions, aiming at comparative analysis of results. The empirical scheme for annotating the functions of the keyword terms selected by the author consisted of carefully reading the title and abstract for a comprehensive understanding of the context of each keyword, and then determining the corresponding category, according to the meaning assigned by the author in the context of the abstract and title. The same categories used by Lu et al. (2019) were adopted: 1) research theme; 2) research method; 3) research object; 4) research area; 5) data; and 6) other.

To understand the function representations of each category, the following descriptions were detailed by Lu et al. (2019, p. 407):

1 Research Topic (T): Problems or topics discussed in research articles.
2 Research Method (M): Methods or solutions used in research articles, including theories, bibliometric indicators, algorithms, mathematical formulas, models, etc. For example, "Bradford's law", "h-index", "PageRank algorithm", "Hall model".
3 Research Object (O): The object that the research studied, including people, group, organization, materials or objects.
4 Research Area (A): The academic area or background of the article, for instance, “bibliometrics”, “physics”, “science of science” and “library and information science (LIS)”.
5 Data (D): The dataset used in the study or the data created by the study, for example, “APS dataset”, “X corpus” or “Web of Science”, etc.
6 Other (OT): Cannot be included in the previous categories.

Each keyword selected by the author is annotated in a term function; on the other hand, a term function may have several keywords assigned, which may represent the regularity of the term functions of the keywords chosen by the author in an article.

The annotation process of functions and use of categories to characterize functions was implemented by the researcher of this investigation and her scientific initiation scholarship student together. In the case of the Information Science area, the annotation process was carried out before the Education area, because the researcher is an expert in LIS and, specifically, in Knowledge Organization and Representation, which proved to be a facilitator for the identification of functions in another area of knowledge outside the knowledge domain.

The procedure was followed to annotate the functions corresponding to each keyword, accompanied by careful reading of the title and abstract for a comprehensive understanding of the context of each keyword and then determining the corresponding category, using the corresponding code, as follows:

a) Example from the area of Information Science
Data citation theory: a review of scientific production in Latin America (Oliveira et al., 2022)

Abstract:
This is bibliographical research, of qualitative nature, which sought to identify the state of the art regarding the theory of data citation in the scientific production conducted in Latin America. To this end, expressions were established in Portuguese, English and Spanish regarding the aforementioned theme, which were used to explore the following databases, repositories and
search engines: Brazilian Digital Library of Theses and Dissertations, OasisBR, La referencia, Redalyc, Networked Digital Library of Theses and Dissertations, Portal de Periódicos Capes, Google Scholar, SciELO and Brapeci (Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação). After analyzing the retrieved papers, only those discussing the topic of citing research data in depth were considered, aiming at contributing to reflection on a theory of data citation, totaling 19 papers. The investigation concluded that there is a significant absence of papers, in Latin America, concerning the theory of data citation, however, papers were identified that, although not referring to a theory properly, offer significant contributions to the topic of citation of research data and can serve as a basis for developing studies on data citation theory. It was also found that Brazil stood out in the production of research on citing research data, and out of the 19 papers analyzed in this research, 17 were Brazilian productions.

**Keywords: 4**
Open Science; OT
Information Science; A
Open Data; O
Data citation theory; T

b) **Example from the area of Education**

Poverty and Resilience in the Narratives of EJA Students in Homeless Situations (Souza; Vieira, 2022)

**Abstract**

Poverty and Resilience in the Narratives of EJA Students in Homeless Situations. This article relates the concepts of poverty and resilience, in the study of homeless students enrolled in Youth and Adult Education (EJA). The study, the result of a Master's thesis, contextualizes the exclusion experienced by this population in the Federal District and presents, in qualitative research, the narratives as a way of bringing the voices and making these subjects visible, since they are made invisible socially and in a census manner. As a result, the students' narratives dialogue with data from the last Census carried out in the FD, recognizing, in the words of these students, subjects of rights capable of narrating, reflecting and re-signifying their human and social condition from their life histories.

**Keywords: 4**
Youth and Adult Education; A
Homeless People; O
Poverty; T
Resilience; T

After annotating the term functions and determining the corresponding categories, Tables were created with the quantitative presentation of the term functions and the position of the terms in relation to the functions in order to support the results, analysis and discussion.

**4 RESULTS AND DISCUSSION**

Keyword assignment is one of the three initial elements, after the title and abstract, that accompany the text of a journal article or event paper. The function of these three elements is precisely the representation of the text content. Each one has its own characteristics and creation rules with different purposes for representing and organizing knowledge, for intellectual access to catalogues, bibliographies and databases (Hjørland, 2003). Therefore, these representations are determinants of the visibility and various quality measures of the specific article (Costa; Moura, 2013).

In this sense, journals and events include specific guidelines for the inclusion of each element in the text. The analysis of the guidelines for authors on keywords in the
area of Information Science, as shown in Chart 1, demonstrates formal requests regarding the minimum and maximum number, between 3 and 5 keywords, both in the case of the events and the journals Transinformação (2022) and Em Questão (2023). ENANCIB (2022a, 2022b) indicates the use of the Brazilian Thesaurus of Information Science to control vocabulary or indicate more specific keywords that are not included; in turn, the journals do not have recommendations for the use of controlled vocabulary. Regarding the type of keywords, the journal Transinformação (2022) recommends that keywords describe the content and that they should be different than the title words. As for the journal Em Questão (2023), it recommends the keywords should not contain metaphorical, abstract or subjective terms.

**Chart 1.** Guidance for authors on assigning keywords to events and journals in the area of Information Science

<table>
<thead>
<tr>
<th>ENANCIB- SIB- W2 - Knowledge Organization and Representation</th>
<th>TRANSINFORMAÇÃO (2022)</th>
<th>EM QUESTÃO (2023)</th>
<th>PERSPECTIVAS EM CIÊNCIA DA INFORMAÇÃO (2022)</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Keywords: include between 3 (three) and 5 (five) keywords, preferably according to the Brazilian Thesaurus of Information Science by the authors Lena Vânia Ribeiro Pinheiro and Helena Dodd Ferrez, available at <a href="http://www.ibict.br/publicacoes-e-institucionais/tesauros-brasileiro-de-ciencia-da-informacao-1">http://www.ibict.br/publicacoes-e-institucionais/tesauros-brasileiro-de-ciencia-da-informacao-1</a>, utilizing specific terms, when these are not covered by the aforementioned thesaurus” (ENANCIB, 2022a, 2022b).</td>
<td>“Keywords or descriptors must be written in capital letters, separated by periods. At the end, it should contain 3 to 5 keywords that exactly describe the content of the paper. We recommend that the indicated keywords are not present in the title” (Transinformação, 2022).</td>
<td>“The title, abstract and keywords must be clear and objective and designate what the text actually produced, avoiding metaphorical, abstract and subjective terms. Enter three to five keywords” (Em Questão, 2022).</td>
<td>There are no specific guidelines for authors to assign keywords.</td>
</tr>
</tbody>
</table>

**Source:** Prepared by the author

In the area of Education, the guidelines for authors on keywords, according to Chart 2, consist of formal requests referring to the maximum number of 5 keywords, without indicating the minimum, for the journals Educação e Realidade (2022) and Educação & Sociedade (2022). At the ANPED National Meeting (WG 22 - Environmental Education, 2021), no specific guidance was found for authors on assigning keywords. In the journal Educação e Realidade, the guidelines indicate a maximum of five keywords in Portuguese and in English, while in the journal Educação & Sociedade, they recommend to include a maximum of five keywords to identify the content of the article, separated by periods, presented in English and Spanish, with the possibility of the Spanish version being replaced by the French version.
In both areas, there are no specific guidelines regarding the use of functions that keywords could represent in journals and events. However, the keyword analysis verified whether the recommendations of the journals Transinformação and Em Questão, in the area of Information Science, as well as Educação e Realidade and Educação & Sociedade, in the area of Education, were attended to by the authors.

The quantitative analysis of keywords, carried out in ENANCIB papers and in articles from the three journals revealed that an average of three to five keywords was maintained, with a total of 364 keywords assigned to 86 articles and papers, with an average of 4.2 keywords for each text, according to Chart 3. In Education, 573 keywords were assigned to 151 articles and papers, with an average of 3.7 keywords for each text, according to Chart 4, with an average of less than 4 keywords. The average numbers are almost similar both in events and in journals, with a slight decrease for ENANCIB.

### Chart 3. Number of papers and average keywords in articles and complete papers in Information Science

<table>
<thead>
<tr>
<th>Event/Journal</th>
<th>Articles/Papers</th>
<th>Keywords</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENANCIB – WG-2 – Knowledge</td>
<td>24</td>
<td>91</td>
<td>3.8</td>
</tr>
<tr>
<td>Organization and Representation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transinformação</td>
<td>30</td>
<td>129</td>
<td>4.3</td>
</tr>
<tr>
<td>Em Questão</td>
<td>25</td>
<td>113</td>
<td>4.5</td>
</tr>
<tr>
<td>Perspectivas em Ciência da Informação</td>
<td>7</td>
<td>33</td>
<td>4.4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>86</strong></td>
<td><strong>364</strong></td>
<td><strong>4.2</strong></td>
</tr>
</tbody>
</table>

Source: Prepared by the author

### Chart 4. Number of papers and average keywords in articles and complete papers in Education

<table>
<thead>
<tr>
<th>Event/Periodical</th>
<th>Articles/Papers</th>
<th>Keywords</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANPED - WG 22 - Environmental</td>
<td>30</td>
<td>110</td>
<td>3.6</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Prepared by the author
Such results correspond to those obtained by Gonçalves (2008) in the analysis of articles from Social Sciences journals, as well as in the analysis of Psychology journals and Communication and Information Science journals, by Fujita, Agustin Lacruz and Terra (2022) on theses and dissertations, by Terra et al. (2021). When analyzing the guidelines for authors on keywords, the Journal of Informetrics, used in the sample by Lu et al. (2019), they recommend using a maximum of six keywords, in addition to avoid general and plural terms and multiple concepts with conjunctions. Thus, the average number of keywords per article in the sample of 693 articles from the Journal of Informetrics was 4.77, therefore higher than the averages for the journals and events analyzed in this study, according to Charts 5 and 6, which meant an increase in the percentage of functions analyzed, as seen in Tables 7 and 8.

The determination of the function corresponding to the keyword took as a parameter the descriptions of the categories based on Lu et al. (2019, p. 407), which allowed to group the keywords of articles from the three ENANCIB journals and papers into the respective categories, as exemplified in Charts 5 and 6, in relation to the grouping of keywords from articles in the two journals and papers from the ANPED National Meeting in the area of Education:

**Chart 5.** Categories of keyword functions assigned to articles in the journal Transinformação in the area of Information Science

<table>
<thead>
<tr>
<th>Functions/categories</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Research Topic (T): Issues or topics discussed in research articles.</td>
<td>Intellectual structure; Science mapping; Co-authorship networks; Open access; Archival description; Digital Preservation; Individual perception; Metadata; Thematic processing of information; Data citation theory; Multicultural; Choice of journal; Open access; Automatic indexing; Ontology; Ordinances; Author entity; Personal authors; Information diffusion; Scientific development; Maturity; Bibliometrics; Audiovisual communication; Scientific communication; Scientific knowledge; Access to information; Article Processing Charge; Journals; Intellectual property; Licensed digital books; User studies;</td>
</tr>
<tr>
<td>31 keywords</td>
<td></td>
</tr>
<tr>
<td>2 Research Method (M): Methods or solutions used in research articles, including theories, bibliometric indicators, algorithms, mathematical formulas, models, etc. For example, “Bradford’s law”, “h-index”, “PageRank algorithm”, “Hall model”.</td>
<td>Journal co-citation analysis; Data integration; Bibliometric study; Evolution mechanism; Empirical analysis; Comparative research; Thematic Specialization; Software evaluation; Bibliometric analysis; Citation analysis; Categorization; Co-citation analysis; Co-word analysis; Analysis of bibliographic references; Bibliometric indicators; Coword analysis; Statistic; Data envelopment analysis; Multi-criteria decision making; Classification</td>
</tr>
<tr>
<td>20 keywords</td>
<td></td>
</tr>
<tr>
<td>3 Research Object (O): The object that the research studied, including people, group, organization, materials or objects.</td>
<td>Scientific data repository; Digital information; Sick user group; Motivation; Graffiti; Open data; Advanced manufacturing industry; Innovation networks; Scientific information; Scientific journals; Grobib; PDFMiner.six; PDFAct; PDF-extract; PDFExtract; Federal Educational Institutions; Collaboration; Co-authorship; Multi-authorship; Scientific production; Bibliographic standards; National standards; Web pages; Social responsibility; Researchers; Audiovisual documents; Scientific</td>
</tr>
<tr>
<td>37 keywords</td>
<td></td>
</tr>
<tr>
<td>4 Research Area (A): The academic area or formation of the article, for example, “bibliometrics”, “physics”, “science of science” and “library and information science (LIS)”.</td>
<td>Production; Capes; Development agencies; Open access; Peer-reviewed journals; Scientific journals; Linked data; Linked open data; Linked enterprise data; Information behavior; Information practice</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>15 keywords</td>
<td>Library and information science; Diplomastics; Cataloguing; Information Organization; Information Science; Documentation; Information; Information representation; Normalization; Knowledge management; Scientometrics; Informetrics; Bibliometrics; Librarianship; Terminology</td>
</tr>
<tr>
<td>5 Data (D): The dataset used in the study or the data created by the study, e.g. “APS dataset”, “X corpus” or “Web of Science”, etc.</td>
<td>Lattes platform; Directory of Open Access Journals; Lattes platform; Fair Data; Cultural heritage; Metadata; Study plans; Semantic Web; Web of Science; Data; Semantic Web;</td>
</tr>
<tr>
<td>11 keywords</td>
<td></td>
</tr>
<tr>
<td>6 Other (OT): Cannot be included in the previous categories.</td>
<td>Ibero-American research (geographical location); Resource for knowledge; Knowledge sharing; Arts; Documents; Open Science; Technologic innovation; Science Policy; Scientific research; R&amp;D; Level; Brazil; Scientific activity; Data envelopment analysis.</td>
</tr>
<tr>
<td>15 keywords</td>
<td></td>
</tr>
<tr>
<td>Total: 129 keywords</td>
<td></td>
</tr>
</tbody>
</table>

**Source:** Prepared by the author

It is observed that, in Chart 5, taken as an example from the Information Science area, the keywords “Open Access” and “Bibliometrics” were annotated in more than one function, not by the same article, though, by different articles, because it depended on the context in which they occurred. This diversity of functions, although small, was repeated in the case of other journals and in ENANCIB.

On the other hand, the terminological diversity obtained by the sample of 364 keywords can be observed, as there are few repeated terms in the total number of articles, as were only the cases of “Scientific production” and “Plataforma Lattes”, which occurred twice, in different articles in the journal Transinformação, no term in articles in the journal Perspectivas em Ciência da Informação and the terms “Bibliometrics” and “Mediação da Informação”, which were repeated once again, in the journal Em Questão. However, in ENANCIB, the most representative terms from WG-2, “Knowledge Organization and representation”, “Knowledge Organization”, “Information Representation”, “Information Organization”, “Information Science”, were the terms repeated more than twice, indicating that the authors demonstrate, through their keywords, an understanding of the terminology specific to their field of expertise. This characteristic is less prevalent in journals, as they do not have thematic focus on a particular specialty, as was the case in WG-2 of ENANCIB.
Chart 6. Categories of keyword functions assigned to articles in the journal Educação e Realidade in the area of Education

<table>
<thead>
<tr>
<th>Functions/categories</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Research Topic (T): Issues or topics discussed in research articles.</td>
<td>Family; Dual Career; Schooling; Distance education; Poverty; Resilience; Inequality; Privileges; Children’s cultures; School; Exhibition spaces; Privilege Walk; Social inequality; Meaningful learning; Emotional dimensions; Teacher training; Early access program; Professor/Bureaucrat; Discretionary; Paulo Freire; Decoloniality; Black Population Education; Implementation; Institutional implementation arrangements; Arboviruses; Scientific Dissemination; Satisfaction; Competencies; Covid-19; Pedagogical practices; Artistic Practices; Analytical procedures; Representation; Disabilities; State of art; Child education; Space-Time Mental disability; Epistemic injustice; Social model of disability; Training policy; Liberalism; Nihilism; Amor fati; Perspectivism; Philosophy of Education; Evil of the File; Basic Education; School; Democracy; Institutional Racism; Affirmative actions; Racial quotas; Public policy; Intervention research; Institutional analysis; University governance; Higher education; Institutional assessment; Occupations; Social process Games; Civic-Military Schools; School Music Education; Postmodernity; Parresia; Educational policies; Public action instruments; Full-time; Associativism; Edutubers; Teaching identities; innovation; Culture; School Culture; Teaching practices; Pedagogical Training; Educational financing; Primary and Secondary Education; Education policy; History of education; Textbook; University Education; Higher education; Intercultural education; Intercultural theory; Yogytksky; Teaching; Teacher tools; Schooling; Education; Transformation; Transformative education; Historical-cultural theory; Child education; Development; Child education; Aesthetic education; Artistic sign; Emotions; Human development; Interculturality; Historical-cultural theory; Critical Multiculturalism; Inclusive Special Education; Human rights; Biography; Life-work; Way of existing; Gestures; Deleuze Class; Guattari; Guidance; Graduate studies; Educational research; Childishness; Discursive field; Transcription; Philosophy of difference; Curriculum; WG Curriculum; Lesson; Writing and reading; Sandra Mara Corazza; Translation; Transcription.</td>
</tr>
<tr>
<td>2 Research Method (M): Methods or solutions used in research articles, including theories, bibliometric indicators, algorithms, mathematical formulas, models, etc. For example, &quot;Bradford’s law&quot;, &quot;h-index&quot;, &quot;PageRank algorithm&quot;, &quot;Hall model&quot;.</td>
<td>Button Regulation; Agency Theory; Historical-Cultural Theory. Historical-Cultural Theory; Historical-Cultural Theory.</td>
</tr>
<tr>
<td>3 Research Object (O): The object that the research studied, including people, group, organization, materials or objects.</td>
<td>School; Soccer; Teacher training; Homeless people; Professional education; Covid-19; Children: Contemporary Arts; School; Motivations; Street-level bureaucracy; Indigenous people; Post abolition; Black musicians; PNAIC; Board games; Emergency remote education; Michael Foucault; Continuing education; Inclusive education; PIBID; Cinema and Education; Teacher training; Church and Neo-Pentecostal; University; Self-management; Catholic University; Secondary movement; Geometry; Gestures; Militarization of schools; Aufklärung Kantiana; New High School; Modern Culture Club; YouTube; Didactic culture; Engineering teachers; Mozambican Education; ZDP; Education in Vygotsky; Vygotsky; Child psychology; Emotions; Vygotsky; Vygotsky; Sandra Mara Corazza; Sandra.</td>
</tr>
</tbody>
</table>
The area of Education, as exemplified in Chart 6, referring to the journal Educação e Realidade, in addition to presenting a difference in vocabulary, which clearly demonstrate terminology specific to their domain, presented more repetition of keywords highlighted in bold, in the aforementioned Table, also using personal names assigned as keywords – “Sandra Maria Corazza”, “Paulo Freire”, “Vygostsky”, “Roland Barthes”, “Michael Foucault” and “Guattari” – which indicate contributions of theoretical approaches or important interdisciplinary knowledge, adopted in the analyzed scientific articles. This characteristic is highlighted by Gonçalves (2008) as typical of the area of Social Sciences, in which proper names serve as keywords. The same patterns of occurrence of keywords were observed in the other journal analyzed and in the scientific papers from ANPED National Meeting, where the theme of the Working Group, “Environmental Education” is predominant in the assignment of keywords, whose terminological diversity is highlighted.

In the analysis of keyword patterns selected by the author, the distribution frequency of the functions of keyword terms was obtained, as seen in Charts 7 (Information Science) and 8 (Education), in which the two functions, “Research topic” and “Research object”, have a higher frequency index in relation to the other functions. The functions with the lowest frequencies in Information Science, in ascending order, are “Data”, followed by “Research area”, “Other” and “Research method”, while those with the lowest frequencies, in ascending order in Education are “Data”, “Research method”, “Research area” and “Other”. In general, the frequencies found between the two areas have little difference, but the low frequency of certain keywords in the categories “Research Method” and “Research Data” in the area of Education draws attention, which may point out another trend or guidance for authors on keyword assignment or even another pattern of research development determining the textual structure of scientific texts, influential in verifying the functions of keywords.
The functions with the highest percentages in the area of Information Science present, according to Chart 9, averages equal to the number of articles, that is, an average of 1.19 keyword terms, in the function of “Research topic”, and an average of 1.28 per article, in the function “Research object”, considering the sample of 86 articles (Chart 3). It is also observed that there is no significant difference between the frequency percentages of ENANCIB and other journals, while in the area of Education, the average number of keywords in the functions “Research topic” and “Research object”, with greater frequency, present, respectively, the averages of 2.04 keywords and 1.02, as demonstrated in Chart 10. This difference between the averages of the function “Research topic” between the areas possibly indicates a positioning of authors in the area of Education regarding the representation of problems or research topics, because they assigned more keywords to this function. This is an aspect that outlines differences in research between areas of knowledge and their representation practices.
The results of the keyword averages in the area of Information Science are lower than the averages obtained in the investigation by Lu et al. (2019), which obtained an average of 2.19 per article in the function “Research Topic”, followed by the “Research method” function with an average number of 1.90 per article. On the other hand, the average for the Education area, in the “Research topic” function, is close to the average for the same function in the aforementioned study. We can consider that the keywords analyzed in the sample of 86 articles in the Information Science area reached the highest average in the “Research object” function and, in the Education area, the highest average was obtained by the “Research topic” function. This result reveals that the two main functions of the samples analyzed in both areas are “Research object” and “Research topic”.

Considering the study by Lu, Li, Zhifeng and Cheng, which found articles from the journal Informetrics with the highest averages for the functions “Research topic” and “Research method”, we can infer that greater relevance is given to the three functions – “Research topic”, “Research object” and “Research method” – during keyword assignment.
5 CONCLUSION

In this article, we examine the functions adopted by authors when assigning keywords, aiming at analyzing the existence of keyword patterns, from the perspective of the function that the term assumes, within the context of the article whose content is being represented. This research, therefore, considered the scientific author as a domain expert indexer, given the need to develop guidance for authors on the indexing process in self-archiving systems for scientific production, such as institutional repositories, scientific journals, events and filling in curriculums.

The findings from the analysis of keyword patterns selected by the author identified that the most frequent are represented by the functions “Research topic”, “Research method” and “Research object”, while the least frequent are those represented by sequence, in ascending order, of the functions “Data”, followed by “Research area” and “Other”. The functions are decisive for assessing the relevance of the keywords assigned by the authors, regarding the thematic representation of the contents, proving that they are in agreement with the title and abstract, in the triad of essential elements of representation content of scientific articles.

It is noteworthy that the method of function annotation, used by the researcher, was rigorously applied and observed, in addition to the function pattern, irregularities in relation to vocabulary control, more than one keyword in the same function and keywords assigned without no articulation with the context of the abstract or title. This last irregularity seemed to reveal an opportunity for representation on a related topic, to be studied in future research. However, specificity of keywords is noted, as well as more generic terms, which is consistent with the representation proposals of the analyzed articles, sufficiently to verify the articulation of the assigned keywords with the most significant content of the articles.

Regarding the method of function annotation, there remains uncertainty about the “Other” category, which fails to specify the representative characteristics of keywords that should be assigned to this function. This is due to the category’s lack of clear, objective definition describing the features of keywords that do not fit other categories but might serve a different unforeseen function. In essence, the “Other” function should not be a function, as it lacks specific characteristics for representing keywords.

It should also be noted that the comparative study of the functions adopted in two different areas of knowledge, Information Science and Education, was relevant, as it demonstrated that there are different patterns in keyword assignment, proven by the highest averages obtained, respectively, in the functions “Research object” and “Research topic”, possibly determined by the type of research carried out in each area. Another aspect to be highlighted between the areas is the great terminological diversity obtained, which must be considered when updating controlled vocabularies in each area, as this demonstrates, above all, research trends in each area. The division into functions could also help in understanding the meanings of keywords.

It is concluded that the main contribution of this investigation is the improvement of guidelines for authors, by adopting the proposal for keyword functions, through categories. In further investigations, it will be important to observe ordering patterns of these functions using natural language and controlled language, that is, the order of the set of keywords by function order – Topic, Object, Method, etc. –, as well as studying functions adopted in other areas of knowledge, such as Biological Sciences or Exact Sciences, in order, obviously, to guide authors from different domains of expertise, considering that they are expert indexers.
REFERENCES


WIVES, L. K.; OLIVEIRA, J. P. M. de. *Aplicação de técnicas de descoberta de conhecimento em textos na inteligência competitiva*. *Cadernos de Informática*, Porto...