



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Fabício Dantas Bispo

**Controle terminológico no combate a crimes de entorpecentes na Polícia  
Federal do Brasil: proposta metodológica**

Florianópolis  
2023

Fabício Dantas Bispo

**Controle terminológico no combate a crimes de entorpecentes na Polícia  
Federal do Brasil: proposta metodológica**

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Mestre em Ciência da Informação.

Orientador: Prof. Dra. Luciane Paula Vital

Florianópolis  
2023

Bispo, Fabricio Dantas

Controle terminológico no combate a crimes de entorpecentes na Polícia Federal do Brasil : proposta metodológica / Fabricio Dantas Bispo ; orientador, Luciane Paula Vital, 2023.

106 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências da Educação, Programa de Pós-Graduação em Ciência da Informação, Florianópolis, 2023.

Inclui referências.

1. Ciência da Informação. 2. Vocabulário Controlado. 3. Organização do Conhecimento. 4. Crimes de Entorpecentes. I. Vital, Luciane Paula . II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Ciência da Informação. III. Título.

Fabricio Dantas Bispo

**Controle terminológico no combate a crimes de entorpecentes na Polícia Federal do Brasil:  
proposta metodológica**

O presente trabalho em nível de Mestrado foi avaliado e aprovado, em 30 de novembro de 2023, pela banca examinadora composta pelos seguintes membros:

Prof. Igor Soares Amorim, Dr.

UFPE - Universidade Federal de Pernambuco

Profa. Graziela Martins de Medeiros, Dra.

UFSC – Universidade Federal de Santa Catarina

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Ciência da Informação



---

Coordenação do Programa de Pós-Graduação



---

Profa. Dra. Luciane Paula Vital

Florianópolis, 2023

"Agradeço primeiramente a Deus, fonte de toda sabedoria, por me guiar  
nesta jornada desafiadora e enriquecedora de pesquisa."

## **AGRADECIMENTOS**

À Polícia Federal, minha sincera gratidão por proporcionar-me o desafio que serviu de base e inspiração para esta pesquisa, fortalecendo meu compromisso e dedicação ao longo deste percurso.

À Universidade Federal de Santa Catarina (UFSC), instituição que sempre promoveu a excelência acadêmica e me deu o suporte necessário para o desenvolvimento deste trabalho.

Em especial, à Professora Luciane Vital, cuja paciência, sabedoria e orientação foram essenciais para guiar-me neste projeto, moldando não apenas esta pesquisa, mas também minha trajetória acadêmica.

À minha querida esposa, agradeço por cada momento de apoio, compreensão e ajuda. Seus conselhos e sua presença constante foram meu porto seguro nos momentos mais turbulentos desta jornada.

Aos meus filhos e meus pais, que me impulsionam diariamente a ser uma pessoa e um profissional melhor. É por vocês que busco incansavelmente aprender, crescer e superar cada obstáculo.

*Consider a future device ... in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory (Bush, 1945).*

## RESUMO

No cenário pós Segunda Guerra Mundial, a informação assumiu uma relevância sem precedentes, refletida nos desafios atuais de gestão da informação na Polícia Federal brasileira. Esta pesquisa se propõe a investigar esses desafios e aprimorar a recuperação de dados relacionados a crimes de entorpecentes. Com base na previsão de uma explosão informacional e na necessidade de sistemas robustos para sua gestão, este estudo adota o uso de Sistemas de Organização do Conhecimento (SOC) e propõe etapas para elaboração de um controle terminológico. Para consolidação desses termos, foram empregadas técnicas de mineração de texto, bem como a análise de termos presentes em vocabulários já existentes, entrevista e questionários online. Esse esforço tem como finalidade aprimorar os processos de busca, indexação e assegurar uma recuperação de informação eficaz. Com a complexidade e especificidades do domínio em foco, a pesquisa evidencia a necessidade de modernização e controle terminológico, ressaltando a relevância do Vocabulário Controlado como ferramenta essencial para enfrentar os desafios informacionais da Polícia Federal.

**Palavras-chave:** Vocabulário Controlado; Crimes de Entorpecentes; Indexação.

## ABSTRACT

In the post-World War II era, information has gained unprecedented significance, a trend reflected in the current challenges of information management within the Brazilian Federal Police. This study investigates these challenges, specifically focusing on enhancing the retrieval of data related to drug-related crimes. Anticipating an information explosion and the need for robust systems to manage it, this research utilizes Knowledge Organization Systems (KOS) and outlines steps for developing a terminological control. The consolidation of terms involved techniques such as text mining, analysis of existing vocabularies, interviews, and online questionnaires. This effort aims to improve search processes, indexing, and ensure efficient information retrieval. Given the complexity and specificities of the focus domain, this research highlights the need for modernization and terminological control, emphasizing the significance of a Controlled Vocabulary as an essential tool to address the informational challenges faced by the Federal Police.

**Keywords:** Controlled Vocabulary; Drug Trafficking Crimes; Indexing.

## LISTA DE FIGURAS

Figura 1 - Relacionamento entre ontologia, thesaurus, taxonomia, índice e dicionário .....	34
Figura 2 - Gráfico de complexidade das Linguagens Documentárias – ANSI/NISO Z39.19-2005 .....	34
Figura 3 - Vocabulário Jurídico Supremo Tribunal Federal .....	37
Figura 4 - Organograma da PF .....	41
Figura 5 - Diversidade de dados analisados numa operação policial.....	43
Figura 6 - Percurso da metodologia da Pesquisa.....	46
Figura 7 - Painel de visualização de dados .....	53
Figura 8 - Amostra de tela do formulário .....	56
Figura 9 - Outra tela do formulário .....	57
Figura 10 - Painel do QlikSense.....	60
Figura 11 - Termos minerados - painel do QlikSense .....	61
Figura 12 - Órgãos dos especialistas .....	69
Figura 13 - Experiência dos especialistas .....	69

## LISTA DE QUADROS

Quadro 1 - Expressões utilizadas por ORCRIM (Organização criminosa) .....	44
Quadro 2 - Etapas realizadas no script part1.py .....	50
Quadro 3 - Etapas realizadas no script part2.py .....	50
Quadro 4 - Etapas realizadas no script classegramatical.py .....	52
Quadro 5 - Amostra dos termos na fase da Entrevista.....	54
Quadro 6 - Termos coletados em vocabulários controlados .....	59
Quadro 7 - Termos consolidados .....	66
Quadro 8 - Termos sinônimos consolidados através dos formulários eletrônicos.....	70
Quadro 9 - Amostra dos termos consolidados na pesquisa .....	73
Quadro 10 – Quadro Síntese .....	73

## LISTA DE TABELAS

Tabela 1 - Amostra de termos minerados nos arquivos – substantivos .....	62
Tabela 2 - Amostra de termos minerados nos arquivos – verbo .....	63
Tabela 3 - Amostra de termos minerados nos arquivos – adjetivo.....	64

**LISTA DE ABREVIATURAS E SIGLAS**

CGI	Coordenação-Geral de Inteligência
DICOR	Diretoria de Investigação e Combate ao Crime Organizado
DIP	Diretoria de Inteligência Policial
DSI	Divisão de Gestão de Informações e Sistemas de Empresas de São Paulo
FGV-EAESP	Fundação Getúlio Vargas - Escola de Administração de Inteligência
NA	Notas
ORC	Organização e Representação do Conhecimento
ORCRIM	Organização Criminosa
PF	Polícia Federal
PNL	Processamento de Linguagem Natural
RI	Recuperação da Informação
SINARM	Sistema Nacional de Armas
SOC	Sistemas de Organização do Conhecimento
TE	Termo específico
TG	Termo genérico
TR	Termo relacionado
TSTF	Tesouro Jurídico do Supremo Tribunal Federal
USE / UP	Equivalência

## SUMARIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>19</b>
<b>2</b>	<b>OBJETIVOS</b> .....	<b>21</b>
2.1	OBJETIVO GERAL .....	21
2.2	OBJETIVOS ESPECÍFICOS.....	21
<b>3</b>	<b>JUSTIFICATIVA</b> .....	<b>22</b>
<b>4</b>	<b>CIÊNCIA DA INFORMAÇÃO</b> .....	<b>26</b>
4.1	RECUPERAÇÃO DA INFORMAÇÃO .....	27
4.2	SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO.....	29
<b>4.2.1</b>	<b>Ontologias</b> .....	<b>30</b>
<b>4.2.2</b>	<b>Tesauros</b> .....	<b>31</b>
<b>4.2.3</b>	<b>Taxonomia</b> .....	<b>33</b>
<b>4.2.4</b>	<b>Anéis de sinônimos</b> .....	<b>35</b>
<b>4.2.5</b>	<b>Vocabulário Controlado</b> .....	<b>36</b>
4.3	POLÍCIA FEDERAL DO BRASIL .....	39
<b>4.3.1</b>	<b>Características dos dados analisados na Polícia Federal</b> .....	<b>42</b>
<b>5</b>	<b>PROCEDIMENTOS METODOLOGICOS</b> .....	<b>45</b>
5.1	ANÁLISE DE VOCABULÁRIOS EXISTENTES .....	46
5.2	MINERAÇÃO DE TEXTO .....	47
<b>5.2.1</b>	<b>Seleção de Corpus</b> .....	<b>48</b>
5.3	ENTREVISTA COM ESPECIALISTA.....	53
5.4	FORMULÁRIOS ELETRÔNICOS.....	55
<b>6</b>	<b>RESULTADOS</b> .....	<b>58</b>
6.1	COLETA DE TERMOS EM VOCABULÁRIOS EXISTENTES.....	58
6.2	COLETA DE TERMOS NA MINERAÇÃO DE TEXTO .....	60
6.3	TERMOS PROVENIENTES DA ENTREVISTA .....	66
6.4	TERMOS PROVENIENTES DOS FORMULÁRIOS ELETRÔNICOS .....	69
6.5	VOCABULÁRIO ESPECÍFICO SOBRE ENTORPECENTES NO CONTEXTO DA POLÍCIA FEDERAL.....	72
<b>7</b>	<b>CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS</b> .....	<b>76</b>
	<b>REFERÊNCIAS</b> .....	<b>78</b>

<b>APÊNDICE A – LISTA DOS TERMOS CONSOLIDADOS NO PROJETO DE PESQUISA.....</b>	<b>82</b>
<b>APÊNDICE B - TERMOS CONSOLIDADOS APÓS MINERAÇÃO DE TEXTO.....</b>	<b>89</b>
<b>APÊNDICE C – CÓDIGO FONTE PART1.PY .....</b>	<b>92</b>
<b>APÊNDICE D– CÓDIGO FONTE PART2.PY .....</b>	<b>93</b>
<b>APÊNDICE E – CÓDIGO FONTE CLASSEGRAMATICAL.PY.....</b>	<b>95</b>
<b>APÊNDICE F – IMAGENS DO FORMULÁRIO ELETRÔNICO .....</b>	<b>96</b>
<b>ANEXO A – QUADRO DA BASE JURÍDICA PF.....</b>	<b>108</b>

## 1 INTRODUÇÃO

Após a Segunda Guerra Mundial, a informação foi um insumo estudado e relevante em diversas disciplinas e ciências já estabelecidas. Alguns momentos marcantes sobre a importância da informação foram apontados no artigo de Vannevar Bush, durante a Segunda Guerra Mundial, quando constatou a relevância da informação em um evento crítico e sensível como uma guerra (Bush, 1945).

No artigo "As We May Think" (1945), Vannevar Bush registrou a importância da catalogação de informações e o crescente volume de conhecimento gerado pela humanidade. Bush (1945) previu a explosão informacional e destacou a necessidade de criar repositórios específicos para tratar e processar as informações catalogadas. Bush ressaltou que, em vez de confiar apenas na memória das pessoas, era crucial desenvolver sistemas eficientes para gerenciar informações.

Consider a future device for individual use, which is a sort of mechanized private file and library. It needs a name, and, to coin one at random, 'memex' will do. A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory (Bush, 1945).

Nesta citação, Bush apresenta o conceito de *memex* como um dispositivo do futuro que poderia ajudar a gerenciar a crescente quantidade de informações, atuando como um repositório eficiente e mecanizado para armazenar e consultar informações rapidamente.

No período pós Segunda Guerra Mundial, a informação e os processos relacionados à produção e recuperação da informação ganharam destaque entre os estudiosos da época, principalmente devido à colaboração entre os Países Aliados, como França, Reino Unido, Estados Unidos e União Soviética. Nesse contexto, surgiram estudos para aprimorar o entendimento sobre a informação e as fases correlatas, evoluindo conceitos já existentes e impulsionando o desenvolvimento de novas ciências e tecnologias com o objetivo de armazenar e disponibilizar informações de forma eficiente (Araújo, 2006).

Nesse sentido, a evolução dos conceitos de documento e informação no contexto da Ciência da Informação tem sido constante ao longo dos anos,

especialmente com o crescimento da internet e a informatização dos processos. Esses conceitos são frequentemente estudados e analisados na área, refletindo o desenvolvimento contínuo e a adaptação às mudanças tecnológicas e às novas demandas da sociedade na busca e no gerenciamento de informações.

Com o avanço tecnológico e a explosão informacional, o volume de documentos e informações tem aumentado significativamente em todos os setores, incluindo a administração pública e os setores que atuam na segurança pública. Essa expansão demanda um tratamento adequado desses documentos, a fim de garantir uma gestão eficiente da informação e permitir o acesso e a recuperação ágil e precisa dos dados necessários para a tomada de decisões e a execução das atividades do setor público.

Dentro desse contexto, os órgãos federais do Brasil, como a Polícia Federal, têm buscado modernizar e automatizar seus processos por meio do uso de recursos tecnológicos. Essa transformação visa suprir a redução de recursos humanos e atender às crescentes demandas internas, aprimorando a prestação de serviços à sociedade. Essa necessidade de eficiência ocorre mesmo diante das restrições orçamentárias e das limitações de vagas em concursos públicos, já que a equipe econômica busca constantemente melhorias nos serviços prestados pelos órgãos públicos.

Nesta pesquisa, abordaremos os desafios enfrentados pela Polícia Federal na gestão da informação, destacando a importância dos conceitos da Ciência da Informação e a aplicação de Sistemas de Organização do Conhecimento (SOC) para a recuperação da informação. Além disso, apresentaremos a metodologia proposta para a construção de um vocabulário controlado, considerando o contexto específico da Polícia Federal. Ao longo do texto, discutiremos os diferentes tipos de SOC e sua relevância no contexto específico da Polícia Federal do Brasil e no domínio escolhido: "crimes de entorpecentes".

Para tanto, essa pesquisa apresenta a seguinte questão norteadora: "Como desenvolver uma proposta para a construção de um controle terminológico especializado em crimes de tráfico de entorpecentes, considerando as necessidades e expectativas da Polícia Federal do Brasil?"

## 2 OBJETIVOS

Os objetivos dessa pesquisa estão divididos em geral e específicos.

### 2.1 OBJETIVO GERAL

Desenvolver um controle terminológico através de Vocabulário Controlado de termos relacionados aos crimes de tráfico de entorpecentes no âmbito da Polícia Federal brasileira.

### 2.2 OBJETIVOS ESPECÍFICOS

- a) Levantar a diversidade terminológica relacionada à crimes de entorpecentes nas bases de dados da Polícia Federal;
- b) Mapear Vocabulários Controlados relacionados ao domínio, disponíveis em fontes externas;
- c) Consolidar uma amostra de termos em um Vocabulário Controlado como projeto piloto;
- d) Propor uma metodologia para elaboração do Vocabulário Controlado contexto da Polícia Federal.

### 3 JUSTIFICATIVA

Esta pesquisa foi concebida levando em consideração as experiências profissionais do autor na Polícia Federal e a falta de eficiência em alguns métodos de recuperação de informação na instituição.

Com base na experiência e observações do autor desta pesquisa, é perceptível que, em relação ao aumento das demandas internas, a atualização da infraestrutura computacional e a busca constante por avanços tecnológicos são essenciais para lidar com o volume crescente de dados apreendidos na Polícia Federal. Outro desafio no ambiente tecnológico é a diversidade de dados presentes, incluindo dados estruturados originários de bancos de dados internos e externos da instituição, bem como dados não estruturados obtidos a partir de dispositivos apreendidos. Essa percepção destaca a importância de abordar esses desafios na pesquisa e no desenvolvimento de soluções eficientes para a gestão da informação na Polícia Federal.

De forma oposta à redução de investigadores na Polícia Federal, no âmbito das operações de polícia judiciária ou das operações de inteligência do órgão, registra-se acréscimo exponencial no volume de dados apreendidos pela instituição provenientes de dispositivos de armazenamento computacional ou de equipamentos eletrônicos, por exemplo: discos rígidos, *pendrives*, cartões de memória, dispositivos móveis (aparelhos celulares e tablets), armazenamento em nuvem, dentre outros. Esse crescimento de dados, desafia a organização nas etapas de processamento e de análise desse material pelos investigadores.

Segundo a 26ª pesquisa anual do uso de Tecnologia da Informação de 2015, realizada pelo Centro de Tecnologia de Informação Aplicada da FGV-EAESP, existem no Brasil cerca de 306 milhões de dispositivos conectados à internet, nos quais 154 milhões são telefones smartphones e, a maioria dos usuários tem migrado a sua utilização de computadores notebooks para smartphones e tablets (Velho, 2016).

Considera-se que a constante escassez de recursos humanos, a necessidade de maior eficiência dos serviços prestados pela Polícia Federal, o aumento das demandas dos órgãos e as dificuldades de tratamento e de processamento das informações, se torna primordial aos órgãos públicos investirem na melhoria do mapeamento de processos, potencializar os mecanismos de busca, criar um histórico

organizacional com as informações já processadas e definir metodologia própria na produção de conhecimentos internos, utilizando conceitos da Ciência da Informação.

Baseados em Velho (2016) e nas experiências do autor pode-se observar que no âmbito da Polícia Federal, tanto nas atividades decorrentes de polícia judiciária ou de atividades de inteligência do órgão, são utilizados os dados oriundos de mídias de armazenamento ou dispositivos apreendidos. E alguns dados, ainda, são oriundos de sistemas de informação disponíveis aos investigadores como sistemas cadastrais, sistemas de investigação ou inteligência, sistemas de outros órgãos. Em outros termos, o volume de dados disponível a um policial é significativo e complexo, principalmente pela diversidade das fontes de informações.

Em relação aos dados oriundos dos sistemas de informação, a Polícia Federal possui uma importante ferramenta web de indexação, que utiliza a tecnologia *Microsoft Fast Enterprise Search*<sup>1</sup>, que possibilita aos investigadores uma pesquisa textual única em todas as bases consolidadas na ferramenta, nos moldes dos grandes buscadores existentes na Internet, como o Google, Yahoo e Bing. Porém, um usuário desta ferramenta, para ter uma busca mais abrangente ou complexa sobre um determinado assunto, precisa fazer diversas pesquisas utilizando vários termos relacionados ao assunto desejado, pois na ferramenta não existe nenhum controle terminológico implementado.

Velho (2016) aborda o processo de armazenamento de dados e o fluxo de solicitações de perícias na Polícia Federal, relacionado a bens apreendidos. No contexto do procedimento investigativo, a autoridade policial responsável solicita ao setor forense do órgão a extração e a indexação dos dados computacionais obtidos durante as apreensões. Posteriormente, esses dados extraídos são analisados pelos investigadores envolvidos no caso. É fundamental destacar que todo o processo forense é conduzido de acordo com uma metodologia e técnica bem estabelecidas, visando preservar a cadeia de custódia das provas. Para isso, a Polícia Federal utiliza uma ferramenta de extração e indexação desenvolvida especificamente para o órgão, garantindo a integridade e a confiabilidade dos dados analisados.

---

<sup>1</sup> A *Microsoft Fast Enterprise Search* é uma solução de busca corporativa avançada e escalável desenvolvida pela Microsoft. A tecnologia oferece recursos de indexação e pesquisa eficientes, permitindo que as organizações localizem informações relevantes em grandes volumes de dados de várias fontes e formatos. A plataforma utiliza algoritmos sofisticados, para classificar e recuperar documentos com base em sua relevância para consultas específicas. Além disso, a *Fast Enterprise Search* oferece recursos de personalização, análise e integração com outras soluções da *Microsoft*, como o *SharePoint*.

Por possuir demandas muito particulares, que não estão disponíveis nas ferramentas existentes no mercado, a Polícia Federal utiliza um aplicativo próprio chamado de Indexador e Processador de Evidências Digitais (IPED). O programa trabalha em conjunto com diversos outros aplicativos de código livre, como o próprio TSK (The Sleuth Kit). O IPED foi desenvolvido por uma equipe de Peritos Criminais Federais, também utilizado por outros órgãos, como a polícia técnico-científicas estaduais, policiais civis, Ministério Público e Receita Federal do Brasil (Velho, 2016).

A utilização de técnicas de controle terminológico é de suma importância para a indexação e recuperação de informações em grandes volumes de dados. Nesse sentido, os Sistemas de Organização do Conhecimento (SOC), como Vocabulários Controlados, Taxonomias, Tesouros e Ontologias, são essenciais para a padronização dos termos utilizados e redução da ambiguidade. Na Polícia Federal, a aplicação dessas técnicas pode ser útil para aumentar a precisão das buscas em casos de tráfico de entorpecentes, por exemplo, permitindo que termos sinônimos sejam considerados na recuperação da informação. Dessa forma, a evolução das ferramentas de busca textual internas da Polícia Federal, com a utilização de módulos e conceitos de processamento de linguagem natural, aliada aos SOC, pode trazer maior eficiência na gestão da informação e contribuir para a resolução de casos mais complexos.

O trabalho da Polícia Federal é voltado para diversas temáticas, dentre as quais se destaca o combate aos crimes de tráfico de entorpecentes, que assumem grande importância no contexto atual. Além disso, a instituição dedica esforços à investigação de crimes financeiros, corrupção, crimes cibernéticos, contrabando, tráfico de armas e de pessoas, terrorismo entre outros.

A Polícia Federal, para compreender e prevenir as atividades criminosas, realiza estudos constantes sobre as organizações criminosas (Orcrims), incluindo o escopo de atuação de cada uma, análise das evoluções das atividades, estudo da estrutura organizacional e dos atores envolvidos. Esses estudos são realizados através de fontes abertas, sistemas disponíveis e dados obtidos através de apreensões policiais.

Diante do exposto, existem diversas informações sobre as organizações criminosas, sejam dados oriundos de materiais apreendidos ou de sistemas de informação, entretanto, os resultados obtidos com as ferramentas de análise, principalmente nas ferramentas de busca textual e de análise de vínculos, não dependem somente da utilização da tecnologia implementada e sim dos

conhecimentos prévios e das experiências dos investigadores que utilizam o sistema. Ou seja, dependendo do usuário pode-se ter resultados mais superficiais ou resultados mais complexos sobre um mesmo assunto.

Para exemplificar a problemática, na utilização de ferramenta textual atual da Polícia Federal, tem-se resultados relevantes e desiguais entre dois usuários com perfis, experiências profissionais e históricos sociais distintos que formulem uma busca na base, pois na ferramenta não existem controles linguísticos implementados, como mencionado anteriormente. Especialistas recomendam que sejam utilizados controles linguísticos nas ferramentas de busca para potencializar os resultados relevantes, como: Vocabulário Controlado, Taxonomia (termos de sinônimos hierarquizados), Tesouros (vocabulários de sinônimos hierarquizados com termos associados sobre um determinado domínio de conhecimento) e Ontologias (Rautenberg, 2008).

Conseqüentemente, fundamentados nas abordagens teóricas da Ciência da Informação estudadas sugere-se que a falta de utilização de controles linguísticos e de não estruturação linguística dos termos relacionados, torna-se mais dispendiosa a definição de metodologia na produção de conhecimento, além da automatização de processos e a criação de legado e de histórico organizacional. A elaboração de um Sistema de Organização de Conhecimento que envolva o registro do conhecimento empírico dos investigadores e a aplicação de controle terminológico, pode ser essencial para aumentar a efetividade das buscas textuais e análises de conexões mais complexas, bem como para a implementação de uma inteligência artificial adequada às ferramentas utilizadas.

Esta pesquisa tem como objetivo desenvolver uma proposta para a construção de um vocabulário controlado especializado em crimes de tráfico de entorpecentes, levando em consideração as necessidades e expectativas da Polícia Federal do Brasil. O estudo irá abordar questões relacionadas às características que devem compor o vocabulário controlado, visando melhorar a qualidade dos serviços prestados pela instituição na recuperação de informações.

## 4 CIÊNCIA DA INFORMAÇÃO

Para uma compreensão mais abrangente da pesquisa, é importante entender a evolução do conceito de informação ao longo do século passado até o surgimento da ciência da informação e suas subáreas. A necessidade de tomada de decisões precisas, rápidas e assertivas pelos profissionais foi impulsionada pela revolução industrial e pela pós Segunda Guerra Mundial, conforme sugerido por Capurro e Hjørland (2007) e Bush (1945). Após o período de guerra, os profissionais responsáveis pelo tratamento e produção de informações foram cada vez mais demandados para aperfeiçoar os mecanismos de recuperação das informações e desenvolvimento dos SOC.

De acordo com Capurro e Hjørland (2007), a informação é um insumo essencial para a produção eficiente de conhecimento em uma sociedade competitiva, sendo primordial a utilização de tecnologias especializadas desde o processo de armazenamento e recuperação de informações. A importância do entendimento do termo 'informação' é fundamental para compreender a origem e desenvolvimento da Ciência da Informação. De acordo com Capurro e Hjørland (2007), a palavra informação é usada para descrever uma perspectiva específica sobre o conhecimento e a comunicação. Esta perspectiva inclui características como novidade e relevância, que são fundamentais para o processo de transformação da informação em conhecimento. Portanto, a compreensão da definição de informação proposta pelos autores é fundamental para a pesquisa em Ciência da Informação. "A transformação da informação em conhecimento ocorre quando novas informações são relacionadas ao conhecimento existente, e este conhecimento é aplicado em novas situações" (Capurro; Hjørland, 2007).

Com base nas ideias apresentadas pelos autores Capurro e Hjørland (2007), é possível compreender que o termo "informação" é fundamental para a Ciência da Informação e precisa ser caracterizado. A transformação da informação em conhecimento é um processo que envolve a relação entre novas informações e conhecimentos pré-existentes, que são aplicados em novas situações (Capurro; Hjørland, 2007).

Nesse contexto, como dito, é importante destacar que a Ciência da Informação se fortaleceu como área de estudo após a Segunda Guerra Mundial, período marcado pela necessidade de tomada de decisões rápidas e precisas por parte de profissionais

responsáveis pelo tratamento das informações e produção de conhecimento (Araújo et al., 2018). Esse fortalecimento levou a uma consolidação da Ciência da Informação como área de estudo multidisciplinar, que se utiliza de conhecimentos da Biblioteconomia, Documentação, Teoria Matemática da Informação e Cibernética, entre outras áreas (Ortega, 2004).

De acordo com Ortega (2004), a Ciência da Informação é uma área multidisciplinar que tem como objetivo compreender o processo de produção, armazenamento, recuperação e disseminação da informação. A recuperação de informação, por sua vez, é um processo que envolve a busca por informações relevantes e precisas em meio a um grande volume de dados (Araújo et al., 2018), e é um dos focos desse estudo.

Neste contexto de pesquisa, é importante diferenciar e esclarecer o uso dos termos-chave. O termo “documento” é compreendido como um suporte físico ou eletrônico que pode conter “dados”, que, por sua vez, são conjuntos brutos de informações sem contexto. A “informação”, por outro lado, é a interpretação e organização desses dados, fornecendo significado e contexto. À medida que a informação é processada e interpretada pelo usuário dentro de um domínio específico, ela pode se transformar em “conhecimento”.

#### 4.1 RECUPERAÇÃO DA INFORMAÇÃO

A recuperação da informação para a Ciência da Informação é o processo pelo qual o usuário da informação consegue buscar documentos relacionados aos termos de pesquisa em uma base de dados a partir da representação documental realizada. Para Saracevic (1995), a recuperação da informação pode ser considerada a vertente tecnológica da Ciência da Informação e é resultado da relação desta com a Ciência da Computação.

Para Saracevic (1995), a recuperação da informação é uma das principais áreas de estudo da Ciência da Informação, tendo em vista a necessidade de lidar com a crescente quantidade de informações disponíveis. Segundo o autor, o objetivo da recuperação da informação é tornar as informações relevantes e úteis para o usuário, permitindo que este encontre a informação que necessita com facilidade. Essa ideia está relacionada ao conceito do “*Memex*” proposto por Bush (1945), que consiste em um sistema de armazenamento e recuperação de informações pessoais que permitiria

aos indivíduos aumentarem sua capacidade de conhecimento e decisão. Para Saracevic (1995) *“Information retrieval is concerned with the representation, organization, and manipulation of information content for the purpose of effective and efficient retrieval and dissemination”*.

Essa definição de recuperação da informação, proposta por Saracevic (1995), destaca a importância da representação e organização da informação para que ela possa ser recuperada de forma eficiente e eficaz.

Ao longo da evolução dos conceitos da Ciência da Informação, tanto as mudanças de estrutura de informação, como as mudanças no fluxo de informação foram sempre notadas no status tecnológico das atividades de armazenamento e transferência da informação, na relação das pesquisas em Ciência da Informação com os usuários e com outros elementos envolvidos (Barreto, 1999).

A utilização de recursos tecnológicos eficientes para organização, armazenamento, indexação e recuperação de informação se tornam essenciais para resultados de pesquisas mais assertivas. Para Oliveira e Araújo (2012) o sucesso na recuperação da informação seria utilizar controle terminológicos, principalmente no processo de indexar e recuperar documentos que apresentem relevância para o usuário.

O controle terminológico é essencial na indexação de documentos, assegurando consistência e precisão na representação da informação. Esse controle é particularmente relevante em áreas onde a precisão terminológica é crítica. Além disso, o controle terminológico facilita a interoperabilidade entre diferentes sistemas de informação, promovendo uma compreensão eficiente das informações indexadas por usuários de diferentes áreas (Oliveira; Araújo, 2012).

A garantia da satisfação dos usuários é um aspecto fundamental no processo de recuperação de informação, que pode ser alcançado através da construção de linguagens documentárias eficientes e do uso de técnicas de indexação e controle terminológico. O sucesso na recuperação de informação está diretamente relacionado à utilização de recursos linguísticos, como os Vocabulários Controlados, que garantem a precisão e a padronização dos termos utilizados na indexação e recuperação de documentos (Oliveira; Araújo, 2012).

O sucesso na recuperação da informação está diretamente relacionado à utilização de recursos linguísticos, como os Vocabulários Controlados, que

garantem a precisão e a padronização dos termos utilizados na indexação e recuperação de documentos (Oliveira; Araújo, 2012).

Na citação acima, é possível notar que para o sucesso na recuperação de informações, é necessário o uso de recursos linguísticos que padronizem os termos utilizados na indexação e recuperação de documentos. Nesse sentido, os Sistemas de Organização do Conhecimento (SOC) são fundamentais, pois se caracterizam por serem conjuntos de termos padronizados que representam conceitos dentro de uma determinada área de conhecimento.

A utilização de recursos linguísticos como vocabulários controlados e sistemas de organização do conhecimento potencializa os resultados de pesquisa ao garantir uma indexação mais precisa e eficiente. Isso se traduz em uma recuperação de informação mais acurada, onde os usuários conseguem encontrar documentos que são verdadeiramente relevantes para suas necessidades de pesquisa. A precisão na indexação e recuperação de informações não apenas economiza tempo e recursos, mas também enriquece a experiência de pesquisa, proporcionando um acesso mais direto ao conhecimento necessário. Além disso, a padronização terminológica ajuda na construção de bases de dados mais robustas e integradas, facilitando pesquisas multidisciplinares e a descoberta de novas conexões entre diferentes áreas do conhecimento.

## 4.2 SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO

A informação é a base para produzir conhecimento, que quando chega ao cérebro impacta os neurônios (Currás, 2010). Por isso, é importante que os sistemas automatizados de informações gerem resultados mais relevantes, precisos e complexos para que o usuário consiga produzir de fato o conhecimento.

O conceito de conhecimento muda quando começam a ser usados os computadores e a informática aplicada aos processos informativo-documentários. Fala-se, inclusive, de “base de conhecimento”, quando, na realidade, trata-se somente de uma base ou armazém de dados acumulados na memória do computador (Currás, 2010).

A importância de organizar o conhecimento, classificando as informações e utilizando os conceitos de Processamento de Linguagem Natural (PLN) na recuperação da informação mais eficiente, são utilizadas por vários especialistas ao

longo do tempo, em diversas áreas, como por exemplo: biologia - identificação de relações entre doenças e suas causas, sintomas e tratamentos e filosofia - categorização e classificação de conceitos filosóficos a partir de diferentes fontes textuais (Almeida, 2017).

A recuperação da informação é hoje um dos grandes desafios da Ciência da Informação. A fragmentação do conhecimento trouxe para a indexação o desafio constante de se adaptar às novas formas de gerenciar a massa gigantesca de dados acumulados nas bases de dados. O uso de palavras-chave é fundamental para esta recuperação, e a linguística entra como processo na recuperação dessa informação, resolvendo problemas de representação de termos no sistema utilizado pelos usuários (Mendonça, 2000). É nesse cenário que os SOC, como Vocabulários Controlados, Taxonomias, Tesouros e Ontologias, se tornam valiosos instrumentos para melhorar a eficiência e precisão na recuperação de informações.

SOC são sistemas conceituais semanticamente estruturados que contemplam termos, definições, relacionamentos e propriedades dos conceitos. Na organização e recuperação da informação, os SOC cumprem o objetivo de padronização terminológica para facilitar e orientar a indexação e os usuários. Quanto à estrutura variam de um esquema simples até o multidimensional, enquanto suas funções incluem a eliminação da ambiguidade, controle de sinônimos ou equivalentes e estabelecimento de relacionamentos semânticos entre conceitos (Carlan; Medeiros, 2011).

Os Sistemas de Organização do Conhecimento (SOC) são ferramentas fundamentais para aprimorar a indexação, especialmente em sistemas robustos de recuperação de informação (RI), onde melhoram a compreensão e a organização dos dados. Entre os principais tipos de SOC utilizados em ambientes digitais, destacam-se as Ontologias, os Tesouros e as Taxonomias (Rautenberg, 2008). Essas ferramentas auxiliam na estruturação e padronização das informações, facilitando o acesso e a recuperação de dados em sistemas de RI. A seguir serão descritos os principais tipos de SOCs.

#### **4.2.1 Ontologias**

As Ontologias já estão sendo utilizadas na Recuperação de Informação há bastante tempo, entretanto normalmente essas Ontologias eram criadas previamente

e manualmente, e utilizadas para expandirem as consultas dos usuários (Gottschalg-Duque, 2005).

Para Boccato (2011), as Ontologias são modelos conceituais específicos, para representar as relações complexas entre objetos, incluindo as regras e axiomas não presentes nas redes semânticas.

Dessa forma, a ontologia é um sistema crucial para representar, formalizar e compartilhar conhecimento de forma eficiente. No campo da Engenharia de Ontologias, existem diversas metodologias de desenvolvimento, cada uma com suas particularidades em relação ao ciclo de vida das ontologias.

Na literatura acadêmica, destaca-se a relevância de combinar diferentes metodologias. Ao analisar várias abordagens, levando em conta fatores como simplicidade, expressividade e funcionalidade, é viável identificar elementos cruciais para o desenvolvimento de ontologias. Desse modo, uma nova metodologia é estabelecida, abrangendo atividades, tarefas e indicações de ferramentas para o processo de construção de ontologias. Essa metodologia é validada por meio de estudos exploratórios e/ou aplicados (Rautenberg, 2008).

A criação de ontologias pode ser complexa, conforme abordado por Rautenberg (2008) e Gottschalg-Duque (2005). Esses autores destacam que o processo envolve conceitos de controle linguístico, como o uso de vocabulários controlados, a fim de garantir maior precisão e eficiência na representação do conhecimento. Segundo Rautenberg (2008), a elaboração de ontologias exige atenção aos detalhes linguísticos e conceituais. Já Gottschalg-Duque (2005) ressalta a importância do controle linguístico para o desenvolvimento bem-sucedido de ontologias.

O uso de controles linguísticos, conforme destacado por Gottschalg-Duque (2005) e Rautenberg (2008), é fundamental para assegurar a precisão e eficiência na representação do conhecimento. Além disso, as ontologias são capazes de expressar regras e axiomas não presentes nos tesouros, o que confere maior riqueza e detalhamento às relações entre os objetos, conforme mencionado por Boccato (2011). Portanto, apesar das dificuldades inerentes à construção de ontologias, é inegável que, quando comparadas aos tesouros, elas proporcionam uma representação do conhecimento mais abrangente e robusta, sendo fundamentadas no controle linguístico e na riqueza de detalhes conceituais.

#### **4.2.2 Tesouros**

De acordo com Currás (2010), uma das primeiras definições do termo tesouro (Currás, 2010, p. 96), diz que são: "... uma lista autorizada, que pode conduzir o usuário de um conceito a outro, por meio de relações heurísticas ou intuitivas. A lista pode-se usar manual ou mecanicamente para atribuir cabeçalhos de indexação."

Os Tesouros, além de sistemas de organização do conhecimento, podem ser classificados em mais duas vertentes, como linguagem de indexação e como linguagem documentária. Como linguagem de indexação, os tesouros são linguagens pós coordenadas, sendo uma lista de termos registrados de forma estruturada com termos relacionados, sendo eles gerais e específicos. O Tesouro é uma linguagem cuja característica se dá pela especificidade que existe entre os termos e suas relações. É utilizado para representar e recuperar informações de forma estrutural (Jesus, 2018).

No âmbito terminológico da representação do conhecimento, os Tesouros e as Ontologias são modelos abordados pela Ciência da Informação. O Tesouro é uma linguagem documentária caracterizada pela complexidade existente no relacionamento entre os termos que comunicam o conhecimento especializado. A Ontologia é um modelo de representação do conhecimento, por vezes empregado como Sistema de Organização do Conhecimento, que, a exemplo do Tesouro, é utilizada para representar e recuperar informação por meio de estruturas conceituais. As Ontologias possibilitam compartilhar uma visão de determinado campo do conhecimento, compartilhar uma forma de pensar de determinado assunto, proporcionando um mapa semântico e uma estrutura conceitual de um domínio específico por meio de um vocabulário comum (Sales, 2009).

Para haver interoperabilidade entre as linguagens documentárias é primordial a utilização de padrões internacionais na documentação, as normas ISO 25964-1 (parte 1) e ISO 25964-2 (parte 2) se destacam por definirem a criação e manutenção de Tesouros e de Vocabulários Controlados, bem como a interoperabilidade entre eles. (ISO, 2011; ISO,2013)

Considerando a necessidade de regras para o desenvolvimento de sistemas de informação utilizando os Tesouros, a norma ISO 25964-1 define: "the transition is necessary for compatibility with any electronic information management platform" (ISO, 2011, p. 3). Nessa parte a norma aborda tanto o desenvolvimento como a manutenção de Tesouros. A norma ISO 25964-1 (2011) registra que o Tesouro tem

como objetivo direcionar o indexador e o pesquisador para a escolha do mesmo termo e respectivamente o mesmo conceito e para que isso ocorra, o Tesouro deve listar todos os conceitos que podem ser utilizados para a recuperação da informação de um determinado domínio (Jesus, 2018), assim como, suas variações.

### **4.2.3 Taxonomia**

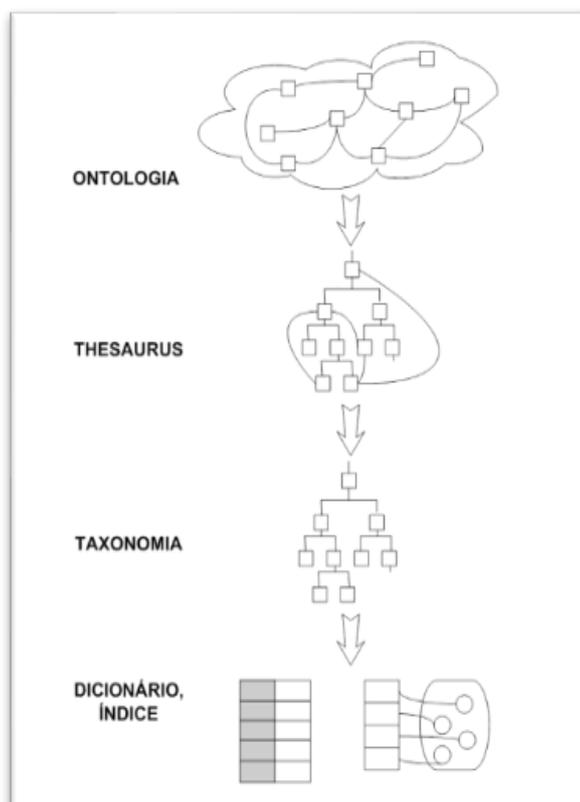
A taxonomia pode ser considerada um instrumento de classificação de informações que pode ser adotado em sistema de recuperação de informação permitindo classificar, alocar, recuperar e comunicar informações de forma lógica internamente em um sistema (Sousa; Araújo Júnior, 2013).

A taxonomia, vinculada à classificação de documentos, tem como finalidade estruturar e sistematizar o conjunto de assuntos correspondentes às atividades desempenhadas pelas áreas de trabalho de uma dada organização, a fim de garantir aos tomadores de decisão e usuários em geral, o acesso mais rápido aos documentos e às informações necessárias ao processo decisório da instituição (Sousa; Araújo Júnior, 2013).

Na obra de Santos (2018), aborda a complexidade e a importância da taxonomia em sistemas de organização da informação, sobretudo por meio de vocabulários controlados. A taxonomia é descrita como uma estrutura lógica e hierárquica de conceitos que facilita a organização, ordenação, recuperação e comunicação da informação. Estes sistemas de organização, que se baseiam em listas de termos estruturados, buscam eliminar ambiguidades, controlar sinônimos e estabelecer relações lógico-semânticas entre os termos para garantir eficiência na recuperação de informações. Nos dias de hoje, a taxonomia além de estruturar e classificar informações, desempenha um papel crucial na era digital ao estruturar websites e plataformas de conteúdo com estabelecimento de relações semânticas entre os produtos. Assim, ao categorizar e estabelecer relações semânticas entre diferentes partes de um conteúdo online, a taxonomia facilita a usabilidade e a experiência do usuário em ambientes digitais.

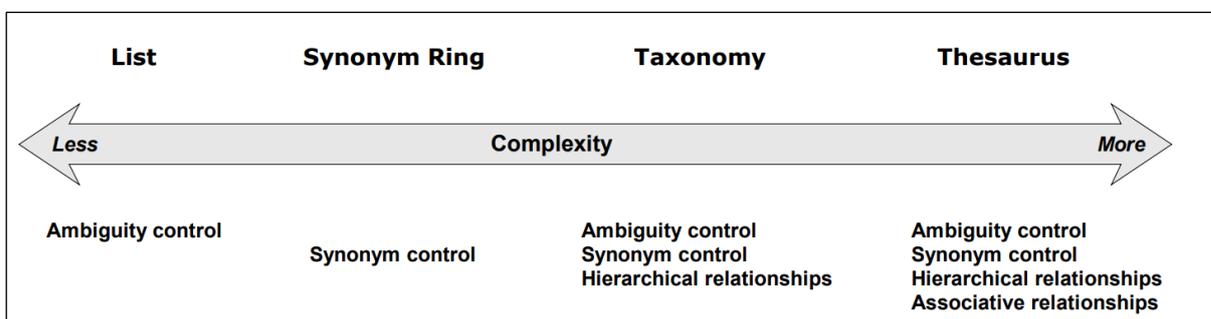
Em síntese, nas figuras 1 e 2 são representados os sistemas de organização do conhecimento por nível de complexidade de cada um, sendo a taxonomia um dos sistemas mais simples e a ontologia um dos mais complexo, tanto para ser utilizado, bem como, ser desenvolvido.

Figura 1 - Relacionamento entre ontologia, thesaurus, taxonomia, índice e dicionário



Fonte: Librelotto et al., 2005

Figura 2 - Gráfico de complexidade das Linguagens Documentárias – ANSI/NISO Z39.19-2005



Fonte: ANSI/NISO Z39.19-2005

De acordo com o Gráfico de Complexidade das Linguagens Documentárias – ANSI/NISO Z39.19-2005 acima, que inicia com listas simples, que apresentam um conjunto de termos relacionados para controlar a ambiguidade. Depois, temos os

anéis de sinônimos, que conectam palavras com significados semelhantes para melhorar a recuperação de informações. A complexidade cresce com as taxonomias, que organizam os termos em hierarquias e categorias. Finalmente, os tesouros que são representados no gráfico acima possuem uma complexidade maior, pois envolvem não apenas hierarquias, mas também relações semânticas, como termos relacionados e termos preferenciais. Essa progressão na complexidade dos SOC auxilia na organização e recuperação de informações, proporcionando maior precisão e eficiência dependendo do contexto de desenvolvimento (ANSI/NISO, 2005; Rautenberg, 2008).

Tesouro, taxonomia, vocabulário controlado e sistemas de classificação são SOC que cumprem diferentes objetivos mas todos tem como base listas de termos controlados, com mais ou menos complexidade na estrutura hierárquica e nas relações conceituais (Carlan; Medeiros, 2011).

#### **4.2.4 Anéis de sinônimos**

Os anéis de sinônimos representam uma abordagem mais simples em sistemas de organização do conhecimento (SOC). Ao contrário de SOC mais complexos, como ontologias ou tesouros, os anéis de sinônimos concentram-se principalmente na associação de termos sem estabelecer relações conceituais profundas. Nesses sistemas, grupos de termos semelhantes ou sinônimos são agrupados, permitindo que os usuários encontrem informações relevantes, mesmo que diferentes palavras ou expressões sejam utilizadas para se referir ao mesmo conceito. (ANSI/NISO, 2005)

Os anéis de sinônimos desempenham um papel crucial na recuperação de informações ao proporcionar uma correspondência flexível entre termos e consultas de pesquisa. Para Esses sistemas simplificados facilitam a busca, uma vez que não exigem a definição rigorosa de hierarquias ou relações complexas entre conceitos. (Carlan; Medeiros, 2011). Em vez disso, eles oferecem uma associação direta entre termos semelhantes, ampliando assim as possibilidades de recuperação de informações relevantes, mesmo quando os usuários usam vocabulários diferentes para se referir ao mesmo tópico.

Embora os anéis de sinônimos possam ser considerados SOC mais simples, essa abordagem desempenha um papel importante em casos específicos em que a simplicidade é uma vantagem. São especialmente úteis quando a necessidade de

relações conceituais profundas é mínima e a prioridade é fornecer uma correspondência eficiente entre termos e consultas de pesquisa. Portanto, a escolha entre anéis de sinônimos e SOC mais complexos dependerá das necessidades específicas do domínio e das metas da recuperação de informações no contexto em questão.

#### **4.2.5 Vocabulário Controlado**

Conforme Moreira (2015), no contexto dos estudos de organização e representação do conhecimento (ORC), há diversos instrumentos de organização e recuperação da informação, como sistemas de classificação, tesouros, taxonomias e ontologias. Esses instrumentos, considerados como sistemas de organização do conhecimento, têm em comum o uso de controle linguístico. No entanto, o SOC denominado "vocabulário controlado" difere dos outros, como ontologias, tesouros e taxonomias, por não abordar relações conceituais complexas e abrangentes.

Para efetividade da recuperação da informação, sem ruídos de comunicação entre os usuários e ativos informacionais, é importante utilizarmos técnicas ou ferramentas que contribuam para padronizar a terminologia na representação da informação. A utilização de instrumentos que garantam a uniformidade das informações, como os tesouros e os vocabulários controlados, por exemplo, proporciona maior qualidade dos termos para a indexação, e dessa forma são extremamente valiosos (Lancaster, 2004).

Vocabulários Controlados são essenciais no processo de organização, representação e recuperação da informação e importantes para mitigar os ruídos na comunicação entre os sistemas de informação e os usuários, pois sem os recursos de padronização terminológica na representação documentária o fluxo de comunicação pode se tornar ineficiente (Moreira, 2015).

Os vocabulários controlados normalmente são utilizados na indexação, no processo de padronização dos conceitos para a representação da informação, atuando como agente mediador entre a linguagem natural contida nos documentos e a linguagem do sistema de informação. Contudo, os vocabulários controlados também podem favorecer a busca de informação a partir de sua disponibilização ao usuário, considerando que isso contribuiria para sua autonomia no processo recuperação da informação, tendo em vista que as bibliotecas possam disponibilizar esse recurso no catálogo online onde se realizam buscas a partir da iniciativa e necessidade dos usuários (Cruz, 2022).

Na área jurídica existem iniciativas de sistemas de organização do conhecimento com a utilização do recurso de controles terminológicos, seja para a recuperação de informação pela indexação de termos ou utilização de SOC como vocabulário controlado, anéis de sinônimo, tesouro ou ontologia.

Há um projeto de um Tesouro Jurídico do Supremo Tribunal Federal (TSTF) com termos relacionados ao tráfico de entorpecentes, domínio desta pesquisa. O TSTF se enquadra como sistema de organização do conhecimento denominado Tesouro, por possuir relações semânticas e estrutura hierárquica.

Considerando a norma ISO 25964 e Davanzo (2016), os tesouros utilizam o controle terminológico por meio das seguintes siglas (ou códigos) para melhor compreensão das relações conceituais: USE / UP (equivalência); TG (termo genérico); TE (termo específico); TR (termo relacionado) e NA (notas).

A figura 3 mostra a pesquisa do termo “Droga” e o resultado no tesouro do Supremo Tribunal Federal – STF, utilizando algumas das siglas mencionadas, o que demonstra a necessidade de vocabulário controlado como base para o desenvolvimento das relações semânticas.

Figura 3 - Vocabulário Jurídico Supremo Tribunal Federal

DROGA

NOTA:

Art. 1º, parágrafo único da Lei 11.343/2006. Usar para os casos regidos por essa Lei. Para casos regidos pela Lei 6.368/1976 usar o descritor SUBSTÂNCIA ENTORPECENTE

TE

- ACETONA
- COCAÍNA
- CRACK
- LANÇA PERFUME
- MACONHA

TR

- CLORETO DE ETILA
- DEPENDÊNCIA FÍSICA
- DEPENDÊNCIA PSÍQUICA
- DEPENDÊNCIA QUÍMICA
- DEPENDÊNCIA TOXICOLÓGICA
- EMBALAGEM
- EXAME TOXICOLÓGICO

Fonte: <https://portal.stf.jus.br/jurisprudencia/tesouro/pesquisa.asp> (2023)

Geralmente, nas áreas que possuem termos especializados, exemplo área jurídica e médica, utilizam instrumento de controle de termos através de um

vocabulário controlado para facilitar o entendimento e as buscas de usuários que não são familiarizados com termos específicos e esse vocabulário pode ser base para utilização de sistemas de organização do conhecimento mais robustos como tesouros e ontologias.

Um vocabulário controlado é essencialmente uma lista de termos autorizados. Em geral, o indexador somente pode atribuir a um documento termos que constem da lista adotada pela instituição para a qual trabalha. Comumente, no entanto, o vocabulário controlado é mais do que uma mera lista, inclui, em geral, uma forma de estrutura semântica. Essa estrutura destina-se especialmente, a: 1. Controlar sinônimos, optando por única forma padronizada, com remissivas de todas as outras; 2. Diferenciar homógrafos. Por exemplo, PERU país e PERU ave; e 3. Reunir ou ligar termos cujos significados apresentem uma relação mais estreita em si. Dois tipos de relações são identificados explicitamente: as hierárquicas e as não-hierárquicas (ou associativas) (Lancaster, 2004).

Os vocabulários controlados são divididos em três tipos: esquema de classificação bibliográfica (como a Classificação Decimal de Dewey), listas de cabeçalhos de assunto e tesouros. Os três tipos apresentam os termos tanto na forma alfabética quanto sistemática, por exemplo, na classificação bibliográfica o arranjo alfabético é secundário, se estrutura em forma de um índice que remete para o arranjo principal considerando a hierarquia, no tesouro o arranjo explícito dos termos é alfabético com estrutura implícita hierárquica e com comportamentos remissivos (Lancaster, 2004).

Para Lancaster (2004), “o vocabulário controlado reduz a diversidade da terminologia. Além disso, ao ligar semanticamente termos que tenham relação entre si, ajuda o usuário a identificar todos os termos que seriam necessários para realizar uma busca completa”.

Entendemos que o vocabulário controlado não é apenas um conjunto de termos padronizados, mas sim uma ferramenta essencial que serve como alicerce para qualquer SOC. Ao longo dos anos, a quantidade e a diversidade das informações armazenadas cresceram exponencialmente. Esse crescimento acelerado destacou a necessidade de um controle terminológico mais rígido, especialmente num contexto como o da Polícia Federal, onde a precisão e a clareza da informação podem ser cruciais para a resolução de casos.

No contexto dos objetivos específicos deste mestrado, propomos o desenvolvimento de um controle terminológico robusto e abrangente, alinhado à constante evolução dos termos relacionados aos crimes de tráfico de entorpecentes.

Primeiramente, buscaremos levantar a diversidade terminológica presente nas bases de dados da Polícia Federal, compreendendo assim a complexidade do domínio em questão. Em seguida, iremos mapear Vocabulários Controlados relevantes provenientes de fontes externas. Com base nesse conhecimento, consolidaremos uma amostra de termos em um Vocabulário Controlado, estabelecendo um projeto piloto. Por fim, elaboraremos uma metodologia específica para a construção e manutenção contínua do Vocabulário Controlado, especialmente adaptada às necessidades da Polícia Federal.

### 4.3 POLÍCIA FEDERAL DO BRASIL

A Polícia Federal do Brasil (PF), antigo Departamento de Polícia Federal (DPF), órgão permanente de Estado, organizado e mantido pela União, foi fundado na hierarquia e disciplina. Com execução orçamentária e gestões administrativa e financeira descentralizadas, integrante da estrutura básica do Ministério da Segurança Pública, tem por finalidade exercer as competências previstas no § 1º do art. 144 da Constituição Federal, e demais dispositivos legais e regulamentares pertinentes (Brasil, 2023).

A PF também atua na segurança pública para a preservação da ordem pública e da incolumidade das pessoas, bem como dos bens e interesses da União, exercendo atividades de polícia marítima, aeroportuária e de fronteiras, repressão ao tráfico de entorpecentes, contrabando e descaminho. De acordo com o artigo 144, parágrafo 1º, da Constituição Federal, é instituída por lei como órgão permanente, organizado e mantido pela União e estruturado em carreira (Brasil, 2023).

Art. 144. A segurança pública, dever do Estado, direito e responsabilidade de todos, é exercida para a preservação da ordem pública e da incolumidade das pessoas e do patrimônio, através dos seguintes órgãos:

I - polícia federal;

II - polícia rodoviária federal;

III - polícia ferroviária federal;

IV - polícias civis;

V - polícias militares e corpos de bombeiros militares.

VI - polícias penais federal, estaduais e distrital.

§ 1º A polícia federal, instituída por lei como órgão permanente, organizado e mantido pela União e estruturado em carreira, destina-se a:

I - apurar infrações penais contra a ordem política e social ou em detrimento de bens, serviços e interesses da União ou de suas entidades autárquicas e empresas públicas, assim como outras infrações cuja prática tenha

repercussão interestadual ou internacional e exija repressão uniforme, segundo se dispuser em lei;  
II - prevenir e reprimir o tráfico ilícito de entorpecentes e drogas afins, o contrabando e o descaminho, sem prejuízo da ação fazendária e de outros órgãos públicos nas respectivas áreas de competência;  
III - exercer as funções de polícia marítima, aeroportuária e de fronteiras;  
IV - exercer, com exclusividade, as funções de polícia judiciária da União.  
(Brasil,1988)

Além das atribuições constitucionais da Polícia Federal, existem outras definidas em leis e decretos, como: Combate ao terrorismo; Garantir a segurança dos Chefes de Estado estrangeiros e Chefes de Organismos Internacionais em visita ao Brasil; Prevenir e reprimir os crimes cibernéticos; Combate à pedofilia; Ser a representante exclusiva da Interpol no Brasil, reprimindo o crime internacional e busca por foragidos internacionais; Prevenir e reprimir os crimes praticados contra os povos indígenas; Repressão ao desvio de recursos públicos; Controlar e gerenciar o SINARM (Sistema Nacional de Armas) e as armas de fogo em poder da população; Reprimir crimes ambientais e contra o patrimônio histórico; Exercer a regulação, controle, e a fiscalização de toda a atividade de segurança privada no Brasil; Implantar e manter o banco de dados de perfis genéticos de criminosos; Controle e fiscalização de produtos químicos de natureza controlada; Reprimir os crimes previdenciários; Investigar e reprimir violações de Direitos Humanos; Reprimir o sequestro, cárcere privado e extorsão mediante sequestro, se o agente foi impelido por motivação política ou quando praticado em razão da função pública exercida pela vítima; e Reprimir o furto, roubo ou receptação de cargas, inclusive bens e valores, transportadas em operação interestadual ou internacional, quando houver indícios da atuação de quadrilha ou bando em mais de um Estado da Federação; Investigar e reprimir os crimes políticos; Investigar e reprimir o crime de lavagem de dinheiro; Cooperar com outras instituições de segurança pública, fornecendo as informações de Inteligência policial que forem pertinentes; Prevenir, investigar e reprimir quaisquer outros crimes de forma subsidiária às Polícias Cíveis estaduais, quando solicitada; Fazer a segurança dos candidatos presidenciais no período das eleições (Brasil, 2023).

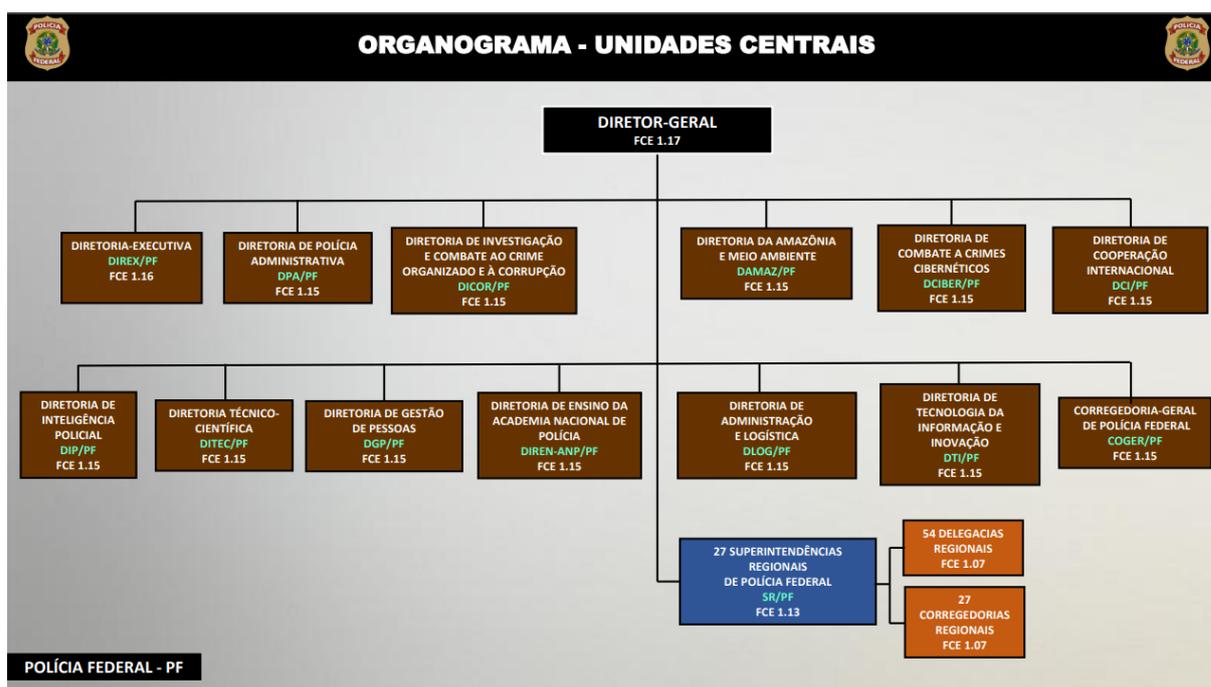
No site governamental da Polícia Federal<sup>2</sup> é possível consultar algumas normas e legislações que são a base jurídica da estrutura organizacional e das competências da Polícia Federal, listados no Anexo A - Base jurídica PF.

---

<sup>2</sup> Site da Polícia Federal - <https://www.gov.br/pf/>

Dentro da complexa estrutura organizacional da Polícia Federal, composta por treze diretorias especializadas, conforme ilustrado na Figura 4, destacam-se a Diretoria de Investigação e Combate ao Crime Organizado – DICOR e Diretoria de Inteligência Policial – DIP na construção de sistemas de informações especializados em consolidação e análise de dados provenientes a área de tráfico de entorpecente. A DICOR desempenha um papel fundamental, dirigindo, planejando, coordenando, controlando e avaliando a atividade de investigação em uma ampla gama de crimes, abrangendo desde Crimes Eleitorais, contra a Fazenda Pública, Contrabando e Sonegação Fiscal, até Delitos Previdenciários, Financeiros, Lavagem de Dinheiro, Corrupção, Tráfico de Drogas, Crimes Violentos, Tráfico de Armas, contra o Patrimônio da União, Roubo de Cargas, Roubo a Bancos e crimes praticados por Facções Criminosas (Brasil, 2023).

Figura 4 - Organograma da PF



Fonte: Brasil, 2023.

Além disso, a DIP tem a importante responsabilidade de integrar a coleta, armazenamento, análise e apresentação dos dados de estatística criminal e produtividade operacional, provenientes dos sistemas de informação da Polícia Federal de várias temáticas. Nesse contexto, a Coordenação-Geral de Inteligência (CGI) e a Divisão de Gestão de Informações e Sistemas de Inteligência (DSI) da

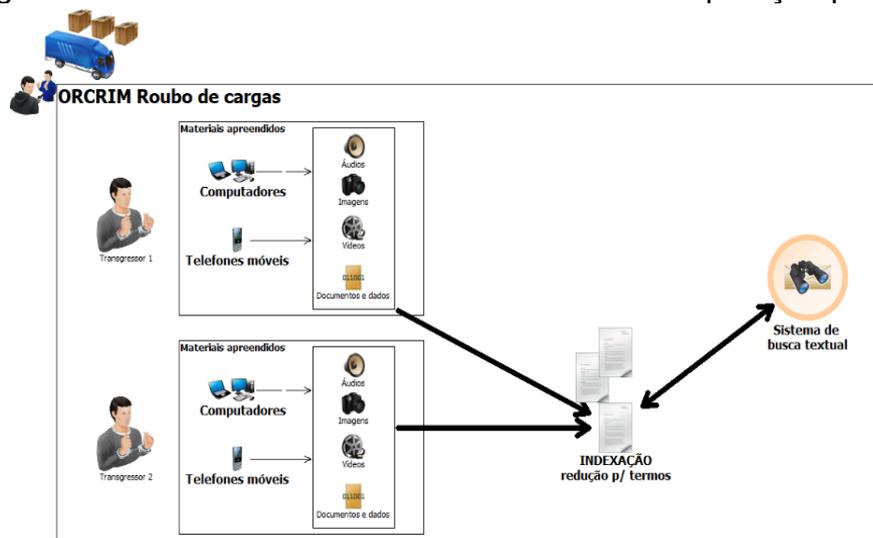
Diretoria de Inteligência Policial (DIP) desempenham um papel estratégico nessa análise multifacetada. Essas unidades buscam criar meios para que os policiais da inteligência e da DICOR possam acessar e trabalhar com dados especializados, com ênfase na área de tráfico de drogas. Essa cooperação é essencial para garantir que informações valiosas não se percam e que os objetivos da Polícia Federal sejam alcançados de maneira eficaz (Brasil, 2023).

#### **4.3.1 Características dos dados analisados na Polícia Federal**

Nesta pesquisa, o objetivo é explorar as características e funções dos Sistemas de Organização do Conhecimento (SOC) e desenvolver um vocabulário controlado especializado para atender às necessidades da Polícia Federal, especificamente da DICOR e da DIP. Para isso, será coletada uma amostra de termos do domínio escolhido e consolidados em uma lista, aplicáveis à casuística da instituição. Conforme destacado na literatura da Ciência da Informação (Rautenberg, 2008; Gottschalg-Duque, 2005; Mendonça, 2000), a implementação de SOC, como um vocabulário controlado especializado, pode potencializar os resultados de pesquisas em sistemas de busca textual orientados por indexação de termos, contribuindo para uma maior eficiência e precisão na recuperação de informações no contexto da Polícia Federal.

A diversidade de dados trabalhados na Polícia Federal torna a recuperação de informação mais complexa. Por exemplo, ao investigar crimes de uma Orcrim relacionada a crimes de patrimônio, especificamente roubo de cargas de caminhões, diversos materiais podem ser apreendidos em uma operação policial, como computadores, notebooks, aparelhos de celulares e outros dispositivos de armazenamento. Todos os equipamentos passarão por perícia para extração dos conteúdos armazenados, desde um simples anexo em um e-mail até um áudio gravado e transmitido no aplicativo de mensageria instantânea. Após essa extração, serão realizados processos de transcrição de áudios e vídeos e, por fim, todo o conteúdo será reduzido a termos para os investigadores utilizarem as ferramentas de busca textual. Nesse contexto, Gesteira (2015) destaca que "é nessa hora que a indexação de conteúdo, disponibilizando palavras-chave sobre os assuntos presentes em cada um desses áudios, possibilita uma busca por assuntos de interesse".

Figura 5 - Diversidade de dados analisados numa operação policial



Fonte: Autor

De acordo com Gesteira (2015), os dados das operações de inteligência policial costumam ser dinâmicos e peculiares, tornando essencial a coleta e sistematização eficientes das informações. O investigador responsável por analisar e definir o fluxo das informações relacionadas a uma organização criminosa (Orkrim) deve levar em consideração os aspectos sociais dessa organização.

Parte da análise envolve a organização dos dados e documentos obtidos, apreendidos ou produzidos durante uma operação, e a indexação se torna um componente relevante nesse processo. A indexação eficiente, aliada a sistemas de recuperação de informação (RI) adequados, está diretamente associada à eficácia da investigação, auxiliando na recuperação e acesso às informações sobre a Orkrim investigada (Mendonça, 2000). Dessa forma, proporciona um melhor entendimento e acompanhamento das atividades criminosas em questão, contribuindo para a efetividade das investigações policiais no Brasil.

No universo das atividades criminosas, há uma grande variabilidade de termos, incluindo gírias e termos informais, que são utilizados em diferentes contextos. Essa diversidade linguística, somada às diferenças regionais do Brasil, cria barreiras adicionais para as investigações federais, uma vez que cada região pode empregar termos e significados distintos para descrever situações semelhantes.

A ausência de Sistemas de Organização do Conhecimento (SOC) na indexação e na busca de informações pode tornar o processo de recuperação de informação bastante dispendioso e até ineficiente para os investigadores. Sem o auxílio de um SOC, o investigador precisaria conhecer todos os termos sinônimos e os termos

relacionados ao termo de busca desejado, o que dificultaria o acesso a informações relevantes e prejudicaria a eficácia das investigações. Como exemplo, o Quadro 1 apresenta algumas expressões e seus respectivos significados relacionados a atividades criminosas, evidenciando as dificuldades decorrentes da não padronização da linguagem.

Quadro 1 - Expressões utilizadas por ORCRIM (Organização criminosa)

Expressão	Possível significado
Lombardi	Caminhão que não pode circular em vias com fiscalização
Carro de bandido	Caminhão roubado
Desembolar	Resolver, desenrolar
Caíram	Foram presos
Passei uma folha	Entregou dinheiro
Os Homens	A polícia
Vai me dar	Vai me entregar para a polícia
Lombrou	Deu errado. A polícia apareceu

Fonte: Adaptado de Gesteira (2015)

Considerando o contexto apresentado e a obra de Gesteira (2015), a indexação de conteúdo ao utilizar os recursos de SOC com descritores especializados nos arquivos e áudios indexados, possibilitaria buscas por assuntos de interesse, de forma precisa. Assim, o investigado poderia buscar no sistema de recuperação de informação, por exemplo, pelo termo “licitação” e os resultados poderiam apresentar áudios presentes naquele universo que versassem sobre o referido termo e suas variáveis.

Para Bispo e Vital (2023):

A recuperação de informação é um aspecto central na Ciência da Informação e assume uma importância significativa quando lidamos com grandes volumes de dados, particularmente em entidades como a Polícia Federal. A busca textual e a indexação de termos, sendo técnicas amplamente utilizadas, facilitam a extração de informações relevantes desses extensos e diversificados bancos de dados. Esta estratégia permite uma análise mais rápida e precisa das análises, tornando-se um elemento crucial na eficiência do trabalho policial.

Com base nas observações de Gesteira (2015) e na ênfase dada por Bispo e Vital (2023) sobre a relevância da recuperação de informação em contextos com grandes volumes de dados, como é o caso das entidades policiais, fica evidente a necessidade de métodos eficientes de indexação e busca. As técnicas de busca textual e indexação se mostram essenciais para garantir a rapidez e precisão na análise de informações em tais contextos. Na seção a seguir, abordaremos os aspectos metodológicos do projeto de pesquisa que busca desenvolver e aprimorar tais técnicas no contexto descrito.

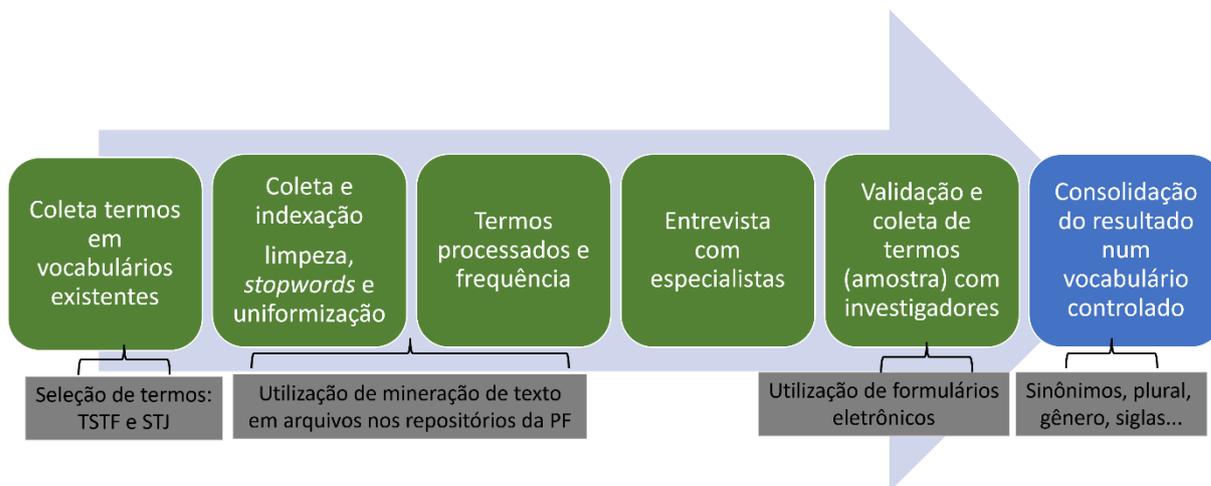
## **5 PROCEDIMENTOS METODOLOGICOS**

Considerando o acompanhamento e os estudos contínuos da Polícia Federal na temática dos crimes relacionados ao tráfico de entorpecentes: origem, evolução da estrutura organizacional, geolocalização e atividades criminosas, esta pesquisa propõe método exploratório, baseado em levantamento documental nos repositórios de documentos dos sistemas disponíveis na instituição. O levantamento documental tem o objetivo de extrair termos para criação de vocabulário especializado inicial sobre o assunto, além de identificar os vocabulários controlados da área jurídica que possuam relação com a temática.

A pesquisa se caracteriza como um estudo de caso, em que será utilizado como instrumento de coleta de dados formulários eletrônicos e entrevista. A coleta de dados será realizada com policiais federais e outros policiais de órgãos parceiros da Polícia Federal que conheçam a temática analisada. O intuito é extrair mais termos oriundos dos conhecimentos implícitos desses profissionais, das experiências vivenciadas.

Para desenvolver um vocabulário controlado a partir da análise de um vocabulário existente, mineração de texto, uso de formulários e entrevistas com especialistas, serão realizadas as seguintes etapas:

Figura 6 - Percurso da metodologia da Pesquisa



Fonte: Do autor (2023)

A seguir serão explicadas cada uma das etapas indicadas na figura acima.

### 5.1 ANÁLISE DE VOCABULÁRIOS EXISTENTES

Ao elaborar um vocabulário controlado para um determinado domínio, é essencial analisar os vocabulários já existentes. Isso permite identificar termos e conceitos relevantes presentes em documentos, recursos online, glossários ou outras fontes de informação. A análise de vocabulários preexistentes facilita a identificação de termos, conceitos e frases importantes, proporcionando uma base sólida para o desenvolvimento de um novo vocabulário controlado adequado às necessidades específicas do domínio em questão.

A seleção criteriosa de termos é um componente crucial para a construção de um vocabulário controlado eficiente. Conforme discutido por Smit e Kobashi (2003), a identificação e escolha de termos e conceitos relevantes são fundamentais para estabelecer uma base sólida para o vocabulário. Ao selecionar termos que refletem adequadamente o conteúdo dos documentos e as necessidades dos usuários, é possível garantir que o vocabulário controlado facilite a busca e recuperação de informações de maneira eficaz e intuitiva.

Na primeira fase do processo de levantamento, iniciou-se a investigação com um conjunto específico de termos-chave relacionados ao domínio do tráfico de

entorpecentes. A seleção desses termos-chave foi cuidadosamente ponderada, levando em consideração sua relevância direta para a área de estudo em questão. Optamos por utilizar os termos 'DROGA', 'TRÁFICO', 'ENTORPECENTE', 'MACONHA' e 'COCAÍNA', bem como a palavra 'CRIMES'. Essa escolha não foi arbitrária, mas sim baseada na natureza do nosso objeto de pesquisa, que se concentra na análise de crimes relacionados ao tráfico e ao uso de substâncias ilícitas. Esses termos-chave foram investigados individualmente nos vocabulários controlados disponíveis no STF - TSTF<sup>3</sup>, no Senado Federal<sup>4</sup> e no STJ<sup>5</sup>.

Em cada um desses vocabulários controlados, o autor buscou os termos inicialmente identificados, avaliando todos os termos resultantes relacionados aos termos-chave, bem como as definições disponíveis. O processo de busca foi cuidadosamente realizado para cada termo em cada um dos referidos SOC, no intuito do processo ter uma abrangência e uma precisão na coleta de termos relacionados.

Após a coleta inicial, o autor avaliou os termos que retornaram como resultados. Este processo de avaliação foi conduzido com base em conhecimento empírico, além de considerar as descrições fornecidas pelos próprios vocabulários controlados. Esse passo permitiu a identificação de termos relevantes que poderiam não ter sido coletados na busca inicial.

Por fim, todos os termos coletados e avaliados foram consolidados numa lista única, referida como "Termos coletados nesses vocabulários controlados" apresentada na seção dos Resultados.

## 5.2 MINERAÇÃO DE TEXTO

A mineração de texto se apresenta como uma abordagem fundamental na construção de uma indexação eficiente e na seleção de termos relevantes para um domínio específico. Nesta pesquisa serão empregadas técnicas de mineração de texto, conforme destacado por Xu *et al.* (2015), para extrair termos e conceitos relevantes a partir de grandes volumes de dados textuais.

---

<sup>3</sup> Disponível em: <https://portal.stf.jus.br/jurisprudencia/tesauro/pesquisa.asp>

<sup>4</sup> Disponível em: <https://www2.senado.leg.br/bdsf/item/id/532112>

<sup>5</sup> Disponível em: <https://scon.stj.jus.br/SCON/thesaurus/>

Assim, ao utilizar as técnicas de mineração de texto espera-se obter uma base sólida de termos e conceitos que contribuirão para a criação de um vocabulário controlado eficaz e abrangente, facilitando a recuperação e o acesso às informações relacionadas ao domínio estudado.

A mineração de texto proporciona extrações de bloco de dados ocultos, sendo utilizado em todas as tecnologias modernas de recuperação de informação e é utilizada amplamente pela área de informática, porém as empresas de todas as áreas se beneficiam com esta mineração, aumentando a capacidade de exploração de conhecimentos potenciais (Rajab Asaad, 2021).

Para a ciência de dados, existem diferenças entres os termos *Data Mining* e *Text Mining*, de forma resumida o *Data Mining* é a mineração de dados estruturados, já o termo *Text Mining* abrangeria a mineração de textos não estruturados ou estruturados (Rezende, 2005).

A mineração de texto compreende basicamente em três etapas: a) Pré-processamento Textual; b) Processamento Textual; e c) Pós-processamento textual. Na primeira etapa são abordados aspectos relativos à coleta de dados textuais, a definição de escopo, normalização do texto, processamento de linguagem natural e preparação para processamento em si. Na segunda etapa são realizadas as técnicas de mineração de texto em si desde a catalogação, a indexação e as demais técnicas a serem implementadas na mineração. É uma técnica bastante utilizada em banco de dados de grandes proporções, em que o auxílio de uma ferramenta automatizada se torna imprescindível para que padrões sejam estabelecidos (Rezende, 2005).

### **5.2.1 Seleção de Corpus**

Na área da mineração de texto, um "corpus" é um conjunto estruturado de documentos textuais, selecionado de forma criteriosa, para possibilitar análises linguísticas e estatísticas eficientes. Nesse contexto, um corpus bem definido se torna uma fonte de dados valiosa, que permite aplicar técnicas de mineração de texto de forma eficiente e desenvolver pesquisas relacionadas à linguagem, recuperação de informações e elaboração de vocabulários controlados.

Nesta fase do projeto, os algoritmos foram ajustados na linguagem Python, incorporando técnicas de mineração de texto. Optamos por analisar uma ampla variedade de mais de 140.000 documentos oficiais recebidos ou gerados pela Polícia

Federal. Esses documentos abrangiam diversos tipos documentais, como ofícios, memorandos, notícias, informações externas e relatórios restritos, que foram amplamente tramitados entre órgãos públicos do Brasil. A escolha por esses tipos documentais específicos não foi aleatória, mas sim estratégica. Esses documentos estavam disponíveis na base de dados da Inteligência da PF e estavam estruturados de forma a favorecer a mineração de texto. A variedade desses tipos documentais nos permitiu obter direcionamentos abrangentes e valiosos, enquanto o padrão de estruturação desses documentos facilitou o processo de extração e análise de informações relevantes para o nosso estudo.

Para facilitar a mineração, os documentos foram previamente processados e convertidos no formato de texto (extensão de arquivos .txt). Embora fosse possível utilizar técnicas como lematização e clusterização para separação por contexto de domínio, o foco do projeto foi direcionado aos arquivos relacionados à temática de tráfico de entorpecentes.

Para otimizar o controle de erros, aprofundar a análise e melhorar a eficiência na fase de mineração, foram aplicados no projeto três algoritmos *Python*. Estes são: *part1.py* (Apêndice C), *part2.py* (Apêndice D) e *classegramatical.py* (Apêndice E). Os códigos fonte destes algoritmos estão detalhadamente descritos nos apêndices mencionados.

O algoritmo *part1.py* foi desenvolvido com o objetivo de facilitar a análise e identificação de termos de interesse, termos resultantes na coleta nos vocabulários controlados existentes, em um conjunto específico de arquivos da Polícia Federal. Em cenários de pesquisa, onde a presença e a frequência de determinados termos em documentos podem ser cruciais, este tipo de automação é extremamente útil. Ao automatizar essa busca, é possível economizar tempo, reduzir erros humanos e obter uma visão quantitativa mais clara da presença desses termos nos documentos. No contexto do script, os termos de interesse são pré-definidos e armazenados em um arquivo separado. A ferramenta então busca em um diretório especificado, identifica e conta a ocorrência desses termos em cada documento, e, finalmente, registra os resultados em arquivos parametrizados CSV para fácil referência e análise, arquivos que serão usados no outro algoritmo. O Quadro 2 ilustra as etapas do script *part1.py*.

Quadro 2 - Etapas realizadas no script part1.py

Etapa	Descrição
Importação de Bibliotecas	Carrega as bibliotecas necessárias para as operações do script.
Detecção de Codificação	Assegura a leitura correta dos arquivos independente de sua codificação.
Leitura de Termos de Interesse	Lê os termos de um arquivo de entrada denominado "termos_de_interesse.txt".
Processamento do Diretório	Busca em um diretório especificado, identificando a presença e a frequência dos termos nos documentos.
Registro dos Resultados	Gera dois arquivos CSV de saída: "nomes_arquivos_interesse.csv", que lista os arquivos que contêm os termos, e "resultados_termos_interesse.csv", que detalha a frequência de cada termo nos documentos.
Execução Principal	Coordena as ações do script, integrando as etapas de leitura, processamento e registro.

Fonte: Do autor (2023)

Após a execução do script part1.py, onde os arquivos contendo termos de interesse foram identificados e as frequências dos termos de interesse registradas, a próxima etapa da mineração de texto é conduzida pelo script part2.py. O objetivo principal deste segundo script é refinar e aprofundar a análise dos documentos de interesse. Utilizando técnicas avançadas como TF-IDF (*Term Frequency-Inverse Document Frequency*), o script é capaz de identificar e ponderar a importância relativa de cada termo nos documentos. Além disso, o script também leva em consideração um vocabulário personalizado e uma lista expandida de stopwords para otimizar a qualidade dos resultados. Os arquivos gerados por part1.py servem como entrada para part2.py, estabelecendo assim uma conexão direta e sequencial entre as duas partes do processo. O Quadro 3 abaixo lista as etapas do part2.py.

Quadro 3 - Etapas realizadas no script part2.py

Etapa	Descrição
1. Preparação	Carregamento das bibliotecas e definição das funções auxiliares. Configuração das <i>stopwords</i> e carregamento do modelo de linguagem <i>spacy</i> para o português.

2. Leitura de Vocabulário e <i>Stopwords</i>	Leitura do vocabulário personalizado (dicionário do português sem acentos) e das <i>stopwords</i> ampliadas a partir dos arquivos br-sem-acentos.txt e stopWords_ampliado_pt.txt.
3. Obtenção de Arquivos de Interesse	A partir do arquivo nomes_arquivos_interesse.csv, obtêm-se a lista de nomes dos arquivos de interesse identificados no part1.py.
4. Processamento de Texto	Realiza o pré-processamento de cada arquivo, removendo acentos, pontuação e espaços extras.
5. Extração de Termos	Uso do <i>CountVectorizer</i> para contar a frequência dos termos e do <i>TfidfVectorizer</i> para calcular o TF-IDF.
6. Geração de Resultados dos Termos	Grava os termos minerados e suas estatísticas no arquivo resultados_termos_minerados.csv. Durante este processo, cada termo é verificado contra o vocabulário personalizado para determinar se pertence ou não a ele.
7. Estatísticas TF-IDF	Compila e grava as estatísticas de TF-IDF para os termos minerados no arquivo termos_minerados_tfidf.csv, considerando também sua presença no vocabulário personalizado.
8. Estatísticas Gerais dos Termos	Gera estatísticas gerais para os termos minerados e grava no arquivo estatisticas_termos_minerados.csv, mostrando informações como frequência, IDF e TF-IDF.

Fonte: Do autor (2023)

Dessa forma, o script part2.py complementa e amplia o trabalho iniciado pelo part1.py, fornecendo uma análise mais detalhada e enriquecida dos documentos de interesse.

A implementação do algoritmo part2.py incorpora uma funcionalidade crucial para aprimorar a qualidade dos termos minerados: a verificação contra um vocabulário personalizado da língua portuguesa. Esta verificação, realizada com a *flag* "no vocabulário personalizado", serve como um filtro, visando restringir os resultados a termos reconhecidos e relevantes para o idioma. Dessa forma, a análise se torna mais precisa, excluindo termos estrangeiros e palavras que, embora possam ser formadas por combinações válidas de letras, não têm significado real no português brasileiro.

O script classegramatical.py serve como um refinamento adicional no processo de mineração de texto. Após a identificação inicial dos termos com os scripts part1.py e part2.py, é essencial garantir que os termos resultantes sejam pertinentes à língua portuguesa e tenham significado real. Este script é projetado especificamente para esse propósito. Esse algoritmo se concentra em filtrar os termos com base em um dicionário personalizado da língua portuguesa e, em seguida, categoriza cada termo

válido de acordo com sua classe gramatical, facilitando uma análise posterior mais aprofundada e específica. O Quadro 4 abaixo lista as etapas do `classegramatical.py`.

Quadro 4 - Etapas realizadas no script `classegramatical.py`

Etapa	Descrição
Carregar Dicionário e Termos Minerados	Carrega o dicionário da língua portuguesa <code>br-sem-acentos.txt</code> . Lê os termos do arquivo <code>resultados_termos_minerados.csv</code> .
Identificação da Classe Gramatical	Utilizando o <code>spaCy</code> , cada termo é processado para determinar sua classe gramatical.
Filtragem de termos fora do dicionário	Termos que não constam no dicionário carregado são descartados, garantindo a preservação de termos relevantes.
Gravação de Termos Válidos	Os termos validados, juntamente com suas classes gramaticais são gravados no arquivo <code>resultados_termos_minerados_portugues.csv</code> .

Fonte: Do autor (2023)

No processo de mineração de texto abrangendo cerca de 140 mil arquivos da PF, foram empregadas três etapas metodológicas essenciais para a análise em uma pesquisa: a primeira, conduzida pelo algoritmo `part1.py`, identificou os termos de interesse, gerando uma pré-seleção dos arquivos que seriam submetidos à extração e contabilização completa de termos. Posteriormente, o `part2.py` aperfeiçoou essa seleção, contabilizando os termos e eliminando aqueles não alinhados ao vocabulário padrão da língua portuguesa. Na etapa conclusiva, o algoritmo `classegramatical.py` categorizou os termos remanescentes de acordo com sua classe gramatical, culminando em uma lista otimizada e relevante de termos para a pesquisa.

Alguns termos minerados nos documentos serão listados na próxima seção. Para melhor análise dos textos minerados, foram importados os arquivos de saída na ferramenta de *business intelligence* *QlikSense*, a figura 7 ilustra o painel de visualização. Onde mostra que 142.086 arquivos foram alvos das pesquisas pelos termos de interesse, no entanto, apenas 53.046 deles foram submetidos ao processamento de mineração de texto. Dentre esses 53.046 arquivos, identificamos 85 termos de interesse e realizamos a mineração de 9.633 termos, excluindo *stopwords* e termos que não estavam em português. É importante destacar que o termo de interesse mais frequente encontrado foi “droga”.



base de termos será continuamente alimentada e refinada, com base nas contribuições de outros especialistas e usuários.

A técnica da entrevista, representa um dos métodos mais empregados no campo da pesquisa, sendo instrumental na captura de reações, sentimentos e hábitos do entrevistado. Essa capacidade de capturar nuances é particularmente vital quando se busca um entendimento profundo de um domínio especializado, como o combate ao tráfico de entorpecentes. Neste estudo, ao dialogar com um policial especialista, foi possível aproveitar as vantagens desse método, esclarecendo terminologias e conceitos específicos e compreendendo a relevância de termos do domínio. A técnica de entrevista e o método do grupo podem revelar lacunas na interação entre os participantes e o objeto de estudo, sendo este um alerta para os pesquisadores estarem cientes e preparados para mitigar (Baptista; Cunha, 2007).

A entrevista teve duração de aproximadamente 60 minutos e foi estruturada de forma interativa. A dinâmica estabelecida permitiu a análise conjunta dos termos identificados nas etapas anteriores. Os termos coletados manualmente e aqueles provenientes da mineração de texto foram submetidos à avaliação do especialista. Durante o processo, cada termo foi discutido de forma detalhada, proporcionando ao especialista a oportunidade de compartilhar perspectivas, sugerir sinônimos, indicar novos termos e descartar aqueles considerados não relevantes. Diversos termos novos e sinônimos foram registrados, enquanto alguns foram classificados como não relevantes. Para ilustrar parte desse processo, o Quadro 5 apresenta uma pequena amostra dos termos consolidados pelo especialista, que foram coletados durante esta fase da entrevista. Essa amostra serve como uma prévia do tipo de termos discutidos e validados. A lista completa, juntamente com seus respectivos conceitos, será detalhada na seção de resultados, no Quadro 7 - 'Termos consolidados do subitem 6.3 - TERMOS PROVENIENTES DA ENTREVISTA'. Este quadro 7 fornece uma visão completa dos termos consolidados e suas definições, fornecendo uma compreensão abrangente do resultado desta etapa do estudo.

#### Quadro 5 - Amostra dos termos na fase da Entrevista

<b>Termos: Bagulho, Granel, Tijolo, Pó, Doce, Pedra, Bala, Beck, Pilantra, Cana, Fita, Dura</b>
---

Fonte: Do autor (2023)

## 5.4 FORMULÁRIOS ELETRÔNICOS

Para garantir uma construção colaborativa do vocabulário controlado, é importante envolver especialistas do assunto e usuários. Para tanto, a criação de formulários eletrônicos é uma opção viável para coletar *feedback* dos *stakeholders*. De acordo com Lancaster (2003), o sucesso de uma indexação depende da interação dos especialistas do assunto com a técnica de indexação.

Os questionários online, conforme realçado por Baptista e Cunha (2007), se beneficiam da disponibilidade constante e da coleta acelerada de dados, características essenciais para uma abordagem colaborativa na construção de vocabulários controlados. Esta digitalização maximiza a participação de especialistas e usuários, facilitando a consolidação dos termos.

A utilização de questionários pode ser estratégica para obter consenso entre especialistas. Em contextos como da pesquisa, essa abordagem contribui para refinar e validar os termos propostos, assegurando relevância e precisão no vocabulário desenvolvido (Baptista; Cunha, 2007).

Os formulários serão desenvolvidos de forma a permitir que especialistas, usuários e outros envolvidos no domínio de interesse possam contribuir com sugestões de termos, sinônimos, regionalismos, além de validar os termos já consolidados. Foram selecionados os termos de relevância alta para preenchimento de sinônimos e alguns termos para validação da relevância pelos especialistas. Essa abordagem permite que o conhecimento empírico, como gírias e termos informais, seja consolidado no vocabulário.

Para a elaboração dos formulários eletrônicos, optou-se pelo *Google Forms* em virtude de sua interface intuitiva, acessibilidade ampla e eficiência comprovada na coleta de dados. Esta ferramenta proporciona uma rápida interação com os *stakeholders*, alinhando-se perfeitamente aos objetivos estabelecidos para a coleta e refinamento dos termos. Foi realizada novamente uma análise manual pelo pesquisador, consolidando os termos sinônimos antes do envio dos formulários eletrônicos. Ao término deste documento, serão anexadas as telas do formulário utilizado, nas figuras 8 e 9 estão ilustradas as imagens do formulário, visando exemplificar claramente o instrumento empregado na pesquisa.

Figura 8 - Amostra de tela do formulário

Seção 2 de 7

**1. Substantivos - Organizações, Papéis e Pessoas** ✕ ⋮

Descrição (opcional)

**1.1. Enriquecimento de Vocabulário Controlado em Tráfico de Entorpecentes**

Liste todos os **sinônimos e gírias** que você conhece para do termos abaixo, **separando** cada um por **vírgula**

⋮

(OPCIONAL) Conheça sinônimo ou gíria para o termo **PRIMEIRO COMANDO DA CAPITAL**: PCC ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **COMANDO VERMELHO**: CV?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **FAMILIA DO NORTE**: FDN ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **FLAGRANTE DELITO**: Flagrante ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **GERENTE DO TRÁFICO**: Chefe ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **INTEGRANTE**: ?

Texto de resposta curta

Fonte: Do autor (2023)

Figura 9 - Outra tela do formulário

Dentro do contexto criminal, especialmente relacionado ao tráfico de entorpecentes e atividades ilícitas, por favor, **marque os termos que você NÃO considera relevantes para construção de vocabulário controlado.**

	Não relevante
AGENTE INFILTRADO: - Pessoa que se infiltra em u...	<input type="radio"/>
CONFRONTO: - Encontro violento, muitas vezes ent...	<input type="radio"/>
DENÚNCIA: - Alegações ou acusações, muitas veze...	<input type="radio"/>
GRAU DE ENVOLVIMENTO: - Extensão da participaç...	<input type="radio"/>
INDIVÍDUO: - Pessoa ou entidade.	<input type="radio"/>
INQUÉRITO: - Investigação ou exame oficial.	<input type="radio"/>
JUSTIÇA: - Sistema legal ou equidade.	<input type="radio"/>
LÍDER DE ORGANIZAÇÃO: Líder - Indivíduo que lider...	<input type="radio"/>
ORGANIZAÇÃO CRIMINOSA: Facção, ORCRIM - Org...	<input type="radio"/>
POLÍCIAS: - Forças ou agências encarregadas de m...	<input type="radio"/>
PEÃO: - Pessoa que executa ordens, geralmente ref...	<input type="radio"/>
QUADRILHAS: Bando, gangue - Grupos criminosos.	<input type="radio"/>
RISCO: - Possibilidade de perigo ou dano.	<input type="radio"/>
SUSPEITA: - Dúvida ou desconfiança.	<input type="radio"/>
TIRA: Policial - Policial.	<input type="radio"/>
VIGILÂNCIA: - Ato de observar ou monitorar.	<input type="radio"/>
X9:Dedo-duro, Informante - Pessoa que entrega ou i...	<input type="radio"/>

(OPCIONAL) Em relação aos **termos relevantes acima** quer adicionar sinônimos?

Modelos de resposta: *Termo chave: Sinônimos*  
**X9: Dedo-duro, Informante**  
**GERENTE: Chefe**

Texto de resposta longa

Antes a seção? Continuar para a próxima seção

Fonte: Do autor (2023)

Na próxima seção os resultados obtidos com a pesquisa serão discutidos em profundidade.

## 6 RESULTADOS

Esta pesquisa visou desenvolver um controle terminológico específico para o domínio dos crimes de tráfico de entorpecentes na Polícia Federal brasileira, um campo onde a precisão e a clareza da linguagem são cruciais. A necessidade de um Vocabulário Controlado neste setor surge da complexidade e diversidade terminológica inerente ao âmbito do direito penal e das operações policiais. A iniciativa de criar um vocabulário controlado tem como objetivo principal fornecer uma ferramenta que auxilie na precisão da comunicação e na eficiência da recuperação de informações dentro da Polícia Federal. Este esforço se alinha ao objetivo geral da pesquisa, promovendo uma padronização linguística que pode impactar positivamente as operações e investigações da instituição.

O resultado desta pesquisa culminou na proposição de uma metodologia para a criação de Vocabulários Controlados no contexto da Polícia Federal. Esta metodologia seguiu as etapas delineadas pelo projeto de pesquisa: coleta de termos em vocabulários existentes em fontes abertas, coleta de termos através de mineração de texto em documentos disponíveis, e por fim, validação e coleta de termos através de entrevistas e formulários eletrônicos com especialistas. O Vocabulário Controlado desenvolvido não apenas reflete a terminologia específica utilizada pela Polícia Federal no combate ao tráfico de entorpecentes, mas também estabelece um modelo para futuras iniciativas de padronização terminológica em outras áreas de atuação da segurança pública.

Os resultados são apresentados de acordo com o tipo de coleta de dados realizada.

### 6.1 COLETA DE TERMOS EM VOCABULÁRIOS EXISTENTES

No quadro 6 é apresentada uma lista de termos existentes nos SOC do STF - TSTF, do Senado Federal e do STJ, provenientes dos termos-chave, servindo como um importante recurso de referência para dissertação. São 90 termos identificados como importantes no domínio estudado.

Quadro 6 - Termos coletados em vocabulários controlados

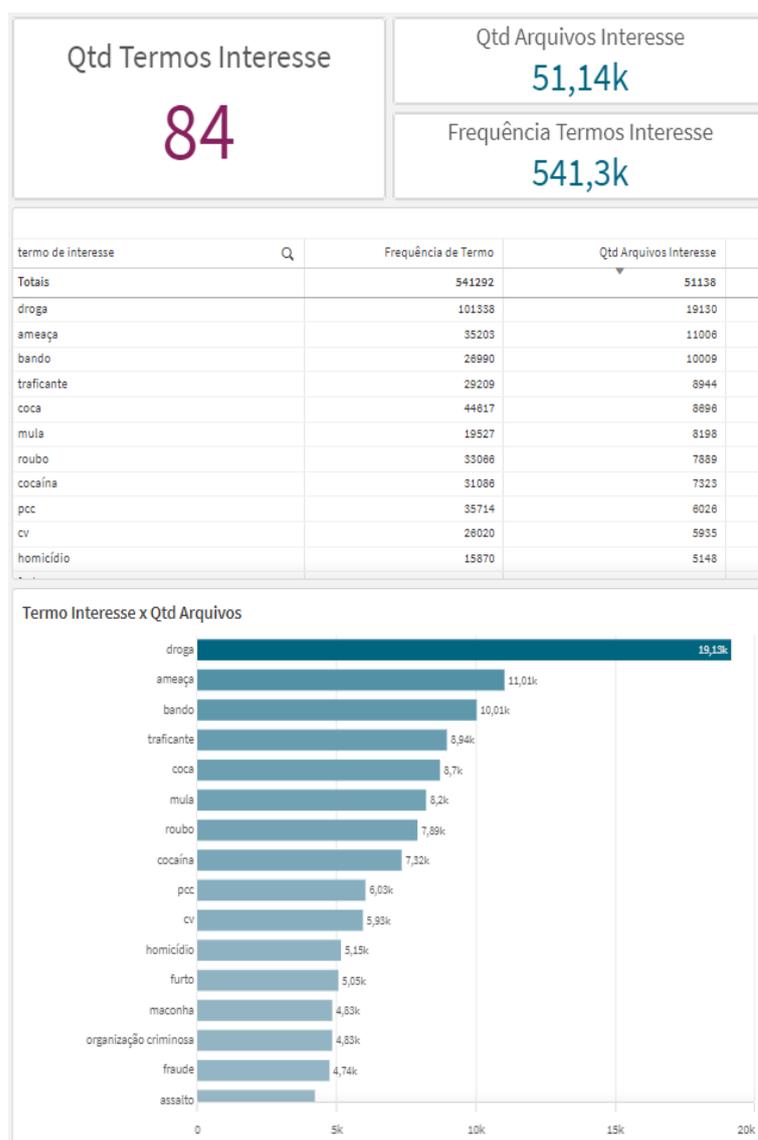
AÇÃO CONTROLADA	CRIME HEDIONDO	ESQUEMA CRIMINOSO
AGENTE INFILTRADO	CRIME INAFIANÇÁVEL	ORCRIM
ALCALOIDE	CRIME QUALIFICADO	ORGANIZAÇÃO CRIMINOSA
AMEAÇA	CV	PCC
APROPRIAÇÃO INDÉBITA	DEPENDÊNCIA FÍSICA	PERÍCIA
ASSALTO	DEPENDÊNCIA PSÍQUICA	POSSE DE ENTORPECENTES
ASSOCIAÇÃO CRIMINOSA	DEPENDÊNCIA QUÍMICA	PRIMEIRO COMANDO DA CAPITAL
ASSOCIAÇÃO PARA O TRÁFICO	DEPENDÊNCIA TOXICOLÓGICA	PRINCÍPIO DO NON OLET
BALANÇA DE PRECISÃO	DROGA	MONITORAMENTO DA AÇÃO
BANDO	DROGA SINTÉTICA	MULA
CÁRCERE PRIVADO	ECSTASE	PSICOTRÓPICO
CLORETO DE ETILA	EXTORSÃO	RECEPTAÇÃO
COCA	EXTORSÃO MEDIANTE SEQUESTRO	RIXA
COCAÍNA	FAMILIA DO NORTE	ROUBO
COLABORAÇÃO PREMIADA	FDN	SUBSTÂNCIA TÓXICA
COLEGIADO	FLAGRANTE DELITO	TRAFICANTE
COMANDO VERMELHO	FRAUDE	TRAFICANTE DE DROGAS
CONLUIO	FRAUDE PROCESSUAL	TRAFICANTE DE ENTORPECENTES
CONSTRANGIMENTO ILEGAL	FURTO	TRÁFICO DE DROGAS PRIVILEGIADO
CRACK	GRAU DE ENVOLVIMENTO	TRÁFICO DE ENTORPECENTES
CRIME ANTECEDENTE	GRUPO ARMADO CIVIL	TRÁFICO INTERESTADUAL DE DROGAS
CRIME CONTRA A PESSOA	HOMICÍDIO	TRÁFICO INTERNACIONAL DE DROGAS
CRIME CONTRA A VIDA	LANÇA PERFUME	TRÁFICO INTERNO DE DROGAS
CRIME CONTRA O PATRIMÔNIO	LEI DE TÓXICOS	USURPAÇÃO
CRIME DE DANO	LESÃO CORPORAL	VICIADO
CRIME DOLOSO CONTRA A VIDA	LÍDER DE ORGANIZAÇÃO	DEPENDÊNCIA DE DROGA
DESACATO	MACONHA	ALUCINÓGENO
SKUNK	ÓPIO	DROGA ALUCINATÓRIA
PSICOTRÓPICO	LSD	DROGA ILÍCITA
HEROÍNA	ESTUPEFACIENTE	NARCÓTICO

Fonte: Do autor (2023)

## 6.2 COLETA DE TERMOS NA MINERAÇÃO DE TEXTO

No âmbito da mineração de texto, para a busca na base documental da PF foram utilizados os termos consolidados na fase anterior como filtro, determinando assim quais arquivos realmente seriam submetidos ao processo de extração e contabilização de termos. Conforme ilustrado na Figura 10, a busca por conteúdo concentrou-se em 84 termos previamente identificados como relevantes. Este processo resultou na identificação de mais de 51 mil arquivos contendo ao menos um dos termos de interesse. Tais arquivos, por sua vez, foram integralmente processados por meio da mineração de texto. A figura 10 ilustra o painel no *QlikSense* utilizado para análise dos dados.

Figura 10 - Painel do QlikSense



Fonte: Do autor (2023)

Na sequência do processo, os números foram ilustrados na figura 11, após a aplicação de filtros de *stopwords*, validação das palavras de acordo com o vocabulário do português do Brasil, e uma seleção restritiva focada nos arquivos que continham os termos de interesse (aproximadamente 51 mil), os documentos foram meticulosamente processados. Este procedimento intensivo resultou na identificação de mais de 74 mil termos distintos, demonstrando a amplitude e profundidade da análise realizada.

Figura 11 - Termos minerados - painel do QlikSense

termo minerado	Q	Classe Gramatical	Q	Frequência S
Qtd Termos Minerados		Qtd Arquivos Minerados		
74,96k		51,14k		
		Frequência Termos Interesse		
		541,3k		
Totais				42797411
nao		Interjeição		514338
federal		Adjetivo		296422
policia		Verbo		281588
sao		Substantivo		223184
inteligencia		Substantivo		189885
la		Substantivo próprio		189179
rua		Substantivo		149558
data		Substantivo		149470
brasil		Substantivo próprio		147961
estado		Verbo		144796
policial		Adjetivo		143318
rio		Substantivo		140489
silva		Substantivo próprio		140378
documento		Substantivo		134889
nome		Substantivo		131300
acesso		Substantivo		110980
tambem		Verbo		107062
local		Adjetivo		108901
dia		Substantivo		108675
sobre		Preposição		103469
paulo		Substantivo próprio		102395
informacoes		Substantivo		100034
ja		Substantivo próprio		99101
ate		Preposição		98391
jose		Substantivo		94222
dados		Substantivo		94109
empresa		Substantivo		93276
seguranca		Substantivo		91552
anos		Substantivo		90913
sendo		Verbo auxiliar		90516
lei		Substantivo		87448
santos		Substantivo		83679
ha		Substantivo		82771
onde		Pronome		81928
deste		Preposição		79739
policiais		Substantivo		79272
pesoas		Substantivo		78602

Fonte: Do autor (2023)

Ao utilizar a classificação gramatical disponibilizada pela biblioteca *Spacy* que, é importante ressaltar, não assegura categorizações infalíveis, mas que se mostrou valiosa para uma análise em larga escala, as tabelas 1, 2 e 3 destacam os termos mais frequentemente identificados no processo de mineração de texto. Essas representações visuais proporcionam uma visão aprofundada da relevância e predominância de certos termos em nosso conjunto de dados analisado.

Tabela 1 - Amostra de termos minerados nos arquivos – substantivos

Termo minerado	Classe Gramatical Spacy	Frequências	Termo minerado	Classe Gramatical Spacy	Frequências
são	Substantivo	223184	informação	Substantivo	60249
inteligência	Substantivo	169665	processo	Substantivo	57462
rua	Substantivo	149558	ministério	Substantivo	57309
data	Substantivo	149470	Antônio	Substantivo	56173
rio	Substantivo	140489	filho	Substantivo	55960
documento	Substantivo	134889	entrada	Substantivo	55045
nome	Substantivo	131300	país	Substantivo	54728
acesso	Substantivo	110980	após	Substantivo	54152
dia	Substantivo	106675	prisão	Substantivo	53628
informações	Substantivo	100034	delegacia	Substantivo	53430
José	Substantivo	94222	centro	Substantivo	53098
dados	Substantivo	94109	parte	Substantivo	52829
empresa	Substantivo	93276	casa	Substantivo	52571
segurança	Substantivo	91552	página	Substantivo	52014
anos	Substantivo	90913	divulgação	Substantivo	50878
lei	Substantivo	87448	veículo	Substantivo	50863
santos	Substantivo	83879	ano	Substantivo	50216
ha	Substantivo	82771	fato	Substantivo	50203
policiais	Substantivo	79272	sigilo	Substantivo	50152
pessoas	Substantivo	78602	Souza	Substantivo	50110
publico	Substantivo	76326	forma	Substantivo	49827
difusão	Substantivo	76221	superintendência	Substantivo	49352
janeiro	Substantivo	75719	los	Substantivo	49022

cidade	Substantivo	73080	sistema	Substantivo	48531
conhecimento	Substantivo	72072	bairro	Substantivo	46730
drogas	Substantivo	69834	sul	Substantivo	46166
departamento	Substantivo	68517	João	Substantivo	45825
tipo	Substantivo	68146	pessoa	Substantivo	45778
maria	Substantivo	63865	três	Substantivo	44942
grupo	Substantivo	63290	referência	Substantivo	44879
operação	Substantivo	63267	situação	Substantivo	44444
crime	Substantivo	63042	armas	Substantivo	42796
dias	Substantivo	62550	ocorrência	Substantivo	42328
telefone	Substantivo	61608	período	Substantivo	42268
assunto	Substantivo	61582			

Fonte: Do autor (2023)

Tabela 2 - Amostra de termos minerados nos arquivos – verbo

termo minerado	Classe Gramatical Spacy	Frequências	termo minerado	Classe Gramatical Spacy	Frequências
polícia	Verbo	261588	apreensão	Verbo	28694
estado	Verbo	144796	pedido	Verbo	28502
também	Verbo	107062	agência	Verbo	26755
diz	Verbo	71442	nascido	Verbo	25182
região	Verbo	63485	apresentou	Verbo	23920
reservado	Verbo	51276	fica	Verbo	23637
ter	Verbo	49138	via	Verbo	22200
preso	Verbo	46749	pergunta	Verbo	21649
publica	Verbo	45754	placa	Verbo	21489
além	Verbo	43360	partir	Verbo	21278
conta	Verbo	42434	celular	Verbo	20912
possui	Verbo	41746	visto	Verbo	20316
estão	Verbo	40980	aguardando	Verbo	20163
disse	Verbo	39423	conhecido	Verbo	19850
pode	Verbo	38985	caracteriza	Verbo	19803
pereira	Verbo	38683	autorizada	Verbo	19794
vista	Verbo	38244	informado	Verbo	19683

consta	Verbo	37228	reclusão	Verbo	19541
então	Verbo	34646	vara	Verbo	19471
delegado	Verbo	31288	ri	Verbo	19293
estados	Verbo	30760	avenida	Verbo	19206
fazer	Verbo	30402	respondeu	Verbo	19076
Colômbia	Verbo	30330	república	Verbo	18654
penitenciária	Verbo	30290	realizada	Verbo	18622
havia	Verbo	29395	fazenda	Verbo	18468

Fonte: Do autor (2023)

Tabela 3 - Amostra de termos minerados nos arquivos – adjetivo

termo minerado	Classe Gramatical Spacy	Frequências	termo minerado	Classe Gramatical Spacy	Frequências
federal	Adjetivo	296422	presente	Adjetivo	32908
policial	Adjetivo	143318	lima	Adjetivo	31137
local	Adjetivo	106901	criminal	Adjetivo	30773
relatório	Adjetivo	73791	social	Adjetivo	30518
total	Adjetivo	69097	nova	Adjetivo	30357
serviço	Adjetivo	67770	presidio	Adjetivo	30079
justiça	Adjetivo	67429	vítima	Adjetivo	29636
tráfico	Adjetivo	61968	estadual	Adjetivo	28120
regional	Adjetivo	61883	capital	Adjetivo	27915
número	Adjetivo	61088	brasileira	Adjetivo	27791
penal	Adjetivo	59850	criminosa	Adjetivo	27212
nacional	Adjetivo	59102	anterior	Adjetivo	26836
oliveira	Adjetivo	55025	boa	Adjetivo	26038
brasileiro	Adjetivo	52832	alvo	Adjetivo	25934
grande	Adjetivo	51607	residente	Adjetivo	25403
civil	Adjetivo	51217	confidencia l	Adjetivo	25323
anexo	Adjetivo	46138	seguintes	Adjetivo	24848
área	Adjetivo	45084	reais	Adjetivo	24517
gerais	Adjetivo	41190	mesma	Adjetivo	24442
responsável	Adjetivo	38259	maior	Adjetivo	24362
santa	Adjetivo	38028	saída	Adjetivo	24158
internacional	Adjetivo	37332	bom	Adjetivo	22469

possível	Adjetivo	36281	histórico	Adjetivo	21401
geral	Adjetivo	35017	machado	Adjetivo	20832
funcional	Adjetivo	34349	atual	Adjetivo	20599

Fonte: Do autor (2023)

Após o processo inicial de mineração de texto e a identificação dos termos por meio da classificação gramatical, foi imperativo realizar uma análise mais criteriosa e subjetiva para garantir a relevância e pertinência dos termos no contexto da pesquisa. Considerando um profundo conhecimento sobre o assunto e a importância de cada termo, o pesquisador fez uma avaliação detalhada dos 300 (trezentos) termos mais frequentes. Esse processo subjetivo não apenas refinou a seleção de termos, mas também assegurou que os termos listados são verdadeiramente representativos do domínio em estudo. Os resultados desta avaliação, juntamente com os termos originados da mineração de texto, apresentados no Apêndice B.

Na busca por estabelecer um vocabulário controlado de um domínio qualquer, é crucial que se considere os termos empregados no cotidiano, que vão desde expressões técnicas até gírias regionais e abreviações comuns. Na fase de mineração de texto foram analisados aproximadamente 140 mil arquivos formais da PF. No entanto, como consequência inerente da formalidade destes documentos, a pesquisa concentrou-se majoritariamente em termos técnicos e oficiais, deixando de lado as nuances linguísticas empregadas no dia a dia.

Dessa forma, foi delimitado o escopo de termos na etapa, excluindo expressões regionais, gírias e outras formas linguísticas que oferecem uma visão mais rica e diversificada do tema. Estes termos, muitas vezes, são cruciais para entender a dinâmica real dos crimes relacionados a drogas e suas peculiaridades em diferentes regiões ou contextos.

Para mitigar esta lacuna identificada durante a mineração de texto, recorreremos a etapas subsequentes: entrevistas com especialistas e a aplicação de formulários eletrônicos. Ambas as abordagens foram concebidas para captar a diversidade e riqueza linguística que os registros formais podem ter omitido, assegurando uma compreensão mais holística e representativa do vocabulário relacionado a crimes de entorpecentes no Brasil.

### 6.3 TERMOS PROVENIENTES DA ENTREVISTA

A realização de entrevista com especialista é uma etapa essencial para entender melhor um tema, especialmente em áreas específicas como os crimes de entorpecentes. Conversar com um profissional que tem ampla experiência no assunto possibilita ao pesquisador perceber detalhes e aspectos que podem não emergir em pesquisas mais amplas ou que são realizadas estritamente com o uso de computadores e baseadas em documentos e bibliografia. Por meio dessa entrevista, foi possível verificar se os termos identificados em outras partes da pesquisa são pertinentes e se são comumente utilizados na prática da área em estudo.

Além de nos ajudar a entender melhor os termos já coletados, os especialistas podem apontar palavras e expressões que são usadas em diferentes regiões do Brasil ou que são gírias comuns na área, complementando a coleta. Essa informação é muito importante para criar um vocabulário que realmente reflète como as pessoas falam e escrevem sobre o assunto. Ao término da entrevista, o especialista proporciona avaliação qualitativa dos termos. Essas contribuições valiosas, juntamente com os *insights* adquiridos durante as discussões, estão sistematizadas na Quadro 7.

No Quadro 7, apresentamos os 80 termos resultantes da pesquisa, categorizados por classes gramaticais e respectiva definição informada na entrevista. Essa organização oferece uma visão estruturada do vocabulário relacionado a crimes de entorpecentes no Brasil, refletindo a riqueza linguística e o conhecimento adquirido com especialista durante a entrevista. O Quadro 7 é um recurso fundamental que auxilia na análise e compreensão do vocabulário controlado desta etapa.

Quadro 7 - Termos consolidados

**Verbos:**

Bolar - enrolar um cigarro de maconha.

Cheirar - inalar drogas em pó, como cocaína.

Injetar - usar uma seringa para administrar drogas.

Fumar - consumir drogas através da inalação de fumaça.

Engolir - consumir drogas por via oral.

Estourar - ação policial para deter traficantes e apreender drogas.

Esticar - diluir drogas para aumentar a quantidade.

Batizar – misturar droga com outra substância.

Arregar - pagar propina a policiais para evitar a repressão ao tráfico de drogas.

Apurar - obter lucro com a venda de drogas.

Cobrar - exigir o pagamento de drogas vendidas.

Capotar - ser preso ou detido pela polícia.

#### **Adjetivos:**

Pesado - traficante envolvido em atividades violentas.

Careta - pessoa que não usa drogas.

Barato - efeito de uma droga no usuário.

Chapado - estado de intoxicação por drogas.

Noiado - usuário de drogas, geralmente crack, com aparência descuidada.

Malhação - uso de drogas injetáveis.

X9 - informante, geralmente alguém que fornece informações à polícia.

Overdose - ingestão excessiva de drogas, causando efeitos adversos ou morte.

Viciado - pessoa dependente de drogas.

Cagueta - informante que denuncia atividades criminosas, como tráfico de drogas.

#### **Substantivos:**

Traficante - alguém que vende drogas ilícitas.

Usuário - pessoa que consome drogas ilícitas.

Narcotráfico - comércio ilegal de drogas.

Cocaína - droga estimulante derivada das folhas de coca.

Crack - forma de cocaína em pedra fumável.

Maconha - droga feita a partir da planta de Cannabis.

Heroína - droga opiácea derivada da morfina.

Metanfetamina - droga estimulante sintética.

LSD - droga alucinógena, também conhecida como ácido.

Ecstasy - droga sintética com efeitos estimulantes e alucinógenos.

Plantão - local onde as drogas são vendidas.

Biqueira - ponto de venda de drogas.

Mocó - esconderijo onde as drogas são guardadas.

Muqui - esconderijo de drogas.

Seda - papel usado para enrolar cigarros de maconha.

Bagulho - gíria para drogas ou objetos ilegais.

Granel - grande quantidade de drogas.

Tijolo - quantidade compactada de drogas, geralmente maconha.

Pó - gíria para cocaína.

Doce - gíria para LSD.

Soltura - liberação da prisão relacionada ao tráfico de drogas.

Pedra - gíria para crack.

Bala - gíria para ecstasy.

Beck - gíria para cigarro de maconha.

Pilantra - pessoa desonesta, geralmente envolvida em atividades criminosas.

Cana - prisão ou detenção.

Fita - atividade criminosa, como roubo ou tráfico de drogas.

Dura - abordagem policial.

Traficar - ato de vender ou distribuir drogas ilegalmente.

Passar - entregar drogas a alguém.

Pixote - jovem envolvido no tráfico de drogas.

Caô - mentira ou situação enganosa.

Camelo - vendedor ambulante de drogas.

12 - gíria para polícia.

Fim de linha - lugar onde as drogas são vendidas.

Mão na massa - envolvimento direto com atividades criminosas.

Pipa - pequena quantidade de drogas, geralmente maconha.

Papelote - embalagem de drogas em pó, como cocaína ou heroína.

Fininho - cigarro de maconha enrolado bem fino.

Baseado - outro termo para cigarro de maconha.

Carreirinha - fileira de droga em pó, como cocaína, preparada para ser cheirada.

Jaque - roubo de drogas ou dinheiro de traficantes.

Lolo - gíria para inalante ou lança-perfume.

Quebrar - ato de dividir ou fracionar drogas para venda.

Descolar - conseguir drogas, geralmente através de compra.

Respingo - efeito colateral negativo do uso de drogas.

Cachimbo - dispositivo usado para fumar crack.

Narco - abreviação de narcotráfico ou traficante de drogas.

Fornecedor - pessoa que fornece drogas aos traficantes.

Cartel - organização criminosa envolvida no tráfico de drogas.

Fronteira - limite entre países, frequentemente associado ao tráfico de drogas.

Lavagem de dinheiro - processo de ocultar a origem ilícita de fundos obtidos com o tráfico de drogas.

Perna - pessoa responsável por transportar drogas.

Boca de fumo - local onde drogas são vendidas, geralmente a céu aberto.

Grilagem - falsificação de documentos relacionados ao tráfico de drogas.

Infiltração - infiltração de um agente da polícia em uma organização de tráfico de drogas.

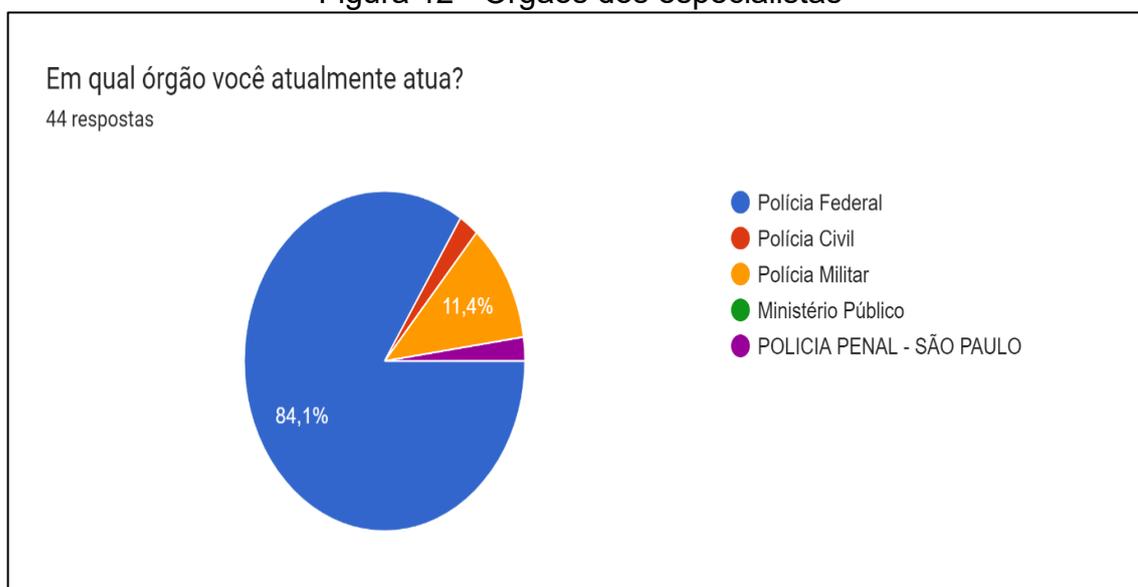
QG - quartel-general do tráfico de drogas.

Resgate - pagamento feito para libertar alguém envolvido no tráfico de drogas.

## 6.4 TERMOS PROVENIENTES DOS FORMULÁRIOS ELETRÔNICOS

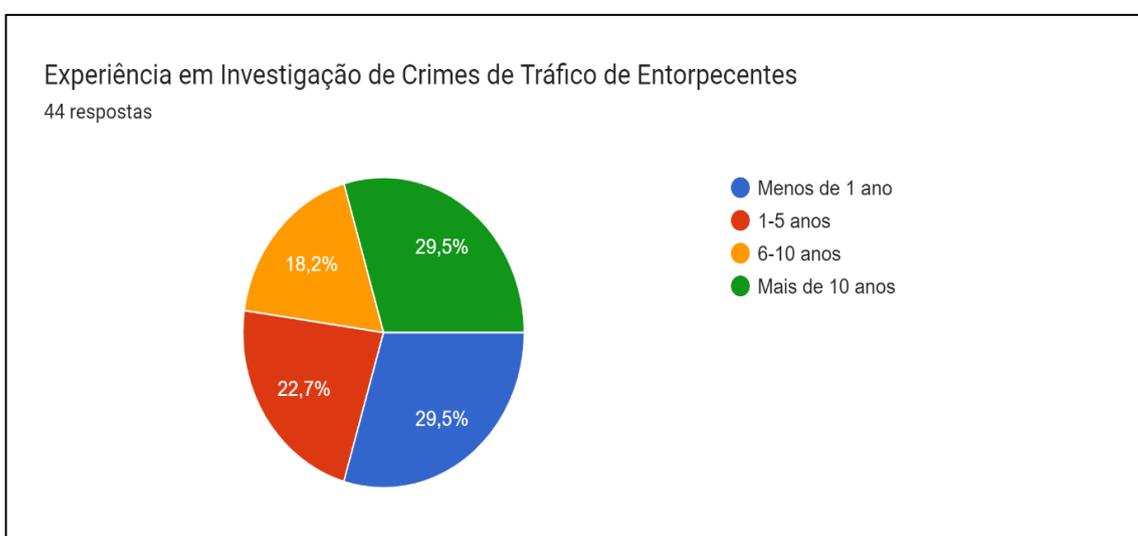
A partir dos formulários eletrônicos, foram coletadas 44 (quarenta e quatro) respostas significativas. As figuras 12 e 13 delineiam o perfil dos policiais que contribuíram, destacando aspectos cruciais de suas características e ambientes profissionais. Essas visualizações enriquecem nossa compreensão sobre a diversidade e a representatividade dos participantes no âmbito da pesquisa.

Figura 12 - Órgãos dos especialistas



Fonte: Do autor (2023)

Figura 13 - Experiência dos especialistas



Fonte: Do autor (2023)

Os especialistas que contribuíram com os formulários forneceram uma valiosa gama de sinônimos e terminologias específicas de sua prática profissional. Esta troca direta através dos formulários reforça a importância da colaboração entre pesquisadores e profissionais especializados, permitindo a captura de conhecimento empírico frequentemente inacessível por outros meios. Cada sugestão foi cuidadosamente avaliada pelo pesquisador, levando em consideração sua relevância e aplicabilidade no intuito de não ficar um formulário muito extenso. Os termos ratificados e consolidados são apresentados no Quadro 8, evidenciando o enriquecimento e a precisão do vocabulário adotado, fruto dessa interação direta com os especialistas no campo.

#### Quadro 8 - Termos sinônimos consolidados através dos formulários eletrônicos

ACUSAR: apontar; botar na reta; caguetar; dedar; denunciar; delatar; dar o serviço; incriminar; indicar 171: estelionatário; fraudador; picareta; pilantra; golpista; rato; malandro
AMEAÇA: chantagem; pressão; enquadro; esculacho; aviso; jurado; decretado; ideia; susto
COCAÍNA: Pó; Farinha; Hollywood; Coca; branquinha; peixe; escama; brilho; mistura; branca; açúcar
COMANDO VERMELHO: CV; falange vermelha; comando;
CRACK: Pedra; Zulu; mármore; amarela; brita; rapadura
DÍVIDA: pendura; débito; calote
DROGA SINTÉTICA: química; larica; rebite; droga do amor; azul; doce; bala; gringa; meta; cristal; balinha; comprimido; adesivo
DROGA: tóxico; barato; coisa; coisinha; bagulho; parada; caixa; mercadoria; entorpecente; produto; carga; material
ECSTASE: bala; doce; azul; pílula
EXPLOSIVO: artefato; bananas; bomba; c4; foguete; dinamite
FAMILIA DO NORTE: FDN
FAZER: Matar; passar; empurrar; concluir; dar sumiço; apagar; zerar; cancelar; resolver; quebrar; assassinar
FLAGRANTE DELITO: Flagrante; flafu; BO; bronca
FUGIR: picar a mula; meter o pé; vazar; sumir; desaparecer; dar o pinote
FUZIL: AK; peça pesada; bico; bicudo; longa; fura
FURTO: 155; ganho; mão
GERENTE DO TRÁFICO: Chefe; comandante; patrão; dono; cabeça; frente
HEROÍNA: pico; injetável; picada
HOMICÍDIO: lona; caixão; passamento; CPF cancelado
INTEGRANTE: soldado; funcionário; faccionado; irmã; mano; parça; batizado; chegado; membro
LANÇA PERFUME: Éter; loló; cheirinho; lança; LP

LEVAR: Portar; transportar; trazer; carregar; conduzir; fretar; descer; subir  
 LSD: ácido; doce; adesivo; figurinha; selo; selinho  
 MACONHA: Dendê; Prensado; beck; beque; verde; erva; mato; fumo; camarão; baseado; preto; capim  
 MUNIÇÃO: tiro; bala; chumbo; pipoco; caroço; aço; ponta; caramelo; cartucho; azeitona  
 PADRINHO: protetor  
 PISTOLA: quadrada; peça; arma curta; canhão; ferro; berro; preta; cromada; cano; ferramenta  
 POSSE DE ENTORPECENTES: 28; vapor; mocosado; bolso cheio  
 PRIMEIRO COMANDO DA CAPITAL: PCC; partido; 15.3.3; comenado; família  
 PRENDER: grampear; pegar; cair; ganchar; segurar  
 REVÓLVER: oitão; ferro; berro; ferramenta; canela seca; pau de foro; chuteira; calçado  
 SKUNK: camarão; flor; solta  
 TER: possuir; tá na mão  
 TESTEMUNHA: cagueta; x9; declarante  
 TRAFICANTE: Narco; Fornecedor; chefe; chegado; meliante; bandido  
 TRAFICAR: Passar; Vender; fornecer; descolar; empurrar; subir; descer  
 TRÁFICO DE DROGAS: Tráfico; correria; corre; caminhada; atividade; comércio  
 USUÁRIO: Nóia; Zé droguinha; Viciado; Bico; peba; cliente; craquento; rato; farinheiro; fuminho; doido; vagabundo; dependente; noiado; playboy

Fonte: Do autor (2023)

Ao analisar os dados fornecidos pelos especialistas, é notável a ampla gama de sinônimos associados a certos termos, particularmente aqueles que denotam verbos relacionados ao tráfico e nomes que representam substâncias entorpecentes. Esta diversidade linguística evidencia a complexidade e o dinamismo do linguajar usado no mundo do crime. Por exemplo, termos relacionados ao tráfico, como "ACUSAR" e "FAZER", bem como substâncias como "MACONHA" e "COCAÍNA", apresentaram uma rica variedade de sinônimos. Esta abundância sugere que os criminosos frequentemente adaptam e renovam sua linguagem para se manterem um passo à frente das autoridades. Notavelmente, alguns termos apresentaram menos variações, indicando talvez uma aceitação mais uniforme ou um entendimento comum dentro dessa comunidade. Estas observações reforçam a relevância de manter um vocabulário atualizado e de compreender a evolução constante da terminologia neste campo específico.

## 6.5 VOCABULÁRIO ESPECÍFICO SOBRE ENTORPECENTES NO CONTEXTO DA POLÍCIA FEDERAL

Neste estudo focado no tráfico de entorpecentes no âmbito da Polícia Federal, reconheceu-se o desafio inerente ao mapeamento terminológico, dada a constante evolução das gírias, regionalismos e codificações empregadas pelos criminosos. Estes métodos de comunicação são frequentemente desenvolvidos para confundir e ocultar atividades ilícitas, tornando a tarefa de criar um Vocabulário Controlado consistentemente preciso e atualizado um projeto complexo. Dessa forma, foi essencial a estruturação metódica deste projeto de pesquisa, considerando os diferentes aspectos e fontes de coleta de termos.

A primeira fase envolveu um levantamento intensivo na diversidade terminológica das bases de dados da Polícia Federal, buscou entender o vasto espectro de termos e expressões relacionadas ao tráfico nos documentos disponíveis. Em seguida, focamos em Vocabulários Controlados já existentes em fontes externas, permitindo um enriquecimento contextual dos dados iniciais. Adicionalmente, com a aplicação da mineração de texto e a incorporação de dados provenientes de entrevistas e questionários eletrônicos, conseguimos uma abordagem abrangente, realista e profunda do tema.

Ao concluir estas fases de pesquisa, obtivemos uma coleta substancial e diversificada de termos, culminando na consolidação de uma amostra representativa para um Vocabulário Controlado piloto. Os termos de maior relevância e aplicabilidade foram priorizados, assegurando que o vocabulário resultante estivesse alinhado às necessidades específicas da Polícia Federal no combate ao tráfico de entorpecentes.

Além disso, o processo revelou uma vasta gama de sinônimos, resultando em uma estrutura que se assemelha aos "Anéis de Sinônimos". Esse tipo de Sistema de Organização do Conhecimento (SOC) se caracteriza pela agrupação de termos sinônimos, permitindo uma recuperação de informação mais rica e abrangente. Desta maneira, nossa pesquisa não se limitou a um Vocabulário Controlado, mas estendeu-se para uma estrutura mais complexa que valoriza a diversidade terminológica.

Os resultados salientam a identificação de 151 termos de vital importância relacionados ao tema proposto. Durante as distintas etapas do projeto, consolidamos um conjunto de 210 termos sinônimos, com destaque para as fases de entrevistas e preenchimento de formulários eletrônicos, enriquecidos pela participação direta de

especialistas. No quadro a seguir, apresentamos uma amostra de 15 termos extraídos desse resultado. Contudo, para uma visão completa e detalhada de todos os termos consolidados neste projeto de pesquisa, lista integral registrada no Apêndice A.

Quadro 9 - Amostra dos termos consolidados na pesquisa

Termo	Descrição	Sinônimos
171	Estelionato ou enganação. Referência ao código penal brasileiro.	estelionatário; fraudador; picareta; pilantra; golpista; rato; malandro
USUÁRIO	Pessoa que consome drogas.	Nóia; Zé droguinha; Viciado; Bico; peba; cliente; craquento; rato; farinheiro; fuminho; doido; vagabundo; dependente; noiado; playboy
GERENTE DO TRÁFICO	Pessoa que controla um ponto de venda de drogas.	Chefe; comandante; patrão; dono; cabeça; frente
X9	Pessoa que entrega ou informa sobre atividades ilícitas.	Dedoduro; Informante
TRAFICANTE	Pessoa que vende drogas ilícitas.	Narco; Fornecedor; chefe; chegado; meliante; bandido
MACONHA	Droga depressiva derivada da planta Cannabis sativa.	Dendê; Prensado; beck; beque; verde; erva; mato; fumo; camarão; baseado; preto; capim
CRACK	Droga derivada da cocaína, consumida em forma de pedra.	Pedra; Zulu; mármore; amarela; brita; rapadura
COCAÍNA	Droga estimulante derivada da planta de coca.	Pó; Farinha; Hollywood; Coca; branquinha; peixe; escama; brilho; mistura; branca; açúcar
MUNIÇÃO	Material usado em armas para atirar.	tiro; bala; chumbo; pipoco; caroço; aço; ponta; caramelo; cartucho; azeitona
TRAFICAR	Ação de vender drogas ilicitamente.	Passar; Vender; fornecer; descolar; empurrar; subir; descer
ACUSAR	Alegar que alguém cometeu um delito.	apontar; botar na reta; caguetar; dedar; denunciar; delatar; dar o serviço; incriminar; indicar
PRENDER	Detiver, colocar sob custódia. Refere-se ao ato policial de detenção.	grampear; pegar; cair; ganchar; segurar
FUGIR	Escapar, evadir. Pode se referir a suspeitos evitando a detenção.	picar a mula; meter o pé; vazar; sumir; desaparecer; dar o pinote
FAZER	Executar, realizar uma ação. No contexto de crime, pode se referir a produção ou venda de drogas.	Matar; passar; empurrar; concluir; dar sumiço; apagar; zerar; cancelar; resolver; quebrar; assassinar
LEVAR	Transportar algo. Em tráfico, pode indicar transporte de substâncias ilegais.	Portar <sup>1</sup> ; transportar; trazer; carregar; conduzir; fretar; descer; subir

Fonte: Do autor (2023)

Ao longo deste estudo, foram conduzidas etapas cautelosas de pesquisa, desde o levantamento inicial na base de dados da Polícia Federal até a incorporação de *insights* de especialistas e a aplicação da mineração de texto. Esta abordagem resultou na identificação e categorização de uma variedade rica de termos e sinônimos pertinentes ao tráfico de entorpecentes. Contudo, é fundamental salientar a natureza dinâmica e evolutiva do domínio estudado. Novos termos, gírias e codificações surgem regularmente, refletindo mudanças nas táticas e comunicações criminosas. Portanto, é imprescindível que este Vocabulário Controlado seja periodicamente revisado e atualizado, garantindo sua relevância e eficácia na assistência às operações da Polícia Federal.

Este quadro síntese oferece uma visão consolidada das fases metodológicas realizadas na criação do Vocabulário Controlado direcionado ao contexto do tráfico de entorpecentes na Polícia Federal brasileira. Cada fase, desde a análise inicial de vocabulários existentes até a consolidação do projeto piloto, incluiu uma avaliação qualitativa contínua dos termos pelo pesquisador. Esta avaliação foi crucial para aprimorar a compreensão e a representação acurada da terminologia utilizada neste domínio específico, assegurando que os resultados obtidos em cada etapa fossem refinados e melhorados para as fases subsequentes. A metodologia empregada abrangeu tanto análises quantitativas quanto qualitativas, incorporando a expertise de profissionais da área e o uso de tecnologias de mineração de texto. O quadro a seguir não apenas detalha o processo e os resultados de cada etapa, mas também inclui reflexões sobre estes resultados, evidenciando o impacto e a relevância do Vocabulário Controlado desenvolvido.

A interação constante entre a avaliação qualitativa dos termos consolidados e as etapas metodológicas garantiu a eficácia e a precisão do vocabulário final proposto neste projeto de pesquisa, refletindo um entendimento profundo e dinâmico do tema abordado, e podendo ser aplicada a mesma metodologia para os em setores ou Órgãos congêneres à Policial Federal que queiram consolidar um vocabulário controlado.

Quadro 10 – Quadro síntese

FASE DA PESQUISA	MÉTODO INICIAL	RESULTADOS OBTIDOS	CONSIDERAÇÕES
Análise de Vocabulários Existentes	Revisão documental de fontes como STF, Senado Federal e STJ.	Identificação de <b>90 termos</b> e aplicabilidade no contexto legal.	Relevância de termos pré-existentes na construção de uma base sólida. A partir de <b>6 (seis) termos-chave</b> .
Mineração de Texto	Análise computacional de documentos disponíveis ( <b>140.000 arquivos</b> ).	Coleta de <b>215 termos únicos</b> , localizados nos documentos.	Descoberta de termos específicos e atualizados, refletindo a realidade do tráfico.
Entrevistas com Especialistas	Sessões de entrevista com especialista	Validação de termos, adição de <b>80 termos especializados</b> , e gírias.	Importância da experiência prática no enriquecimento do vocabulário.
Formulários Eletrônicos para Especialistas	Distribuição de questionários eletrônicos para especialistas.	Confirmação de <b>210 termos sinônimos</b> e variações terminológicas.	Ampliação e diversificação do vocabulário controlado..
Consolidação de Vocabulário Controlado - Projeto Piloto	Análise e síntese dos dados coletados nas fases anteriores.	Criação de um Vocabulário Controlado com <b>151 termos-chave e 210 termos sinônimos</b> .	Efetivação de um recurso valioso para uso prático e futuro aprimoramento.

Fonte: Do autor (2023)

## 7 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

A implementação de um vocabulário controlado e especializado na Polícia Federal transcende a simples utilidade para os investigadores atuais; representa um investimento para o futuro. Em um cenário onde criminosos frequentemente empregam uma variedade de termos e gírias para ocultar suas atividades ilícitas, um vocabulário adaptativo e atualizado é essencial para decifrar e antecipar suas estratégias. A criação e validação de tal vocabulário apresenta seus desafios, no entanto, os benefícios potenciais para a Polícia Federal, em termos de precisão e eficácia nas investigações, fazem deste um esforço imperativo. Este instrumento não somente potencializaria as investigações contemporâneas, mas também se estabelecerá como um recurso inestimável para a evolução da inteligência policial.

Ao refletir sobre a evolução da pesquisa, é notável o progresso desde as etapas iniciais até a finalização. Iniciando com um conjunto focado de 6 (seis) termos-chave pesquisados em vocabulários controlados disponíveis em fontes externas, a pesquisa avançou para a identificação de 151 (cento e cinquenta e um) termos essenciais, devidamente conceituados no domínio estudado, mais 210 (duzentos e dez) termos sinônimos após a entrevista e o preenchimento de formulários eletrônicos. Este crescimento demonstra a profundidade e a abrangência da abordagem adotada neste estudo, pode-se verificar nos termos listados no Apêndice A.

Uma das limitações encontradas ao longo desta pesquisa foi a inexistência de vocabulários controlados disponíveis em formatos parametrizados ou bancos de dados relacionais prontos para *download*. Esta condição não proporcionou uma fase inicial de coleta automatizada, a abordagem utilizada não utilizou tecnologia ou algoritmos de inteligência artificial para uma coleta inicial de termos nos vocabulários existentes, a pesquisa teve que se apoiar predominantemente na coleta manual dos termos. Esse método, por sua vez, baseou-se fortemente no conhecimento empírico do pesquisador, apresentando desafios adicionais e potencialmente limitando a abrangência inicial do levantamento.

No processo de mineração de texto, ao analisar aproximadamente 140 mil arquivos formais e não estruturados da Polícia Federal, obtivemos resultados significativos quanto à terminologia técnica e oficial. No entanto, dada a natureza estritamente formal desses documentos, percebemos uma lacuna na captação de expressões regionais, gírias e outras nuances linguísticas comumente empregadas

na rotina dos criminosos. Para um aprofundamento, especialmente em termos de gírias e regionalismos, seria essencial incluir uma gama mais ampla de textos informais no escopo da pesquisa. Esta inclusão poderia proporcionar uma compreensão mais rica e detalhada da diversidade linguística relacionada ao tema em questão. Sobre esta etapa pode-se utilizar também outras estratégias de clusterização, de lematização e de coleta de termos compostos pois podem trazer resultados mais precisos.

Nas etapas de entrevista com especialista e preenchimento de formulários eletrônicos, evidenciou-se uma notável eficácia na captação de novos termos e sinônimos. A interação direta com os experts não apenas enriqueceu nosso repertório terminológico, mas também solidificou a importância da validação dos termos previamente levantados. Além disso, os formulários eletrônicos mostraram-se como uma ferramenta promissora não só para a coleta inicial, mas também para a contínua validação e atualização dos termos. Essa abordagem dinâmica assegura que o vocabulário controlado permaneça fidedigno e reflita fielmente a realidade e as nuances do domínio estudado.

Para trabalhos futuros e avanços na área, sugere-se primeiramente o desenvolvimento de um método para disponibilizar o vocabulário controlado em uma estrutura relacional ou parametrizada, com vistas à sua evolução para uma estrutura de tesouro. Tal estrutura permitiria uma organização mais sofisticada e uma busca mais intuitiva, possibilitando uma recuperação de informações mais rica e dinâmica, alinhada às tendências modernas de organização do conhecimento, podendo ser acoplada em ferramentas de indexação textual.

Além disso, é recomendável aprimorar as etapas de entrevista e preenchimento de formulários eletrônicos, tornando-as ainda mais robustas e capazes de captar uma gama mais ampla de termos em qualquer domínio estudado na Polícia Federal ou outros setores congêneres. Na fase de mineração de texto, a inclusão de arquivos diversificados, bem como a transcrição de áudios, arquivos informais, poderia proporcionar uma visão mais holística, capturando termos mais informais e rotineiros. Esta abordagem ampliada garantiria um vocabulário que abrange não apenas termos técnicos, mas também expressões utilizadas no cotidiano, refletindo mais fielmente a realidade da linguagem em uso.

## REFERÊNCIAS

ARAÚJO, C. A. A. Ciência da Informação, Ciência Social Aplicada: um estudo à luz do pensamento de Gurvitch. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, n. esp., 1º sem. 2006. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/15369>. Acesso em: 21 abr. 2023.

ARAÚJO, C. A. et al. **O que é Ciência da Informação**. Belo Horizonte: KMA, 2018. Disponível em: <http://arquivistica.fci.unb.br/wp-content/uploads/tainacan-items/476350/819502/15958-73972-2-PB.pdf>. Acesso em: 21 abr. 2023.

BARRETO, A. A. Os destinos da ciência da informação: entre o cristal e a chama. **DataGramZero**, v. 0, n. 0, 1999. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/4086>. Acesso em: 06 jan. 2023.

BAPTISTA, S.G; CUNHA, MB. Estudo de usuários: visão global dos métodos de coleta de dados. **Perspectivas em ciência da informação**, SciELO Brasil, 2007. Disponível em: <https://www.scielo.br/j/pci/a/h6HP4rNKxTby9VZzgzp8qGQ/>. Acesso em: 12 out. 2023.

BISPO, F. D.; VITAL, L. P. O uso de vocabulário controlado para o combate ao crime organizado na Polícia Federal. **Ciência da Informação e Ciências Policiais**, [S. l.], v. 4, 2023. Disponível em: <https://anis.pro-metrics.org/index.php/a/article/view/58>. Acesso em: 12 out. 2023

BOCCATO, V. R. C. Os sistemas de organização do conhecimento nas perspectivas atuais das normas internacionais de construção. InCID: **Revista de Ciência da Informação e Documentação**, [S. l.], v. 2, n. 1, p. 165-192, 2011. DOI: 10.11606/issn.2178-2075.v2i1p165-192. Disponível em: <https://www.revistas.usp.br/incid/article/view/42340>. Acesso em: 22 dez. 2022.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasil, 1988. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.htm](http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm) Acesso em: 12 jan. 2023.

BRASIL. Ministério da Justiça e Segurança Pública. Gabinete do Ministro. **Portaria nº 155, de 27 de setembro de 2018**. Aprova o Regimento Interno da Polícia Federal. Disponível em: <https://www.gov.br/pf/pt-br/acao-a-informacao/institucional/regimento-interno-da-policia-federal-2018>. Acesso em: 27 jan. 2023.

BUSH, V. *As we may think*. **The Atlantic Monthly**, v.176, n. 1, p. 101-108, 1945. Disponível em: <http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>. Acesso em: 24 jan. 2023.

CAPURRO, R.; HJORLAND, B. et. al. O conceito de informação. **Perspectivas em Ciência da Informação**, v. 12, n. 1, 2007. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/33134>. Acesso em: 31 dez. 2022.

CARLAN, E.; MEDEIROS, M. Sistemas de Organização do Conhecimento na visão da Ciência da Informação. **Revista Ibero-Americana de Ciência da Informação**, v. 4, n. 2, p. 53–73, 2011.

CRUZ, M. C. A.; FERNEDA, E.; FUJITA, M. S. L. A disponibilização de vocabulário controlado aos usuários para a recuperação da informação. **Revista Ibero-Americana de Ciência da Informação**, v. 15, p. 266-282, 2022.

CURRÁS, E. **Ontologias, taxonomia e tesouros em teoria de sistemas e sistemática**. Brasília: Thesaurus, 2010.

DAVANZO, L. **Vocabulário controlado para arquivos: análise de viabilidade e proposta**. Dissertação (Mestrado em Ciência da Informação) Universidade Estadual Paulista, Marília, 2016. Disponível em: [https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/Davanzo\\_l\\_me\\_mar.pdf](https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/Davanzo_l_me_mar.pdf) Acesso em: 21 jan. 2023.

GESTEIRA, A. M. R. B. **Organização da informação em contexto dinâmico e sensível: tesouro de inteligência policial**. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/186518>. Acesso em: 06 fev. 2023.

GOTTSCHALG-DUQUE, C. Sirilico uma proposta para um sistema de recuperação de informação baseado em teorias da lingüística computacional e Ontologia. **Perspectivas em Ciência da Informação**, v. 10, n. 2, 2005. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/35761>. Acesso em: 06 jan. 2023.

*INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 25964-1: Information and documentation: thesauri and interoperability with other vocabularies: part 1:thesauri for information retrieval*. Genebra: ISO, 2011.

*INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 25964-2: Information and documentation: Thesauri and interoperability with other vocabularies, Part 2: Interoperability with other vocabularies*. Genebra: ISO, 2013.

JESUS, R. C. **Construção e manutenção de tesouros brasileiros: análise na perspectiva das normas internacionais**. 2018. 84 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista / Faculdade de Filosofia e Ciências, Marília, 2018. Disponível em: <http://hdl.handle.net/11449/155854>. Acesso em 22 jan. 2023.

LANCASTER F. W. **Indexação e resumos: teoria e prática**. 2. ed. Brasília: Briquet de Lemos. 2004.

LIBRELOTTO, G. R.; RAMALHO, J.C.; HENRIQUES, P.R.; **Representação de conhecimento na semantic web.**, repositorium.sdum.uminho.pt, 2005.

MENDONÇA, E. S. A linguística e a ciência da informação: estudos de uma interseção. **Ciência da Informação**, v. 29, n. 3, 2000. DOI: 10.18225/ci.inf..v29i3.873 Acesso em: 03 jan. 2023.

MOREIRA, W; FUJITA, M. S. L.; DAVANZO, L, PIOVEZAN, L. B. **Vocabulário controlado para a representação documental em arquivos correntes da UNESP**. In: SEMINÁRIO CIENTÍFICO ARQUIVOLOGIA E BIBLIOTECONOMIA, 4., 2015, Marília, SP. Anais...Marília: UNESP, 2015. Disponível em: <https://www.marilia.unesp.br/Home/Eventos/2015/seminariodearquivologiaebiblioteconomia/moreiraw.-fujita-m.s.l.-davanzo-l.-piovezan-l.b.pdf> Acesso em: 04 jan. 2023.

RAJAB ASAAD, R.; MASOUD ABDULHAKIM, R. *The Concept of Data Mining and Knowledge Extraction Techniques*. **Qubahan Academic Journal**, [S. l.], v. 1, n. 2, p. 17–20, 2021. DOI: 10.48161/qaj.v1n2a43. Disponível em: <https://journal.qubahan.com/index.php/qaj/article/view/43>. Acesso em: 6 fev. 2023.

RAUTENBERG, S., TODESCO, J. L., STEIL, A. V., GAUTHIER, F. A. O. Uma metodologia para o desenvolvimento de Ontologias. **RECEN - Revista Ciências Exatas e Naturais**, vol. 10, n. 2, 2008, pp. 237–262. Disponível em: <https://revistas.unicentro.br/index.php/RECEN/article/view/711>. Acesso em: 01 fev. 2023.

REZENDE, S. O. **Mineração de Dados**. São Leopoldo, 2005. Disponível em: <http://www.addlabs.uff.br/enia%20site/dw/mineracaodedados.pdf>. Acesso em: 29 jun. 2023.

SALES, R.; CAFÉ, L. Diferenças entre tesouros e Ontologias. **Perspectivas em Ciência da Informação**, v. 14, n. 1, p. 99-116, 2009. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/35503>. Acesso em: 01 fev. 2023.

SANTOS, T. H. N. A taxonomia e a folksonomia na representação da informação de fotografias. **Perspectivas em Ciência da Informação**, v. 23, n. 1, p. 89-103, 2018. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/35063> . Acesso em: 27 out. 2023.

SARACEVIC, T. *Interdisciplinary nature of information science*. **Ciência da Informação**, Brasília, v.24, n.1, p.36-41, 1995.

SMIT, Johanna W. Ilhemina; KOBASHI, Nair Yumiko. **Como elaborar vocabulário controlado para aplicação em arquivos**. São Paulo: Arquivo do Estado, Imprensa Oficial, 2003. 56 p. correto

SOUSA, R. T. B.; ARAÚJO JÚNIOR, R. H. A classificação e a taxonomia como instrumentos efetivos para a recuperação da informação arquivística. **Ciência da Informação**, v. 42, n. 1, 2013. DOI: 10.18225/ci.inf.v42i1.1400 Acesso em: 02 fev. 2023.

OLIVEIRA, D. A.; ARAÚJO, R. F. Construção de linguagens documentárias em sistemas de recuperação da informação: a importância da garantia do usuário. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 17, n. 34, p. 17-30, 2012. DOI: 10.5007/1518-2924.2012v17n34p17 Acesso em: 06 fev. 2023.

ORTEGA, C. D. Surgimento e consolidação da documentação: subsídios para compreensão da história da ciência da informação no Brasil. **Perspectivas em Ciência da Informação**, n. Especial, p. 59-79, 2009. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/101806>. Acesso em: 06 fev. 2023.

VELHO, J. A. (org.) **Tratado de Computação Forense**. Millennium, Campinas, 2016.

UNISIST, Princípios de indexação. **Revista da Escola de Biblioteconomia da UFMG**, v. 10, n. 1, 1981. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/73723>. Acesso em: 06 fev. 2023.

XU, G., WU, Z., LI, G. *et al.* *Improving contextual advertising matching by using Wikipedia thesaurus knowledge*. **Knowl Inf Syst** 43, 599–631 (2015). Disponível em: <https://doi.org/10.1007/s10115-014-0745-z> Acesso em: 22 fev. 2023.

## APÊNDICE A – LISTA DOS TERMOS CONSOLIDADOS NO PROJETO DE PESQUISA

Fase	Relevância		Termo	Descrição	Classe Gramatical	Sinônimos
Mineração	Alta		ILEGAL	Contrário à lei ou proibido.	adjetivo	
Mineração	Alta		CRIMINOSA	Descreve atividades ou comportamentos ilegais.	adjetivo	
Mineração	Alta		PENAIS	Ligado a punições e ao sistema penal, frequentemente usado em contextos de julgamento ou detenção por crimes de drogas.	adjetivo	
Mineração	Alta		ELETRÔNICO	Relacionado a dispositivos ou meios digitais, muitas vezes usado em contextos de cybercrime ou transações digitais.	adjetivo	
Mineração	Alta		ARMADAS	Relacionado a grupos armados envolvidos no tráfico ou crime organizado.	adjetivo	
mineração	Alta		JUDICIAL	Relacionado ao sistema de justiça e processos judiciais.	adjetivo	
mineração	Alta		POLICIAL	Relacionado às forças da ordem ou investigações.	adjetivo	
mineração	Alta		CRIMINAL	Relativo a atos ilegais ou atividades criminosas.	adjetivo	
mineração	Alta		FINANCEIRO	Relativo a dinheiro ou gestão de ativos.	adjetivo	
mineração	Alta		PENITENCIÁRIO	Relativo a prisões ou sistemas de detenção.	adjetivo	
mineração	Alta		GRAVE	Sério ou crítico, muitas vezes usado para descrever crimes ou situações.	adjetivo	
especialista	Alta		NARCOTRÁFICO	Comércio ilícito de drogas.	Substantivo	Traficância
especialista	Alta		171	Estelionato ou enganação. Referência ao código penal brasileiro.	Substantivo	estelionatário; fraudador; picareta; pilantra; golpista; rato; malandro
especialista	Alta		LAVAGEM DE DINHEIRO	Processo de tornar dinheiro ilícito em "limpo".	Substantivo	Lavagem
especialista	Alta		CANA	Prisão ou detenção policial.	Substantivo	Tranca; Sufoco
especialista	Alta		DESINTOXICAÇÃO	Processo de limpeza do organismo de substâncias tóxicas ou drogas.	Substantivo	
especialista	Alta		USUÁRIO	Pessoa que consome drogas.	Substantivo	Nóia; Zé droguinha; Viciado; Bico; peba; cliente; craquento; rato; farinha; fuminho; doído; vagabundo; dependente; noiado; playboy
especialista	Alta		GERENTE DO TRÁFICO	Pessoa que controla um ponto de venda de drogas.	Substantivo	Chefe; comandante; patrão; dono; cabeça; frente
especialista	Alta		X9	Pessoa que entrega ou informa sobre atividades ilícitas.	Substantivo	Dedoduro; Informante
especialista	Alta		PEÃO	Pessoa que executa ordens, geralmente referindo-se a membros inferiores no tráfico.	Substantivo	
especialista	Alta		TRAFICANTE	Pessoa que vende drogas ilícitas.	Substantivo	Narco; Fornecedor; chefe; chegado; meliante; bandido
especialista	Alta		PADRINHO	Pode se referir a um protetor ou líder no mundo do crime.	Substantivo	protetor
especialista	Alta		TIRA	Policial.	Substantivo	Policial
especialista	Alta		MACONHA	Droga depressiva derivada da planta Cannabis sativa.	Substantivo	Dendê; Prensado; beck; beque; verde; erva; mato; fumo; camarão; baseado; preto; capim

especialista	Alta		CRACK	Droga derivada da cocaína, consumida em forma de pedra.	Substantivo	Pedra; Zulu; mármore; amarela; brita; rapadura
especialista	Alta		COCAÍNA	Droga estimulante derivada da planta de coca.	Substantivo	Pó; Farinha; Hollywood; Coca; branquinha; peixe; escama; brilho; mistura; branca; açúcar
especialista	Alta		BIQUEIRA	Lugar onde drogas são vendidas.	Substantivo	Fim de linha; Boca de fumo
especialista	Alta		DÍVIDA	Montante que alguém deve, frequentemente relacionado ao tráfico.	Substantivo	pendura; débito; calote
especialista	Alta		CARGA	Quantidade significativa de droga destinada à venda.	Substantivo	Droga
mineração	Alta		DETENÇÃO	Ato de reter ou prender alguém.	substantivo	
mineração	Alta		EXTORSÃO	Ato de obter algo, especialmente dinheiro, através de ameaças ou violência.	substantivo	
mineração	Alta		SEQUESTRO	Ato de reter alguém contra sua vontade, muitas vezes por resgate.	substantivo	
mineração	Alta		DELITOS	Crimes ou ofensas.	substantivo	Crimes
mineração	Alta		BUSCAS	Ação de procurar ou pesquisar.	substantivo	
mineração	Alta		ARMAMENTO	Armas em geral.	substantivo	
mineração	Alta		IDENTIDADE	Características ou informação que distinguem uma pessoa.	substantivo	
mineração	Alta		DOCUMENTAÇÃO	Conjunto de documentos, muitas vezes relacionados à identidade ou atividade criminal.	substantivo	
mineração	Alta		EXPLOSIVO	Materiais que podem detonar.	substantivo	artefato; bananas; bomba; c4; foguete; dinamite
mineração	Alta		MUNIÇÃO	Material usado em armas para atirar.	substantivo	tiro; bala; chumbo; pipoco; caroço; aço; ponta; caramelo; cartucho; azeitona
mineração	Alta		MONITORAMENTO	Observação ou rastreamento contínuo.	substantivo	
mineração	Alta		MANDADOS	Ordens judiciais.	substantivo	Pedido
mineração	Alta		PISTOLA	Tipo de arma curta.	substantivo	quadrada; peça; arma curta; canhão; ferro; berro; preta; cromada; cano; ferramenta
mineração	Alta		FUZIL	Tipo de arma longa.	substantivo	AK; peça pesada; bico; bicudo; longa; fura
mineração	Alta		DENÚNCIA	Alegações ou acusações, muitas vezes de atividade ilegal.	substantivo	
mineração	Alta		VIGILÂNCIA	Ato de observar ou monitorar.	substantivo	
mineração	Alta		SUSPEITA	Dúvida ou desconfiança.	substantivo	
mineração	Alta		CONFRONTO	Encontro violento, muitas vezes entre forças policiais e criminosos.	substantivo	
mineração	Alta		POLÍCIAS	Forças ou agências encarregadas de manter a ordem e aplicar a lei.	substantivo	
mineração	Alta		QUADRILHAS	Grupos criminosos.	substantivo	Bando; gangue
mineração	Alta		INQUÉRITO	Investigação ou exame oficial.	substantivo	
mineração	Alta		INTEGRANTE	Membro de um grupo.	substantivo	soldado; funcionário; faccionado; irmã; mano; parça; batizado; chegado; membro

mineração	Alta		ALVO	Objetivo ou foco específico.	substantivo	
mineração	Alta		INDIVÍDUO	Pessoa ou entidade.	substantivo	
mineração	Alta		VÍTIMA	Pessoa que sofreu algum dano ou prejuízo.	substantivo	
mineração	Alta		TESTEMUNHA	Pessoa que vê ou tem conhecimento de um evento.	substantivo	cagueta; x9; declarante
mineração	Alta		RISCO	Possibilidade de perigo ou dano.	substantivo	
mineração	Alta		ADVOGADO	Profissional que representa ou aconselha em questões legais.	substantivo	
mineração	Alta		JUSTIÇA	Sistema legal ou equidade.	substantivo	
mineração	Alta		APREENSÕES	Captura ou confisco de algo.	substantivo	
mineração	Alta		SUBSTÂNCIA	Material específico, muitas vezes referindo-se a drogas.	substantivo	Droga
tesauros	Alta		CONLUIO	Acordo secreto entre partes para fins ilícitos.	Substantivo	Acordo
tesauros	Alta		ASSOCIAÇÃO PARA O TRÁFICO	Agrupamento de indivíduos com o objetivo de praticar atividades de tráfico de drogas.	Substantivo	
tesauros	Alta		RECEPTAÇÃO	Ato de adquirir ou possuir algo que foi obtido através de um crime.	Substantivo	
tesauros	Alta		LESÃO CORPORAL	Ato de causar dano físico a alguém.	Substantivo	
tesauros	Alta		DESACATO	Ato de desrespeitar autoridades ou funcionários públicos.	Substantivo	
tesauros	Alta		FRAUDE	Ato de enganar ou trapacear para obter benefício.	Substantivo	
tesauros	Alta		AMEAÇA	Ato de expressar a intenção de causar dano ou prejuízo a alguém.	Substantivo	chantagem; pressão; enquadro; esculacho; aviso; jurado; decretado; ideia; susto
tesauros	Alta		CONSTRANGIMENTO ILEGAL	Ato de forçar alguém a fazer algo contra sua vontade usando ameaça ou violência.	Substantivo	
tesauros	Alta		HOMICÍDIO	Ato de matar outro ser humano.	Substantivo	lona; caixão; passamento; CPF cancelado
tesauros	Alta		TRÁFICO DE DROGAS	Ato de produzir, vender ou distribuir drogas ilegais.	Substantivo	Tráfico; correria; corre; caminhada; atividade; comércio
tesauros	Alta		TRÁFICO DE PESSOAS	Ato de recrutar, transportar ou hospedar pessoas para exploração.	Substantivo	
tesauros	Alta		EXTORSÃO MEDIANTE SEQUESTRO	Ato de sequestrar alguém e exigir pagamento em troca da libertação.	Substantivo	Sequestro
tesauros	Alta		FURTO	Ato de tomar posse de algo sem permissão, sem usar violência.	Substantivo	155; ganho; mão
tesauros	Alta		ASSALTO	Ato de usar violência ou ameaça para roubar alguém.	Substantivo	
tesauros	Alta		CRIME HEDIONDO	Categoria de crimes considerada especialmente grave pela legislação.	Substantivo	
tesauros	Alta		LEI DE TÓXICOS	Conjunto de leis que regulam a produção, venda e consumo de drogas.	Substantivo	
tesauros	Alta		CRIME ANTECEDENTE	Crime anterior que fundamenta a acusação de lavagem de dinheiro.	Substantivo	Ficha
tesauros	Alta		CRIME QUALIFICADO	Crime com agravantes que aumentam a pena.	Substantivo	

tesauros	Alta		CRIME DOLOSO CONTRA A VIDA	Crime onde há intenção de matar.	Substantivo	
tesauros	Alta		CRIME INAFIANÇÁVEL	Crime para o qual não se pode conceder fiança.	Substantivo	
tesauros	Alta		CRIME DE DANO	Crime que resulta em dano à propriedade alheia.	Substantivo	
tesauros	Alta		CRIME CONTRA O PATRIMÔNIO	Crimes que afetam bens, como roubo e furto.	Substantivo	
tesauros	Alta		CRIME CONTRA A PESSOA	Crimes que afetam diretamente a pessoa, como homicídio e lesão corporal.	Substantivo	
tesauros	Alta		CRIME CONTRA A VIDA	Crimes que resultam em morte ou risco de morte.	Substantivo	
tesauros	Alta		REDUÇÃO A CONDIÇÃO ANÁLOGA À DE ESCRAVO	Explorar alguém em condições degradantes ou jornadas exaustivas.	Substantivo	
tesauros	Alta		AÇÃO CONTROLADA	Operação policial que permite que um ato ilícito se complete para identificar todos os envolvidos.	Substantivo	
tesauros	Alta		PRINCÍPIO DO NON OLET	Princípio legal que diz que o dinheiro não tem cheiro, não importando sua origem ilícita.	Substantivo	
tesauros	Alta		CÁRCERE PRIVADO	Restrição da liberdade de alguém sem autorização legal.	Substantivo	
tesauros	Alta		PRISÃO	Restrição da liberdade de um indivíduo.	Substantivo	Cana
tesauros	Alta		ESTUPEFACIENTE	Substância que pode causar estupor ou diminuir a sensibilidade.	Substantivo	
tesauros	Alta		POSSE DE ENTORPECENTES	Ter em sua posse substâncias proibidas ou controladas.	Substantivo	28; vapor; mocosado; bolso cheio
tesauros	Alta		RECLUSÃO	Tipo de pena privativa de liberdade.	Substantivo	
tesauros	Alta		ASSOCIAÇÃO CRIMINOSA	União de duas ou mais pessoas para a prática de crimes.	Substantivo	
tesauros	Alta		FRAUDE PROCESSUAL	Uso de meios ilícitos para interferir ou alterar um processo judicial.	Substantivo	
tesauros	Alta		COLABORAÇÃO PREMIADA	Acordo entre a justiça e o réu para fornecer informações em troca de benefícios legais.	Substantivo	
tesauros	Alta		COLEGIADO	Grupo de indivíduos que tomam decisões conjuntas.	Substantivo	
tesauros	Alta		GRUPO ARMADO CIVIL	Grupo de pessoas não militares que possuem armas.	Substantivo	
tesauros	Alta		BALANÇA DE PRECISÃO	Instrumento usado para pesar substâncias, frequentemente usado no tráfico para dosagem de drogas.	Substantivo	
tesauros	Alta		ESQUEMA CRIMINOSO	Plano ou operação realizada para cometer crimes.	Substantivo	Esquema
tesauros	Alta		GRAU DE ENVOLVIMENTO	Extensão da participação de um indivíduo em uma atividade ou crime.	Substantivo	
tesauros	Alta		LÍDER DE ORGANIZAÇÃO	Indivíduo que lidera ou dirige uma organização, incluindo organizações criminosas.	Substantivo	Líder

tesauros	Alta		AGENTE INFILTRADO	Pessoa que se infiltra em uma organização ou grupo para obter informações ou evidências.	Substantivo	
tesauros	Alta		FLAGRANTE DELITO	Situação em que uma pessoa é pega no momento em que está cometendo um crime.	Substantivo	Flagrante; flafu; BO; bronca
tesauros	Alta		COMANDO VERMELHO	Uma das principais facções criminosas do Brasil, envolvida no tráfico de drogas.	Substantivo	CV; falange vermelha; comando;
tesauros	Alta		PRIMEIRO COMANDO DA CAPITAL	Uma das principais facções criminosas do Brasil.	Substantivo	PCC; partido; 15.3.3;comenado; família
tesauros	Alta		PERÍCIA	Análise técnica realizada por especialistas em determinado campo.	Substantivo	
tesauros	Alta		ALCALOIDE	Classe de compostos orgânicos, muitos dos quais são usados como drogas psicoativas.	Substantivo	
tesauros	Alta		CLORETO DE ETILA	Composto químico volátil usado ocasionalmente como inalante recreativo.	Substantivo	
tesauros	Alta		USO DE ENTORPECENTES	Consumo de substâncias proibidas ou controladas.	Substantivo	
tesauros	Alta		DEPENDÊNCIA TOXICOLÓGICA	Dependência causada por substâncias tóxicas.	Substantivo	
tesauros	Alta		LSD	Droga alucinógena sintética.	Substantivo	ácido; doce; adesivo; figurinha; selo; selinho
tesauros	Alta		ÓPIO	Droga analgésica derivada da papoula.	Substantivo	
tesauros	Alta		DROGA SINTÉTICA	Droga fabricada através de processos químicos, não derivada de plantas.	Substantivo	química; larica; rebite; droga do amor; azul; doce; bala; gringa; meta; cristal; balinha; comprimido; adesivo
tesauros	Alta		HEROÍNA	Droga opioide derivada do ópio.	Substantivo	pico; injetável; picada
tesauros	Alta		ECSTASE	Droga sintética que altera a percepção e o humor.	Substantivo	bala; doce; azul; pílula
tesauros	Alta		FAMILIA DO NORTE	Facção criminosa que opera no norte do Brasil.	Substantivo	FDN
tesauros	Alta		DEPENDÊNCIA FÍSICA	Necessidade do corpo por uma substância após uso contínuo.	Substantivo	
tesauros	Alta		DEPENDÊNCIA QUÍMICA	Necessidade geral por uma substância devido à uso contínuo.	Substantivo	
tesauros	Alta		DEPENDÊNCIA PSÍQUICA	Necessidade mental ou emocional por uma substância.	Substantivo	
tesauros	Alta		COCA	Planta da qual é extraída a cocaína.	Substantivo	
tesauros	Alta		LANÇA PERFUME	Solução de éter e clorofórmio, muitas vezes inalada como droga.	Substantivo	Éter; loló; cheirinho; lança; LP
tesauros	Alta		PSICOTRÓPICO	Substância que afeta o funcionamento da mente.	Substantivo	
tesauros	Alta		DROGA	Substância que altera o funcionamento normal do corpo ou mente.	Substantivo	tóxico; barato; coisa; coisinha; bagulho; parada; caixa; mercadoria; entorpecente; produto; carga; material
tesauros	Alta		SKUNK	Variedade potente de maconha.	Substantivo	camarão; flor; solta
Formulário	-		REVÓLVER	Arma letal	Substantivo	oitão; ferro; berro; ferramenta; canela seca; pau de fogo; chuteira; calçado
especialista	Alta		TRAFICAR	Ação de vender drogas ilegalmente.	Verbo	Passar; Vender; fornecer; descolar; empurrar; subir; descer

mineração	Alta		INDICIAR	Acusar formalmente de um crime.	verbo	
mineração	Alta		TOMAR	Adquirir, assumir controle. Pode se referir ao controle de territórios por gangues.	verbo	Controlar
mineração	Alta		ATUAR	Agir, ter um papel. Pode se referir à como um indivíduo age dentro de uma organização criminosa.	verbo	
mineração	Alta		ACUSAR	Alegar que alguém cometeu um delito.	verbo	apontar; botar na reta; caguetar; dedar; denunciar; delatar; dar o serviço; incriminar; indicar
mineração	Alta		ABORDAR	Aproximar-se com um propósito. Em polícia, pode indicar ação de parar e questionar.	verbo	
mineração	Alta		PROCURAR	Buscar, tentar encontrar. Pode se referir a busca por suspeitos ou informações.	verbo	
mineração	Alta		INFORMAR	Comunicar, notificar. Em investigações, usado para relatar descobertas ou situações.	verbo	
mineração	Alta		SEGUIR	Continuar, acompanhar. Em contextos policiais, pode indicar vigilância ou perseguição.	verbo	
mineração	Alta		ENCONTRAR	Descobrir, deparar-se com algo. Usado em investigações para indicar descobertas.	verbo	
mineração	Alta		PRENDER	Detiver, colocar sob custódia. Refere-se ao ato policial de detenção.	verbo	grampear; pegar; cair; ganchar; segurar
mineração	Alta		FUGIR	Escapar, evadir. Pode se referir a suspeitos evitando a detenção.	verbo	picar a mula; meter o pé; vazar; sumir; desaparecer; dar o pinote
mineração	Alta		VERIFICAR	Examinar, assegurar-se. Pode se referir a verificar a autenticidade de informações ou provas.	verbo	
mineração	Alta		INVESTIGAR	Examinar, pesquisar. Usualmente se refere ao processo policial de busca de provas.	verbo	
mineração	Alta		FAZER	Executar, realizar uma ação. No contexto de crime, pode se referir a produção ou venda de drogas.	verbo	Matar; passar; empurrar; concluir; dar sumiço; apagar; zerar; cancelar; resolver; quebrar; assassinar
mineração	Alta		ENVOLVER	Incluir, implicar. Pode indicar participação em atividades criminosas.	verbo	
mineração	Alta		APURAR	Investigar, determinar a verdade.	verbo	
mineração	Alta		ACOMPANHAR	Ir junto, observar. Pode se referir a monitoramento de suspeitos.	verbo	Vigiar; observar
mineração	Alta		VER	Observar, perceber. Em investigações, pode se referir ao ato de testemunhar algo.	verbo	
mineração	Alta		SOLICITAR	Pedir, requerer. Pode indicar um pedido formal por informações ou apoio.	verbo	
mineração	Alta		CONTINUAR	Persistir, seguir. Pode se referir a investigações ou operações em andamento.	verbo	
mineração	Alta		TER	Possuir, manter. No contexto de drogas, pode indicar posse de substâncias ilegais.	verbo	possuir; tá na mão
mineração	Alta		PRATICAR	Realizar, executar. Pode indicar a prática de atividades criminosas.	verbo	
mineração	Alta		IDENTIFICAR	Reconhecer, nomear. Usado em investigações para reconhecer suspeitos ou provas.	verbo	

mineração	Alta		APREENDER	Tomar posse, confiscar. Usualmente se refere a apreensão de drogas ou bens.	verbo	
mineração	Alta		LEVAR	Transportar algo. Em tráfico, pode indicar transporte de substâncias ilegais.	verbo	Portar1; transportar; trazer; carregar; conduzir; fretar; descer; subir

## APÊNDICE B - TERMOS CONSOLIDADOS APÓS MINERAÇÃO DE TEXTO

Relevância	Termo	Classe Gramatical	Descrição
Alta	armadas	adjetivo	Relacionado a grupos armados envolvidos no tráfico ou crime organizado.
Alta	criminal	adjetivo	Relativo a atos ilegais ou atividades criminosas.
Alta	criminosa	adjetivo	Descreve atividades ou comportamentos ilegais.
Alta	eletrônico	adjetivo	Relacionado a dispositivos ou meios digitais, muitas vezes usado em contextos de cybercrime ou transações digitais.
Alta	financeiro	adjetivo	Relativo a dinheiro ou gestão de ativos.
Alta	grave	adjetivo	Sério ou crítico, muitas vezes usado para descrever crimes ou situações.
Alta	ilegal	adjetivo	Contrário à lei ou proibido.
Alta	judicial	adjetivo	Relacionado ao sistema de justiça e processos judiciais.
Alta	penais	adjetivo	Ligado a punições e ao sistema penal, frequentemente usado em contextos de julgamento ou detenção por crimes de drogas.
Alta	penitenciário	adjetivo	Relativo a prisões ou sistemas de detenção.
Alta	policial	adjetivo	Relacionado às forças da ordem ou investigações.
Alta	advogado	substantivo	Profissional que representa ou aconselha em questões legais.
Alta	alvo	substantivo	Objetivo ou foco específico.
Alta	ameaças	substantivo	Indicações ou possibilidades de perigo ou dano.
Alta	apreensões	substantivo	Captura ou confisco de algo.
Alta	armamento	substantivo	Armas em geral.
Alta	assaltos	substantivo	Atos de violência com objetivo de roubo.
Alta	buscas	substantivo	Ação de procurar ou pesquisar.
Alta	confronto	substantivo	Encontro violento, muitas vezes entre forças policiais e criminosos.
Alta	delitos	substantivo	Crimes ou ofensas.
Alta	denúncia	substantivo	Alegações ou acusações, muitas vezes de atividade ilegal.
Alta	detenção	substantivo	Ato de reter ou prender alguém.
Alta	documentação	substantivo	Conjunto de documentos, muitas vezes relacionados à identidade ou atividade criminal.
Alta	esquema	substantivo	Plano ou estratégia, muitas vezes ilegal.
Alta	explosivos	substantivo	Materiais que podem detonar.
Alta	extorsão	substantivo	Ato de obter algo, especialmente dinheiro, através de ameaças ou violência.
Alta	facções	substantivo	Grupos, muitas vezes referindo-se a grupos criminosos organizados.
Alta	fuzis	substantivo	Tipo de arma longa.
Alta	identidade	substantivo	Características ou informação que distinguem uma pessoa.
Alta	indivíduo	substantivo	Pessoa ou entidade.
Alta	inquérito	substantivo	Investigação ou exame oficial.
Alta	integrante	substantivo	Membro de um grupo.
Alta	justiça	substantivo	Sistema legal ou equidade.
Alta	líderes	substantivo	Indivíduos no comando ou que dirigem grupos.
Alta	maconha	substantivo	Droga geralmente ilícita, obtida da planta Cannabis.
Alta	mandados	substantivo	Ordens judiciais.
Alta	monitoramento	substantivo	Observação ou rastreamento contínuo.
Alta	munição	substantivo	Material usado em armas para atirar.
Alta	pistolas	substantivo	Tipo de arma curta.
Alta	polícias	substantivo	Forças ou agências encarregadas de manter a ordem e aplicar a lei.
Alta	prisões	substantivo	Locais onde pessoas são mantidas sob custódia ou o ato de prender alguém.
Alta	quadrilhas	substantivo	Grupos criminosos.
Alta	risco	substantivo	Possibilidade de perigo ou dano.
Alta	sequestro	substantivo	Ato de reter alguém contra sua vontade, muitas vezes por resgate.
Alta	substância	substantivo	Material específico, muitas vezes referindo-se a drogas.
Alta	suspeita	substantivo	Dúvida ou desconfiança.
Alta	testemunha	substantivo	Pessoa que vê ou tem conhecimento de um evento.
Alta	tráfico	substantivo	Comércio ilegal, muitas vezes referindo-se a drogas.
Alta	vigilância	substantivo	Ato de observar ou monitorar.
Alta	vítima	substantivo	Pessoa que sofreu algum dano ou prejuízo.
Alta	abordar	verbo	Aproximar-se com um propósito. Em polícia, pode indicar ação de parar e questionar.
Alta	acompanhar	verbo	Ir junto, observar. Pode se referir a monitoramento de suspeitos.
Alta	acusar	verbo	Alegar que alguém cometeu um delito.
Alta	apreender	verbo	Tomar posse, confiscar. Usualmente se refere a apreensão de drogas ou bens.
Alta	apurar	verbo	Investigar, determinar a verdade.
Alta	atuar	verbo	Agir, ter um papel. Pode se referir à como um indivíduo age dentro de uma organização criminosa.
Alta	continuar	verbo	Persistir, seguir. Pode se referir a investigações ou operações em andamento.
Alta	encontrar	verbo	Descobrir, deparar-se com algo. Usado em investigações para indicar descobertas.
Alta	envolver	verbo	Incluir, implicar. Pode indicar participação em atividades criminosas.
Alta	fazer	verbo	Executar, realizar uma ação. No contexto de crime, pode se referir a produção ou venda de drogas.
Alta	fugir	verbo	Escapar, evadir. Pode se referir a suspeitos evitando a detenção.
Alta	identificar	verbo	Reconhecer, nomear. Usado em investigações para reconhecer suspeitos ou provas.
Alta	indiciar	verbo	Acusar formalmente de um crime.
Alta	informar	verbo	Comunicar, notificar. Em investigações, usado para relatar descobertas ou situações.
Alta	investigar	verbo	Examinar, pesquisar. Usualmente se refere ao processo policial de busca de provas.
Alta	levar	verbo	Transportar algo. Em tráfico, pode indicar transporte de substâncias ilegais.
Alta	praticar	verbo	Realizar, executar. Pode indicar a prática de atividades criminosas.
Alta	prender	verbo	Detiver, colocar sob custódia. Refere-se ao ato policial de detenção.
Alta	procurar	verbo	Buscar, tentar encontrar. Pode se referir a busca por suspeitos ou informações.
Alta	seguir	verbo	Continuar, acompanhar. Em contextos policiais, pode indicar vigilância ou perseguição.
Alta	solicitar	verbo	Pedir, requerer. Pode indicar um pedido formal por informações ou apoio.

Alta	ter	verbo	Possuir, manter. No contexto de drogas, pode indicar posse de substâncias ilegais.
Alta	tomar	verbo	Adquirir, assumir controle. Pode se referir ao controle de territórios por gangues.
Alta	ver	verbo	Observar, perceber. Em investigações, pode se referir ao ato de testemunhar algo.
Alta	verificar	verbo	Examinar, assegurar-se. Pode se referir a verificar a autenticidade de informações ou provas.
Média	comercial	adjetivo	Relativo a comércio, pode ser usado em contextos de tráfico ou venda.
Média	administrativa	adjetivo	Relacionado a procedimentos ou funções administrativas, muitas vezes em contextos governamentais ou de organização.
Média	comercial	adjetivo	Pertencente ou relacionado ao comércio.
Média	econômica	adjetivo	Relativo à economia ou finanças.
Média	estadual	adjetivo	Pertencente ou relativo a um estado específico.
Média	estrangeiros	adjetivo	Originário de ou pertencente a outro país.
Média	federal	adjetivo	Pertencente ou relativo à nação como um todo.
Média	fiscal	adjetivo	Relativo a impostos ou taxas.
Média	grande	adjetivo	De extenso tamanho ou importância.
Média	internacional	adjetivo	Relativo a ou envolvendo dois ou mais países.
Média	irregular	adjetivo	Não seguindo um padrão ou norma; muitas vezes usado em contextos de atividades ilegais.
Média	legal	adjetivo	De acordo com a lei.
Média	maior	adjetivo	Superior em tamanho ou quantidade.
Média	municipal	adjetivo	Pertencente ou relativo a uma cidade ou município.
Média	nacional	adjetivo	Pertencente ou relativo a uma nação.
Média	operacional	adjetivo	Relativo à operação ou execução de tarefas.
Média	político	adjetivo	Relacionado a política ou governo.
Média	possíveis	adjetivo	Capazes de ocorrer ou se tornar realidade. Pode ser usado para avaliar cenários ou suspeitas em investigações.
Média	principal	adjetivo	Mais importante ou proeminente.
Média	regional	adjetivo	Relativo a uma área geográfica específica.
Média	responsável	adjetivo	Que tem a obrigação de responder por algo ou alguém.
Média	acompanhamento	substantivo	Monitorização ou seguimento de algo ou alguém.
Média	administração	substantivo	Gestão ou controle de algo.
Média	aparelhos	substantivo	Equipamentos ou dispositivos.
Média	área	substantivo	Região ou espaço específico.
Média	autos	substantivo	Documentos oficiais, geralmente relacionados a processos.
Média	auxílio	substantivo	Ajuda ou assistência.
Média	batalhão	substantivo	Unidade militar.
Média	capacidade	substantivo	Habilidade ou potencial para algo.
Média	carteira	substantivo	Documento de identificação ou objeto para guardar tais documentos.
Média	cartuchos	substantivo	Recipiente que contém munição.
Média	cenário	substantivo	Ambiente ou contexto.
Média	complexo	substantivo	Algo composto de muitas partes.
Média	condições	substantivo	Estados ou circunstâncias.
Média	conduta	substantivo	Comportamento ou maneira de agir.
Média	consumo	substantivo	Uso ou ingestão de algo.
Média	conteúdo	substantivo	O que está contido ou incluído em algo.
Média	coordenação	substantivo	Organização ou gestão de várias partes ou elementos.
Média	dano	substantivo	Prejuízo ou lesão.
Média	determinação	substantivo	Decisão ou resolução.
Média	disparo	substantivo	Ação de atirar com uma arma.
Média	disposição	substantivo	Arranjo ou organização.
Média	efetivo	substantivo	Força real ou concreta.
Média	espaço	substantivo	Lugar ou região.
Média	exame	substantivo	Análise ou avaliação.
Média	exercício	substantivo	Prática ou realização de uma atividade.
Média	exterior	substantivo	Regiões ou países fora de um ponto de referência.
Média	favela	substantivo	Comunidade geralmente carente.
Média	força	substantivo	Potência ou vigor.
Média	formação	substantivo	Criação ou desenvolvimento de algo.
Média	fornecimento	substantivo	Ato de fornecer ou entregar algo.
Média	funcionário	substantivo	Pessoa empregada por uma organização.
Média	fundos	substantivo	Recursos financeiros.
Média	gerente	substantivo	Indivíduo responsável por administrar ou dirigir.
Média	guarnição	substantivo	Força militar ou policial em uma determinada localidade.
Média	imóvel	substantivo	Propriedade que não pode ser movida.
Média	inciso	substantivo	Divisão ou subseção de um documento legal.
Média	instituições	substantivo	Organizações ou entidades com propósitos específicos.
Média	liderança	substantivo	Ação ou posição de liderar.
Média	ligações	substantivo	Conexões ou relacionamentos.
Média	localidade	substantivo	Área ou região específica.
Média	membro	substantivo	Parte de um grupo ou organização.
Média	mercadorias	substantivo	Itens ou produtos.
Média	moradores	substantivo	Pessoas que residem em um lugar específico.
Média	moto	substantivo	Veículo motorizado de duas rodas.
Média	movimentação	substantivo	Ação ou processo de mover ou ser movido.
Média	negócios	substantivo	Atividades comerciais ou transações.
Média	nível	substantivo	Grau ou extensão.
Média	núcleo	substantivo	Centro ou parte essencial.
Média	patrulhamento	substantivo	Ação de vigiar ou monitorar uma área.

Média	perícia	substantivo	Exame técnico ou especializado.
Média	política	substantivo	Estratégias ou diretrizes.
Média	presídio	substantivo	Prisão ou instituição correcional.
Média	problema	substantivo	Questão ou dificuldade.
Média	procedimentos	substantivo	Métodos ou etapas para fazer algo.
Média	produção	substantivo	Criação ou fabricação de algo.
Média	proprietário	substantivo	Pessoa que possui algo.
Média	proteção	substantivo	Ação de proteger ou ser protegido.
Média	regime	substantivo	Sistema ou método de governo.
Média	relatório	substantivo	Descrição ou registro detalhado.
Média	resistência	substantivo	Oposição ou luta contra algo.
Média	responsabilidade	substantivo	Obrigação ou dever.
Média	resultados	substantivo	Conclusões ou produtos de um esforço.
Média	rota	substantivo	Caminho ou direção.
Média	ruas	substantivo	Vias públicas em uma cidade ou vila.
Média	técnica	substantivo	Método ou procedimento.
Média	telefones	substantivo	Dispositivos para comunicação à distância.
Média	território	substantivo	Área ou região.
Média	trabalhos	substantivo	Esforços ou tarefas.
Média	tratamento	substantivo	Maneira de lidar ou cuidar de algo.
Média	tribunal	substantivo	Corte ou órgão de julgamento.
Média	verdade	substantivo	Realidade ou fato comprovado.
Média	viaturas	substantivo	Veículos, muitas vezes referindo-se a veículos policiais.
Média	afirmar	verbo	Declarar, assegurar. Pode se referir a declarações durante investigações.
Média	apontar	verbo	Indicar, mostrar. Pode se referir a identificação de suspeitos ou locais.
Média	apresentar	verbo	Mostrar, indicar. Usado em contextos legais para apresentar provas ou suspeitos.
Média	atender	verbo	Responder, cuidar. Em situações policiais, pode se referir a atender uma ocorrência.
Média	conduzir	verbo	Levar, guiar. Pode se referir a levar um suspeito sob custódia.
Média	conseguir	verbo	Obter sucesso em algo. Em investigações, pode indicar obtenção de provas.
Média	constatar	verbo	Confirmar, observar como verdadeiro.
Média	dar	verbo	Conceder, fornecer. Pode se referir ao ato de fornecer informações ou drogas.
Média	deixar	verbo	Permitir, abandonar. Pode indicar o ato de soltar um suspeito ou deixar uma cena.
Média	determinar	verbo	Decidir, estabelecer. Em contextos legais, pode se referir a dar ordens.
Média	dizer	verbo	Comunicar, expressar algo. Em contextos de investigação, pode se referir a declarações de suspeitos.
Média	efetuar	verbo	Realizar, concluir. Em investigações, pode se referir a prisões ou operações realizadas.
Média	encaminhar	verbo	Direcionar, enviar. Em contextos legais, pode se referir a enviar um caso para julgamento.
Média	entrar	verbo	Ingressar em algum lugar. Pode indicar a entrada em áreas de venda ou produção de drogas.
Média	estar	verbo	Indica estado ou condição. Pode ser usado para descrever a condição de um suspeito ou local.
Média	ficar	verbo	Permanecer, estar em algum estado. Pode se referir ao estado de alerta ou detenção de um suspeito.
Média	haver	verbo	Existir, ocorrer. Pode se referir à existência de provas ou suspeitos.
Média	ir	verbo	Deslocar-se de um lugar para outro. Pode indicar deslocamento para locais de venda de drogas.
Média	manter	verbo	Sustentar, continuar com algo. Em detenções, pode indicar manter alguém sob custódia.
Média	passar	verbo	Mover-se, transmitir. Pode se referir a transação de drogas ou informações.
Média	poder	verbo	Ter capacidade ou autoridade. Em contextos legais, pode se referir a autorizações.
Média	realizar	verbo	Concretizar, fazer acontecer. Pode se referir a operações policiais.
Média	receber	verbo	Obter, aceitar. Pode se referir a recebimento de informações, drogas ou subornos.
Média	sair	verbo	Deixar um lugar. Pode se referir a suspeitos deixando a cena do crime.
Média	ser	verbo	Indica identidade ou característica. Usado para classificar indivíduos ou substâncias em investigações.
Média	tentar	verbo	Esforçar-se, procurar. Pode se referir a tentativas de detenção ou obtenção de informações.
Média	trabalhar	verbo	Exercer função, atuar. Pode se referir a atividades dentro do crime organizado.
Média	tratar	verbo	Lidar, discutir. Pode se referir a negociações ou tratamento de informações.
Média	viver	verbo	Existir, levar a vida. Em contextos criminosos, pode se referir a estilo de vida ou atividades diárias.
Baixa	atual	adjetivo	Do momento presente; pode ser usado para se referir a situações ou tendências contemporâneas.
Baixa	civil	adjetivo	Relativo a cidadãos, em oposição a militar ou policial.
Baixa	constante	adjetivo	Ocorrendo frequentemente ou sem interrupção.
Baixa	corporal	adjetivo	Relativo ao corpo humano.
Baixa	corrente	adjetivo	Que está acontecendo agora; comum.
Baixa	funcional	adjetivo	Relativo à função ou operação de algo.
Baixa	geral	adjetivo	Abrangente ou comum.
Baixa	local	adjetivo	Relativo a uma área ou região específica.
Baixa	necessário	adjetivo	Indispensável ou essencial.
Baixa	peçoal	adjetivo	Pertencente ou relativo a um indivíduo.
Baixa	restrito	adjetivo	Limitado, geralmente referindo-se a informações ou áreas de acesso controlado.
Baixa	rural	adjetivo	Relativo ao campo ou áreas não urbanas.
Baixa	âmbito	substantivo	Alcance ou extensão.
Baixa	capital	substantivo	Cidade principal ou recursos financeiros.
Baixa	imagem	substantivo	Representação visual.
Baixa	interesse	substantivo	Atenção ou curiosidade sobre algo.
Baixa	maneira	substantivo	Método ou forma de fazer algo.
Baixa	sociedade	substantivo	Comunidade ou grupo de pessoas.

## APÊNDICE C – CÓDIGO FONTE PART1.PY

```

import os
import re
import csv
import glob
import string
import chardet
import nltk
from nltk.corpus import stopwords
from collections import defaultdict

nltk.download('stopwords')
stop_words = set(stopwords.words('portuguese'))

# Função para abrir arquivos com a codificação correta
def open_file_with_correct_encoding(file_path):
    with open(file_path, "rb") as file:
        result = chardet.detect(file.read())
        encoding = result["encoding"]
        return open(file_path, "r", encoding=encoding)

# Função para ler termos de um arquivo
def read_terms(file_path):
    with open_file_with_correct_encoding(file_path) as file:
        terms = [term.strip().lower() for term in file.readlines()]
    return terms

# Função para processar um diretório e identificar arquivos com termos de interesse e suas frequências
def process_directory(directory_path, terms):
    files_with_terms = defaultdict(list)
    term_freq_results = []

    for file_name in glob.glob(os.path.join(directory_path, "*")):
        with open_file_with_correct_encoding(file_name) as file:
            content = file.read().lower()

            for term in terms:
                term = term.translate(str.maketrans("", "", string.punctuation))
                freq = content.count(term)

                if freq > 0:
                    files_with_terms[file_name].append(term)
                    term_freq_results.append((file_name, term, freq))

    return files_with_terms, term_freq_results

# Função para escrever os resultados em arquivos CSV
def write_csv_files(files_with_terms, term_freq_results):
    with open("nomes_arquivos_interesse.csv", "w", encoding="utf-8", newline="") as csvfile:
        csvwriter = csv.writer(csvfile)
        for file_name in files_with_terms:
            csvwriter.writerow([os.path.basename(file_name)])

    with open("resultados_termos_interesse.csv", "w", encoding="utf-8", newline="") as csvfile:
        csvwriter = csv.writer(csvfile)
        csvwriter.writerow(["nome arquivo", "termo de interesse", "frequencia"])
        for row in term_freq_results:
            csvwriter.writerow([os.path.basename(row[0]), row[1], row[2]])

# Função principal que lê os termos de interesse, processa o diretório e escreve os resultados em arquivos CSV
def main(directory_path, term_file_path):
    terms = read_terms(term_file_path)
    files_with_terms, term_freq_results = process_directory(directory_path, terms)
    write_csv_files(files_with_terms, term_freq_results)

if __name__ == "__main__":
    directory_path = "C:/temp/mestrado/PesquisaTermoArquivos/arquivos"
    script_dir = os.path.dirname(os.path.abspath(__file__))
    term_file_path = os.path.join(script_dir, "termos_de_interesse.txt")
    main(directory_path, term_file_path)

```

## APÊNDICE D– CÓDIGO FONTE PART2.PY

```

import os
import re
import csv
import string
import chardet
import spacy
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import CountVectorizer
from collections import defaultdict
from unidecode import unidecode

stop_words = set(stopwords.words('portuguese'))
nlp = spacy.load('pt_core_news_sm')

def is_in_spacy_vocab(word):
    return nlp.vocab.has_vector(word)

def is_in_custom_vocab(word, custom_vocab):
    return word in custom_vocab

# Adicione a função para obter a classe gramatical da palavra
def get_pos_tag(word):
    doc = nlp(word)
    return doc[0].pos_

def open_file_with_correct_encoding(file_path):
    with open(file_path, "rb") as file:
        result = chardet.detect(file.read())
        encoding = result["encoding"]
        return open(file_path, "r", encoding=encoding)

# Lê as stopwords personalizadas de um arquivo
def read_stopwords(file_path):
    with open(file_path, "r", encoding="utf-8") as file:
        custom_stopwords = [word.strip().lower() for word in file.readlines()]
    return custom_stopwords

# Pré-processa o texto removendo pontuação e espaços extras
def preprocess_text(text):
    text = unidecode(text) # Adicione esta linha para remover acentos e caracteres especiais
    text = re.sub(r'\s+', ' ', text)
    text = text.translate(str.maketrans("", "", string.punctuation)).lower()
    return text

# Processa os arquivos de interesse e calcula a contagem de termos e TF-IDF
def process_files(directory_path, interest_files, stopword_list):
    documents = []
    for file_name in interest_files:
        file_path = os.path.join(directory_path, file_name)
        with open_file_with_correct_encoding(file_path) as file:
            content = file.read()
            preprocessed_content = preprocess_text(content)
            documents.append((file_name, preprocessed_content))

    count_vectorizer = CountVectorizer(stop_words=list(stopword_list))
    count_matrix = count_vectorizer.fit_transform([doc[1] for doc in documents])

    tfidf_vectorizer = TfidfVectorizer(stop_words=list(stopword_list))
    tfidf_matrix = tfidf_vectorizer.fit_transform([doc[1] for doc in documents])
    terms = tfidf_vectorizer.get_feature_names_out()

    return documents, terms, count_matrix, tfidf_matrix, tfidf_vectorizer

def write_mined_terms_csv(documents, terms, count_matrix, tfidf_matrix):
    with open("resultados_termos_minerados.csv", "w", encoding="utf-8", newline="") as csvfile:
        csvwriter = csv.writer(csvfile)
        csvwriter.writerow(["arquivo de interesse", "termo minerado", "frequencia do termo no arquivo (TF-IDF)", "frequencia simples"])

    for i, (file_name, _) in enumerate(documents):
        count_row = count_matrix.getrow(i)
        tfidf_row = tfidf_matrix.getrow(i).toarray()[0]
        nonzero_indices = count_row.nonzero()[1]
        for idx in nonzero_indices:
            term = terms[idx]
            count_freq = count_row[0, idx]
            tfidf_freq = tfidf_row[idx]

```

```

        csvwriter.writerow([os.path.basename(file_name), term, tfidf_freq, count_freq])

# Escreve as estatísticas dos termos minerados em um arquivo CSV
def write_term_stats_tfidf_csv(terms, count_matrix, tfidf_matrix, vectorizer, custom_vocab):
    term_count_freq = count_matrix.sum(axis=0)
    term_tfidf_corpus = tfidf_matrix.sum(axis=0).A1
    term_in_files_count = (count_matrix > 0).sum(axis=0)

    with open("termos_minerados_tfidf.csv", "w", encoding="utf-8", newline=") as csvfile:
        csvwriter = csv.writer(csvfile)
        # Adicione a coluna 'classe gramatical' e 'no vocabulário personalizado' no cabeçalho
        csvwriter.writerow(["termo minerado", "qtd de arquivos presente termo minerado", "tfidf corpus", "frequencia simples nos arquivos - somatorio",
"classe gramatical", "no vocabulário personalizado"])

        for term, in_files_count, tfidf_corpus, count_freq in zip(terms, term_in_files_count, term_tfidf_corpus, term_count_freq):
            if is_in_custom_vocab(term, custom_vocab): # Verifica se o termo está no custom_vocab
                pos_tag = get_pos_tag(term)
                in_custom_vocab = True # Já sabemos que o termo está no custom_vocab
                csvwriter.writerow([term, in_files_count, tfidf_corpus, count_freq, pos_tag, in_custom_vocab])
                print(f"Termo inserido: {term}")

# Escreve as estatísticas dos termos minerados em um arquivo CSV
def write_term_statistics_csv(terms, tfidf_matrix, vectorizer):
    term_freq = tfidf_matrix.sum(axis=0).A1
    term_data = []

    for term, freq, idf in zip(terms, term_freq, vectorizer.idf_):
        tfidf = freq * idf
        term_data.append((term, freq, idf, tfidf))

    with open("estatisticas_termos_minerados.csv", "w", encoding="utf-8", newline=") as csvfile:
        csvwriter = csv.writer(csvfile)
        csvwriter.writerow(["termo", "frequencia nos arquivos de interesse", "IDF", "TF-IDF"])

    for row in term_data:
        csvwriter.writerow(row)

def read_interest_files(interest_files_csv):
    with open(interest_files_csv, "r", encoding="utf-8") as file:
        reader = csv.reader(file)
        interest_files = [row[0] for row in reader]
    return interest_files

def read_portuguese_dictionary(file_path):
    with open(file_path, "r", encoding="utf-8") as file:
        custom_dictionary = {word.strip().lower() for word in file.readlines()}
    return custom_dictionary

# Adicione prints na função main_part2
def main_part2(directory_path, interest_files_csv, stopword_file_path, custom_vocab_path):
    custom_vocab = read_portuguese_dictionary(custom_vocab_path)
    stopword_list = set(stopwords.words('portuguese') + read_stopwords(stopword_file_path))
    interest_files = read_interest_files(interest_files_csv)
    documents, terms, count_matrix, tfidf_matrix, vectorizer = process_files(directory_path, interest_files, stopword_list)
    write_mined_terms_csv(documents, terms, count_matrix, tfidf_matrix)
    print("Escrevendo arquivo termos_minerados_tfidf.csv") # Adicione esta linha
    write_term_stats_tfidf_csv(terms, count_matrix, tfidf_matrix, vectorizer, custom_vocab)
    write_term_statistics_csv(terms, tfidf_matrix, vectorizer)

if __name__ == "__main__":
    directory_path = "C:/temp/mestrado/PesquisaTermoArquivos/arquivos/"
    interest_files_csv = "C:/Users/bispo.fdb/OneDrive - Polícia Federal/PCF Bispo/mestrado/python/tfidf/nomes_arquivos_interesse.csv"
    stopword_file_path = "C:/Users/bispo.fdb/OneDrive - Polícia Federal/PCF Bispo/mestrado/python/tfidf/stopWords_ampliado_pt.txt"
    custom_vocab_file_path = "C:/Users/bispo.fdb/OneDrive - Polícia Federal/PCF Bispo/mestrado/python/tfidf/br-sem-acentos.txt" # Substitua pelo
caminho do arquivo br-sem-acentos.txt
    main_part2(directory_path, interest_files_csv, stopword_file_path, custom_vocab_file_path)

```

## APÊNDICE E – CÓDIGO FONTE CLASSEGRAMATICAL.PY

```

import csv
import spacy

nlp = spacy.load('pt_core_news_sm')

def read_dictionary(file_path):
    with open(file_path, "r", encoding="utf-8") as file:
        dictionary_words = file.read().splitlines() # Assumes one word per line
    return set(dictionary_words) # Use a set for faster lookup

def read_mined_terms(file_path):
    print(f"Lendo termos minerados de {file_path}...")
    with open(file_path, "r", encoding="utf-8") as file:
        reader = csv.reader(file)
        next(reader) # Skip the header
        mined_terms = [row[1] for row in reader] # Assumes the term is in the second column
    print(f"Primeiros 10 termos: {mined_terms[:10]}")
    return mined_terms

def get_grammatical_info(term):
    doc = nlp(term)
    if len(doc) > 1:
        return None # Skip phrases, we want individual words
    token = doc[0]
    return token.pos_ # Return its Part-Of-Speech tag

def process_terms(mined_terms, dictionary_words):
    print("Processando termos minerados...")
    valid_terms = {}
    for term in mined_terms:
        if term not in dictionary_words:
            print(f"O termo '{term}' não está no dicionário e será ignorado.")
            continue
        pos = get_grammatical_info(term)
        if pos: # If the term is valid and has a POS tag
            valid_terms[term] = pos
            print(f"Termo: {term}, Classe Gramatical: {pos}")
    print(f"Processamento completo. Encontrados {len(valid_terms)} termos válidos.")
    return valid_terms

def write_valid_terms_to_csv(valid_terms_file, valid_terms):
    with open(valid_terms_file, "w", encoding="utf-8", newline="") as file:
        writer = csv.writer(file)
        writer.writerow(["termo", "classe gramatical"]) # Write the header
        for term, pos in valid_terms.items():
            writer.writerow([term, pos])

def main(dictionary_file, mined_terms_file, valid_terms_file):
    dictionary_words = read_dictionary(dictionary_file)
    mined_terms = read_mined_terms(mined_terms_file)
    valid_terms = process_terms(mined_terms, dictionary_words)
    write_valid_terms_to_csv(valid_terms_file, valid_terms)

if __name__ == "__main__":
    dictionary_file = "br-sem-acentos.txt"
    mined_terms_file = "resultados_termos_minerados.csv"
    valid_terms_file = "resultados_termos_minerados_portugues.csv"
    main(dictionary_file, mined_terms_file, valid_terms_file)

```

## APÊNDICE F – IMAGENS DO FORMULÁRIO ELETRÔNICO



**Programa de Pós-Graduação em Ciência da Informação**

Seção 1 de 7

## Validação e Enriquecimento de Vocabulário Controlado em Tráfico de Entorpecentes

Prezado(a) Especialista,

Sou o Fabrício Bispo, policial federal e mestrando em Ciência da Informação pela UFSC (Universidade Federal de Santa Catarina). Estou conduzindo um estudo dedicado à construção de um vocabulário controlado específico para o domínio de crimes de tráfico de entorpecentes. A aspiração é que, eventualmente, os frutos deste trabalho possam ser incorporados em sistemas de busca textual de órgãos públicos, aprimorando a recuperação de informações vitais neste segmento.

O trabalho que realizei até o momento compreende uma pesquisa em tesouros online, mineração de texto em conjuntos de dados substanciais e consultas diretas a especialistas na área. A sua expertise é fundamental para validar, complementar e enriquecer este acervo de termos.

Para facilitar sua contribuição, dividi os termos em **5 categorias**:

1. Substantivos - Organizações, Papéis e Pessoas;
2. Substantivos - Crimes e Atividades Ilegais;
3. Substantivos - Substâncias, Drogas e Efeitos;
4. Verbos - Ações e Processos;
5. Adjetivos e Substantivos - Características e Materiais.

Este agrupamento tem por objetivo simplificar o processo de preenchimento e revisão. **Estima-se que o tempo de resposta do formulário seja entre 11 a 15 minutos.**

Dentro do formulário, será solicitado que você valide os termos apresentados, sugira sinônimos quando achar pertinente e, caso identifique alguma lacuna, adicione novos termos que julgue relevantes. Sua colaboração será essencial para a eficácia e precisão deste projeto, beneficiando uma gama de profissionais e pesquisadores no campo.

Agradeço profundamente por dedicar parte do seu tempo a este estudo e por contribuir com sua valiosa experiência.

Atenciosamente,

**Fabrício Bispo**  
[fabricao.bispo@gmail.com](mailto:fabricao.bispo@gmail.com)

---

Este formulário está coletando automaticamente os e-mails de todos os participantes. [Alterar configurações](#)

**Em qual órgão você atualmente atua? \***

Polícia Federal

Polícia Civil

Polícia Militar

Ministério Público

Outros...

**Experiência em Investigação de Crimes de Tráfico de Entorpecentes \***

Menos de 1 ano

1-5 anos

6-10 anos

Mais de 10 anos

Seção 2 de 7

**1. Substantivos - Organizações, Papeis e Pessoas**

Descrição (opcional)

**1.1. Enriquecimento de Vocabulário Controlado em Tráfico de Entorpecentes**Liste todos os **sinônimos e gírias** que você conhece para do termos abaixo, separando cada um por vírgula(OPCIONAL) Conheça sinônimo ou gíria para o termo **PRIMEIRO COMANDO DA CAPITAL**: PCC ?

Texto de resposta curta

:::

(OPCIONAL) Conheça sinônimo ou gíria para o termo **COMANDO VERMELHO**: CV?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **FAMILIA DO NORTE**: FDN ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **FLAGRANTE DELITO**: Flagrante ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **GERENTE DO TRÁFICO**: Chefe ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **INTEGRANTE**: ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **PADRINHO**: ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **TESTEMUNHA**: ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **TRAFICANTE**: Narco, Fornecedor ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **ADVOGADO**: ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **USUÁRIO**: Nôia, Zé droguinha, Viciado, Bico?

Texto de resposta curta

## 1.2. Validação de termos de Vocabulário Controlado em Tráfico de Entorpecentes

Descrição (opcional)

Dentro do contexto criminal, especialmente relacionado ao tráfico de entorpecentes e atividades ilícitas, por favor, **marque os termos que você NÃO considera relevantes para construção de vocabulário controlado.**

	Não relevante
AGENTE INFILTRADO: - Pessoa que se infiltra em u...	<input type="radio"/>
CONFRONTO: - Encontro violento, muitas vezes ent...	<input type="radio"/>
DENÚNCIA: - Alegações ou acusações, muitas veze...	<input type="radio"/>
GRAU DE ENVOLVIMENTO: - Extensão da participaç...	<input type="radio"/>
INDIVÍDUO: - Pessoa ou entidade.	<input type="radio"/>
INQUÉRITO: - Investigação ou exame oficial.	<input type="radio"/>
JUSTIÇA: - Sistema legal ou equidade.	<input type="radio"/>
LÍDER DE ORGANIZAÇÃO: Líder - Indivíduo que lider...	<input type="radio"/>
ORGANIZAÇÃO CRIMINOSA: Facção, ORCRIM - Org...	<input type="radio"/>
POLÍCIAS: - Forças ou agências encarregadas de m...	<input type="radio"/>
PEÃO: - Pessoa que executa ordens, geralmente ref...	<input type="radio"/>
QUADRILHAS: Bando, gangue - Grupos criminosos.	<input type="radio"/>
RISCO: - Possibilidade de perigo ou dano.	<input type="radio"/>
SUSPEITA: - Dúvida ou desconfiança.	<input type="radio"/>
TIRA: Policial - Policial.	<input type="radio"/>
VIGILÂNCIA: - Ato de observar ou monitorar.	<input type="radio"/>
X9:Dedo-duro, Informante - Pessoa que entrega ou i...	<input type="radio"/>

(OPCIONAL) Em relação aos **termos relevantes acima** quer adicionar sinônimos?

Modelos de resposta: *Termo chave: Sinônimos*

**X9: Dedo-duro, Informante**

**GERENTE: Chefe**

Texto de resposta longa

Após a seção 2 Continuar para a próxima seção

Seção 3 de 7

**2. Substantivos - Crimes e atividades ilegais**

Descrição (opcional)

**2.1. Enriquecimento de Vocabulário Controlado em Tráfico de Entorpecentes**Liste todos os **sinônimos e gírias** que você conhece para do termos abaixo, separando cada um por vírgula

(OPCIONAL) Conhece sinônimo ou gíria para o termo 171: estelionatário, fraudador?

Texto de resposta curta  
.....(OPCIONAL) Conhece sinônimo ou gíria para o termo **AMEAÇA**?Texto de resposta curta  
.....(OPCIONAL) Conhece sinônimo ou gíria para o termo **FURTO** ?Texto de resposta curta  
.....(OPCIONAL) Conhece sinônimo ou gíria para o termo **HOMICÍDIO**?Texto de resposta curta  
.....(OPCIONAL) Conhece sinônimo ou gíria para o termo **POSSE DE ENTORPECENTES**?Texto de resposta curta  
.....(OPCIONAL) Conhece sinônimo ou gíria para o termo **TRÁFICO DE DROGAS**: TráficoTexto de resposta curta  
.....**2.2. Validação de termos de Vocabulário Controlado em Tráfico de Entorpecentes**

Descrição (opcional)

Dentro do contexto criminal, especialmente relacionado ao tráfico de entorpecentes e atividades ilícitas, por favor, **marque os termos que você NÃO considera relevantes para construção de vocabulário controlado.**

	Não relevante
AÇÃO CONTROLADA: - Operação policial que permi...	<input type="radio"/>
ARMAMENTO: - Armas em geral.	<input type="radio"/>
ASSALTO: - Ato de usar violência ou ameaça para r...	<input type="radio"/>
ASSOCIAÇÃO CRIMINOSA: - União de duas ou mais...	<input type="radio"/>
ASSOCIAÇÃO PARA O TRÁFICO: - Agrupamento de ...	<input type="radio"/>
CÁRCERE PRIVADO: - Restrição da liberdade de alg...	<input type="radio"/>
CONSTRANGIMENTO ILEGAL: - Ato de forçar algué...	<input type="radio"/>
CRIME CONTRA A PESSOA: - Crimes que afetam dir...	<input type="radio"/>
CRIME CONTRA A VIDA: - Crimes que resultam em ...	<input type="radio"/>
CRIME CONTRA O PATRIMÔNIO: - Crimes que afeta...	<input type="radio"/>
CRIME DE DANO: - Crime que resulta em dano à pr...	<input type="radio"/>
CRIME DOLOSO CONTRA A VIDA: - Crime onde há i...	<input type="radio"/>
CRIME HEDIONDO: - Categoria de crimes considera...	<input type="radio"/>
CONLUIO: Acordo - Acordo secreto entre partes par...	<input type="radio"/>
CRIME ANTECEDENTE: Ficha - Crime anterior que f...	<input type="radio"/>
CRIME INAFIANÇÁVEL: - Crime para o qual não se ...	<input type="radio"/>
CRIME QUALIFICADO: - Crime com agravantes que ...	<input type="radio"/>
DELITOS: Crimes - Crimes ou ofensas.	<input type="radio"/>
DESACATO: - Ato de desrespeitar autoridades ou fu...	<input type="radio"/>
ESTUPEFACIENTE: - Substância que pode causar e...	<input type="radio"/>
EXTORSÃO: - Ato de obter algo, especialmente dinh...	<input type="radio"/>
EXTORSÃO MEDIANTE SEQUESTRO: Sequestro - At...	<input type="radio"/>
FRAUDE: - Ato de enganar ou trapacear para obter ...	<input type="radio"/>
FRAUDE PROCESSUAL: - Uso de meios ilícitos para ...	<input type="radio"/>
LAVAGEM DE DINHEIRO: Lavagem - Processo de to...	<input type="radio"/>
LEI DE TÓXICOS: - Conjunto de leis que regulam a p...	<input type="radio"/>
LESÃO CORPORAL: - Ato de causar dano físico a al...	<input type="radio"/>
NARCOTRÁFICO: Tráficação - Comércio ilícito de dr...	<input type="radio"/>
RECEPÇÃO: - Ato de adquirir ou possuir algo que...	<input type="radio"/>
RECLUSÃO: - Tipo de pena privativa de liberdade.	<input type="radio"/>
REDUÇÃO A CONDIÇÃO ANÁLOGA À DE ESCRAVO: ...	<input type="radio"/>
TRÁFICO DE PESSOAS: - Ato de recrutar, transporta...	<input type="radio"/>
PRINCÍPIO DO NON OLET: - Princípio legal que diz ...	<input type="radio"/>
PRISÃO: Cans - Restrição da liberdade de um indivi...	<input type="radio"/>

(OPCIONAL) Em relação aos **termos acima** quer adicionar sinônimos?

Modelos de resposta: *Termo chave: Sinônimos*

**Prisão: Cana, Jaula**

**171: estelionatário, fraudador**

Texto de resposta longa

Após a seção 3 Continuar para a próxima seção

Seção 4 de 7

### 3. Substantivos - Substâncias, Drogas e Efeitos

Descrição (opcional)

#### 3.1. Enriquecimento de Vocabulário Controlado em Tráfico de Entorpecentes

Liste todos os **sinônimos e gírias** que você conhece para do termos abaixo, separando cada um por vírgula

(OPCIONAL) Conhece sinônimo ou gíria para o termo **COCAÍNA**: Pó, Farinha, Hollywood, Coca ?

Texto de resposta curta

(OPCIONAL) Conhece sinônimo ou gíria para o termo **CRACK**: Pedra, Zulu ?

Texto de resposta curta

(OPCIONAL) Conhece sinônimo ou gíria para o termo **DÍVIDA**: ?

Texto de resposta curta

(OPCIONAL) Conhece sinônimo ou gíria para o termo **DROGA**: ?

Texto de resposta curta

(OPCIONAL) Conhece sinônimo ou gíria para o termo **DROGA SINTÉTICA**: ?

Texto de resposta curta

(OPCIONAL) Conhece sinônimo ou gíria para o termo **ECSTASE**: ?

Texto de resposta curta

(OPCIONAL) Conhece sinônimo ou gíria para o termo **HEROÍNA**: ?

Texto de resposta curta

(OPCIONAL) Conhece sinônimo ou gíria para o termo **LANÇA PERFUME**: Éter ?

Texto de resposta curta

(OPCIONAL) Conhece sinônimo ou gíria para o termo **LSD**: ?

Texto de resposta curta

(OPCIONAL) Conhece sinônimo ou gíria para o termo **MACONHA**: Dendê, Prensado ?

Texto de resposta curta

(OPCIONAL) Conhece sinônimo ou gíria para o termo **SKUNK** : ?

Texto de resposta curta

### 3.2. Validação de termos de Vocabulário Controlado em Tráfico de Entorpecentes

Descrição (opcional)

Dentro do contexto criminal, especialmente relacionado ao tráfico de entorpecentes e atividades ilícitas, por favor, **marque os termos que você NÃO considera relevantes para construção de vocabulário controlado.**

	Não relevante
ALCALOIDE: - Classe de compostos orgânicos, mui...	<input type="radio"/>
APREENSÕES: - Captura ou confisco de algo.	<input type="radio"/>
BIQUEIRA: Fim de linha, Boca de fumo - Lugar onde ...	<input type="radio"/>
CARGA: Droga - Quantidade significativa de droga d...	<input type="radio"/>
CLORETO DE ETILA: - Composto químico volátil us...	<input type="radio"/>
DEPENDÊNCIA FÍSICA: - Necessidade do corpo por ...	<input type="radio"/>
DEPENDÊNCIA PSÍQUICA: - Necessidade mental ou...	<input type="radio"/>
DEPENDÊNCIA QUÍMICA: - Necessidade geral por u...	<input type="radio"/>
DEPENDÊNCIA TOXICOLÓGICA: - Dependência cau...	<input type="radio"/>
ÓPIO: - Droga analgésica derivada da papoula.	<input type="radio"/>
PERÍCIA: - Análise técnica realizada por especialist...	<input type="radio"/>
PSICOTRÓPICO: - Substância que afeta o funciona...	<input type="radio"/>
SUBSTÂNCIA: Droga - Material específico, muitas ve...	<input type="radio"/>
USO DE ENTORPECENTES: - Consumo de substânc...	<input type="radio"/>

(OPCIONAL) Em relação aos **termos relevantes acima** quer adicionar sinônimos?

Modelos de resposta: *Termo chave: Sinônimos*

**Substância: Droga**

**Crack: Pedra, Zulu**

Texto de resposta longa

Seção 5 de 7

#### 4. Verbos - Ações e Processos

Descrição (opcional)

##### 4.1. Enriquecimento de Vocabulário Controlado em Tráfico de Entorpecentes

Liste todos os sinônimos e gírias que você conhece para do termos abaixo, separando cada um por vírgula

(OPCIONAL) Conheça sinônimo ou gíria para o termo **ACUSAR**: ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **FAZER**: Matar ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **FUGIR**: ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **LEVAR**: Portar, transportar, trazer ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **PRENDER**: ?

Texto de resposta curta

...

(OPCIONAL) Conheça sinônimo ou gíria para o termo **TER**: Possuir ?

Texto de resposta curta

(OPCIONAL) Conheça sinônimo ou gíria para o termo **TRAFICAR**: Passar, Vender?

Texto de resposta curta

Dentro do contexto criminal, especialmente relacionado ao tráfico de entorpecentes e atividades ilícitas, por favor, **marque os termos que você NÃO considera relevantes para construção de vocabulário controlado.**

	Não relevante
ABORDAR: - Aproximar-se com um propósito. Em p...	<input type="radio"/>
ACOMPANHAR: Vigiar, observar - Ir junto, observar. ...	<input type="radio"/>
APREENDER: - Tomar posse, confiscar. Usualmente...	<input type="radio"/>
APURAR: - Investigar, determinar a verdade.	<input type="radio"/>
ATUAR: - Agir, ter um papel. Pode se referir a como ...	<input type="radio"/>
CONTINUAR: - Persistir, seguir. Pode se referir a inv...	<input type="radio"/>
ENCONTRAR: - Descobrir, deparar-se com algo. Usa...	<input type="radio"/>
ENVOLVER: - Incluir, implicar. Pode indicar participa...	<input type="radio"/>
IDENTIFICAR: - Reconhecer, nomear. Usado em inv...	<input type="radio"/>
INDICIAR: - Acusar formalmente de um crime.	<input type="radio"/>
INFORMAR: - Comunicar, notificar. Em investigação...	<input type="radio"/>
INVESTIGAR: - Examinar, pesquisar. Usualmente se ...	<input type="radio"/>
PRATICAR: - Realizar, executar. Pode indicar a práti...	<input type="radio"/>
PROCURAR: - Buscar, tentar encontrar. Pode se refe...	<input type="radio"/>
SEGUIR: - Continuar, acompanhar. Em contextos po...	<input type="radio"/>
SOLICITAR: - Pedir, requerer. Pode indicar um pedid...	<input type="radio"/>
TOMAR: Controlar - Adquirir, assumir controle. Pode...	<input type="radio"/>
VER: - Observar, perceber. Em investigações, pode s...	<input type="radio"/>
VERIFICAR: - Examinar, assegurar-se. Pode se referi...	<input type="radio"/>
ARMADAS: - Relacionado a grupos armados envolv...	<input type="radio"/>

(OPCIONAL) Em relação aos **termos relevantes acima** quer adicionar sinônimos?

Modelos de resposta: *Termo chave: Sinônimos*

**Traficar: Passar, Vender**

**Ter: Possuir**

Texto de resposta longa

Seção 6 de 7

**5. Adjetivos e substantivos - Características e materiais** ✕ ⋮

Descrição (opcional)

---

**5.1. Enriquecimento de Vocabulário Controlado em Tráfico de Entorpecentes**

Liste todos os **sinônimos e gírias** que você conhece para do termos abaixo, separando cada um por vírgula

(OPCIONAL) Conhece sinônimo ou gíria para o termo **EXPLOSIVO** ?

Texto de resposta curta

---

(OPCIONAL) Conhece sinônimo ou gíria para o termo **MUNIÇÃO**: tiro, bala ?

Texto de resposta curta

---

(OPCIONAL) Conhece sinônimo ou gíria para o termo **PISTOLA** ?

Texto de resposta curta

---

(OPCIONAL) Conhece sinônimo ou gíria para o termo **REVÓLVER** ?

Texto de resposta curta

---

(OPCIONAL) Conhece sinônimo ou gíria para o termo **FUZIL** ?

Texto de resposta curta

---

**5.2. Validação de termos de Vocabulário Controlado em Tráfico de Entorpecentes**

Descrição (opcional)

Dentro do contexto criminal, especialmente relacionado ao tráfico de entorpecentes e atividades ilícitas, por favor, **marque os termos que você NÃO considera relevantes para construção de vocabulário controlado.**

	Não relevante
ARMADAS: - Relacionado a grupos armados envolv...	<input type="radio"/>
ARMAMENTO: - Armas em geral.	<input type="radio"/>
BALANÇA DE PRECISÃO: - Instrumento usado para ...	<input type="radio"/>
BUSCAS: - Ação de procurar ou pesquisar.	<input type="radio"/>
CANA:Tranca, Sufoco - Prisão ou detenção policial.	<input type="radio"/>
COLABORAÇÃO PREMIADA: - Acordo entre a justic...	<input type="radio"/>
COLEGIADO: - Grupo de indivíduos que tomam deci...	<input type="radio"/>
CRIMINAL: - Relativo a atos ilegais ou atividades cri...	<input type="radio"/>
CRIMINOSA: - Descreve atividades ou comportame...	<input type="radio"/>
DESINTOXICAÇÃO: - Processo de limpeza do organ...	<input type="radio"/>
DOCUMENTAÇÃO: - Conjunto de documentos, muit...	<input type="radio"/>
ESQUEMA CRIMINOSO:Esquema - Plano ou operaç...	<input type="radio"/>
FINANCEIRO: - Relativo a dinheiro ou gestão de ativ...	<input type="radio"/>
GRAVE: - Sério ou crítico, muitas vezes usado para ...	<input type="radio"/>
GRUPO ARMADO CIVIL: - Grupo de pessoas não mil...	<input type="radio"/>
IDENTIDADE: - Características ou informação que d...	<input type="radio"/>
ILEGAL: - Contrário à lei ou proibido.	<input type="radio"/>
JUDICIAL: - Relacionado ao sistema de justiça e pr...	<input type="radio"/>
MANDADOS: Pedido - Ordens judiciais.	<input type="radio"/>
MONITORAMENTO: - Observação ou rastreamento ...	<input type="radio"/>
PENAS: - Ligado a punições e ao sistema penal, fre...	<input type="radio"/>
PENITENCIÁRIO: - Relativo a prisões ou sistemas d...	<input type="radio"/>

(OPCIONAL) Em relação aos **termos acima** quer adicionar sinônimos?

Modelos de resposta: *Termo chave: Sinônimos*

**Cana: Prisão, Sufoco**

**Munição: Bala, Tiro**

Texto de resposta longa

Seção 7 de 7

**Coleta de novos termos** v ⋮

Descrição (opcional)

⋮

Valorizamos muito a sua contribuição em nos ajudar a criar um vocabulário controlado completo e preciso. Sabemos que cada pessoa tem uma perspectiva única e pode conhecer termos ou expressões que não capturamos.

**Por favor, se conhece algum termo ou expressão relevante que não foi mencionado até agora, compartilhe conosco!** Sua expertise é fundamental para aprimorar nosso trabalho. Modelos de resposta: [Termo chave](#); [Sinônimos](#) - [Descrição](#)

**CANA:Tranca, Sufoco - Prisão ou detenção policial.**

Texto de resposta longa

---

**Obrigado por dedicar seu tempo e por ajudar a melhorar nosso repositório de termos!**

A ideia é fazer o respondente sentir que sua contribuição é valiosa e única, incentivando-o a participar ativamente.

## ANEXO A – QUADRO DA BASE JURÍDICA PF

### Constituição Federal

### Código de Processo Penal

Lei nº 10.357, de 27 de dezembro de 2001 - Estabelece normas de controle e fiscalização sobre produtos químicos que direta ou indiretamente possam ser destinados à elaboração ilícita de substâncias entorpecentes, psicotrópicas ou que determinem dependência física ou psíquica, e dá outras providências.

Lei nº 10.826, de 22 de dezembro de 2003 - Dispõe sobre registro, posse e comercialização de armas de fogo e munição, sobre o Sistema Nacional de Armas – Sinarm, define crimes e dá outras providências.

Lei nº 13.445, de 24 de maio de 2017 - Institui a Lei de Migração.

Lei nº 13.460, de 26 de junho de 2017 - Dispõe sobre participação, proteção e defesa dos direitos do usuário dos serviços públicos da administração pública.

Lei nº 7.102, de 20 de junho de 1983 - Dispõe sobre segurança para estabelecimentos financeiros, estabelece normas para constituição e funcionamento das empresas particulares que exploram serviços de vigilância e de transporte de valores, e dá outras providências.

Lei nº 9.017, de 30 de março de 1995 - Estabelece normas de controle e fiscalização sobre produtos e insumos químicos que possam ser destinados à elaboração da cocaína em suas diversas formas e de outras substâncias entorpecentes ou que determinem dependência física ou psíquica, e altera dispositivos da Lei nº 7.102, de 20 de junho de 1983, que dispõe sobre segurança para estabelecimentos financeiros, estabelece normas para constituição e funcionamento de empresas particulares que explorem serviços de vigilância e de transporte de valores, e dá outras providências.

Decreto nº 4.262, de 10 de junho de 2002 - Regulamenta a Lei nº 10.357, de 27 de dezembro de 2001, que estabelece normas de controle e fiscalização sobre produtos químicos que direta ou indiretamente possam ser destinados à elaboração ilícita de substâncias entorpecentes, psicotrópicas ou que determinem dependência física ou psíquica, e dá outras providências.

Decreto nº 9.845, de 25 de junho de 2019 - Regulamenta a Lei nº 10.826, de 22 de dezembro de 2003, para dispor sobre a aquisição, o cadastro, o registro e a posse de armas de fogo e de munição.

Decreto nº 9.846, de 25 de junho de 2019 - Regulamenta a Lei nº 10.826, de 22 de dezembro de 2003, para dispor sobre o registro, o cadastro e a aquisição de armas e de munições por caçadores, colecionadores e atiradores.

Decreto nº 9.847, de 25 de junho de 2019 - Regulamenta a Lei nº 10.826, de 22 de dezembro de 2003, para dispor sobre a aquisição, o cadastro, o registro, o porte e a comercialização de armas de fogo e de munição e sobre o Sistema Nacional de Armas e o Sistema de Gerenciamento Militar de Armas.

[Decreto nº 9.199, de 20 de novembro de 2017](#) - Regulamenta a Lei nº 13.445, de 24 de maio de 2017, que institui a Lei de Migração.

[Decreto nº 1.983, de 14 de agosto de 1996](#) - Institui, no âmbito do Departamento de Polícia Federal do Ministério da Justiça e da Diretoria-Geral de Assuntos Consulares, Jurídicos e de Assistência a Brasileiros no Exterior do Ministério das Relações Exteriores, o Programa de Modernização, Agilização, Aprimoramento e Segurança da Fiscalização do Tráfego Internacional e do Passaporte Brasileiro (PROMASP), e aprova o Regulamento de Documentos de Viagem.

[Decreto nº 89.056, de 24 de novembro de 1983](#) - Regulamenta a Lei nº 7.102, de 20 de junho de 1983, que "dispõe sobre segurança para estabelecimentos financeiros, estabelece normas para constituição e funcionamento das empresas particulares que exploram serviços de vigilância e de transporte de valores e dá outras providências".

[Portaria nº 10-MJSP, de 16 de abril de 2019](#) - Estabelece normas e procedimentos para a implantação e funcionamento do Sistema de Controle e Fiscalização de Produtos Químicos (SIPROQUIM 2) no âmbito da Polícia Federal.

[Portaria nº 240-MJSP, de 12 de março de 2019](#) - Estabelece procedimentos para o controle e a fiscalização de produtos químicos e define os produtos químicos sujeitos a controle pela Polícia Federal.

[Portaria nº 577-MJSP, de 5 de junho de 2019](#) - Prorroga a data de entrada em vigor da Portaria nº 240, de 12 de março de 2019, que estabelece procedimentos para o controle e a fiscalização de produtos químicos e define os produtos químicos sujeitos a controle pela Polícia Federal.

[Portaria nº 155, de 27 de dezembro de 2018 \(Regimento Interno\)](#)

[Portaria DG/PF nº 16.145, de 26 de abril de 2022](#) - Define as circunscrições das superintendências regionais e das delegacias de Polícia Federal.

[Portaria DG/PF nº 16.797, de 10 de novembro de 2022](#) - Altera a Portaria DG/PF nº 16.145, de 26 de abril de 2022, que define as circunscrições das superintendências regionais e das delegacias de Polícia Federal.

[Instrução Normativa nº. 013/2005-DG/DPF, de 15 de Junho de 2005](#) - Define as competências específicas das unidades centrais e descentralizadas do Departamento de Polícia Federal e as atribuições de seus dirigentes. (Alterada pela Instrução Normativa Nº 39/2010-DG/DPF, de 5 de Outubro de 2010)

Portaria DG/PF nº 16.145, de 26 de abril de 2022