

LEVANTAMENTO TERMINOLÓGICO SOBRE CRIMES DE TRÁFICO DE ENTORPECENTES

TERMINOLOGICAL ANALYSIS ON NARCOTICS TRAFFICKING CRIMES

Luciane Paula Vital ^a
Fabrício Dantas Bispo ^b

RESUMO

Objetivo: levantar e selecionar termos relacionados a crimes de entorpecentes no âmbito da Polícia Federal do Brasil. A pesquisa se caracteriza como exploratória, pesquisa documental e estudo de caso quanto aos objetivos. **Metodologia:** Foram definidas quatro etapas metodológicas: Levantamento de termos em vocabulários controlados correlatos, Mineração dos termos em documentos oficiais da PF, Validação e ampliação dos termos com especialistas e Consolidação do levantamento terminológico. **Resultados:** foram identificados 90 termos relevantes em Sistemas de Organização do Conhecimento correlatos analisados. Na mineração de texto foram analisados aproximadamente 140 mil documentos formais da Polícia Federal, resultando em 151 termos relacionados ao tema. **Conclusão:** Ao final das etapas, consolidou-se 210 termos e seus sinônimos. Essa pesquisa é um primeiro passo no levantamento terminológico que irá culminar em um vocabulário controlado mais robusto. A interação direta com especialistas na temática enriqueceu o repertório terminológico e solidificou a importância da validação dos termos.

Descritores: Organização do Conhecimento. Vocabulário controlado. Crimes de tráfico.

1 INTRODUÇÃO

No ambiente digital, a diversidade de dados produzidos, incluindo dados estruturados originários de bancos de dados internos e externos às organizações, bem como dados não estruturados, contribui para o desafio do gerenciamento. Porém, da mesma forma que as tecnologias digitais impactam a

^a Doutora em Ciência da Informação da Universidade Federal de Santa Catarina (UFSC). Docente no Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Santa Catarina (UFSC), Florianópolis, Brasil, Email: luciane.vital@ufsc.br.

^b Mestre em Ciência da Informação pela Universidade Federal de Santa Catarina (UFSC). Florianópolis, Brasil, Email: bispo.fdb@pf.gov.br

produção de dados e informação, também são essenciais para lidar com eles. Essa pesquisa foi realizada no âmbito da Polícia Federal (PF) do Brasil que, considerando a escassez de pessoal, a necessidade de maior eficiência dos serviços prestados, o aumento das demandas dos órgãos governamentais e as dificuldades de tratamento e de processamento das informações, necessita de constante revisão e aperfeiçoamento de seus processos. Os órgãos públicos precisam investir 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 para responder às demandas sociais.

O controle terminológico é de suma importância para a indexação e recuperação de informações em grandes volumes de dados. Os Sistemas de Organização do Conhecimento (SOCs) são essenciais para a padronização dos termos utilizados, redução das ambiguidades e construção de relacionamentos semânticos em áreas específicas do conhecimento. A Polícia Federal possui uma ferramenta web de indexação que utiliza a tecnologia Microsoft Fast Enterprise Search e possibilita aos investigadores uma pesquisa textual única em todas as bases consolidadas na ferramenta, nos moldes dos grandes buscadores existentes na Internet. Porém, o usuário desta ferramenta, ao fazer uma busca mais específica ou complexa sobre um assunto, precisa realizar diversas pesquisas utilizando diferentes termos relacionados ao assunto desejado. Isso se deve pelo fato de que a ferramenta não possui controle terminológico implementado, gerando problemas na recuperação da informação e demora, comprometendo a resposta às demandas existentes na instituição.

O trabalho da Polícia Federal é voltado para diversas frentes de investigação, dentre as quais se destaca o combate aos crimes de tráfico de entorpecentes, que assume grande importância no contexto atual. No processo investigativo, os dados analisados são derivados de diferentes fontes e formatos, tornando o trabalho de análise complexo. O sucesso na extração e uso dos dados não depende somente da utilização da tecnologia, mas também dos conhecimentos prévios e da experiência dos investigadores que utilizam o sistema. Ou seja, dependendo do usuário pode-se ter resultados mais

superficiais ou resultados mais completos sobre um mesmo assunto. A elaboração de um SOC que englobe o registro do conhecimento empírico dos investigadores e seu controle terminológico, se mostra essencial para aumentar a efetividade das buscas textuais e análises de conexões entre as informações, bem como para a implementação de uma inteligência artificial adequada às ferramentas utilizadas.

Na construção de um SOC é preciso, inicialmente, o levantamento terminológico, termos e conceitos, sua validação, para posterior estabelecimento de relacionamentos semânticos, aumentando a complexidade envolvida. Nesse contexto, esta pesquisa tem como objetivo levantar e selecionar termos relacionados a crimes de entorpecentes no âmbito da Polícia Federal do Brasil.

2 VOCABULÁRIO CONTROLADO

Os SOC's 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. Conforme Moreira *et al.* (2015), no contexto dos estudos de Organização e Representação do Conhecimento (ORC), há diferentes tipos de SOC, como sistemas de classificação, tesouros, taxonomias e ontologias. Esses sistemas têm em comum o levantamento e o controle terminológico em áreas específicas do conhecimento humano. De acordo com Cruz, Ferneda e Fujita (2022, p. 271) “O vocabulário controlado consiste em uma ferramenta que visa o controle dos termos utilizados em um sistema de informação, ou seja, há uma estrutura construída para representar conceitos atribuídos aos documentos.” Para Lancaster (2004, p.19), “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”. O vocabulário controlado não é apenas um conjunto de termos estruturados, mas uma ferramenta essencial que serve como base para a construção de diferentes SOC.

Shintaku; Sabbag; Costal; Meneses (2021, p. 18) afirmam que, “Para o estabelecimento dos termos e conceitos, o controle de vocabulário é essencial,

evitando a ocorrência de fenômenos linguísticos, como: ambiguidade, plurissignificação, polissemia e homonímia, entre outros.”. Para a efetividade da recuperação da informação, sem ruídos de comunicação entre os usuários e os ativos informacionais, é importante a utilização de técnicas ou ferramentas que contribuam para padronizar a terminologia na representação da informação. A utilização de instrumentos que garantam uma base linguística compartilhada proporciona maior qualidade dos termos para a indexação e, conseqüentemente, para a recuperação das informações (Lancaster, 2004).

Com as tecnologias digitais, a quantidade e a diversidade das informações armazenadas cresceram exponencialmente. Esse crescimento acelerado destacou a necessidade de um controle terminológico mais rígido, especialmente em um contexto como o da PF, onde a precisão e a clareza da informação podem ser cruciais para a resolução de crimes.

3 PROCEDIMENTOS METODOLÓGICOS

A pesquisa se caracteriza como exploratória quanto aos fins e uma pesquisa documental e estudo de caso quanto aos objetivos. Para a coleta de dados serão utilizados como instrumentos o formulário eletrônico e entrevista. O levantamento terminológico foi realizado inicialmente em uma base documental da PF, utilizada para extrair uma amostra de termos que serão a base para o corpus inicial da pesquisa, além de vocabulários controlados da área jurídica que possuam relação com a temática. Posteriormente, a validação dos termos foi realizada com policiais federais que atuam na área de combate ao crime de entorpecentes, conhecedores da temática.

A metodologia proposta está organizada em quatro etapas, Figura 1, que são: Levantamento de termos em vocabulários controlados correlatos, Mineração dos termos em documentos oficiais da PF, Validação e ampliação dos termos com especialistas e Consolidação do levantamento terminológico.

Figura 1 – Etapas do levantamento terminológico



Fonte: Dos autores (2024).

A seguir serão apresentados os procedimentos realizados para atingir o objetivo de cada uma das etapas.

a) Levantamento de termos em vocabulários controlados correlatos

A seleção criteriosa de termos é um componente crucial para a construção de um vocabulário controlado. 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.

Na primeira fase do processo de levantamento, a metodologia inclui uma investigação com um conjunto específico de termos-chave relacionados ao domínio do tráfico de entorpecentes. A seleção desses termos-chave é realizada levando em consideração a relevância direta para a área de estudo, baseada na natureza do objeto de pesquisa, crimes relacionados ao tráfico de entorpecentes. Os termos selecionados para a construção da amostra foram: cocaína, crimes, droga, entorpecente, maconha e tráfico. Esses termos foram buscados nos tesouros dos seguintes órgãos: Supremo Tribunal Federal (STF), Senado Federal e Superior Tribunal de Justiça (STJ). Em cada um dos tesouros utilizados, buscou-se os termos identificados, avaliando também os termos relacionados e as definições disponíveis. Os termos coletados foram analisados com base no conhecimento empírico dos autores, permitindo a identificação de termos relevantes que não haviam sido coletados no levantamento inicial. Por fim, todos os termos coletados e avaliados foram consolidados em uma lista única denominada "Termos coletados em vocabulários controlados" apresentada adiante nos resultados.

b) Mineração de texto em documentos oficiais da PF

Nesta pesquisa são 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. 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 compreende, de forma geral, 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. Na mineração de texto um corpus é um conjunto estruturado de documentos textuais, selecionado de forma criteriosa, que possibilita análises linguísticas e estatísticas eficientes. Foram selecionados em torno de 140.000 documentos oficiais recebidos ou gerados pela Polícia Federal. Esses documentos são de diversos tipos, como ofícios, memorandos, notícias, informações externas e relatórios restritos, que foram amplamente tramitados entre órgãos públicos do Brasil. Os documentos estavam disponíveis na base de dados da Inteligência da PF e estão estruturados de forma a favorecer a mineração de texto. A variedade de tipos documentais permite obter informações mais abrangentes, para além de um setor da instituição. Para facilitar a mineração, os documentos foram previamente processados e convertidos no formato de texto (extensão de arquivos .txt).

Para otimizar o controle de erros, aprofundar a análise e melhorar a eficiência na fase de mineração, foram aplicados no projeto três algoritmos Python, são eles: *part1.py*, *part2.py* e *classegramatical.py*.

O algoritmo *part1.py* foi desenvolvido com o objetivo de facilitar a análise e identificação de termos de interesse, termos resultantes da coleta nos

vocabulários controlados existentes. Ao automatizar essa busca, é possível economizar tempo, reduzir erros humanos e obter uma visão quantitativa mais clara da presença dos 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.

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. 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 em língua estrangeira 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.

Para melhor análise dos textos minerados, foram importados os arquivos de saída na ferramenta de business intelligence QlikSense, a Figura 2 ilustra o painel de visualização. 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 estes, identificou-se 85 termos de interesse e realizou-se 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”.

Figura 2 – Painel de dados



Fonte: Dos autores (2023)

c) Validação e ampliação dos termos com especialistas

Durante a fase de validação dos termos identificados na coleta manual e na mineração de texto, optou-se por entrevistar um especialista no domínio de interesse. O recurso à entrevista com apenas um especialista se deu por restrições de tempo e pelo fato de ser uma amostra, ressaltando que a validação pelo especialista é uma etapa complementar, agregando valor ao processo de

coleta e refinamento, mas não será a única fonte de validação.

Essa entrevista teve como objetivo não apenas validar os termos identificados nas etapas anteriores, mas também consolidar o conhecimento empírico, capturando gírias, termos informais e outras nuances que poderiam não estar evidentes nas bases de dados examinadas, para posterior análise de um grupo maior de profissionais. O especialista entrevistado é policial federal que ingressou na PF em 2005 e possui vasta experiência em investigações de tráfico de entorpecentes nas regiões Norte e na fronteira com o Paraguai, no Estado do Paraná, atualmente, o servidor está lotado em Brasília/DF.

A entrevista teve duração de aproximadamente 60 minutos e foi estruturada de forma interativa, em que os 85 termos minerados foram apresentados e analisados pelo profissional. 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. Nesse processo foram acrescentados sinônimos, incluindo gírias e expressões regionais que enriqueceram o levantamento terminológico inicial.

d) Consolidação do levantamento terminológico

Para garantir uma construção colaborativa do vocabulário controlado é fundamental a participação ampla de especialistas do assunto e usuários. Para tanto, a criação de formulários eletrônicos é uma opção viável para coleta de um número maior de colaborações. 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. 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 foram desenvolvidos de forma a permitir que especialistas, usuários e outros envolvidos no domínio de interesse possam ampliar a contribuição com sugestões de termos, sinônimos, regionalismos, além de validar os termos já consolidados. Foram selecionados para compor o formulário

online os termos de alta relevância, solicitando o para preenchimento de sinônimos e validação da relevância. 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. O formulário foi enviado para profissionais que integram a PF e outras forças de segurança como polícias Civil e Militar, além do Ministério Público Federal, que trabalham diretamente no combate aos crimes de entorpecentes. A Figura 3 é uma amostra do formulário enviado para análise, em que os termos estão apresentados e é solicitada a complementação, quando for o caso.

Figura 3 – Formulário

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 os termos abaixo, separando cada um por vírgula

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

Texto de resposta curta

...

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

Texto de resposta curta

(OPCIONAL) Conhece sinônimo ou gíria para o termo **FAMÍLIA DO NORTE: FDN** ?

Texto de resposta curta

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

Texto de resposta curta

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

Texto de resposta curta

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

Texto de resposta curta

Fonte: Dos autores (2024)

Na próxima seção os resultados obtidos com a pesquisa serão apresentados e discutidos.

4 RESULTADOS

Os resultados serão apresentados seguindo as quatro etapas da metodologia: Levantamento de termos em vocabulários controlados correlatos, Mineração de texto em documentos oficiais da PF, Validação com especialistas e Consolidação do levantamento terminológico.

a) Levantamento de termos em vocabulários controlados correlatos

No quadro 1 é apresentada uma lista de termos existentes nos tesouros do STF - TSTF, do Senado Federal e do STJ, provenientes dos termos-chave, servindo como um importante recurso de referência. Foram identificados 90 termos considerados importantes no domínio estudado.

Quadro 1 – Termos coletados em vocabulários controlados

AÇÃO CONTROLADA	CRIME HEDIONDO	ESQUEMA CRIMINOSO
AGENTE INFILTRADO	CRIME INAFIANÇÁVEL	ORCRIM
ALCALOIDE	CRIME QUALIFICADO	ORGANIZAÇÃO
AMEAÇA	CV	CRIMINOSA
APROPRIAÇÃO	DEPENDÊNCIA FÍSICA	PCC
INDÉBITA	DEPENDÊNCIA	PERÍCIA
ASSALTO	PSÍQUICA	POSSE DE
ASSOCIAÇÃO	DEPENDÊNCIA	ENTORPECENTES
CRIMINOSA	QUÍMICA	PRIMEIRO COMANDO
ASSOCIAÇÃO PARA	DEPENDÊNCIA	DA CAPITAL
O TRÁFICO	TOXICOLÓGICA	PRINCÍPIO DO NON
BALANÇA DE	DROGA	OLET
PRECISÃO	DROGA SINTÉTICA	MONITORAMENTO DA
BANDO	ECSTASE	AÇÃO
CÁRCERE PRIVADO	EXTORSÃO	MULA
CLORETO DE ETILA	EXTORSÃO MEDIANTE	PSICOTRÓPICO
COCA	SEQUESTRO	RECEPTAÇÃO
COCAÍNA	FAMÍLIA DO NORTE	RIXA
COLABORAÇÃO	FDN	ROUBO
PREMIADA	FLAGRANTE DELITO	SUBSTÂNCIA TÓXICA
COLEGIADO	FRAUDE	TRAFICANTE
COMANDO	FRAUDE PROCESSUAL	TRAFICANTE DE
VERMELHO	FURTO	DROGAS
CONLUÍO	GRAU DE	TRAFICANTE DE
CONSTRANGIMENTO	ENVOLVIMENTO	ENTORPECENTES
ILEGAL	GRUPO ARMADO CIVIL	TRÁFICO DE DROGAS
CRACK	HOMICÍDIO	PRIVILEGIADO
CRIME	LANÇA PERFUME	TRÁFICO DE
ANTECEDENTE	LEI DE TÓXICOS	ENTORPECENTES
CRIME CONTRA A	LESÃO CORPORAL	TRÁFICO

PESSOA CRIME CONTRA A VIDA CRIME CONTRA O PATRIMÔNIO CRIME DE DANO CRIME DOLOSO CONTRA A VIDA DESACATO SKUNK PSICOTRÓPICO HEROÍNA	LÍDER DE ORGANIZAÇÃO MACONHA ÓPIO LSD ESTUPEFACIENTE NARCÓTICO	INTERESTADUAL DE DROGAS TRÁFICO INTERNACIONAL DE DROGAS TRÁFICO INTERNO DE DROGAS USURPAÇÃO VICIADO DEPENDÊNCIA DE DROGA ALUCINÓGENO DROGA ALUCINATÓRIA DROGA ILÍCITA
---	---	---

Fonte: Dos autores (2023).

Esses termos serão utilizados como filtros na próxima etapa, da mineração de texto.

b) Mineração de texto em documentos oficiais da PF:

A busca na base de inteligência da PF se concentrou 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.

Na sequência do processo, Figura 6, após a aplicação de filtros de *stopwords*, validação das palavras de acordo com o vocabulário na língua portuguesa 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 6 – Termos minerados - painel do QlikSense



Fonte: Dos autores (2024)

Ao utilizar a classificação gramatical disponibilizada pela biblioteca *Spacy* que é importante ressaltar, não assegura categorizações infalíveis, se mostrou valiosa para uma análise em larga escala, as tabelas 1, 2 e 3 destacam os vinte termos mais frequentes identificados no processo de mineração de texto.

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

Fonte: Dos autores (2024).

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

Fonte: Dos autores (2024).

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	confidencial	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

Fonte: Dos autores (2024).

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 para garantir a relevância e pertinência dos termos no contexto da pesquisa. Esse processo analítico não apenas refinou a seleção de termos, mas também assegurou que os termos listados são verdadeiramente representativos

do domínio em estudo. Foram analisados os trezentos termos mais frequentes na etapa da mineração.

Na fase de mineração de texto foram analisados aproximadamente 140 mil documentos formais da PF. No entanto, como consequência inerente da formalidade destes documentos, o resultado concentrou-se majoritariamente em termos técnicos e oficiais, deixando de lado as variações linguísticas existentes, como expressões regionais, gírias e outras, 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 entorpecentes e suas peculiaridades em diferentes regiões e contextos.

Como já apresentado na metodologia, para mitigar esta lacuna identificada durante a mineração de texto, recorreremos às etapas subsequentes: entrevista e aplicação de formulários com especialistas.

c) Validação dos termos com especialistas

Por meio da entrevista foi possível verificar se os termos identificados nas etapas precedentes da pesquisa são pertinentes e comumente utilizados na prática profissional do campo em estudo. O especialista entrevistado possibilitou a avaliação qualitativa dos termos identificados. As contribuições dos especialistas totalizam 44 respostas, juntamente com os conhecimentos advindos das discussões, foram sistematizados. 84% dos respondentes integram a PF, 11,4% a polícia militar e os 4,6% restantes são servidores da polícia civil e polícia penal do Estado de SP. Quanto ao tempo de experiência profissional, o gráfico 1 aponta que são diversas, 52,2% têm menos de 1 ano até 5 anos de experiência, e 47,7% têm mais de 6 anos de experiência profissional.

Gráfico 1: Tempo de experiência profissional



Fonte: Dos autores (2023).

No quadro 2 é apresentada uma amostra dos termos resultantes da pesquisa e as respectivas definições. Essa organização oferece uma visão estruturada do vocabulário relacionado a crimes de entorpecentes no Brasil, refletindo a riqueza e diversidade linguística.

Quadro 2 - Termos consolidados

<p>Bolar - enrolar um cigarro de maconha.</p> <p>Cheirar - inalar drogas em pó, como cocaína.</p> <p>Injetar - usar uma seringa para administrar drogas.</p> <p>Fumar - consumir drogas através da inalação de fumaça.</p> <p>Engolir - consumir drogas por via oral.</p> <p>Estourar - ação policial para deter traficantes e apreender drogas.</p> <p>Esticar - diluir drogas para aumentar a quantidade.</p> <p>Fornecedor - pessoa que fornece drogas aos traficantes.</p> <p>Batizar - misturar droga com outra substância.</p> <p>Arregar - pagar propina a policiais para evitar a repressão ao tráfico de drogas.</p> <p>Apurar - obter lucro com a venda de drogas.</p> <p>Cobrar - exigir o pagamento de drogas vendidas.</p> <p>Capotar - ser preso ou detido pela polícia.</p> <p>Pesado - traficante envolvido em atividades violentas.</p> <p>Careta - pessoa que não usa drogas.</p> <p>Barato - efeito de uma droga no usuário.</p> <p>Chapado - estado de intoxicação por drogas.</p>	<p>Granel - grande quantidade de drogas.</p> <p>Tijolo - quantidade compactada de drogas, geralmente maconha.</p> <p>Pó - gíria para cocaína.</p> <p>Doce - gíria para LSD.</p> <p>Soltura - liberação da prisão relacionada ao tráfico de drogas.</p> <p>Pedra - gíria para crack.</p> <p>Bala - gíria para ecstasy.</p> <p>Beck - gíria para cigarro de maconha.</p> <p>Pilantra - pessoa desonesta, geralmente envolvida em atividades criminosas.</p> <p>Cana - prisão ou detenção.</p> <p>Fita - atividade criminosa, como roubo ou tráfico de drogas.</p> <p>Dura - abordagem policial.</p> <p>Traficar - ato de vender ou distribuir drogas ilegalmente.</p> <p>Passar - entregar drogas a alguém.</p> <p>Pixote - jovem envolvido no tráfico de drogas.</p> <p>Caô - mentira ou situação enganosa.</p> <p>Camelo - vendedor ambulante de drogas.</p> <p>12 - gíria para polícia.</p> <p>Fim de linha - lugar onde as drogas são vendidas.</p> <p>Mão na massa - envolvimento direto com</p>
--	---

<p>Noiado - usuário de drogas, geralmente crack, com aparência descuidada.</p> <p>Malhação - uso de drogas injetáveis.</p> <p>X9 - informante, geralmente alguém que fornece informações à polícia.</p> <p>Overdose - ingestão excessiva de drogas, causando efeitos adversos ou morte.</p> <p>Viciado - pessoa dependente de drogas.</p> <p>Cagueta - informante que denuncia atividades criminosas, como tráfico de drogas.</p> <p>Traficante - alguém que vende drogas ilícitas.</p> <p>Usuário - pessoa que consome drogas ilícitas.</p> <p>Narcotráfico - comércio ilegal de drogas.</p> <p>Cocaína - droga estimulante derivada das folhas de coca.</p> <p>Crack - forma de cocaína em pedra fumável.</p> <p>Maconha - droga feita a partir da planta de Cannabis.</p> <p>Heroína - droga opiácea derivada da morfina.</p> <p>Metanfetamina - droga estimulante sintética.</p> <p>LSD - droga alucinógena, também conhecida como ácido.</p> <p>Ecstasy - droga sintética com efeitos estimulantes e alucinógenos.</p> <p>Plantão - local onde as drogas são vendidas.</p> <p>Biqueira - ponto de venda de drogas.</p> <p>Mocó - esconderijo onde as drogas são guardadas.</p> <p>Muqui - esconderijo de drogas.</p> <p>Seda - papel usado para enrolar cigarros de maconha.</p> <p>Bagulho - gíria para drogas ou objetos ilegais.</p> <p>Infiltração - infiltração de um agente da polícia em uma organização de tráfico de drogas.</p> <p>QG - Quartel-general do tráfico de drogas.</p> <p>Resgate - pagamento feito para libertar alguém envolvido no tráfico de drogas.</p>	<p>atividades criminosas.</p> <p>Pipa - pequena quantidade de drogas, geralmente maconha.</p> <p>Papelote - embalagem de drogas em pó, como cocaína ou heroína.</p> <p>Fininho - cigarro de maconha enrolado bem fino.</p> <p>Baseado - outro termo para cigarro de maconha.</p> <p>Carreirinha - fileira de droga em pó, como cocaína, preparada para ser cheirada.</p> <p>Jaque - roubo de drogas ou dinheiro de traficantes.</p> <p>Lolo - gíria para inalante ou lança-perfume.</p> <p>Quebrar - ato de dividir ou fracionar drogas para venda.</p> <p>Descolar - conseguir drogas, geralmente através de compra.</p> <p>Respingo - efeito colateral negativo do uso de drogas.</p> <p>Cachimbo - dispositivo usado para fumar crack.</p> <p>Narco - abreviação de narcotráfico ou traficante de drogas.</p> <p>Cartel - organização criminosa envolvida no tráfico de drogas.</p> <p>Fronteira - limite entre países, frequentemente associado ao tráfico de drogas.</p> <p>Lavagem de dinheiro - processo de ocultar a origem ilícita de fundos obtidos com o tráfico de drogas.</p> <p>Perna - pessoa responsável por transportar drogas.</p> <p>Boca de fumo - local onde drogas são vendidas, geralmente a céu aberto.</p> <p>Grilagem - falsificação de documentos relacionados ao tráfico de drogas.</p>
--	---

Fonte: Dos autores (2024)

Os especialistas que contribuíram com as análises forneceram uma valiosa gama de sinônimos e terminologias específicas de sua prática profissional. Esta troca direta através do formulário reforça a importância da colaboração entre pesquisadores e profissionais especializados, permitindo a captura de conhecimento empírico frequentemente inacessível por outros meios. Os termos ratificados e consolidados são apresentados no Quadro 3,

evidenciando o enriquecimento dos termos levantados.

Quadro 3 - Termos sinônimos consolidados

<p>ACUSAR: apontar; botar na reta; caguetar; dedar; denunciar; delatar; dar o serviço; incriminar; indicar</p> <p>171: estelionatário; fraudador; picareta; pilantra; golpista; rato; malandro</p> <p>AMEAÇA: chantagem; pressão; enquadro; esculacho; aviso; jurado; decretado; ideia; susto</p> <p>COCAÍNA: Pó; Farinha; Hollywood; Coca; branquinha; peixe; escama; brilho; mistura; branca; açúcar</p> <p>COMANDO VERMELHO: CV; falange vermelha; comando;</p> <p>CRACK: Pedra; Zulu; mármore; amarela; brita; rapadura</p> <p>DÍVIDA: pendura; débito; calote</p> <p>DROGA SINTÉTICA: química; larica; rebite; droga do amor; azul; doce; bala; gringa; meta; cristal; balinha; comprimido; adesivo</p> <p>DROGA: tóxico; barato; coisa; coisinha; bagulho; parada; caixa; mercadoria; entorpecente; produto; carga; material</p> <p>ECSTASY: bala; doce; azul; pílula</p> <p>EXPLOSIVO: artefato; bananas; bomba; c4; foguete; dinamite</p> <p>FAMILIA DO NORTE: FDN</p> <p>FAZER: Matar; passar; empurrar; concluir; dar sumiço; apagar; zerar; cancelar; resolver; quebrar; assassinar</p> <p>FLAGRANTE DELITO: Flagrante; flaflu; BO; bronca</p> <p>FUGIR: picar a mula; meter o pé; vazar; sumir; desaparecer; dar o pinote</p> <p>FUZIL: AK; peça pesada; bico; bicudo; longa; fura</p> <p>FURTO: 155; ganho; mão</p> <p>GERENTE DO TRÁFICO: Chefe; comandante; patrão; dono; cabeça; frente</p> <p>HEROÍNA: pico; injetável; picada</p> <p>HOMICÍDIO: lona; caixão; passamento; CPF cancelado</p>	<p>INTEGRANTE: soldado; funcionário; faccionado; irmã; mano; parça; batizado; chegado; membro</p> <p>LANÇA PERFUME: Éter; loló; cheirinho; lança; LP</p> <p>LEVAR: Portar; transportar; trazer; carregar; conduzir; fretar; descer; subir</p> <p>LSD: ácido; doce; adesivo; figurinha; selo; selinho</p> <p>MACONHA: Dendê; Prensado; beck; beque; verde; erva; mato; fumo; camarão; baseado; preto; capim</p> <p>MUNIÇÃO: tiro; bala; chumbo; pipoco; caroço; aço; ponta; caramelo; cartucho; azeitona</p> <p>PADRINHO: protetor</p> <p>PISTOLA: quadrada; peça; arma curta; canhão; ferro; berro; preta; cromada; cano; ferramenta</p> <p>POSSE DE ENTORPECENTES: 28; vapor; mocosado; bolso cheio</p> <p>PRIMEIRO COMANDO DA CAPITAL: PCC; partido; 15.3.3; comenado; família</p> <p>PRENDER: grampear; pegar; cair; ganchar; segurar</p> <p>REVÓLVER: oitão; ferro; berro; ferramenta; canela seca; pau de foro; chuteira; calçado</p> <p>SKUNK: camarão; flor; solta</p> <p>TER: possuir; tá na mão</p> <p>TESTEMUNHA: cagueta; x9; declarante</p> <p>TRAFICANTE: Narco; Fornecedor; chefe; chegado; meliante; bandido</p> <p>TRAFICAR: Passar; Vender; fornecer; descolar; empurrar; subir; descer</p> <p>TRÁFICO DE DROGAS: Tráfico; correria; corre; caminhada; atividade; comércio</p> <p>USUÁRIO: Nóia; Zé droguinha; Viciado; Bico; peba; cliente; craqueto; rato; farinheiro; fuminho; doido; vagabundo; dependente; noiado; playboy</p>
---	--

Fonte: Dos autores (2024).

Ao analisar os termos apresentados pelos especialistas, é notável a ampla gama de sinônimos associados, 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. 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. O grande número de sinônimos 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.

5 CONSIDERAÇÕES FINAIS

A implementação de um vocabulário controlado e especializado na Polícia Federal transcende a utilidade para os investigadores atuais; representa um investimento para profissionais e ações futuras. Em um cenário onde criminosos frequentemente empregam uma variedade de termos e gírias para ocultar suas atividades ilícitas, um vocabulário controlado e atualizado é essencial para decifrar e antecipar suas estratégias, tornando-se fundamental para a inteligência policial. 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. O processo de levantamento terminológico está intrínseco à representação conceitual, mas as metodologias podem ser desenvolvidas em fases distintas, que foi a escolha dessa pesquisa. Inicialmente foi feito o levantamento e seleção de vocábulos relacionados aos crimes de tráfico de entorpecentes para que, posteriormente, os conceitos sejam explicitados e as relações semânticas construídas, em uma nova fase da pesquisa, na busca do SOC mais adequado às necessidades da PF.

A interação direta com especialistas na temática enriqueceu o repertório

terminológico, mas também solidificou a importância da validação dos termos previamente levantados, configurando-se em uma etapa muito importante do processo. Essa abordagem assegura que o vocabulário controlado permaneça fidedigno e reflita fielmente a realidade e as nuances do domínio estudado.

Esse é um primeiro passo no levantamento terminológico que irá culminar em um vocabulário controlado, para isso é preciso desenvolver uma estrutura relacional ou parametrizada, com vistas à sua evolução para uma estrutura de tesauro. Tal estrutura permitiria uma organização mais sofisticada e uma busca mais intuitiva, possibilitando uma recuperação de informações mais rica e dinâmica, alinhada às tecnologias digitais na organização do conhecimento, podendo ser acoplada em ferramentas de indexação textual.

Também recomenda-se aprimorar as etapas de entrevista e validação dos termos com especialistas, tornando-as mais robustas e capazes de captar uma gama mais ampla de termos em outros setores da Polícia Federal ou 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

- BAPTISTA, S. G; CUNHA, M. B. Estudo de usuários: visão global dos métodos de coleta de dados. **Perspectivas em ciência da informação**, [S.l.], 2007. Disponível em: <https://www.scielo.br/j/pci/a/h6HP4rNKxTby9VZgzp8qGQ/>. Acesso em: 12 out. 2023.
- 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**, [S.l.], v. 15, n. 1, p. 266–282, 2022. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/42464>. Acesso em: 15 ago. 2024.
- LANCASTER F. W. **Indexação e resumos**: teoria e prática. 2. ed. Brasília: Briquet de Lemos. 2004.

MOREIRA, W.; FUJITA, M. S. L.; DAVANZO, L.; PIOVEZAN, L. B. Vocabulário controlado para a representação documentária 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.

SHINTAKU, M.; SABBAG, D. M. M.; COSTAL, M.; MENÊSES, R. V. **Guia sobre a construção de tesouros**. Brasília: Ibict, 2021. Disponível em: <https://ridi.ibict.br/bitstream/123456789/1167/4/Guia%20sobre%20a%20Constru%C3%A7%C3%A3o%20de%20Tesouros.pdf>. Acesso em: 31 ago. 2024.

SMIT, J. W. I.; KOBASHI, N. Y. **Como elaborar vocabulário controlado para aplicação em arquivos**. São Paulo: Arquivo do Estado, Imprensa Oficial, 2003. 56 p.

XU, G., WU, Z., LI, G.; CHEN, E. Improving contextual advertising matching by using Wikipedia thesaurus knowledge. **Knowl. Inf. Syst.**, [S.l.], 43, P. 599–631, 2015. Disponível em: <https://doi.org/10.1007/s10115-014-0745-z> Acesso em: 22 nov. 2023.

TERMINOLOGICAL ANALYSIS ON NARCOTICS TRAFFICKING CRIMES

ABSTRACT

Objective: The aim of this research is to collect and select terms related to narcotics crimes within the scope of the Brazilian Federal Police. **Methodology:** The research is characterized as exploratory, documentary research and a case study in terms of its objectives. Four methodological stages were defined: Survey of terms in correlated controlled vocabularies, Mining of terms in official PF documents, Validation and expansion of terms with experts and Consolidation of the terminological survey. **Results:** Results 90 relevant terms were identified in correlated Knowledge Organization Systems analyzed. Text mining analyzed approximately 140,000 formal Federal Police documents, resulting in 151 terms related to the topic. At the end of the steps, 210 terms and their synonyms were consolidated. **Conclusions:** This research is a first step in the terminology survey that will culminate in a more robust controlled vocabulary. Direct interaction with experts in the field enriched the terminological repertoire and solidified the importance of validating the terms.

Descriptors: Knowledge Organization. Controlled vocabulary. Trafficking crimes.

ENCUESTA TERMINOLÓGICA SOBRE DELITOS DE NARCOTRÁFICO

RESUMEN

Objetivo: El objetivo de esta investigación es recopilar y seleccionar términos relacionados con los delitos de estupefacientes en el ámbito de la Policía Federal Brasileña. **Metodología:** La investigación se caracteriza por ser exploratoria, documental y un estudio de caso en cuanto a sus objetivos. Se definieron cuatro etapas metodológicas: Relevamiento de términos en vocabularios controlados relacionados, Minería de términos en documentos oficiales de la PF, Validación y ampliación de términos con expertos y Consolidación del relevamiento terminológico. **Resultados:** Se identificaron 90 términos relevantes en los Sistemas de Organización del Conocimiento relacionados analizados. La minería de textos analizó aproximadamente 140.000 documentos oficiales de la Policía Federal, lo que dio como resultado 151 términos relacionados con el tema. Al final de las etapas, se consolidaron 210 términos y sus sinónimos. **Conclusiones:** Esta investigación es un primer paso en el estudio terminológico que culminará en un vocabulario controlado más robusto. La interacción directa con expertos en la materia enriqueció el repertorio terminológico y solidificó la importancia de validar los términos.

Descriptores: Organización del conocimiento. Vocabulario controlado. Delitos de tráfico.

Recebido em: 09.09.2024

Aceito em: 25.07.2025