



Correspondência dos autores

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

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

Adriana Carla Ribeiro dos Santos¹ 
Thiago Henrique Bragato Barros² 

RESUMO

Introdução: No cenário da Organização do Conhecimento (OC) e no campo arquivístico, o uso de tesouros 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 – Tesouro 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. Tesouro. THESA.

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

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:** This study presents a proposal for the development of a micro-thesaurus aimed at the "Police Inquiries Series" of historical documents from the Public Archives of the State of Pará (APEP), using the software THESA – Applied Semantic Thesaurus as a tool.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.

CRediT

- **Reconhecimentos:** Não aplicável.
- **Financiamento:** Não aplicável.
- **Conflitos de interesse:** Os autores certificam que não têm interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.
- **Aprovação ética:** Não aplicável.
- **Disponibilidade de dados e material:** Os conjuntos de dados gerados e/ou analisados durante o presente estudo estão disponíveis no Repositório Institucional do PPGCIUFPA no site: <https://ppgci.propesp.ufpa.br/index.php/br/>
- **Contribuições dos autores:** Conceituação, Curadoria de Dados, Análise Formal, Investigação, Metodologia, Administração de Projetos, Recursos, Software, Validação, Visualização, Escrita - rascunho original, Escrita – revisão & edição: SANTOS, A.C.R.; Supervisão, Escrita – rascunho original, Escrita – revisão & edição: BARROS, T.H.B.

JITA: ID. Knowledge representation

ODS: 17 - Parcerias e meios de implementação



Artigo submetido ao sistema de similaridade

Submetido em: 31/01/2024 – Aceito em: 24/07/2024 – Publicado em: 27/08/2024

Editor: Gilденir Carolino Santos

1 INTRODUÇÃO

O progresso constante nas áreas de Informática e Comunicação tem impulsionado de maneira significativa a ampliação da produção de conhecimento arquivístico, especialmente no contexto digital. A complexidade deste ambiente, demanda à representação eficaz do conhecimento por meio de instrumentos terminológicos, tornando-se essencial para facilitar o acesso informacional.

Apesar da importância da representação terminológica, a utilização de tesouros funcionais, ferramentas complementares para solucionar problemas de acesso à informação orgânica, é quase inexistente nos arquivos brasileiros. Essa lacuna compromete a eficácia dos sistemas de recuperação de informações arquivísticas, prejudicando a transmissão de conhecimentos acumulados.

Este artigo constitui uma parte dos desdobramentos da dissertação de mestrado, em que Santos (2022) propôs a elaboração de um tesouro destinado à documentação do APEP, utilizando o *software* THESA. O propósito deste artigo é apresentar os resultados obtidos com o micro tesouro, que se configura como um modelo mais minucioso da informação orgânico-funcional, visando aprimorar os procedimentos metodológicos da Organização Conhecimento no âmbito arquivístico.

O *software* THESA, baseado nas diretrizes do *Simple Knowledge Organization System* (SKOS) e nas Normas ISO e NISO de tesouros e vocabulários controlados, foi empregado como ferramenta para a construção do tesouro proposto. Este aplicativo, disponível para em plataforma web gratuita, apresenta-se como uma solução baseada na *Web Semântica*, adequada tanto para fins didáticos em instituições de ensino quanto para aplicações profissionais.

A escolha da Série Inquéritos Policiais do APEP, justifica-se pela sua natureza jurídico-administrativa e pela alta demanda de usuários/pesquisadores. O tesouro visa não apenas atender às necessidades específicas dessa documentação, mas também ressalta a importância dos tesouros como ferramentas complementares para superar desafios de acesso à informação orgânico-funcional.

Além desta seção, o presente artigo é composto por mais quatro seções distintas. A segunda, Os tesouros na Organização do Conhecimento, apresenta uma abordagem teórica e metodológica da Organização do Conhecimento e da aplicação de tesouros no contexto dos arquivos. A terceira seção, trata do "Enquadramento Metodológico", delinea os procedimentos metodológicos adotados durante as etapas referentes à elaboração do micro tesouro da Série Inquéritos Policiais. Já a quarta seção, intitulada "Resultados e Discussão", apresenta de forma detalhada os resultados obtidos por meio da utilização do *Software* THESA. Por fim, a última seção, "Conclusões", destaca os principais contributos alcançados ao longo do estudo e aponta possibilidades para trabalhos futuros.

2 OS TESAUROS NA ORGANIZAÇÃO DO CONHECIMENTO

O "tesouro" é um instrumento essencial no universo da Organização do Conhecimento. Este sistema, muitas vezes subestimado, desempenha um papel vital na classificação e no relacionamento de termos, permitindo uma indexação mais precisa e uma recuperação mais eficiente da informação. Ao explorar o tesouro como um Sistema de Organização do Conhecimento, examinamos como ele contribui para a coerência e a interconexão de conceitos, facilitando a busca e a compreensão de informações em diversas áreas do conhecimento. Portanto, buscamos na literatura da Ciência da Informação e em áreas afins como diferentes autores abordam e aplicam conceitos relacionados à Organização do Conhecimento e tesouros,

a fim de permitir uma compreensão mais abrangente das tendências, desafios e desenvolvimentos na área.

2.1 Organização do Conhecimento: Abordagem teórica-metodológica

A Organização do Conhecimento (OC) mantém uma estreita relação com a Ciência da Informação, consolidando-se ao longo do tempo como um campo interdisciplinar em constante busca por uma identidade definida. Esse esforço por consolidação, diálogo teórico e aplicação é evidenciado na literatura da International Society of Knowledge Organization - ISKO (Guimarães, 2017).

Apesar de sua discussão remontar a períodos antigos, diversos modelos foram concebidos ao longo dos anos. A noção de OC teve seus primórdios por volta de 1900, com contribuições significativas de Charles A. Cutter, W. C. Berwick Sayers e Ernest Cushington Richardson. Em 1929, Henry Bliss publicou o livro *The Organization of Knowledge and the System of the Sciences*, considerado uma das principais contribuições intelectuais no campo da OC (Hjørland, 2007).

Conforme Rowley (1992), a OC se refere aos sistemas que organizam documentos e informações, permitindo sua recuperação pelos usuários quando solicitados. Hjørland (2007) destaca a diversidade de abordagens e teorias que contextualizam os estudos da OC, enfatizando a importância de escolhas bem-informadas no ensino e na pesquisa do campo, com base em sistemas de atividades e teorias científicas fundamentais para organizar o conhecimento.

Dentro do escopo da OC, Campos (1994) destaca a relevância dos princípios teóricos, conceituais e metodológicos inerentes à Teoria da Classificação Facetada, à Teoria do Conceito e à Teoria Geral da Terminologia. Esses componentes desempenham um papel determinante no desenvolvimento de abordagens eficazes para a organização, representação e recuperação da informação e do conhecimento.

Na primeira metade do século XX, foi desenvolvida a Teoria da Classificação Facetada (TCF) por Shiyali Ramamrita Ranganathan. Durante a década de 1930, a TCF foi considerada uma estrutura dinâmica e multidimensional, sendo o termo "faceta" atualmente utilizado nos estudos modernos da teoria da classificação como sinônimo de característica (Barbosa, 1969, p. 16).

A TCF é amplamente citada nos estudos sobre classificação e elaboração de tesouros. O livro *Prolegomena to Library Classification* de Ranganathan apresenta os fundamentos de sua teoria e os princípios para a organização de conceitos hierarquicamente estruturados, considerados uma abordagem universal. Além disso, em textos e estudos mais recentes, a TCF é amplamente citada na elaboração de ontologias, que contribuem para o aprimoramento de sistemas de recuperação da informação no contexto da web semântica (Laipelt, 2015).

Laipelt (2015) ressalta que, embora a TCF se refira à elaboração de sistemas de classificação, que visam, inclusive, à organização física de um acervo, ela é amplamente usada para a elaboração de tesouros, pois utiliza categorias universais que podem ser adaptadas e aplicadas em vários domínios do conhecimento, conforme descreveu a autora.

Também temos a chamada Teoria Geral da Terminologia (TGT), conhecida como a Teoria da Terminologia de Wüster, concebida a partir de uma cuidadosa seleção de termos da engenharia industrial, com o objetivo de normatizar os conceitos linguísticos e intralinguísticos. Seu propósito é eliminar ambiguidades inerentes à linguagem natural, proporcionando clareza na comunicação profissional em âmbitos nacional e internacional. Estruturada com instruções

precisas, a TGT busca garantir a univocidade comunicativa. Entretanto, no contexto da comunicação natural, suas características revelam-se insuficientes (Cabrè, 1999).

Por sua vez, a Teoria do Conceito (TC) é uma metodologia híbrida, combinando métodos dedutivo e indutivo. Desenvolvida na década de 1970 por Dahlberg, a TC foi fortemente influenciada pela Teoria Geral da Terminologia de Wüster e pela Classificação Facetada de Ranganathan.

Dahlberg (1978) conceitua o elemento como uma entidade concreta nos sistemas de classificação, utilizando uma teoria analítica de conceitos para representar o conhecimento. A TC viabiliza a associação de cada conceito a um termo, defendendo, assim, a necessidade de padronização terminológica. Além disso, a TC apresenta contribuições relevantes para a Arquivologia, notadamente no que se refere ao Plano de Classificação, ao fornecer uma estrutura sustentada por termos específicos.

Conforme Dahlberg (1978), os conceitos são classificados como gerais ou individuais, possibilitando a formulação de conceitos referentes a diversas entidades. Cada enunciado verdadeiro representa um elemento do conceito, e a combinação de todas as afirmações verdadeiras compõe o conceito associado a um objeto específico. De acordo com a TC, cada enunciado apresenta um atributo predicável do objeto, chamado de característica no nível do conceito. Muitas vezes, essa hierarquia de características culmina em uma categoria, entendida como o conceito em sua mais ampla extensão.

Ao considerar essas teorias, é possível compreender a Organização do Conhecimento como um domínio dentro de uma perspectiva sistêmica. A OC emerge como uma área que desempenha um papel significativo na criação e desenvolvimento de metodologias para o acesso à informação, incluindo o desenvolvimento de Sistemas de Organização do Conhecimento (SOC), ferramentas que proporcionam uma interpretação organizada e estruturada dos objetos em questão.

2.2 Tesouro enquanto Sistema de Organização do Conhecimento

Pesquisamos uma variedade de definições sobre tesouro, tornando necessário recorrer a um quadro-resumo que apresente algumas dessas definições, conforme normas e autores específicos. A consulta ao Quadro 1 permite uma compreensão abrangente das nuances e abordagens que cercam o conceito de tesouro:

Quadro 1. Definições de tesouro

Definição de tesouro	Autores específicos
"[...] vocabulário controlado e dinâmico abrangendo área específica do conhecimento. Em sua estrutura, patenteia as relações vigentes entre os termos ou descritores – sinonímicas hierárquicas e outras – que, no conjunto, constitui a linguagem de indexação".	(IBICT, 1984, p. 5)
"[...] vocabulário de uma linguagem de indexação controlado e organizado formalmente com objetivo de explicitar as relações a priori entre conceitos (por exemplo, mais genérico que... ou mais específico que...)". Para isso, descreve a linguagem de indexação como: "conjunto controlado de termos extraídos da linguagem	(ISO 2788, 1986)

natural e utilizados para representar de forma breve os assuntos dos documentos”.	
Tesouro é um vocabulário de termos relacionados genérica e semanticamente sobre determinada área de conhecimento.	(Motta, 1987)
“[...] Linguagem documentária dinâmica que contém termos relacionados semântica e logicamente, cobrindo de modo compreensivo um domínio do conhecimento”.	Gomes, 1990, p. 16)
“[...] lista estruturada de conceitos destinados a representar de maneira unívoca o conteúdo dos documentos e das consultas dentro de um sistema documental determinado [...] inclui descritores, não-descritores, relações hierárquicas e de associação e equivalências linguísticas”.	(Van Slype, 1991, p. 23-24)
“[...] linguagem documentária, construída por meio de unidades conceituais, extraídas da linguagem formal de uma área específica do conhecimento científico ou técnico. Sua estrutura sugere a ideia de sistema, visto que os conceitos se relacionam entre si e são representados por termos. Cada termo, por sua vez, possui vinculação com outro termo, por meio de relação de equivalência, de hierarquia ou de associação. O tesouro, utilizado para a organização e recuperação da informação, constitui-se em importante “[...] instrumento de apoio às pesquisas científicas nas áreas de conhecimento”.	(Fujita, 1998)
“[...] linguagem documentária que representa de forma normalizada os conceitos de uma área específica através de termos que se manifestam em estruturas lógico-semânticas”.	(Tálamo; Iara; Kobashi, 1992)
“[...] vocabulário controlado de uma linguagem de indexação, formalmente organizado para explicitar as relações a priori entre conceitos (por exemplo, como genéricas e específicas).	(UNESCO, 1993, p. 14)
“[...] vocabulário especializado, normalizado, pós-coordenado, usado com fins documentários, onde os elementos linguísticos que o compõem, termos simples ou compostos, encontram-se relacionados entre si sintática e semanticamente”.	(Currás, 1998)
“[...] sistema de classificação temática ou facetada, cuja estrutura básica está conformada por uma relação de descritores que representam ou descrevem autoridades ou conteúdos temáticos. [...] por meio de unidades linguísticas, semânticas e suas relações, extraídas da linguagem formal de uma disciplina ou área específica do conhecimento	(Naumis Peña, 2000)

que [...] se torna um instrumento de representação e recuperação da informação”.	
“[...] instrumento apropriado para transmitir conceitos e as relações recíprocas destes, semelhantemente ao que ocorre com os termos expressos na linguagem dos documentos”.	(Dodebei, 2002, p. 67)
“[...] como função – “um instrumento de controle terminológico que permite traduzir a linguagem natural dos documentos, dos indexadores e dos usuários, numa ‘linguagem sistêmica’ mais rígida (linguagem documentária, linguagem do sistema de informação)”. Como estrutura – “um vocabulário controlado e dinâmico de termos relacionados semântica e genericamente, que cobre um campo específico de conhecimentos”.	(ANSI/NISO-Z39.19, 2005, p. 9)
“[...] linguagens de estruturas combinatórias e pós-coordenadas, constituídas de termos – unidades linguísticas provenientes da linguagem de especialidade e da linguagem natural –, denominados de descritores, providos de relações sintático-semânticas, referentes a domínios científicos especializados, possibilitando a representação temática do conteúdo de um documento, bem como a recuperação da informação”.	(Boccatto; Ramalho; Fujita, 2008, p. 201)
Tipo de sistema de organização do conhecimento integrado aos termos analisados e padronizados que mantêm relações semânticas e funcionais entre si. O dicionário de sinônimos é organizado sob forte controle terminológico, a fim de fornecer um instrumento ideal para armazenar e recuperar informações em áreas especializadas. [...]	(Baritè <i>et al.</i> , 2015)

Fonte: (Santos; Barros; Laipelt, 2021)

Em concordância com as definições apresentadas no Quadro 1, é possível ressaltar que o tesauro se configura como um instrumento de relevância significativa e especialmente apropriado para a representação e recuperação de informações em domínios especializados.

O tesauro enquanto Sistema de Organização do Conhecimento são usados para sistematizar conteúdos de conhecimento, representando e relacionando o vocabulário inerente a um domínio particular (Santos; Moreira, 2018). Portanto, o tesauro é um instrumento que aliados às novas Tecnologias de Informação e Comunicação transformam o conteúdo de conhecimento formalizado, identificável e interoperável por meio de máquinas, considerando a dinamicidade da infraestrutura da *web*.

Segundo Vogel e Kobashi (2019), há uma relevância na criação de tesouros funcionais como o instrumento mais apropriado, sobretudo, para arquivos correntes e intermediários, pois, o tesauro tradicional é concebido de maneira temática, sendo mais apropriado para arquivos permanentes, já os tesouros funcionais se revelam ideais para eficaz tratamento da informação nas práticas e metodologias específicas da área.

O tesauro funcional oferece novas possibilidades para acessar informações arquivísticas, promovendo uma maior confiança no sistema. No entanto, para um

desenvolvimento e uso eficazes dos tesouros funcionais, é imprescindível planejar e estabelecer um plano de classificação arquivística, pois o plano reflete as funções e atividades realizadas nas instituições.

Isso implica que o tesouro tradicional pode não ser a escolha mais apropriada para a organização de arquivos. No entanto, é relevante observar que existem tesouros específicos para arquivos, notadamente aqueles de natureza histórica, que enfatizam a categorização temática do conteúdo. Exemplificando, temos o "Tesouro del Patrimonio Histórico Andaluz" em espanhol, o "United Kingdom Archival Thesaurus (UKAT)" em inglês, e os franceses "Thesaurus W: vocabulaire normalisé pour la description des archives administratives locales contemporaines" e "Archives de Cannes". Estes tesouros mencionados geralmente se estruturam em torno de três listas principais: assuntos, instituições e localidades geográficas (Giménez-Chornet e Escrig-Giménez, 2011).

Assim, a utilização de uma linguagem normalizada, alinhada ao contexto e à cultura organizacional, é um dos elementos essenciais para alcançar eficiência nos processos de produção, organização e disseminação da informação. Embora o plano de classificação seja consolidado na Arquivologia, na prática, a necessidade de instrumentos que contribuam para a padronização terminológica é insipiente.

Isso ressalta a importância de utilizar um instrumento de representação e recuperação da informação direcionado aos usuários e pesquisadores, proporcionando maior confiança em suas atividades de pesquisa. Ao construir um tesouro funcional na Arquivologia, é imperativo considerar o plano de classificação, uma vez que o tesouro funcional impacta significativamente nas funções e atividades das instituições.

A compreensão de que o tesouro funcional desempenha um papel primordial no apoio ao arquivista, a OC o destaca como uma ponte de acesso essencial, orientando as estratégias de busca nos sistemas. A construção do tesouro funcional deve ser uma colaboração abrangente que envolve a instituição, produtores de documentos, usuários, arquivistas e outros profissionais capazes de contribuir, adaptando-o às realidades informacionais específicas das instituições. | 8

Apesar da escassez de pesquisas abordando tesouros funcionais na Arquivologia, que exploram a relação entre organização, representação e uso da informação, as normas de descrição na área destacam a necessidade de instrumentos que operem com perspectivas de padronização e controle terminológico.

No entanto, Davanzo (2016) observa a ausência de abordagem nas normas sobre como o tesouro funcional pode ser desenvolvido para contribuir com o tratamento da informação. Além disso, a autora fala que as normas não esclarecem se a Arquivologia pode fazer uso de um tesouro funcional ou conceitual.

Para o arquivista, a necessidade de disponibilizar informações é incontestável, e o sucesso nesse objetivo depende fundamentalmente do uso de instrumentos que contribuam com essa perspectiva. Assim, a utilização do tesouro funcional para orientar a organização, especialmente na recuperação da informação arquivística, não apenas reforça o princípio da proveniência, mas também materializa, em sua estrutura, relações hierárquicas, associativas e de equivalências por meio de descritores normalizados, representando de maneira eficaz as atividades e funções contidas nos documentos e assegurando a preservação da informação orgânico-funcional na memória documentária arquivística.

3 METODOLOGIA

Conforme as diretrizes da metodologia de pesquisa científica estabelecidas por Gil (2002), este estudo é caracterizado da seguinte forma:

- a) Natureza: Considera-se como pesquisa aplicada, uma vez que resultou em aplicações práticas diretas na elaboração de tesouros destinados a arquivos.
- b) Abordagem do Problema: Classifica-se como pesquisa qualitativa, dada a natureza do conhecimento explorado, intrinsecamente difícil de mensurar quantitativamente. O foco reside na compreensão do problema, sem a intenção de quantificar os dados envolvidos.
- c) Objetivos: Enquadra-se como pesquisa explicativa, pois visa examinar a inter-relação entre a Organização do Conhecimento e a Arquivologia, à luz da literatura existente.
- d) Procedimentos Técnicos: Assume a forma de pesquisa bibliográfica, ao lidar com a literatura publicada; e pesquisa experimental, devido à construção de um modelo de tesouro para o domínio da Arquivologia.

O estudo foi conduzido em três fases distintas, conforme proposto por Minayo (2009): a fase exploratória, o trabalho de campo e o tratamento e análise dos dados. O *corpus* da pesquisa consiste em textos da área de Ciência da Informação e disciplinas afins, abordando tópicos como Organização do Conhecimento e Arquivologia.

Dado o escasso número de referências para o tesouro funcional, a pesquisa se concentra exclusivamente em tesouros como instrumentos aplicados à Organização do Conhecimento para recuperação da informação.

A etapa experimental compreendeu a construção de um modelo de tesouro para o domínio arquivístico, baseado no modelo australiano AGIFT, identificado no *site* do *National Archives of Australia* (NAA) e elaborado com o auxílio do *software* THESA. Este modelo foi testado em documentos da Série Inquéritos Policiais do Arquivo Público do Estado do Pará – APEP, visando avaliar sua aplicabilidade e interoperabilidade.

O embasamento teórico para a construção desse modelo é derivado da interdisciplinaridade entre as metodologias da Ciência da Informação e da *Web Semântica*. Este modelo servirá tanto para padronização e controle terminológico quanto para mitigar problemas relacionados ao acesso à informação arquivística.

O *software* THESA segue as diretrizes do *SKOS* (*Simple Knowledge Organization System*) e das Normas ISO e NISO para tesouros e vocabulários controlados. Com uma base na *Web Semântica*, é disponibilizado gratuitamente e pode ser utilizado tanto academicamente, em disciplinas de graduação e pós-graduação, quanto profissionalmente. Este aplicativo permite a criação de um número ilimitado de tesouros, oferecendo a opção de configuração para uso público ou privado (Gabriel Junior; Laipelt, 2017).

É crucial destacar a escolha do *software* devido à sua estrutura, baseada na concepção das relações entre os conceitos, partindo do pressuposto de que um conceito pode ser representado por um termo, uma imagem, um som, um *link* ou qualquer outra forma que possa ser explicitada, conforme descrito pelos seus desenvolvedores.

4 RESULTADOS E DISCUSSÃO

A pesquisa documental se deu no Arquivo Público do Estado do Pará (APEP), Fundo documental da Secretaria de Segurança Pública - Série Documental Inquéritos Policiais, datados de 1956 até 1990, classificados e armazenados em 438 caixas de arquivos.

Foi feita leitura, análise, descrição arquivística e uma taxonomia dos documentos produzidos pelas delegacias da Região Metropolitana de Belém e Interiores do Pará.

Foi construído um roteiro de análise dos documentos, com as seguintes etapas:

- a) Tratamento documental;
- b) Leitura e descrição dos documentos da série inquéritos policiais;
- c) Identificação das temáticas;
- d) Catalogação dos termos com maior frequência;
- e) Construção da taxonomia.
- f) Transferência dos termos para o THESA

O APEP possui o setor de Documentação Permanente que é responsável pelo gerenciamento, controle e tratamento técnico da documentação existente na instituição. Tem como atividades: o recolhimento, separação (tipos documentais, fundos e data limite), além da catalogação e acondicionamento. Portanto, o tratamento documental, leitura e descrição dos inquéritos policiais foi feita neste setor, sob a orientação e supervisão dos funcionários.

O acervo analisado para o *corpus* da pesquisa foi composto por 198 caixas de arquivos. Cada caixa é identificada com o nome do fundo, da série, da delegacia, ano de ocorrência e número da caixa. As caixas contêm um número variado de inquéritos policiais, a menor apresenta três pastas de inquéritos policiais, mais precisamente da Delegacia de Ordem Política e Social - DOPS e a de maior número, com 85 pastas de inquéritos da Delegacia do município de Santa Izabel no Pará.

Segundo Nucci (2016) o Inquérito Policial é um procedimento preparatório da ação penal, de caráter administrativo, conduzido pela polícia judiciária e voltado à colheita preliminar de provas para apurar a prática de uma infração penal e sua autoria. Portanto, a finalidade do inquérito policial é a apuração dos fatos que configuram a ação ou transgressão criminal, estabelecidas na legislação penal brasileira.

Foram catalogados 45 termos com maior frequência nos documentos da Série Inquéritos Policiais, de uma amostragem de 50 caixas de arquivos, conforme apresentado no Quadro 2:

Quadro 2. Catalogação dos termos com maior frequência

	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

Fonte: Elaborado pelos autores (2022)

A partir da seleção específica e da mudança na frequência das palavras, foram determinadas as categorias essenciais para a condução da pesquisa, com o propósito de desenvolver uma taxonomia. Na área da Ciência da Informação, a taxonomia representa um

sistema de estruturação do conhecimento, que classifica e simplifica a obtenção de informações ao organizar esses dados em categorias, estabelecendo normas para as relações hierárquicas que existem entre elas (Barros; Bastos; Santos, 2022).

A taxonomia é definida como:

um vocabulário controlado de uma determinada área do conhecimento e, acima de tudo, um instrumento ou elemento de estrutura que permite alocar, recuperar e comunicar informações dentro de um sistema sob uma premissa lógica (Terra *et al.*, 2005, p. 1).

Desde a leitura e descrição dos documentos da Série Inquéritos Policiais, observou-se como eram registrados os boletins de ocorrências e definiu-se a categorização das classes. Para Campos e Gomes (2008) a categorização é um processo que requer pensar o domínio de forma dedutiva, ou seja, determinar as classes de maior abrangência dentro da temática escolhida.

Partindo dos princípios classificatórios de elaboração de taxonomia, foi construído uma estrutura de relações hierárquicas com as seguintes categorias, conforme é apresentado no Quadro 3:

Quadro 3. Estrutura das Relações Hierárquicas

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.
--	---	--

Fonte: Elaborado pelos autores (2022)

Com as categorias definidas, construímos uma taxonomia da Série Inquéritos Policiais que faz parte do *corpus* da dissertação de mestrado intitulada: “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”, submetida ao Programa de Pós-Graduação da Universidade Federal do Pará – PPGCI/UFPA.

Nesse parâmetro, a taxonomia foi o instrumento escolhido para atender à demanda de maneira sucinta ao classificar as informações contidas na Série Inquéritos Policiais, o que auxilia na representação e na relação dos conceitos por meio de um Sistema de Organização do Conhecimento.

O outro momento, foi transferir os termos classificados para o *software* THESA, para construção efetiva do tesouro.

O THESA representa uma aplicação gratuita baseada na *web*, concebida como uma ferramenta para alunos de graduação em biblioteconomia, especialmente na disciplina de Linguagens Documentárias. Seu propósito é simplificar as tarefas operacionais, permitindo que os estudantes concentrem seus esforços cognitivos e conceituais na modelagem do domínio, visando aprimorar a eficiência do processo (Gabriel Júnior; Laipelt, 2017).

A escolha deste *software*, se fez pela sua estrutura que é baseada na concepção das relações entre os conceitos, ou seja, nele o conceito pode ser simbolizado por um termo, uma imagem, um som, um *link* ou qualquer outra forma que os seus criadores escolham para explicitá-lo.

O THESA¹ pode ser acessado no endereço informado no rodapé e baixado no *GitHub*². A versão 0.20.05.18 do THESA, é possível criar um número ilimitado de tesouros, ou seja, tem uma concepção de múltiplos tesouros e os usuários desses tesouros, podem configurá-lo para uso público ou privado. (Gabriel Júnior; Laipelt, 2017)

Além do mais, o THESA tem como base as normas internacionais de construção de tesouros da ISO e NISO vigentes, assim como, as diretrizes e os requisitos semânticos dos Sistemas de Organização do Conhecimento e da *Web Semântica*.

O micro tesouro da Série Inquéritos Policiais está sendo elaborado para um público em geral, porém editado com acesso restrito até sua fase final. Até o momento, os 45 descritores apresentam: o Termo Geral; Termo Relacionado TR; Definição do Conceito; Nota de escopo/nota de aplicação e um respectivo Grafo.

No que se refere a definição de conceitos, ressaltamos que a documentação da Série Inquéritos Policiais é específica, fundamentada na Legislação Penal Brasileira, de uma determinada época e contexto.

Como descreve, Campos e Gomes (2008), o vocabulário de um tesouro não se constitui por palavras de linguagem natural (discurso), e sim, por meio de um conjunto de

¹ Acesso: <https://www.ufrgs.br/tesauros/index.php/thesa>

² Acesso: <https://github.com/ReneFGJr/Thesa>

termos que consistem em sinais verbais que indicam um conceito em um contexto particular.

Foi feita também, uma análise preliminar dos 45 grafos gerados pelo THESA. Cada grafo apresenta uma representação visual das relações existentes, ou seja, um mapa conceitual das relações entre termos. O quadro 3 é apresentado o quantitativo dos vértices no grafo e mostra o termo relacionado (TR) do descritor:

Quadro 4. Quantitativo dos vértices

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
Ebulho 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

Fonte: Elaborado pelos autores (2022)

O Quadro 3 do quantitativo dos vértices pode ser usado, como um parâmetro de avaliação quantitativa das relações entre termos. O tesauro preliminar apresenta grafos contendo até 8 vértices e grafos de apenas 1 vértice. O que demonstra um indicativo para reavaliar e/ou estabelecer novas relações semânticas para a elaboração do micro tesauro da Série Inquéritos Policiais.

| 15

Observamos também que há uma predominância dos termos nas relações afinidade, significa que quando o usuário for fazer a pesquisa pelo termo que considera mais adequado terá uma boa precisão nos documentos que citam o termo, o que demonstra que o *software* THESA tem um refinamento completo, o que possivelmente caberia uma estrutura ontológica para uma extensão do tesauro para adicionar novas informações.

Estas foram nossas análises preliminares, portanto, os resultados demonstram que o *software* THESA pode ser adaptado e utilizado, como instrumento de controle terminológico em arquivos, podendo ser utilizado por outras instituições. Por intermédio do produto desenvolvido de forma precisa e representativa da Organização do Conhecimento.

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 tesouro que otimize os processos de transferência de informações nos sistemas arquivísticos.

Destaca-se, especialmente, a relevância do micro tesouro 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.

REFERÊNCIAS

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. Disponível em: <http://www.niso.org/standards/resources/Z39-19-2005.pdf> Acesso em: 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. Disponível em: <https://revista.an.gov.br/index.php/revistaacervo/article/view/1812>. Acesso em: 21 jun. 2024.

BOCCATO, V. R. C.; RAMALHO, R. A. S.; FUJITA, M. S. L. A contribuição dos tesouros 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 tesouro-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 Tesouro Conceitual: a categorização como princípio norteador. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 3, 2008. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/23767>. Acesso em: 21 jun. 2024.

CURRÁS, E. **Tesouros**: 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. Disponível em: <http://hdl.handle.net/11449/138956>. Acesso em: 21 jun. 2024.

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

FUJITA, M. S. L. A Estrutura de categorias do tesouro: 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 tesouro semântico aplicado interoperável. **Revista P2P & Inovação**, Rio de Janeiro, RJ, v. 3, n. 2, p.124-145, mar./set. 2017. Disponível em: <https://revista.ibict.br/p2p/article/view/3815>. Acesso em: 26 ago. 2024.

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

GOMES, H. E. (org.). **Manual de elaboração de tesouros 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>. Acesso em: 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 tesouros 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 tesouros**: 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 tesouros**. Rio de Janeiro: SENAI, 1987. 89 p. (Coleção Albano Franco, n. 12).

NAUMIS PEÑA, C. **Los tesouros 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 tesouros 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 tesouros na web semântica. **ConCI: Convergências em Ciência da Informação**, Aracaju, v. 4, p. 1–31, 2021. DOI: 10.33467/conci.v4i.16493. Disponível em: <https://periodicos.ufs.br/conci/article/view/16493>. Acesso em: 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. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/29303>. Acesso em: 26 ago. 2024.

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

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

UNESCO. **Diretrizes para o estabelecimento e desenvolvimento de tesouros 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. Disponível em: <https://ojs.letras.up.pt/index.php/paginasueb/article/view/6370> . Acesso em: 15 jan. 2024.