



Authors' correspondence

¹ Universidade Estadual Paulista
Marília, SP - Brasil
adriana.carla@unesp.br

² Universidade Federal do Rio Grande
do Sul
Porto Alegre, RS - Brasil
seanvogel@gmail.com

Implementation of the Applied Semantic Thesaurus software: a construction of the micro thesaurus of the Police Inquiries Series of the Public Archive of the State of Pará

Adriana Carla Ribeiro dos Santos¹

Thiago Henrique Bragato Barros²

ABSTRACT

Introduction: In the context of Knowledge Organization (KO) and the archival field, the use of thesauri emerges as an essential and complementary tool in addressing the inherent challenges of information access. **Objective:** **Objective:** This study presents a proposal for the development of a micro-thesaurus aimed at the "Police Inquiries Series" of historical documents from the Public Archives of the State of Pará (APEP), using the software THESA – Applied Semantic Thesaurus as a tool. This initiative aimed to create a more comprehensive and detailed model to represent the organic-functional information of the archives. **Methodology:** The research unfolded in three well-defined phases: an initial exploration phase, followed by fieldwork execution, and finally, the analysis and discussion of the obtained data. The research was based on texts from Information Science and related disciplines, focusing particularly on studies that deal with the intersection of KO and Archival Science. **Results:** The study demonstrates that the micro-thesaurus developed for the "Police Inquiries Series" within APEP plays a significant role in the effective management of the controlled vocabulary for this document series. **Conclusion:** The THESA software is an effective tool for terminological control in archives, with the potential for implementation by other archival institutions.

KEYWORDS

Knowledge organization. Archival science Thesaurus. THESA.

Implementação do *software* Tesauro Semântico Aplicado: uma construção do micro tesauro da Série Inquéritos Policiais do Arquivo Público do Estado do Pará

RESUMO

Introdução: No cenário da Organização do Conhecimento (OC) e no campo arquivístico, o uso de tesauros emergem como uma ferramenta essencial e complementar, diante dos desafios inerentes ao acesso à informação. **Objetivo:** apresenta uma proposta de elaboração de um micro tesauro destinado à Série Inquéritos Policiais dos documentos históricos do Arquivo Público do Estado do Pará (APEP), utilizando-se como ferramenta, o *software* THESA – Tesauro Semântico Aplicado. Essa iniciativa objetivou-se em criar um modelo mais abrangente e detalhado para representar a informação orgânico-funcional dos arquivos. **Metodologia:** O desenvolvimento da pesquisa desdobrou-se

em três fases bem definidas: uma etapa inicial de exploração, seguida pela execução do trabalho de campo e, por último, a análise e discussão dos dados obtidos. A pesquisa foi formulada por textos originários do campo da Ciência da Informação e disciplinas afins, concentrando-se especialmente em pesquisas que tratam da interseção entre OC e Arquivologia. **Resultados:** A pesquisa evidencia que o micro tesauro, desenvolvido para a Série Inquéritos Policiais no âmbito do APEP, desempenha um papel substancial na gestão eficaz do vocabulário controlado desta série documental. **Conclusão:** O *software* THESA é uma ferramenta eficaz para o controle terminológico em arquivos, com potencial para ser implementado por outras instituições arquivísticas.

PALAVRAS-CHAVE

Organização do conhecimento. Arquivologia. Tesauro. THESA.

CRediT

- **Acknowledgments:** Not applicable.
- **Funding:** Not applicable.
- **Conflicts of interest:** Authors certify that they have 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:** The datasets generated and/or analyzed during this study are available in the PPGCIUFPA Institutional Repository on the website: <https://ppgci.prospesp.ufpa.br/index.php/br/>
- **Authors' contributions:** Conceptualization, Data Curation, Formal Analysis, Research, Methodology, Project Management, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing: SANTOS, A.C.R.; Supervision, Writing – original draft, Writing – proofreading & editing: BARROS, T.H.B.
- **Image:** Taken from the author's video recording
- **Translation:** Trama Textual: Serv. Trad., Revisão e Legendagem

JITA: ID. Knowledge representation

ODS: 17 - Partnerships and means of implementation



Article submitted to the similarity system

Submitted: 31/01/2024 – Accepted: 24/07/2024 – Published: 15/09/2024

Editor: Gildenir Carolino Santos

1 INTRODUCTION

The constant progress in Information Technology and Communication has significantly driven the expansion of archival knowledge production, especially in the digital context. The complexity of this environment demands the effective representation of knowledge through terminological tools, which has become essential to facilitate information access.

Despite the importance of terminological representation, functional thesauri—complementary tools to solve problems of access to organic information—are almost nonexistent in Brazilian archives. This gap compromises the effectiveness of archival information retrieval systems, hindering the transmission of accumulated knowledge.

This article constitutes part of the developments from a master's dissertation in which Santos (2022) proposed creating a thesaurus for the documentation of APEP using the THESA software. This article aims to present the results obtained from the micro-thesaurus, which serves as a more detailed model of organic-functional information, aiming to improve the methodological procedures of Knowledge Organization within the archival field.

The THESA software, based on the guidelines of the Simple Knowledge Organization System (SKOS) and the ISO and NISO standards for thesauri and controlled vocabularies, was used as a tool for constructing the proposed thesaurus. This software, available on a free web platform, is presented as a solution based on Semantic Web, suitable for both educational purposes in academic institutions and professional applications.

The choice of the Police Inquiries Series of APEP is justified by its legal-administrative nature and the high demand from users/researchers. The thesaurus aims not only to meet the specific needs of this documentation but also to highlight the importance of thesauri as complementary tools to overcome challenges in accessing organic-functional information.

In addition to this section, the present article comprises four other distinct sections. The second section, "Thesauri in Knowledge Organization," presents a theoretical and methodological approach to Knowledge Organization and the application of thesauri in the archival context. The third section addresses the "Methodological Framework," outlining the methodological procedures adopted while creating the micro-thesaurus for the Police Inquiries Series. The fourth section, titled "Results and Discussion," presents the results obtained using THESA software in detail. Finally, the last section, "Conclusions," highlights the main contributions achieved throughout the study and points to possibilities for future work.

| 3

2 THESAURI IN KNOWLEDGE ORGANIZATION

The "thesaurus" is an essential tool in the universe of Knowledge Organization. Often underestimated, this system plays a vital role in the classification and relationship of terms, enabling more precise indexing and more efficient information retrieval. By exploring the thesaurus as a Knowledge Organization System, we examine how it contributes to the coherence and interconnection of concepts, facilitating the search and understanding of information across various fields of knowledge. Therefore, we turn to the literature in Information Science and related areas to explore how different authors address and apply concepts related to Knowledge Organization and thesauri, aiming to provide a more comprehensive understanding of the field's trends, challenges, and developments.

Knowledge Organization (KO) maintains a close relationship with Information Science, gradually consolidating itself as an interdisciplinary field in constant search of a defined identity. This effort towards consolidation, theoretical dialogue, and application is evident in the literature of the International Society of Knowledge Organization (ISKO) (Guimarães, 2017).

Although the discussion dates to ancient periods, several models have been conceived. The notion of KO began around 1900, with significant contributions from Charles A. Cutter, W. C. Berwick Sayers, and Ernest Cushington Richardson. In 1929, Henry Bliss published *The Organization of Knowledge and the System of the Sciences*, considered one of the major intellectual contributions to the field of KO (Hjørland, 2007).

According to Rowley (1992), KO refers to systems that organize documents and information, allowing users to retrieve them when needed. Hjørland (2007) highlights the diversity of approaches and theories contextualizing KO studies, emphasizing the importance of well-informed choices in teaching and research within the field based on systems of activities and scientific theories that are fundamental to organizing knowledge.

Within the scope of KO, Campos (1994) underscores the importance of theoretical, conceptual, and methodological principles inherent in the Faceted Classification Theory, Concept Theory, and General Theory of Terminology. These components play a decisive role in developing practical approaches for the organization, representation, and retrieval of information and knowledge.

Shiyali Ramamrita Ranganathan developed the Faceted Classification Theory (FCT) in the first half of the 20th century. During the 1930s, FCT was regarded as a dynamic and multidimensional framework, with the term "facet" now used in modern classification theory studies as a synonym for "characteristic" (Barbosa, 1969, p. 16).

FCT is widely cited in studies on classification and the development of thesauri. Ranganathan's book *Prolegomena to Library Classification* presents the foundations of his theory and principles for organizing hierarchically structured concepts, which are considered a universal approach. Moreover, in more recent texts and studies, FCT is extensively referenced in the development of ontologies, which contribute to improving information retrieval systems in the context of the Semantic Web (Laipelt, 2015).

Laipelt (2015) emphasizes that, although FCT refers to the development of classification systems that aim, among other things, to organize a collection physically, it is widely used for thesaurus development, as it uses universal categories that can be adapted and applied in various fields of knowledge, as the author described.

There is also the so-called General Theory of Terminology (GTT), known as Wüster's Terminology Theory, conceived from a careful selection of terms in industrial engineering, with the aim of standardizing linguistic and intralinguistic concepts. Its purpose is to eliminate ambiguities inherent to natural language, providing clarity in professional communication at both national and international levels. Structured with precise instructions, GTT seeks to ensure communicative univocity. However, in natural communication, its characteristics prove insufficient (Cabrè, 1999).

Concept Theory (CT), on the other hand, is a hybrid methodology that combines deductive and inductive methods. Developed in the 1970s by Dahlberg, CT was heavily influenced by Wüster's General Theory of Terminology and Ranganathan's Faceted Classification.

Dahlberg (1978) conceptualizes the element as a concrete entity in classification systems, using an analytical theory of concepts to represent knowledge. CT enables the association of each concept with a term, thus advocating for the need for terminological

standardization. Furthermore, CT significantly contributes to Archival Science, particularly regarding the Classification Plan, by providing a structure supported by specific terms.

According to Dahlberg (1978), concepts are classified as general or individual, formulating concepts relating to various entities. Each true statement represents an element of the concept, and the combination of all factual statements constitutes the concept associated with a specific object. According to CT, each statement presents a predicable attribute of the object, called a characteristic at the conceptual level. This hierarchy of characteristics often culminates in a category, understood as the concept in its broadest extension.

Considering these theories, it is possible to understand Knowledge Organization as a domain within a systemic perspective. KO emerges as an area that plays a significant role in creating and developing methodologies for information access, including the development of Knowledge Organization Systems (KOS), tools that provide an organized and structured interpretation of the objects in question.

2.2 Thesaurus as a Knowledge Organization System

We researched a variety of definitions regarding thesauri, making it necessary to refer to a summary table that presents some of these definitions according to specific standards and authors. Consulting Chart 1 provides a comprehensive understanding of the nuances and approaches surrounding the concept of thesaurus.

Chart 1. Definitions of Thesaurus

Definição de tesouro	Autores específicos
"[...] a controlled and dynamic vocabulary encompassing a specific area of knowledge. In its structure, it highlights the relationships between terms or descriptors – hierarchical, synonymous, and others – which, together, constitute the indexing language."	(IBICT, 1984, p. 5)
"[...] the vocabulary of an indexing language that is controlled and formally organized to make explicit the a priori relationships between concepts (e.g., broader than... or more specific than...). It describes the indexing language as: 'a controlled set of terms extracted from natural language and used to briefly represent the subjects of documents.'"	(ISO 2788, 1986)
A thesaurus is a vocabulary of terms that are generically and semantically related to a specific area of knowledge.	(Motta, 1987)
"[...] A dynamic documentary language containing terms that are semantically and logically related, comprehensively covering a domain of knowledge.""	Gomes, 1990, p. 16)
[...] A structured list of concepts intended to univocally represent the content of documents and queries within a specific documentary system [...]	(Van Slype, 1991, p.23-24)

includes descriptors, non-descriptors, hierarchical and associative relationships, and linguistic equivalences."	
"[...] A documentary language constructed through conceptual units extracted from the formal language of a specific area of scientific or technical knowledge. Its structure suggests the idea of a system, as the concepts are related to each other and are represented by terms. Each term, in turn, is linked to another term through relationships of equivalence, hierarchy, or association. The thesaurus, used for the organization and retrieval of information, constitutes an important '[...] tool to support scientific research in knowledge areas.'"	(Fujita, 1998)
"[...] A documentary language that standardizes the concepts of a specific area through terms expressed in logical-semantic structures."	(Táلامo; Iara; Kobashi, 1992)
"[...] A controlled vocabulary of an indexing language, formally organized to make explicit the a priori relationships between concepts (e.g., generic and specific)"	(UNESCO, 1993, p. 14)
"[...] A specialized, standardized, post-coordinated vocabulary used for documentary purposes, where the linguistic elements that compose it, whether simple or compound terms, are syntactically and semantically related to each other."	(Currás, 1998)
"[...] A thematic or faceted classification system, whose basic structure is formed by a relationship of descriptors representing or describing authorities or thematic content. [...] Through linguistic, semantic units and their relationships, extracted from the formal language of a discipline or specific area of knowledge, it becomes a tool for the representation and retrieval of information."	(Naumis Peña, 2000)
"[...] An appropriate tool for transmitting concepts and their reciprocal relationships, similar to what occurs with terms expressed in the language of documents."	(Dodebei, 2002, p. 67)
"[...] As a function – 'an instrument of terminological control that allows the natural language of documents, indexers, and users to be translated into a more rigid "systemic language" (documentary language, information system language).' As a structure – 'a controlled and dynamic vocabulary of terms that are semantically and generically related, covering a specific field of knowledge.'"	(ANSI/NISO-Z39.19, 2005, p. 9)
"[...] Post-coordinated combinatorial structures, consisting of terms – linguistic units derived from	(Bocato; Ramalho; Fujita, 2008, p. 201)

both specialized language and natural language – called descriptors, endowed with syntactic-semantic relationships, referring to specialized scientific domains, allowing thematic representation of document content and information retrieval”	
“A type of knowledge organization system integrated with analyzed and standardized terms that maintain semantic and functional relationships with each other. The thesaurus is organized under strict terminological control to provide an ideal tool for storing and retrieving information in specialized areas. [...]”	(Baritè <i>et al.</i> , 2015)

Source: (Santos; Barros; Laipelt, 2021)

Following the definitions presented in Chart 1, the thesaurus is a highly significant and appropriate tool for representing and retrieving information in specialized domains. As a Knowledge Organization System, thesauri are used to systematize knowledge content, representing and relating the vocabulary inherent to a particular domain (Santos; Moreira, 2018). Therefore, the thesaurus, when combined with new Information and Communication Technologies, transforms formalized knowledge into content that is identifiable and machine-interoperable, considering the dynamic infrastructure of the web.

According to Vogel and Kobashi (2019), functional thesauri are particularly relevant as the most appropriate tool for current and intermediate archives since the traditional thesaurus is thematically conceived, making it more suitable for permanent archives. In contrast, functional thesauri are ideal for effectively treating information within specific practices and methodologies of the field.

The functional thesaurus offers new possibilities for accessing archival information, fostering greater trust in the system. However, it is crucial to plan and establish an archival classification scheme to effectively develop and use functional thesauri, as this plan reflects the functions and activities performed within institutions.

This suggests that the traditional thesaurus may only sometimes be the most appropriate choice for organizing archives. However, it is essential to note that specific thesauri for archives exist, particularly historical ones, which emphasize the thematic categorization of content. Examples include the "Tesouro del Patrimonio Histórico Andaluz" in Spanish, the "United Kingdom Archival Thesaurus (UKAT)" in English, and the French "Thesaurus W: vocabulaire normalisé pour la description des archives administratives locales contemporaines" and "Archives de Cannes." These thesauri are generally structured around three main lists: subjects, institutions, and geographic locations (Giménez-Chornet and Escrig-Giménez, 2011).

Thus, using standardized language aligned with the context and organizational culture is one of the essential elements for achieving efficiency in producing, organizing, and disseminating information. While the classification plan is well-established in Archival Science, in practice, the need for tools that contribute to terminological standardization remains incipient.

This highlights the importance of using a tool for representing and retrieving information targeted toward users and researchers, providing greater confidence in their research activities. When building a functional thesaurus in Archival Science, it is imperative to consider the classification plan, as the functional thesaurus significantly impacts the functions and activities of institutions.

The functional thesaurus is fundamental in supporting Knowledge Organization's (KO) archivist position as a critical access bridge, guiding system search strategies. The construction of the functional thesaurus should be a comprehensive collaboration involving the institution, document producers, users, archivists, and other professionals who can contribute, adapting it to the specific informational realities of institutions.

Despite the scarcity of research addressing functional thesauri in Archival Science, particularly exploring the relationship between organization, representation, and the use of information, description standards in the field emphasize the need for tools that operate with perspectives of standardization and terminological control.

However, Davanzo (2016) notes that the standards lack guidance on how a functional thesaurus can be developed to contribute to information treatment. Furthermore, the author states that the standards must clarify whether Archival Science can use a functional or conceptual thesaurus.

For the archivist, the need to make information available is undeniable, and success in this goal fundamentally depends on using tools that support this perspective. Therefore, using a functional thesaurus to guide an organization, especially in retrieving archival information, not only reinforces the principle of provenance but also materializes hierarchical, associative, and equivalence relationships through standardized descriptors in its structure. This effectively represents the activities and functions of the documents and ensures the preservation of organic-functional information within the archival memory system.

3 METHODOLOGY

According to the research methodology guidelines established by Gil (2002), this study is characterized as follows:

a) **Nature:** It is considered applied research, as it resulted in direct practical applications in developing thesauri intended for archives.

b) **Approach to the Problem:** Given the nature of the explored knowledge, which is inherently difficult to quantify, this research is classified as qualitative. The focus is on understanding the problem without intending to quantify the involved data.

c) **Objectives:** Based on the existing literature, this is classified as explanatory research, as it aims to examine the interrelationship between Knowledge Organization and Archival Science.

d) **Technical Procedures:** The study combines bibliographic research, dealing with published literature, and experimental research due to constructing a thesaurus model for the archival domain.

Minayo (2009) proposed that the study was conducted in three distinct phases: the exploratory phase, the fieldwork, and the data processing and analysis. The research corpus comprises texts from Information Science and related disciplines, addressing topics such as Knowledge Organization and Archival Science.

Given the limited number of references on functional thesauruses, the research focuses exclusively on thesauri as tools for information retrieval applied to knowledge organizations.

The experimental phase involved the construction of a thesaurus model for the archival domain, based on the Australian AGIFT model, identified on the website of the National Archives of Australia (NAA) and developed using the THESA software. This model was tested on documents from the Police Inquiries Series of the Public Archive of the State of Pará (APEP) to evaluate its applicability and interoperability.

The theoretical foundation for constructing this model is derived from the interdisciplinarity between methodologies from Information Science and the Semantic Web. This model will serve both for terminological standardization and control and to address problems related to access to archival information.

The THESA software follows the SKOS (Simple Knowledge Organization System) guidelines and ISO and NISO standards for thesauri and controlled vocabularies. Based on the Semantic Web, it is available for free and can be used academically in undergraduate and graduate courses and professionally. This application allows for the creation of an unlimited number of thesauri, offering configuration options for public or private use (Gabriel Junior; Laipelt, 2017).

It is essential to highlight the choice of software due to its structure, based on the conception of relationships between concepts. This assumes that an idea can be represented by a term, an image, a sound, a link, or any other form that can be explicitly defined as described by its developers.

4 RESULTS AND DISCUSSION

The documentary research was conducted at the Public Archives of the State of Pará (APEP), focusing on the documentary collection of the Public Security Department—Police Inquiry Series, which dates from 1956 to 1990. The collection is classified and stored in 438 archive boxes.

The research involved reading, analysis, archival description, and creating a taxonomy for the documents produced by police stations in the Metropolitan Region of Belém and the interior of Pará.

A document analysis guide was developed, comprising the following steps:

- a) Document processing;
- b) Reading and description of the police inquiry series documents;
- c) Identification of themes;
- d) Cataloging of the most frequent terms;
- e) Construction of a taxonomy;
- f) Transfer of terms to the THESA software.

The APEP has a Permanent Documentation department responsible for the management, control, and technical treatment of the institution's documentation. Its activities include collection, separation (by document type, collections, and time limit), and cataloging and storage. Therefore, the document processing, reading, and description of the police inquiries were carried out in this department under the guidance and supervision of staff.

The collection analyzed for the research corpus comprised 198 archive boxes. Each box is identified with the name of the collection, series, police station, year of occurrence, and box number. The boxes contain various police inquiries, with the smallest box containing three inquiry folders, specifically from the Political and Social Order Department (DOPS), and the largest containing 85 inquiry folders from the police station of the municipality of Santa Izabel in Pará.

According to Nucci (2016), the Police Inquiry is an administrative preparatory procedure for criminal prosecution conducted by the judicial police aimed at the preliminary collection of evidence to investigate the commission of a criminal offense and its perpetrator. Therefore, the purpose of the police inquiry is to investigate the facts that constitute a criminal action or offense, as established in Brazilian criminal law.

A total of 45 terms with the highest frequency were cataloged from the documents of the Police Inquiry Series, based on a sample of 50 archive boxes, as presented in **Chart 2**

Chart 2. Cataloging of the Most Frequent Terms (in Portuguese)

	TERMOS	FREQUÊNCIA
1	Incêndio	05
2	Injúria	05
3	Apropriação indébita	06
4	Chantagem	06
5	Concussão	06
6	Constranger alguém	06
7	Desobediência	06
8	Maus tratos	06
9	Peculato	06
10	Sequestro	06
11	Calúnia	07
12	Desacato	07
13	Difamação	07
14	Rixa	07
15	Bigamia	08
16	Esbulho possessório	08
17	Extorsão	08
18	Prevaricação	08
19	Abandono de incapaz	09
20	Falsa identidade	09
21	Porte de arma	09
22	Falsidade ideológica	10
23	Falsificar documentos	10
24	Ato obsceno	11
25	Contrabando	11
26	Dano	11
27	Violação de domicílio	11
28	Atentado ao pudor	12
29	Corrupção ativa	12
30	Abuso de poder	13
31	Resistência	14
32	Ameaça	15
33	Corrupção passiva	15

34	Abandono de função	16
35	Estelionato	17
36	Estupro	17
37	Sedução	17
38	Posse de entorpecentes	21
39	Receptação	22
40	Roubo	24
41	Tráfico de drogas	33
42	Homicídio culposo	40
43	Homicídio doloso	59
44	Furto	245
45	Lesão corporal	350

Source: Authors (2022)

Based on the specific selection and changes in word frequency, essential categories were determined for conducting the research to develop a taxonomy. In Information Science, taxonomy represents a system for structuring knowledge, which classifies and simplifies information retrieval by organizing data into categories and establishing norms for the hierarchical relationships between them (Barros; Bastos; Santos, 2022).

Taxonomy is defined as:

A controlled vocabulary of a given area of knowledge and, above all, an instrument or structural element that allows for the allocation, retrieval, and communication of information within a system under a logical premise (Terra *et al.*, 2005, p. 1).

| 11

From the reading and description of the documents in the Police Inquiry Series, it was observed how incident reports were recorded and the classification of categories was defined. According to Campos and Gomes (2008), categorization is a process that requires thinking about the domain deductively, that is, determining the broadest classes within the chosen theme.

Based on the classification principles for developing a taxonomy, a hierarchical structure with the following categories was constructed, as presented in **Chart 3**:

Chart 3. Structure of Hierarchical Relationships (in Portuguese)

Categoria	Característica	Relações
1. Crime contra a pessoa:	Abrange os documentos de crimes previstos pelo Código Penal Brasileiro (1940) – CPB. Segundo o CPB, são crimes cometidos que afetam diretamente a pessoa, como bens físicos ou morais, categorizadas como ofensas e ameaças direcionadas a personalidade humana, tais como	Abandono de incapaz; Ameaça; Calúnia; Chantagem; Constranger alguém; Difamação; Homicídio; Injúria; Lesão Corporal; Maus tratos; Sequestro e Rixa.

2. Crime contra o patrimônio:	São os documentos que envolve ação criminosa direcionadas a tentar contra o patrimônio de um indivíduo ou de uma organização.	Apropriação indébita; Dano; Esbulho possessório; Estelionato; Extorsão; Furto; Porte de arma; Receptação; Roubo; Tráfico de drogas; Posse de entorpecentes e Violação de domicílio.
3. Crime contra a dignidade sexual:	São os documentos mais graves de violação dos direitos humanos, são crimes que atingem em sua maioria mulheres e menores de idade.	Estupro; Atentado ao pudor; Sedução; Ato obsceno e Bigamia.
4. Crimes contra a Incolumidade pública	São os documentos classificados com crimes que representam perigo ou algum dano a sociedade como um todo.	Incêndio.
5. Crimes contra a fé pública	São as informações referentes a documentos descritos nos artigos do Código Penal Brasileiro.	Falsificar documentos; Falsidade ideológica e Falsa identidade
6. Crimes contra a administração pública	São todos os documentos relacionados a infrações praticadas por agentes públicos.	Peculato; Concussão; Corrupção ativa; Corrupção Passiva; Contrabando; Prevaricação; Abandono de função; Resistência; Desobediência; Desacato e Abuso de Poder.

Source: Authors (2022)

With the categories defined, we constructed a taxonomy of the Police Inquiry Series, which is part of the corpus of the master's thesis entitled: *"The Application Path of Thesauri in Archival Documents: The creation of the micro-thesaurus of the Police Inquiry Series of the Public Archives of the State of Pará,"* submitted to the Graduate Program at the Federal University of Pará – PPGCI/UFPA.

In this context, taxonomy was chosen to briefly meet the demand by classifying the information in the Police Inquiry Series. Taxonomy helps represent and relate concepts through a Knowledge Organization System.

The next step was to transfer the classified terms into the THESA software to construct the thesaurus effectively.

THESA represents a free web-based application designed as a tool for undergraduate library science students, especially in the course on Documentary Languages. Its purpose is to simplify operational tasks, allowing students to focus their cognitive and conceptual efforts on domain modeling to improve process efficiency (Gabriel Júnior; Laipelt, 2017).

The choice of this software was based on its structure, which is grounded in the conception of relationships between concepts. In THESA, a concept can be represented by a term, an image, a sound, a link, or any other form chosen by its creators to express it.

THESA can be accessed via the link provided in the footer and downloaded from GitHub. In version 0.20.05.18 of THESA, it is possible to create an unlimited number of thesauri, meaning it supports multiple thesauri, and the users of these thesauri can configure it for public or private use (Gabriel Júnior; Laipelt, 2017).

Moreover, THESA is based on the current international standards for thesaurus construction from ISO and NISO, as well as the semantic guidelines and requirements of Knowledge Organization Systems and the Semantic Web.

The Police Inquiry Series micro-thesaurus is being developed for the general public, but access to it is restricted until its final stage. So far, the 45 descriptors include the General Term, Related Term (RT), Concept Definition, Scope Note/Application Note, and corresponding graph.

Regarding the definition of concepts, we emphasize that the documentation of the Police Inquiry Series is specific, grounded in Brazilian Penal Legislation from a particular time and context.

As described by Campos and Gomes (2008), a thesaurus's vocabulary does not consist of words from natural language (discourse) but rather of a set of terms that are verbal signals indicating a concept within a particular context.

A preliminary analysis of the 45 graphs generated by THESA was also conducted. Each graph presents a visual representation of existing relationships, a conceptual map of the relationships between terms. **Chart 4** shows the number of vertices in the graph and indicates the related term (RT) for each descriptor:

Chart 4. Number of Vertices (in Portuguese)

Descritor	Quantidade de vértices
Abandono de função	2
Abandono de incapaz	2
Abuso de poder	1
Ameaça	5
Apropriação indébita	5
Atentado ao pudor	3
Ato obsceno	3
Bigamia	1
Calúnia	1
Chantagem	2
Concussão	1
Constranger alguém	8
Contrabando	4
Corrupção ativa	5
Corrupção passiva	2
Dano	3
Desacato	3
Desobediência	3
Difamação	3
Esbulho possessório	3
Estelionato	4

Estupro	3
Extorsão	2
Falsa identidade	3
Falsidade ideológica	2
Falsificar documentos	2
Furto	4
Homicídio culposo	2
Homicídio doloso	1
Incêndio	1
Injúria	3
Lesão corporal	8
Maus tratos	2
Peculato	5
Porte de arma	1
Posse de entorpecentes	1
Prevaricação	2
Receptação	2
Resistência	2
Rixa	1
Roubo	5
Sedução	5
Sequestro	1
Tráfico de drogas	2
Violação de domicílio	2

Source: Authors (2022)

5 CONCLUSÃO

O propósito deste artigo foi apresentar a proposta de elaboração de um micro tesauro da Série Inquéritos Policiais do Arquivo Público do Estado do Pará. Os procedimentos metodológicos foram de caráter exploratório – apoiados em pesquisa bibliográfica e documental e aplicativo – com auxílio do *software* THESA.

Os resultados destacaram a eficácia do THESA na modelagem conceitual da Série Inquéritos Policiais, evidenciando sua utilidade na criação de modelos mais precisos. Dessa forma, percebe-se que a aplicação do THESA contribui significativamente para a elaboração de um tesauro que otimize os processos de transferência de informações nos sistemas arquivísticos.

Destaca-se, especialmente, a relevância do micro tesauro da Série Inquéritos Policiais como um complemento vital para a recuperação da informação, assim como, finalizado em sua totalidade, beneficiará usuários e pesquisadores, e contribuindo para o controle efetivo do

vocabulário utilizado. Em síntese, este estudo oferece uma valiosa contribuição ao domínio da informação orgânica-funcional e futuras pesquisas com o uso do THESA em novas séries documentais.

REFERENCES

AMERICAN NATIONAL STANDARDS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.19-2005**: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda, Ma: NISO Press, 2005. 184 p. Available at: <http://www.niso.org/standards/resources/Z39-19-2005.pdf> Access on: 15 fev. 2020.

BARITÉ, M. *et al.* **Diccionario de organización del conocimiento**: clasificación, indización, terminología. 6 ad. Montevideo: CSIC, 2015.

BARBOSA, A. **Teoria e prática dos sistemas de classificação bibliográfica**. Rio de Janeiro: Instituto Brasileiro de Bibliografia e Documentação, 1969. 441 p.

BARROS, T.H.B; BASTOS, C. M. C; SANTOS, A. C. R. dos. Sistemas de organização do conhecimento no contexto da arquivologia: aportes metodológicos para seu desenvolvimento. **Acervo**, Rio de Janeiro, RJ, v. 35, n. 2, p. 1-20, 2022. Available at: <https://revista.an.gov.br/index.php/revistaacervo/article/view/1812>. Access on: 21 jun. 2024.

BOCCATO, V. R. C.; RAMALHO, R. A. S.; FUJITA, M. S. L. A contribuição dos tesauros na construção de ontologias como instrumento de organização e recuperação da informação em ambientes digitais. **Ibersid**: revista de sistemas de información y documentación, Zaragoza, v. 2, p. 199–209, 2008. DOI: 10.54886/ibersid.v2i.2235.

CABRÉ, M.T. **La Terminología**: teoria, metodologia e aplicaciones. Barcelona: Editorial Antártida/Empúries, 1993. 529 p.

CAMPOS, M. L de A. **Em busca de princípios comuns na área de representação da informação**: uma comparação entre o método de classificação facetada, o método de tesauro-baseado-em-conceito e a teoria geral da terminologia. 1994, 196 f. Dissertação (Mestrado em Ciência da Informação) - Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ.

CAMPOS, M. L. A.; GOMES, H. E. Metodologia de Elaboração de Tesauro Conceitual: a categorização como princípio norteador. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 3, 2008. Available at: <https://periodicos.ufmg.br/index.php/pci/article/view/23767>. Access on: 21 jun. 2024.

CURRÁS, E. **Tesauros**: manual de construcción y uso. Madrid: Kaher II, 1998.

DAHLBERG. Teoria do conceito. **Ciência da Informação**, Rio de Janeiro, RJ, v. 7, n. 2, p. 101-107, 1978.

DAVANZO, L. **Vocabulário controlado para arquivos**: análise de viabilidade e proposta. 2016. Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista, Marília, SP. Available at: <http://hdl.handle.net/11449/138956>. Access on: 21 jun. 2024.

DODEBEI, V. L. D. **Tesauro**: linguagem de representação da memória documentária. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002.

FUJITA, M. S. L. A Estrutura de categorias do tesauro: modelos de elaboração. **Cadernos do F.F.C.**, Marília, SP, v.7, n.12, p. 107-120, 1998.

GABRIEL JUNIOR, R. F.; LAIPELT, R. C. Thesa: ferramenta para construção de tesauro semântico aplicado interoperável. **Revista P2P & Inovação**, Rio de Janeiro, RJ, v. 3, n. 2, p.124-145, mar./set. 2017. Available at: <https://revista.ibict.br/p2p/article/view/3815>. Access on: 26 ago. 2024.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2002.

GOMES, H. E. (org.). **Manual de elaboração de tesauros monolíngues**. Brasília: Programa Nacional de Bibliotecas de Instituições de Ensino Superior, 1990.

GUIMARÃES, J. A. C. Organização do Conhecimento: passado, presente e futuro sob a perspectiva da ISKO. **Informação & Informação**, Londrina, v. 22, n. 2, p. 84 – 98, maio/ago, 2017. <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/31443>. Access on: 26 ago. 2024.

HJORLAND, B. Semantics and knowledge organization. **Annual Review of Information Science and Technology**, v. 41, p. 367-405, 2007.

IBICT. **Diretrizes para elaboração de tesauros monolíngues**. Brasília, 1984.

ISO 2788. **Guidelines for the establishment and development of monolingual thesauri**. Geneve: International Standard Organization, 1986.

LAIPELT, R. C. F. **Metodologia para seleção de termos equivalentes e descritores de tesauros**: um estudo no âmbito do Direito do Trabalho e do Direito Previdenciário. 2015. Tese (Doutorado) - Programa de Pós-Graduação em Linguística Aplicada, Escola da Indústria Criativa: comunicação, designe linguagens, Universidade do Vale do Rio dos Sinos, São Leopoldo.

MINAYO, M. C. de S. (org.). **Pesquisa social**: teoria, método e criatividade. Petrópolis: Vozes, 2009.

MOTTA, Dilza Fonseca da. **Método relacional como nova abordagem para construção de tesauros**. Rio de janeiro: SENAI, 1987. 89 p. (Coleção Albano Franco, n. 12).

NAUMIS PEÑA, C. **Los tesauros documentales y su aplicación en la información impresa, digital y multimedia**. México: Universidad Nacional Autónoma de Mexico; Buenos Aires: Alfagrama, 2007. 288 p

NUCCI, G. de S. **Manual de processo penal e execução penal**. 13. ed. rev., atual. e ampl. Rio de Janeiro: Forense, 2016.

ROWLEY, J.E. **Organizing knowledge**: na introduction to informatiion retrieval. 2 ed. England: Ashgate, 1992. 509 p.

SANTOS, A.C.R. dos: **A trajetória de aplicação de tesauros em documentos de arquivo**: a criação do micro tesouro da Série Inquéritos Policiais do Arquivo Público do Estado do Pará. 2022. Dissertação (mestrado em Ciência da Informação) – Universidade Federal do Pará, Belém, PA.

SANTOS, A.C. R. dos; BARROS, T. H. B; LAIPELT, R. C. F. Organização do conhecimento e gestão dos tesauros na web semântica. **ConCI**: Convergências em Ciência da Informação, Aracaju, v. 4, p. 1–31, 2021. DOI: 10.33467/conci.v4i.16493. Available at: <https://periodicos.ufs.br/conci/article/view/16493>. Access on: 14 jan. 2024.

SANTOS, J. C. F. dos; MOREIRA, W. SKOS: uma análise sobre as abordagens e suas as aplicações na Ciência da Informação. **Informação & Informação**, Londrina, v. 23, n. 3, p. 362-389, 2018. Available at: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/29303>. Access on: 26 ago. 2024.

TÁLAMO, M. de F. G. M.; LARA, M. L. G. de; KOBASHI, N. Y. Contribuição da terminologia para a elaboração de tesauros. **Ciência da Informação**, Brasília, v.21, n.3, p.197-200, set./dez. 1992. Available at: <https://revista.ibict.br/ciinf/article/view/432>. Access on: 26 ago. 2024.

TERRA, J. C. C. *et al.* Taxonomia: elemento fundamental para a gestão do conhecimento. **Biblioteca Terra Forum Consultores**, 2005. Available at: <http://paginapessoal.utfpr.edu.br/mansano/arquivos/taxonomia.pdf>. Access on: 21 jun. 2024.

UNESCO. **Diretrizes para o estabelecimento e desenvolvimento de tesauros monolíngues**. [2ª ed. rev. por Derek Austin e Peter Dale]; Trad. de Bianca Amaro de Melo; rev. de Lígia Maria Café de Miranda. Brasília: IBICT; SENAI, 1993.

VAN SLYPE, G. **Los lenguajes de indización**: concepción, construcción y utilización en los sistemas documentales. Traducción del frances: Pedro Hípola, Félix de Moya. Madrid: Fundación Germán Sanchez Ruipérez, 1991.

VOGEL, M. J. M.; KOBASHI, N. Y. Tesouro funcional para organização de arquivos administrativos. **Páginas a&b**: arquivos e bibliotecas, Porto, p. 48–62, 2019. Available at: <https://ojs.letras.up.pt/index.php/paginasueb/article/view/6370> . Access on: 15 jan. 2024.