

# EL TESAURO DE LA COMISIÓN DE LA VERDAD

Documento de sistematización



EQUIPO DE LENGUAJE CONTROLADO SISTEMA DE INFORMACIÓN MISIONAL - SIM 2021

#### Tabla de contenido

#### Introducción

- 1. ¿Qué es un Tesauro?
- 2. El Tesauro de la Comisión de la Verdad
- 2.1. Estructura del Tesauro de la Comisión de la Verdad
- 2.1.1. Núcleos, dominios y directorios
- 2.1.2. Relaciones de jerarquía, equivalencia y asociación
- 2.1.3. Las notas del Tesauro
- 3. Etapas de trabajo para el Tesauro de la Comisión de la Verdad
- 3.1. Etapa I (2019 II)
- 3.2. Etapa II (2020 I)
- 3.3. Etapa III (2020 II/ 2021)
- 4. Metodología para la gestión e ingreso de términos en el Tesauro de la Comisión de la

#### Verdad

- 4.1. Sesiones de capacitación con los equipos misionales
- 4.2. Solicitud de información por correo electrónico
- 4.3. Revisión documental
- 4.4. Gestión de información mediante el formato de términos sugeridos
- 4.4.1. Diligenciamiento del formato de términos sugeridos
- 4.4.2. Revisión del formato diligenciado
- 4.4.3. Validación del formato en las reuniones de equipo
- 4.4.4. Integración de los términos nuevos en la herramienta
- 5. Cómo usar el Tesauro de la Comisión de la Verdad

#### Referencias

#### Introducción

El esclarecimiento de los patrones y causas del conflicto armado colombiano implicó para la Comisión de la Verdad la recepción y producción de una cantidad abrumadora de información sobre las víctimas y los victimarios, a saber: testimonios, sentencias, informes, artículos y, en general, todo tipo de documentos y bases de datos. Por supuesto, toda esta información necesitaba ser clasificada, organizada, descrita y analizada para garantizar su correcto manejo por parte de los funcionarios de la entidad y, sobre todo, por la ciudadanía que accederá a ella una vez la Comisión culmine su quehacer. Es aquí donde comienza la labor del *Sistema de Información Misional - SIM*, un área que, mediante el trabajo de sus diferentes equipos, ha logrado catalogar, registrar, estructurar y sistematizar la información asociada al conflicto armado.

En particular, el Equipo de Lenguaje Controlado del SIM es el que apoya el control terminológico de la Comisión, lo que involucra ordenar, depurar, normalizar y establecer los términos relacionados con las actividades investigativas de la entidad. Ahora bien, dado que la ejecución de estas acciones necesitaba de un sistema único y homogéneo de términos, el equipo construyó el *Tesauro de la Comisión de la Verdad*, una herramienta en línea a disposición de los funcionarios de la entidad para que consulten la terminología validada para su labor. Esta herramienta está conformada por 11.363 términos, agrupados en diferentes campos temáticos denominados núcleos, dominios transversales y directorios.

Para generar una mayor claridad sobre el *Tesauro de la Comisión de la Verdad*, el Equipo de Lenguaje Controlado realizó el presente documento que reúne los componentes principales de esta herramienta, así como su proceso de creación y gestión. En concreto, el lector encontrará cinco apartados: los dos primeros ofrecen un breve marco conceptual sobre el Tesauro, es decir, qué es, cuáles son sus características esenciales, cómo es su estructura temática y en qué consisten sus relaciones semánticas. La tercera sección describe las etapas de construcción del *Tesauro de la Comisión de la Verdad*, desde 2019 hasta 2021. Cada una de ellas, además, expone las lecciones aprendidas y las sugerencias que se pueden

implementar para mejorar su proceso de elaboración. La cuarta sección explica detalladamente las estrategias utilizadas por el equipo para gestionar y validar la información que integra el Tesauro. Y, finalmente, el quinto apartado presenta la interfaz de la herramienta, junto con los pasos a seguir para su uso adecuado.

#### 1. ¿Qué es un tesauro?

Es bien sabido que el acceso al conocimiento (o a la información, si se prefiere) mejora notablemente cuando se orienta al usuario con una organización preestablecida o un esquema de clasificación específico. En efecto, aquellos mecanismos que permiten organizar la información, de manera que se facilite su gestión, así como la interacción entre el conocimiento y el usuario, son conocidos como Sistemas de Organización del Conocimiento (SOC) (Martínez y Mendes, 2015). Ahora bien, aunque estos mecanismos comparten el mismo propósito organizativo, cabe resaltar que todos presentan características muy variadas según su funcionalidad. Así pues, una lista de términos no es igual que un glosario, de la misma forma que un diccionario no es lo mismo que un tesauro.

En particular, de acuerdo con Martínez y Mendes (2015, p. 65), un tesauro es un sistema de organización del conocimiento que permite organizar y recuperar el conocimiento registrado de bases de datos, páginas web o de consultas bibliográficas, y que se compone de un vocabulario controlado, una estructura de relaciones semánticas, un orden sistemático de conceptos, una sintaxis poscoordinada y una cobertura limitada a un dominio determinado. A continuación, se explican brevemente cada uno de estos componentes:

- El vocabulario controlado se refiere a que el tesauro aporta consistencia a la información, ya que está conformado por un conjunto de términos que designan los conceptos de una forma apropiada.
- Las relaciones semánticas, por su parte, vinculan los términos del tesauro según su equivalencia, jerarquía o asociación. La relación de equivalencia muestra las variantes de un término y establece aquel que debe preferirse. La relación de jerarquía señala el término genérico que abarca otros específicos. Y, por último, la relación asociativa muestra la conexión entre dos o más términos, ya sea porque pertenecen al mismo ámbito o porque se usan en contextos similares.

- El orden sistemático de conceptos alude al esquema de clasificación que organiza la información según campos temáticos, disciplinas o facetas.
- La sintaxis poscoordinada significa que el usuario del tesauro recupera la información según sus necesidades. Es decir, el tesauro no proporciona un orden de mención de sus elementos.
- Finalmente, un tesauro suele estar limitado a una especialidad, ámbito o campo en particular. En palabras de Martínez (s.f., p. 7), «los tesauros suelen ser lenguajes especializados, hacen referencia a una temática concreta y están construidos para servir a una institución u organización determinada».

#### 2. El Tesauro de la Comisión de la Verdad

El Tesauro de la Comisión para el Esclarecimiento de la Verdad, la Convivencia y la No Repetición - CEV nació de la necesidad de normalizar, actualizar y dar una estructura a los términos técnicos y populares, o de mayor frecuencia de uso, en el análisis documental de la entidad y para contar con un lenguaje común sobre el conflicto armado en Colombia, y los procesos de verdad, justicia, reparación y restitución, producto de la firma del Acuerdo de Paz. En ese sentido, las funciones principales de esta herramienta son: primero, registrar y estandarizar en forma ordenada los términos utilizados por las diferentes áreas de la Comisión; y, segundo, ser el marco de referencia de los procesos de documentación, catalogación y curaduría de la entidad.

Desde el 2019, la Comisión de la Verdad ha administrado el Tesauro (por el momento, para uso corporativo) como un excelente instrumento para que datos cualitativos (palabras, frases, oraciones y notas) se clasifiquen de manera homogénea, propiciando de esta manera la indización de términos, la interoperabilidad y la reutilización. Asimismo, se caracteriza por ser un vocabulario controlado, diseñado con el propósito de organizar, con base en relaciones semánticas, los conceptos y la terminología de las áreas de interés de la Comisión. Cabe agregar que el Tesauro de la Comisión de la Verdad se encuentra en línea y su mayor impacto es el de facilitar el acceso y la visibilidad de los datos, con el objetivo de que estos y sus

.

<sup>&</sup>lt;sup>1</sup>A través de índices.

distintas conexiones semánticas aporten a la precisión en la búsqueda de información en documentos y en herramientas digitales.

#### 2.1. Estructura del Tesauro de la Comisión de la Verdad

#### 2.1.1. Núcleos, dominios y directorios

Teniendo en cuenta que el conflicto armado colombiano debe ser analizado desde diferentes dimensiones que den cuenta de su complejidad, y en respuesta al orden sistemático que caracteriza a todo tesauro, fue necesario organizar el Tesauro de la Comisión de la Verdad en tres campos temáticos para una mejor distribución de la información registrada. Estos campos son los núcleos, los dominios y los directorios. Los dos primeros campos fueron construidos en numerosas sesiones de trabajo con los investigadores de la Dirección de Conocimiento de la Comisión (ver la sección 3.), mientras que los directorios responden tanto a las dinámicas investigativas de la entidad, como a la información gestionada con los equipos misionales, la cual es revisada y validada por el Equipo de Lenguaje Controlado mediante el procedimiento que será explicado en el numeral 4.

Actualmente, el Tesauro de la Comisión de la Verdad cuenta con 10 **núcleos**, los cuales se refieren a los temas o ámbitos centrales analizados por los investigadores en relación con el conflicto armado colombiano. En otras palabras, los núcleos se interesan por describir los actores o participantes del conflicto, las causas de este, y los procesos o dinámicas que se desenvuelven en el marco del mismo. Los **dominios**, por su parte, son 11 en total y tratan los temas transversales a los núcleos centrales, ya que contribuyen para su definición y comprensión. Algunos de estos temas son los hechos victimizantes, las consecuencias o impactos del conflicto armado y las diferentes formas de afrontamiento y resistencia de las víctimas. Por último, el Tesauro está conformado por 17 **directorios**, los cuales contienen, por un lado, información complementaria sobre el conflicto armado (tipos de armas, clases de estructuras armadas) y, por otro, los datos de interés para el quehacer de la Comisión (tipos de poblaciones, organizaciones, hitos de la historia nacional, entre otros).

#### 2.1.2. Relaciones de jerarquía, equivalencia y asociación

El Tesauro de la Comisión de la Verdad está conformado por dos tipos de descriptores o términos: simples y compuestos. Martínez Ferreras (2010, p. 9) explica que los **términos simples** «se emplean cuando el concepto es claro en sí mismo, sin necesidad de añadir ninguna otra palabra», tal como el descriptor *Exilio*. En cambio, los **términos compuestos** son aquellos que generan una mayor especificidad, en tanto que su estructura sintáctica, constituida por frases sustantivas o frases preposicionales, así lo permite. Por ejemplo, el término *Exilio como consecuencia del conflicto armado* presenta un conjunto de palabras conectadas por preposiciones, mientras que el término *Estatus migratorio* es una frase compuesta de dos sustantivos (CNMH, 2018, p. 3).

Estos términos, simples y compuestos, fueron organizados en el Tesauro según tres tipos de relaciones semánticas, las cuales permiten «especificar el significado de un descriptor asignándolo a un campo semántico o microdisciplina determinada» (Martínez Ferreras, 2010, p. 11). A continuación, describimos cada una de ellas:

a. Relación de jerarquía: Se establece cuando «un término es superior o genérico de otro y este otro término es inferior o específico del anterior» (Martínez Ferreras, 2010, p. 12). Como notación, se utiliza TG para el término genérico y TE para los términos específicos. A modo de ejemplo, se puede mencionar el directorio de *Organizaciones* como el término genérico, y las *Organizaciones de víctimas* como el término específico que se subordina al directorio en cuestión:

TG Organizaciones

TE Organizaciones de víctimas

Este tipo de clasificación delimita cada término en un campo temático específico, lo que es fundamental en todo tesauro, dado que «esta estructura clasificatoria distingue al tesauro de las listas alfabéticas de materias o palabras clave, así como de diccionarios y léxicos» (Martínez Ferreras, 2010, p. 12).

**b.** Relación de equivalencia: «[L]a equivalencia semántica entre descriptores se aplica a los sinónimos, términos que poseen significados iguales en diferentes contextos, y a

los cuasi-sinónimos, términos considerados de significado diferente en el uso común, pero tratados como si fueran sinónimos en materia de indización» (Martínez Ferreras, 2010, p. 11). Se emplea la notación USE para el término preferente y UP (usado por) para el término equivalente. Obsérvese que en el siguiente ejemplo el término preferente es *Autoridad tradicional*, mientras que el término *Autoridad ancestral* es el equivalente.

Autoridad ancestral USE Autoridad tradicional Autoridad tradicional UP Autoridad ancestral

c. Relación de asociación: «Cubre las relaciones entre los términos que están fuertemente ligados conceptualmente, pero no jerárquicamente» (CNMH, 2018, p. 7). Por ende, de acuerdo con Martínez Ferreras (2010, p. 13), «[1]a mayoría de las relaciones asociativas se establecen entre términos que comparten alguna característica común, pero que pertenecen a campos semánticos diferentes». La notación adecuada para este tipo de relación es TR. Para ilustrar, se presentan los siguientes términos: Asociación de Cabildos Indígenas del Norte del Cauca - ACIN, que pertenece al directorio de Organizaciones, y Casa de Pensamiento, que hace parte del directorio Programas e Iniciativas Sociales. A continuación, se muestra cómo estos términos establecen una relación de asociación porque no son sinónimos ni cuasi-sinónimos.

Asociación de Cabildos Indígenas TR Casa de Pensamiento

del Norte del Cauca - ACIN

Casa de Pensamiento TR Asociación de Cabildos

Indígenas del Norte del Cauca 
ACIN

#### 2.1.3. Las notas del Tesauro

Para mejorar la comprensión de algunos términos o descriptores que conforman el Tesauro de la Comisión de la Verdad, se han añadido tres tipos de notas (de catalogación, bibliográficas e históricas) que permiten aclarar dichos términos según diferentes propósitos. La **nota de catalogación**, también conocida como nota de definición, «proporciona la definición de un término y es utilizada cuando el mismo puede interpretarse de modo ambiguo o equívoco» (Martínez y Mendes, 2015, p. 117). En el caso del tesauro en cuestión, este tipo de notas se utiliza, por una parte, para describir los campos temáticos que lo componen (es decir, los núcleos, dominios y directorios), toda vez que es necesario aclarar el tipo de información que presentan cada uno de ellos. Por otro lado, las notas de catalogación son útiles para describir los términos que integran algunos de los directorios: por ejemplo, el directorio de *Programas e iniciativas sociales* tiene una nota de catalogación para mostrar de qué trata; pero, a su vez, cada uno de los términos que lo conforman tiene una nota para explicar en qué consiste el programa o la iniciativa respectiva.

Las **notas históricas**, en cambio, cumplen dos funciones: por un lado, permiten identificar los hechos o datos históricos de determinados descriptores, como los que conforman el directorio de *Hitos*; y, por otro lado, ayudan a especificar desde cuándo se utiliza un término preferente, así como cuáles son los términos (equivalentes o preferentes) que solían designar el mismo concepto (Martínez y Mendes, 2015, p. 118). Finalmente, las **notas bibliográficas** son empleadas para insertar la referencia (el documento, el libro, la página web) que apoya la definición del término. En algunos casos se incluye la expresión "Adaptado de" para mostrar que la descripción del término no es una cita textual, sino una paráfrasis del documento de base.

#### 3. Etapas de trabajo para el Tesauro de la Comisión de la Verdad

El proceso de elaboración y consolidación del Tesauro de la Comisión se puede resumir en tres etapas:

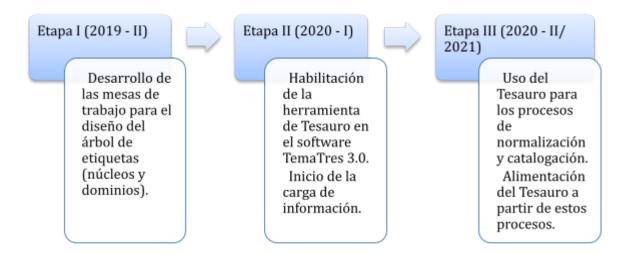


Figura 1. Etapas de construcción del Tesauro de la Comisión de la Verdad

Cabe aclarar que, si bien la primera etapa corresponde a una actividad propia del proceso de transcripción y etiquetado de las entrevistas de la Comisión, se consideró una actividad relevante para la construcción del Tesauro, toda vez que los núcleos y dominios del árbol de etiquetas fueron el primer insumo cargado en dicha herramienta, una vez esta fue habilitada en el software *TemaTres*. También es importante tener en cuenta que la alimentación del Tesauro es una característica fundamental tanto de la segunda como de la tercera etapa. Sin embargo, la segunda fase se destaca porque, además del ingreso de los núcleos y dominios, también se crearon los primeros directorios. Esto le otorgó al Tesauro su estructura temática y, por ende, orientó la carga de información en la tercera etapa. Para mayor claridad, en los siguientes apartados se explicarán en detalle las etapas en cuestión, y se presentarán los aprendizajes y recomendaciones asociados a cada una de ellas.

#### 3.1. Etapa I (2019 - II)

Con el ingreso de Gloria Andrea Córdoba en la Comisión de la Verdad, comenzaron las actividades de lenguaje controlado en la entidad, las cuales se enfocaron, primeramente, en el diseño de un árbol de etiquetas que permitiera la categorización, codificación y análisis de los testimonios recogidos por la Comisión. Así pues, desde agosto hasta noviembre de 2019, la contratista orientó una serie de mesas de trabajo con los diferentes integrantes de los núcleos de investigación y la coordinación de la Subdirección de Conocimiento, en aras de establecer las categorías de análisis para las entrevistas, con base en los 13 mandatos de la entidad. A la

par, el Sistema de Información Misional (SIM) adaptó la herramienta *Dataturks* para que fuera utilizada por los transcriptores en el proceso de transcripción y etiquetado.

Como resultado de estas mesas, se obtuvo un árbol de etiquetas compuesto por 10 núcleos y 11 dominios que apoyaron el etiquetado de los testimonios (un total de 538 etiquetas). Sin embargo, luego de diferentes ejercicios de retroalimentación y la prueba piloto del etiquetado, se determinó que era necesario depurar el árbol de etiquetas para agilizar el proceso de etiquetado, teniendo en cuenta que los transcriptores, además de la tarea de transcribir y etiquetar, habían asumido la misión de llenar las fichas de las entrevistas. Por consiguiente, después de la recategorización de un grupo de etiquetas, la versión final del árbol quedó conformada por 392 etiquetas. Ahora bien, como se indicará en la sección 2.3., la primera información que se carga en el Tesauro (después de su habilitación en febrero de 2020) corresponde a la versión original del árbol de etiquetas. Esto se realizó con miras a mantener el resultado inicial de las mesas de trabajo ya mencionadas.

## Lecciones aprendidas

- Se fortalecieron las habilidades para el trabajo en equipo.
- Se logró una mayor apropiación de la misionalidad de la Comisión de la Verdad.
- Se fortalecieron las habilidades para la construcción de un árbol de etiquetas, con miras a la indización de las entrevistas realizadas por la Comisión.

#### Se sugiere hacer de la misma forma:

 Desarrollar las mesas de trabajo como una estrategia para delimitar las categorías de análisis del proceso de etiquetado. Este ejercicio colaborativo con los equipos de investigación de la Comisión permitió identificar tanto la estructuración temática de las entrevistas como los componentes investigativos centrales de la entidad.

#### Se sugiere hacer diferente:

• (Andrea)

# 3.2. Etapa II (2020 - I)

Luego de la consulta de varias fuentes bibliográficas y la revisión de diferentes programas informáticos, la profesional Gloria Andrea Córdoba determinó que la mejor opción para habilitar el Tesauro de la Comisión de la Verdad era el software *TemaTres 3.0*, dado que, como software libre, cumplía con el interés de la Comisión de garantizar el acceso sin restricciones de los usuarios, para así conseguir un mejor uso y circulación de la información. Sumado a esto, la plataforma en cuestión es flexible y sencilla para la ejecución de acciones como el ingreso de términos, la vinculación de términos asociados y equivalentes, la inclusión de notas y, en algunos casos, la eliminación de términos o la eliminación de relaciones. Del mismo modo, también se destaca por la visualización cómoda de la información, ya sea por campos temáticos o por orden alfabético. Finalmente, el programa cuenta con dos herramientas para la búsqueda de términos: una sección de búsqueda rápida y otra de búsquedas avanzadas.

Así pues, con la ayuda del Equipo de Desarrollo del SIM, se adaptó el software *TemaTres* para la Comisión de la Verdad, y en febrero de 2020 el Tesauro comenzó su funcionamiento. Durante los meses siguientes (hasta junio de 2020), la profesional Gloria Córdoba, la estudiante Lady Velasco (pasante de la Universidad Nacional) y Julio Palacios (del Equipo de Desarrollo), fueron los encargados de ingresar en el Tesauro toda la terminología asociada al quehacer de la entidad. La primera información que fue integrada en la herramienta corresponde a los núcleos y dominios transversales que orientaron la labor investigativa de la Comisión. Posteriormente, se crearon los primeros directorios, cuya información fue tomada de la revisión documental realizada por Gloria Córdoba y la estudiante pasante.

Para junio de 2020, el Tesauro de la Comisión de la Verdad estaba conformado por 11 dominios transversales, 10 núcleos y 10 directorios (*Organizaciones*, *Estructuras armadas*, *Funcionarios públicos*, *Profesiones/oficios*, *Poblaciones*, *Armas*, *Normas*, *Hitos*, *Medios de comunicación* y *Partidos políticos*). A su vez, algunos de estos directorios estaban conformados por ciertas subdivisiones, como los tipos de organizaciones, la clasificación de las estructuras armadas, entre otras. Vale aclarar que el número de núcleos y dominios permanece igual durante las etapas II y III de construcción del Tesauro, toda vez que se trata de una información consolidada con la ayuda de los investigadores de la Comisión. En

cambio, el número de directorios sí aumenta considerablemente durante la tercera etapa, en tanto que las necesidades de catalogación de la entidad así lo requirieron.

#### Lecciones aprendidas

- Se fortalecieron las habilidades de liderazgo, decisión y seguimiento.
- Se fortalecieron las habilidades de búsqueda, selección, ordenamiento y administración de información misional.
- Se ampliaron los conocimientos teóricos relacionados con la creación y gestión del Tesauro como un sistema terminológico.
- Se aprendió el manejo del software *TemaTres 3.0* para la administración del Tesauro de la Comisión de la Verdad.

Se sugiere hacer de la misma forma:

 Usar el software *TemaTres* como herramienta para gestionar la terminología de la Comisión de la Verdad. Igualmente, este software puede ser útil para otras entidades que requieran organizar la terminología propia de su quehacer.

Se sugiere hacer diferente:

(Andrea)

# 3.3. Etapa III (2020 - II / 2021)

Esta última etapa es la más amplia en el proceso de construcción del Tesauro, ya que abarca el segundo semestre del año 2020 y todo el año 2021. En general, la fase se destaca por la alimentación permanente de la herramienta, hasta su consolidación y entrega en noviembre de 2021. Respecto al 2020, es importante resaltar la vinculación en el mes de junio de las profesionales Udiluz Monsalve y Mónica Ramírez, con quienes se logró la conformación del Equipo de Lenguaje Controlado para apoyar a Gloria Córdoba, la líder del grupo, en las actividades de normalización, activación de la Wiki² y administración del Tesauro. Justamente, para garantizar el uso adecuado de esta última herramienta, las dos nuevas

<sup>&</sup>lt;sup>2</sup> Para mayor claridad sobre los procesos de normalización y activación de la wiki, se recomiendan los apartados ... y.... de este documento.

compañeras fueron capacitadas por la líder del equipo y la estudiante Lady Velasco en las funciones del Tesauro, su estructuración temática, los pasos para la creación y eliminación de términos, entre otros aspectos.

Una vez se desarrolló esta capacitación, a partir de julio de 2020 las nuevas integrantes del equipo fueron autorizadas para usar y administrar el Tesauro. Este apoyo fue muy pertinente, dado que, durante el mismo mes, el Equipo de Lenguaje Controlado inició los procesos de normalización y desambiguación de las herramientas del SIM. Esto quiere decir que, por una parte, el Tesauro empezó a funcionar como el marco de referencia para la validación del vocabulario alojado en dichas herramientas y, al mismo tiempo, este ejercicio de validación comenzó a generar nuevos términos para el Tesauro. Dentro de estos procesos de normalización y desambiguación que, durante el 2020, proporcionaron más términos para el Tesauro, cabe mencionar: la recategorización de respuestas "otro, ¿cuál?" en el Módulo de Captura; la homologación de respuestas libres para organizaciones, cargos públicos y actores armados; la normalización de metadatos del buscador del SIM; y, por último, la normalización de profesiones y oficios<sup>3</sup>.

Sumado a lo anterior, en julio de 2020 también inició el uso del Tesauro como marco de referencia para los equipos de Catalogación y de Analítica del SIM, en lo que se refiere a la catalogación de fuentes internas, fuentes externas y la implementación del portal de microdatos. Al igual que los procesos de normalización, el uso del Tesauro para la catalogación de recursos contribuyó sustancialmente a la ampliación del Tesauro mediante la gestión de términos nuevos. Como se verá en el siguiente apartado, este ejercicio de gestión e integración de términos se llevó a cabo mediante el diligenciamiento de un formato de términos sugeridos por parte de los equipos en cuestión. Este formato era validado por el Equipo de Lenguaje Controlado y, como parte de este ejercicio, se acudió de manera permanente a una revisión documental que también suministró más términos para el Tesauro.

Otras actividades relevantes de la administración del Tesauro durante el segundo semestre de 2020 fueron las siguientes: en primer lugar, la líder del Equipo de Lenguaje Controlado

<sup>&</sup>lt;sup>3</sup> Ver apartado ... para ampliar la información sobre los procesos de normalización ejecutados por el Equipo de Lenguaje Controlado.

participó de numerosos encuentros con otros equipos del SIM y de la Comisión para gestionar información, realizar capacitaciones y aclarar dudas sobre la herramienta. En segundo lugar, se redactó el «Protocolo de ingreso de términos en el Tesauro» como una guía para los integrantes del Equipo de Lenguaje Controlado que otorga las pautas para la escritura de términos genéricos y términos subordinados, así como los lineamientos para la redacción y presentación apropiada de las notas de catalogación que describen los términos. En tercer lugar, se construyeron 36 fichas descriptivas: 15 para cada uno de los directorios del Tesauro, 10 para los núcleos y 11 para los dominios. Con este ejercicio, se logró consolidar una definición para cada campo temático, y estas definiciones fueron ingresadas en el Tesauro como notas de catalogación. Finalmente, el equipo inició el proceso de sistematización del Tesauro mediante la construcción del presente documento, con el cual se da cuenta de los contenidos de la herramienta, de su proceso de creación y gestión, de las recomendaciones para su uso, entre otros aspectos.

Para diciembre de 2020, el Tesauro contaba con el mismo número de núcleos y dominios (10 y 11, respectivamente), pero, luego de la creación de 5 nuevos directorios y la aplicación de otros ajustes, la herramienta quedó compuesta por 15 directorios, a saber: *Cartografía*; *Género*; *Lenguas nativas*; *Mandatos, enfoques y objetivos de la Comisión de la Verdad*; *Mecanismos de participación ciudadana*; *Niveles educativos*; *Programas e iniciativas sociales*; *Organizaciones*; *Estructuras armadas*; *Profesiones y oficios*; *Poblaciones*; *Armas*; *Hitos*; *Medios de comunicación* y *Partidos políticos*.

En cuanto al año 2021, la administración del Tesauro se resume en lo siguiente: en primer lugar, el Equipo de Lenguaje Controlado continuó con las reuniones y capacitaciones con los equipos misionales para la gestión de términos. Sin embargo, a diferencia del año anterior, dichos encuentros también estuvieron orientados a la gestión de términos para el diccionario, el proyecto principal del Equipo de Lenguaje Controlado durante 2021. En segundo lugar, los procesos de normalización ejecutados por el equipo continuaron articulados con el Tesauro, bien como marco de referencia, bien como repositorio de términos nuevos sobre el quehacer de la Comisión. Particularmente, en 2021 se extrajeron términos para el Tesauro de un nuevo proceso de homologación de respuestas libres en el Módulo de Captura, de la normalización

para la herramienta DataLake, y de la normalización de los términos asociados a las actividades de investigación de la Dirección de Territorios.

Igualmente, el Equipo de Lenguaje Controlado continuó con la revisión del formato de términos sugeridos para el Tesauro por parte de los equipos misionales. Principalmente, se realizó un trabajo en conjunto con el Equipo de Catalogación para la revisión de los términos extraídos del proceso de Catalogación Colaborativa (fuentes internas de la entidad, fuentes externas recibidas de otras entidades y gestión del portal de microdatos). Para este ejercicio, se prosiguió con la estrategia de validación grupal en las reuniones semanales, la cual permitió resolver y discutir dudas de la terminología propuesta. Finalmente, durante 2021, el equipo se enfocó en la sistematización de todas sus actividades, bajo los lineamientos otorgados por el Equipo de Sistematización de la Comisión. Esto incluyó la reestructuración y culminación del presente documento (iniciado en 2020), así como la actualización mensual del Tesauro en formato Excel; versión que fue elaborada por el Equipo de Catalogación, pero que continuó administrando el Equipo de Lenguaje Controlado. Con este archivo se apoyó la integración del Tesauro en otras herramientas del SIM y, al mismo tiempo, se conservó un respaldo de sus contenidos, en vista de que el software *TemaTres* no permite exportar un archivo con la estructura completa del Tesauro.

A septiembre de 2021, el Tesauro estaba conformado por 10 núcleos y 11 dominios (el mismo número del año anterior), pero se crearon dos nuevos directorios (*Alias de los actores armados y Direcciones de la Comisión de la Verdad*), para un total de 17. Durante los meses de septiembre y octubre se cargaron en la herramienta los últimos grupos de términos (extraídos de todos los procesos mencionados) y, por último, en noviembre de 2021, se procedió a la revisión final, consolidación y entrega del Tesauro para su inclusión en el Museo Virtual de la Comisión.

#### Lecciones aprendidas

 Se fortalecieron las habilidades para el trabajo en equipo entre los miembros del Equipo de Lenguaje Controlado.

- Se fortalecieron las habilidades de comunicación, asertividad y gestión con los otros equipos del SIM y de la Comisión.
- Se fortalecieron las habilidades de liderazgo, decisión y seguimiento, tanto en la líder del equipo, como en todos sus demás integrantes.
- Se fortalecieron las habilidades de búsqueda, selección, ordenamiento y administración de información misional.
- Se ampliaron y consolidaron los conocimientos teóricos relacionados con la creación y gestión del Tesauro como un sistema terminológico.
- Se aprendió y fortaleció el manejo del software *TemaTres 3.0* para la administración del Tesauro de la Comisión de la Verdad.

#### Se sugiere hacer de la misma forma:

- Desarrollar los encuentros semanales del Equipo de Lenguaje Controlado, ya que este espacio fue fundamental para el ejercicio de validación de los términos que integrarían el Tesauro. Con esta actividad, todos los miembros del equipo podían dar sus opiniones y recomendaciones para mejorar la plataforma.
- Desarrollar reuniones con los equipos del SIM y otros equipos de la Comisión para la gestión de términos. De esta manera, se garantizó que la información cargada en el Tesauro correspondiera, en gran medida, con el quehacer misional de todos estos grupos.
- Utilizar un formato para la sugerencia de términos por parte de los equipos misionales, dado que esta estrategia permitió llevar un registro ordenado y claro de los términos propuestos, así como de las observaciones del Equipo de Lenguaje Controlado para los términos sugeridos.

#### Se sugiere hacer diferente:

• Pese a la insistencia de todos los miembros del Equipo de Lenguaje Controlado (especialmente de la líder del equipo) para que todos los equipos de la Comisión utilizaran el Tesauro de manera permanente, la mayoría de estos hicieron caso omiso a dicha solicitud, lo que se vio reflejado en la incongruencia entre el marco de referencia y algunos procesos de la entidad. Por ende, es importante que cualquier entidad que utilice un tesauro como sistema de gestión terminológica, se base en este constantemente para orientar todos los procesos de la organización (ya sean de investigación, de catalogación, de normalización, entre otros).

# 4. Metodología para la gestión e ingreso de términos en el Tesauro de la Comisión de la Verdad

Para la gestión de información y su posterior integración en el Tesauro, el Equipo de Lenguaje Controlado utilizó cuatro procedimientos que le permitieron responder a todas las necesidades de vocabulario de la Comisión. Cabe aclarar que estos procedimientos no fueron excluyentes entre ellos, sino que siempre funcionaron de manera complementaria para que la gestión de términos se desarrollara de manera ordenada y efectiva. A continuación, se explican detalladamente cada uno de ellos:

# 4.1. Sesiones de capacitación con los equipos misionales

En aras de garantizar el uso del Tesauro por parte de los equipos misionales de la Comisión, el proceso de alimentación de esta herramienta se desarrolló a la par de numerosas sesiones de capacitación con las que se formó a los colaboradores de la entidad en aspectos como: los propósitos del Tesauro, su estructura temática (núcleos, dominios y directorios), las relaciones semánticas entre los términos (jerarquía, equivalencia y asociación) y cómo buscar términos en la plataforma. Como resultado de estas capacitaciones, muchos equipos comenzaron a usar el Tesauro de manera adecuada y, sobre todo, se motivaron para entregar información de su quehacer misional. Esta información (en formato Excel) fue remitida a la líder del Equipo de Lenguaje Controlado mediante correo electrónico.

#### 4.2. Solicitud de información por correo electrónico

En algunos casos, para agilizar la gestión de información, solamente se acudió al correo electrónico para solicitar la ayuda de los equipos misionales en la entrega de información para el Tesauro. Por fortuna, varias solicitudes fueron contestadas positivamente, tal como ocurrió en 2020 con el área de Afrontamientos y Resistencias, la cual envió dos tablas en Excel que sumaban más de 800 organizaciones. Estos términos fueron revisados, depurados y

normalizados por el Equipo de Lenguaje Controlado en otra tabla diseñada para tal fin (ver figura 2) y, luego, fueron cargados en el Tesauro según los diferentes tipos de organizaciones que conforman la herramienta.

#### 4.3. Revisión documental

El tercer procedimiento, y uno de los más utilizados por el Equipo de Lenguaje Controlado, es la consulta de páginas web institucionales, de documentos o de bases de datos en línea, para la creación de directorios en el Tesauro o la ampliación de los ya existentes. El ejercicio se complementa con la organización de listados de términos en hojas de cálculo en Drive, para el trabajo colaborativo entre todos los miembros del equipo. Posteriormente, se realizan reuniones de consulta grupal para determinar la relevancia de los términos consultados y, por último, estos son cargados en la plataforma, según su ubicación en las secciones respectivas. A modo de ilustración, en 2020 se diseñó una tabla para registrar y conceptualizar los términos que formarían parte del directorio de *Género*, los cuales están asociados a identidad de género y orientación sexual (ver figura 3). Otros directorios que fueron construidos bajo esta misma dinámica son: *Lenguas nativas*, *Medios de comunicación*, *Niveles educativos y Poblaciones*. También se utilizó la misma estrategia para construir algunos subdirectorios que conforman el directorio de *Organizaciones*, tales como *Agremiaciones*, *Asociaciones indígenas*, *Entidades territoriales locales*, entre otras.

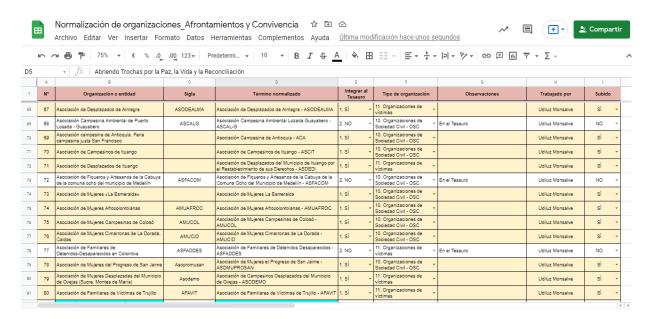


Figura 2. Revisión y normalización de organizaciones ingresadas al Tesauro

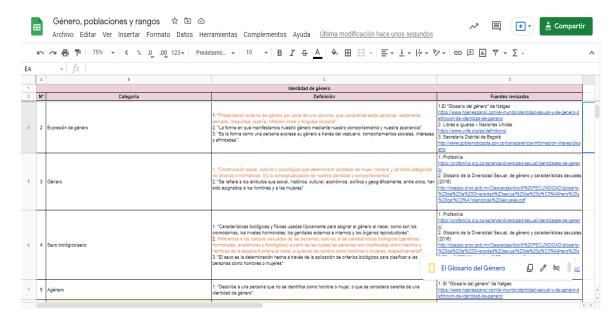


Figura 3. Revisión documental para el directorio de Género

#### 4.4. Gestión de información mediante el formato de términos sugeridos

El último procedimiento de gestión de términos para el Tesauro surgió en 2020 como una estrategia para lograr una mayor articulación entre el Equipo de Lenguaje Controlado y los equipos misionales que aportan información para la plataforma. La base de esta estrategia es un formato, diseñado en Excel, que permite registrar todos los términos sugeridos para el Tesauro, a la vez que se hace un seguimiento de la información consignada. Este formato se utiliza, especialmente, para los términos extraídos del proceso de Catalogación Colaborativa, por lo cual fue construido y ajustado con la ayuda del Equipo de Catalogación del SIM.

En aras de que la gestión de términos se desarrollara de manera ordenada y eficaz, se implementó la ruta de trabajo de la Figura 4, la cual requería de una comunicación constante con el Equipo de Catalogación para acordar plazos de entrega y tiempos de revisión. Así pues, mientras que el diligenciamiento del formato por parte de los catalogadores era permanente, la revisión y validación del formato se realizaba en dos cortes mensuales: un primer corte el día 15 de cada mes y un segundo corte el día 26. Esto significa que el ingreso de términos en el Tesauro era bimensual, a menos de que otras actividades del Equipo de Lenguaje Controlado redujeran el proceso a un solo corte a final de mes.



Figura 4. Ruta de trabajo para la gestión de información del Tesauro

En lo que sigue, se presentarán las características de cada una de estas fases:

# 4.4.1. Diligenciamiento del formato de términos sugeridos

La ruta de trabajo comienza con el diligenciamiento del formato por parte del grupo de catalogadores. Como se observa en la figura 5, el formato está conformado por 12 columnas, pero solo las tres últimas eran diligenciadas por el Equipo de Lenguaje Controlado, cuando se había completado la revisión y validación de los términos sugeridos. Para mayor claridad, se explicarán, uno por uno, los datos solicitados por la tabla.

- Clase de término sugerido: En esta parte se señala si es un término nuevo, si es equivalente de otro ya existente, o si es un término existente que necesita ser ajustado o reubicado a otra sección de la herramienta.
- *Término sugerido*: Se escribe el término propuesto para la herramienta o, en su defecto, el término que debe ser ajustado o reubicado.
- Información relacionada: Se refiere a la información complementaria del término, como los años de ocurrencia de un hecho o los años de funcionamiento de una organización.
- Nota de catalogación: Se recomienda una nota aclaratoria para el término sugerido.
- *Fuente*: Se indican los recursos (documentos, páginas web) de donde se obtuvo el término sugerido o que, en su defecto, apoyarán su validación.
- *Directorio*: Se escribe el nombre del directorio que podría acoger el término sugerido.
- *Términos relacionados*: Se señalan los términos de la herramienta que se asocian al término sugerido.
- Fecha de registro: Se indica la fecha en la que fue sugerido el término.

- Catalogador: Se escribe el nombre de la persona que sugiere el término.
- *Concepto de vocabulario*: El Equipo de Lenguaje Controlado indica si el término sugerido fue integrado, si se encuentra en consulta o si fue rechazado.
- Comentario de vocabulario: Esta parte es para las observaciones del Equipo de Lenguaje Controlado respecto al término sugerido. Por ejemplo, se pueden indicar las razones que justifican el rechazo de un término sugerido.
- Fecha de ingreso en el Tesauro: Es la fecha en la que el término sugerido fue integrado en la herramienta.

# 4.4.2. Revisión del formato diligenciado

La segunda fase del procedimiento es la revisión del formato por parte del Equipo de Lenguaje Controlado, específicamente por el administrador de la herramienta. Esta revisión reúne varias tareas como: 1) comprobar que el término sugerido no se encuentre en el Tesauro; 2) verificar la estructura del término sugerido (guiones, mayúsculas, siglas, etc.), según el «Protocolo de ingreso de términos en el Tesauro»; 3) revisar las fuentes sugeridas (o buscar otras fuentes) para verificar que el término sea acorde con la denominación de la entidad respectiva; 4) construir o ajustar las notas de catalogación, históricas y bibliográficas que requieran los términos sugeridos; y 5) organizar una lista con las sugerencias que requieren una consulta grupal.

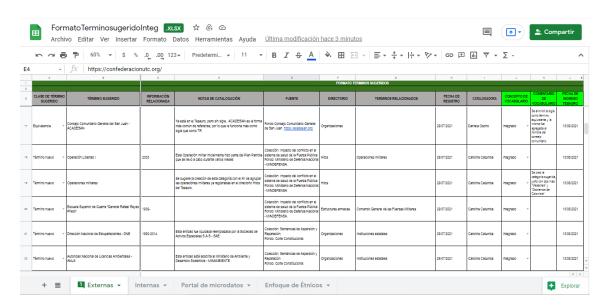


Figura 5. Formato de términos sugeridos para el Tesauro

## 4.4.3. Validación del formato en las reuniones de equipo

Por lo general, algunas sugerencias requerían de la asesoría y aprobación de todos los miembros del Equipo de Lenguaje Controlado, ya que los catalogadores solicitaban ajustes complejos que debían ser discutidos (por ejemplo, la reestructuración de las secciones de la herramienta o la creación de nuevos directorios). El espacio para estas reflexiones eran las reuniones semanales del equipo, en las que se resolvían todas las inquietudes relacionadas con el formato y se validaba la revisión hecha por el administrador de la herramienta.

En algunos casos, fue necesario acudir a otros equipos misionales para la validación de ciertos términos que requerían de una mirada experta, tal como ocurrió en 2021 con un grupo de términos sugeridos por los catalogadores para el directorio de *Estructuras armadas*. En vista de que este directorio está conformado por numerosas subdivisiones (filiaciones, bloques, batallones, entre otros), se solicitó el apoyo de los especialistas en Actores Armados de la Comisión para que validaran la pertenencia de esos términos en una estructura armada específica. Una vez se obtuvo su opinión, los términos fueron revisados nuevamente por el Equipo de Lenguaje Controlado para su carga en la herramienta.

# 4.4.4. Integración de los términos nuevos en la herramienta

Finalmente, luego de la validación en la consulta grupal y/o la aprobación de los especialistas de la Comisión, el último paso del procedimiento es el ingreso en el Tesauro de todos los términos aprobados y de las notas aclaratorias en los términos respectivos. Este paso también incluye la aplicación de los ajustes sugeridos en los términos ya existentes, ya sea su reubicación, la corrección de errores ortotipográficos o la adición de una parte faltante (como una sigla). A la par, se diligencian las tres últimas columnas del formato, según lo explicado en el numeral 4.4.1.

# 5. ¿Cómo usar el Tesauro de la Comisión de la Verdad?

Como se mencionó en los apartados anteriores, el Tesauro de la Comisión de la Verdad es una herramienta en línea a disposición de los funcionarios de esta entidad para el registro y control de los términos asociados al análisis documental producido en sus diferentes áreas. Dado su propósito, es fundamental que sea un instrumento de fácil acceso y, ante todo, de uso

sencillo para la revisión y consulta por parte de sus usuarios. En lo que sigue, se presentarán los pasos que todo usuario debe seguir para ingresar al Tesauro de la Comisión de la Verdad y consultar la información que este contiene:

- Ingrese al enlace <a href="https://sim.comisiondelaverdad.co/">https://sim.comisiondelaverdad.co/</a>. En esta página se encuentran todas las herramientas creadas y gestionadas por el Sistema de Información Misional, incluido el Tesauro de la Comisión.
- 2) Una vez allí, diríjase a la opción Tesauro, la cual se señala en la figura 6 con un círculo rojo. Si desea revisar los video tutoriales para la navegación del Tesauro, puede dar clic en el símbolo de interrogación (?) que se señala con una flecha roja en la figura 6.

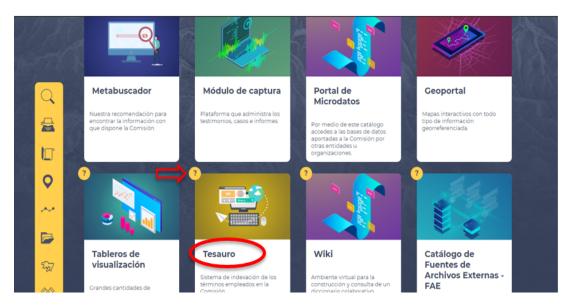


Figura 6. Herramientas del Sistema de Información Misional

3) Cuando sea direccionado a la página principal del Tesauro, encontrará los siguientes componentes: primero, una barra con el buscador de términos (flecha morada en la figura 7); segundo, una opción de búsqueda avanzada (flecha verde), y, por último, una barra con los términos en orden alfabético (flecha roja).



Figura 7. Parte superior del Tesauro de la Comisión de la Verdad

4) Luego de la primera sección, observará los campos temáticos del Tesauro, los cuales también están organizados en orden alfabético. Esto quiere decir que los núcleos, los dominios y los directorios aparecen en forma de lista, según la letra inicial de su denominación. Por ende, el directorio Alias de los actores armados es el primero de la plataforma, seguido de Armas. El último, según el orden alfabético, es el directorio de Programas e iniciativas sociales.



Figura 8. Campos temáticos del Tesauro en orden alfabético (primera parte)

Como puede verse en las figuras 8 y 9, los dominios y los núcleos se agrupan como parte de las letras 'd' y 'n', respectivamente. Además, al final de la página de inicio se repite la barra de letras en orden alfabético (flecha roja en la figura 9).



Figura 9. Campos temáticos del Tesauro en orden alfabético (segunda parte)

- 5) Para buscar un término en el Tesauro de la Comisión de la Verdad, puede proceder de varias maneras:
- Si desea consultar, por ejemplo, el nombre de alguna entidad o institución, diríjase al directorio de *Organizaciones* y despliegue los tipos de organizaciones dando clic en la flecha que se ubica a un costado del directorio (círculo rojo en la figura 10).
   Posteriormente, seleccione el tipo de organización y realice la búsqueda del término en orden alfabético.

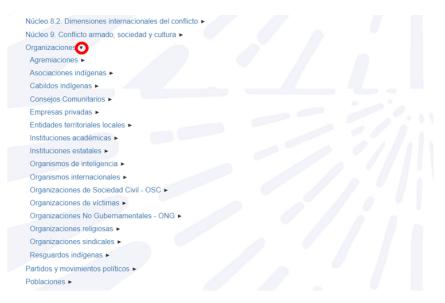


Figura 10. Directorio de Organizaciones en el Tesauro de la Comisión de la Verdad

- Igualmente, puede realizar la búsqueda del término en las barras de letras que se encuentran en la parte superior e inferior del Tesauro. Solo debe escoger la letra inicial de su interés y hallar el término (figura 11).



Figura 11. Barra de búsqueda alfabética

Por supuesto, la barra de búsqueda rápida es la mejor opción para encontrar términos en el Tesauro. Sin embargo, se recomienda hacer la búsqueda con una o dos palabras claves escritas correctamente. Si falta una tilde o, por error, omite alguna letra, el Tesauro no reconocerá el término. Por ejemplo, si escribe la palabra "desaparición" sin tilde en la última sílaba, el buscador no arrojará ningún resultado (ver figura 12). En contraste, observe todos los resultados que despliega el buscador con la palabra escrita en forma adecuada (figura 13).



Figura 12. Búsqueda rápida sin resultados



Figura 13. Búsqueda rápida que arroja resultados

Si presiona la tecla enter después de insertar la palabra en el buscador, podrá observar con más detalle los resultados correspondientes (figura 14).

- La última opción de búsqueda es la Búsqueda avanzada que se encuentra en la parte superior del Tesauro (círculo rojo en la figura 15). Una vez allí, se puede indagar por un término con las palabras exactas, o especificar si se trata de un término equivalente

(no preferido) o de un metatérmino<sup>4</sup>. Además, la herramienta también permite buscar términos dentro de las notas (figura 16).

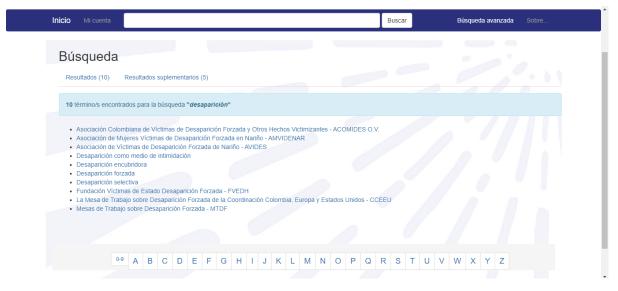


Figura 14. Lista de resultados de la búsqueda rápida

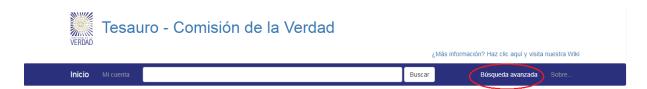
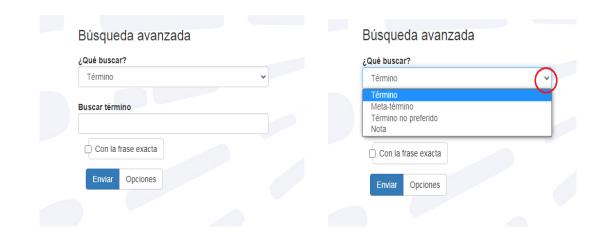


Figura 15. Búsqueda avanzada del Tesauro de la Comisión de la Verdad



<sup>&</sup>lt;sup>4</sup> Los meta-términos permiten clasificar la información del Tesauro sin ser considerados en el conteo de términos de la herramienta (por ejemplo, "Asociaciones de Amazonas"). Sin embargo, para el Tesauro de la Comisión no se identificó ningún término como metatérmino, ya que las clasificaciones son consideradas como parte de la estructura terminológica de la plataforma.

Figura 16. Opciones de la búsqueda avanzada

#### Referencias

Centro Nacional de Memoria Histórica (CNMH). (2018). Tesauro con enfoque diferencial sobre graves violaciones a los DDHH e infracciones al DIH con ocasión del conflicto armado colombiano. Archivo de Derechos Humanos.

Martínez Ferreras, D. (s.f.). *Los tesauros*. Universitat Oberta de Catalunya. http://biblioteca.udgvirtual.udg.mx/jspui/bitstream/123456789/1353/1/Los%20tesauros.pdf

Martínez Tamayo, A. M. y Mendes. P. V. (2015). *Diseño y desarrollo de tesauros*. (Biblioteca Humanidades, 38). Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación. <a href="http://www.libros.fahce.unlp.edu.ar/index.php/libros/catalog/book/68">http://www.libros.fahce.unlp.edu.ar/index.php/libros/catalog/book/68</a>