



PROCESOS DE NORMALIZACIÓN DE LA COMISIÓN DE LA VERDAD

Documento de sistematización



EQUIPO DE LENGUAJE CONTROLADO
SISTEMA DE INFORMACIÓN MISIONAL - SIM
2021

Tabla de Contenido

Tabla de Contenido	2
Introducción.	3
Contexto de los procesos que realiza el Equipo de Lenguaje Controlado (ELC) dentro del Sistema de Información Misional.	4
¿Qué se entiende por Normalización?	6
¿Para qué sirve la Normalización?	7
¿Cómo se realiza la Normalización?	9
Protocolo de Normalización.	10
Sobre las acciones correctas.	15
Las Actividades de Normalización Realizadas por el ELC, Periodo 2020 - 2021.	17
Referencias bibliográficas.	20

Introducción.

Siendo la Normalización uno de los procesos que realiza el Equipo de Lenguaje Controlado (ELC), para el Sistema de Información Misional (SIM), es importante documentar el proceso en sí mismo, dando cuenta de las actividades realizadas desde el mes de febrero hasta noviembre del año 2021.

Teniendo en cuenta lo anterior, este documento presentará una breve explicación sobre lo que se entiende por Normalización y lenguaje controlado, así mismo, plantea la función que tiene este tipo de lenguaje para algunas áreas de interés para, posteriormente, focalizar esta teoría a las actividades realizadas por el SIM para la Comisión para el Esclarecimiento de la Verdad (CEV). Por otro lado, se presentará un protocolo para el desarrollo de las actividades de normalización basadas en el Tesoro de la Comisión (Herramienta desarrollada igualmente por el ELC), con el cual se expone la manera adecuada para realizar las actividades de Normalización solicitadas al ELC. Finalmente, se proponen una serie de lecciones aprendidas a partir de la experiencia como contratista que llevó a cabo las actividades de Normalización con la guía, ayuda y supervisión de los miembros del ELC en el periodo febrero – noviembre de 2021.

Este documento, puede tomarse como punto de partida para el ajuste y mejoramiento de las actividades relacionadas con la Normalización en diferentes instituciones que reciban información de todo tipo que por su naturaleza ameriten una estandarización en diferentes herramientas digitales. Estas reflexiones y propuestas se gestan a partir de la propia experiencia, realizando una evaluación de la actividad individual que hace parte de los procesos de un Equipo (ELC), por ende, dichas reflexiones toman en cuenta las solicitudes y sugerencias que realizaron los miembros del Equipo y que el contratista interiorizó, adaptó e integro a la realización de la Normalización.

Contexto de los procesos que realiza el Equipo de Lenguaje Controlado (ELC) dentro del Sistema de Información Misional.

El Equipo de Lenguaje Controlado (ELC) se encuentra integrado al Sistema de Información Misional (SIM). Desde el ingreso de la líder del equipo Gloria Andrea Córdoba Henao en el 2019, al ELC se le han asignado las tareas de creación y administración de las herramientas Wiki y Tesauro, Herramientas que continúan actualizándose y que por lo menos, el Tesauro hará parte del Legado. Y, finalmente, como proyecto principal el ELC desarrolla el Diccionario Temático.

El proyecto que ocupa el mayor tiempo de trabajo del ELC es, sin duda alguna, es la creación de un diccionario temático sobre diferentes términos relevantes sobre el conflicto armado. Este Diccionario contará con 300 entradas léxicas (Aproximadamente) y con un número igual de casos ilustrativos. Este proyecto hará parte del Legado. Quedará alojado dentro del Museo en el Módulo Explora. Actualmente, parte de los integrantes del ELC se encuentran recolectando la información, casos ilustrativos, definiciones, contextos y demás metadatos necesarios para su posterior discusión e ingreso a la plataforma que está en desarrollo para el proyecto.

Por otro lado, el ELC es administrador de las herramientas Wiki y Tesauro. En lo que respecta a la Wiki, los diferentes trabajadores, contratistas e investigadores pueden unirse a la plataforma y proponer nuevas entradas para alimentar dicha herramienta, lo cual permite su actualización permanente, eso sí dicha herramienta en algún momento cerrará su ciclo. En primera instancia fue administrado por la profesional Mónica Ramírez Cano, quien se retiró del ELC en junio de 2021 y posteriormente fue administrado al nuevo integrante del ELC Andrés Salazar Ávila.

Por su parte, el Tesauro es alimentado a través de diferentes formatos que son revisados por la profesional Udiluz Monsalve Muñoz y discutidos, cuando es necesario, en las reuniones semanales del ELC. Esta herramienta resulta vital para la adecuada realización de la normalización, ya que es la base en la que se apoya dicha actividad. Los términos incluidos

en el Tesauro son términos validados por diferentes equipos misionales de la Comisión (CEV) por lo tanto, la información allí consignada es verídica y se encuentra organizada en directorios.

Sin embargo, el proceso que aquí nos ocupa es la Normalización. Este proceso inició con Andrea Córdoba y posteriormente fueron integradas a algunas actividades a las profesionales Udiluz Monsalve y Mónica Ramírez, en cifras, el ELC logró normalizar 16.778 términos dentro de estas cifras, se resalta la normalización de la opción *Homologación de respuestas a preguntas derivadas* que se encuentra en el Módulo de Captura y la *reasignación de Categorías*, también consignadas en el Módulo. Estas actividades en algunos casos permitieron la identificación de términos nuevos que debían ingresar al Tesauro

Por otro lado, las actividades de Normalización realizadas por el ELC en el periodo febrero – noviembre 2021 fueron asignadas al profesional Alexis Peña Castañeda, éstas estuvieron relacionadas con la validación de etiquetas para apoyar los procesos de automatización del etiquetado de entidades y actividades solicitadas por los equipos misionales para validar y ajustar algunos términos que serán trasladados al Data Lake (Lago de Datos).

¿Qué se entiende por Normalización?

Considerando que las fichas trabajaron, en algunas ocasiones, con preguntas abiertas se estableció un lenguaje natural que debía procesarse, analizarse y organizarse. Teniendo en cuenta estas necesidades resulta necesario un lenguaje controlado. Para algunos autores este lenguaje es conocido como *Indización*, *Homologación*, *Estandarización*, o lenguaje documental.

Para Méndez y Moreiro (1999) es un proceso fundamental que analiza un lenguaje natural y lo caracteriza reteniendo cierta información para asociarlo a unos términos controlados. Para González Sigüero (2017) el lenguaje controlado es “*una versión restringida de un lenguaje natural que ha sido construida para conseguir un objetivo específico*” y dentro de sus alcances es que sea procesable por un ordenador. Por su parte, Gil Leiva (1997) considera pensar en verbos como “*(...) analizar, seleccionar, asignar y convertir para referirnos al proceso de indización.*”. De acuerdo con Gorges van Slype (1991) se debe entender al lenguaje controlado como un sistema de signos que representan un determinado contenido.

Partiendo de lo anterior, se puede afirmar que la Normalización es un proceso que permite analizar el lenguaje natural para determinar una serie de características semánticas o temáticas de dichos términos y de esa manera asignar o seleccionar un único término, que podemos llamar *descriptor*, basándonos en Iglesias (1992) quien sostiene que un *Descriptor* “*Un término o símbolo autorizado y formalizado que figura en un tesoro y que se usa para representar sin ambigüedad, los conceptos contenidos en los documentos.*” que logre rescatar la información de estos diferentes términos. De acuerdo con lo anterior, información que produjo las fichas de la Comisión para el Esclarecimiento de la Verdad (CEV) es propia del lenguaje natural y por ende debe ser normalizada en base al Tesoro cuyo desarrollo permite los diferentes *descriptores* que deben ser usados para realizar de manera adecuada el ejercicio de Normalización.

¿Para qué sirve la Normalización?

Para Gil Urdiciáin (1992), los lenguajes controlados sirven para proveer una cantidad clara y precisa de información, así mismo, para presentar la información en niveles de comprensión apropiados para los usuarios y llevar a cabo estas funciones en el menor tiempo posible.

En la misma línea se expresa Iglesias Maturana (1997), quien considera que los lenguajes controlados son puentes de comunicación entre el texto y el usuario de este, estos “puentes” permiten al usuario navegar entre la información requerida por él de manera clara y específica.

Por su parte, Ramírez Polo (2012) se pueden destacar dos usos principales, por un lado, el lenguaje controlado es útil para los seres humanos en la medida que mejora el análisis de determinados textos y, por otro lado, en lo que respecta a las herramientas computarizadas facilita el procesamiento del lenguaje natural.

Como se ha podido ver, los lenguajes controlados permiten al usuario categorizar la información propia del lenguaje natural para que sus diferentes investigaciones no resulten engorrosas o muy extensas, este tipo de lenguaje al ser limitado permite realizar búsquedas sin ambigüedades. Por otro lado, permiten rescatar y seleccionar información relevante para los investigadores proveniente de diferentes textos y temáticas.

Por su parte, los desarrollos digitales se nutren de la Normalización ya que con su aplicación se puede facilitar el procesamiento de lenguaje natural de información física o digital, realizado a través de los computadores o bien mejorar las capacidades predictivas de Inteligencias Artificiales diseñadas con el fin de recuperar y seleccionar información de manera automática.

Para la CEV resulta importante las actividades propias del proceso de normalización ya que con ellas se pueden establecer datos cuantitativos; los desarrollos de georreferenciación logran nutrirse de manera eficiente. Por otro lado, el lenguaje controlado otorga facilidades

a la hora de catalogar y realizar búsquedas ya que aporta información acotada que los diferentes equipos puedan explotar como ellos consideren adecuado. Dentro de las actividades de etiquetado, cierto es que, si bien no se usa la normalización que el ELC lleva a cabo, sí se pueden observar que la normalización y el lenguaje controlado tienen mucho que ver con el desarrollo de estas actividades.

Desde el punto de vista de los recursos digitales, el lenguaje controlado le sirve a los desarrolladores como base de lenguaje de programación, así mismo sirve para traspasar los datos de una herramienta a otra y a las personas que trabajan con lenguaje natural y que deben integrar datos a desarrollos digitales, les sirve para integrar datos de manera ajustada al lenguaje que ya está establecido por la CEV.

¿Cómo se realiza la Normalización?

A continuación, se presentará un Protocolo de Normalización con el cual se permite observar la manera adecuada para realizar la normalización y aquellas acciones no recomendadas para el desarrollo de esta. De igual manera se presenta de manera breve la estructura del Tesauro, herramienta necesaria para realizar la normalización y finalmente, se podrán observar una lista de términos recurrentes que sirve como una guía para algunos colaboradores de la CEV y algunas acciones que deberían evitar estos.

El Tesauro tiene un orden alfabético y numérico, podrá observar dos divisiones puntuales, la primera, una interfaz principal con botón de inicio, login, barra de búsqueda, botón de búsqueda avanzada (que permite navegar el Tesauro entre términos, metadatos, término no preferido o notas), barra de búsqueda alfabética y un acceso directo a la Wiki:



La segunda parte, está compuesta por todos los directorios y en la parte final una barra de búsqueda alfabética. Estos directorios son de primer nivel, como se podrá observar todos tiene un icono que indica la presencia de subcategorías o términos de segundo nivel:





Una vez acceda al directorio adecuado para normalizar el término requerido podrá observar que el término se desagregará en varias subcategorías. El apartado de Término es el descriptor en sí mismo, Notas, se refiere a aquellas notas aclaratorias del término controlado, pueden ser bibliográficas, históricas o de catalogación y el apartado de Metadatos aporta información sobre la fecha de creación del término controlado en la herramienta, parte del lenguaje de programación (Códigos) y accesos directos a motores de búsqueda.

Protocolo de Normalización.

1. Reconocimiento del material a normalizar.

Se debe reconocer la procedencia de los términos que serán normalizados, esto permite identificar los directorios adecuados con los cuales se realizará un ejercicio de normalización adecuado.

- Por ejemplo, si el término a normalizar es: “yo era docente” y proviene de preguntas relacionadas con actividades, profesiones y similares, el directorio adecuado será Profesiones y oficios > Profesiones y oficios legales, y el término correcto será Profesor (Observará en el término que uno de los Términos no preferidos es educador)

- Si el mismo término a normalizar (yo era docente) proviene de preguntas relacionadas con el tipo de población al que la persona hacía parte, el directorio adecuado es Poblaciones y allí encontrará el término Docentes que será el correcto.

2. Estructura y formato adecuado para normalizar.

Una vez reconocida la proveniencia de los términos a normalizar, traslade los datos a un formato, el Equipo de Lenguaje Controlado (ELC) sugiere trabajar en Hojas de Cálculo del espacio institucional provisto por Google, no obstante, si considera que Excel u otra herramienta organizativa cuya base sean filas y columnas estará bien.

El formato que se propone es sencillo, le ayudará a organizar la información de manera clara y así mismo, para el ELC resultará de gran ayuda para realizar las validaciones pertinentes:

Término original	Proviene de	Reemplazar por	Observaciones	En el Tesouro	Integrar al Tesouro	Fuente
Término original para normalizar	La fuente de la cual proviene el término original	El término normalizado o en base a los directorios del Tesouro	Toda información extra sobre el Término o la justificación de la normalización	Si: para aquellos que se encuentran en el Tesouro. No: para aquellos que no lo están	Término normalizado que no está en el Tesouro y que propone el colaborador	Fuente de la cual extrae la propuesta de término normalizado que no está en el Tesouro.

Titule su documento con el nombre del equipo misional, el nombre del proceso y el mes en que se realiza la normalización para que se reconozca el equipo misional que está realizando la actividad.

- Se debe recordar que todos los procesos de la CEV que requieran normalización de datos deben ajustarse al Tesouro, si el equipo no logra realizar una actividad de

normalización adecuada realice la alerta con el asunto: *Solicitud de Normalización* + el nombre del equipo misional que requiere el apoyo al ELC en los correos electrónicos:

udiluz.monsalve@comisiondelaverdad.co,
anderson.pena@comisiondelaverdad.co,
andres.salazar@comisiondelaverdad.co.

con copia a gloria.cordoba@comisiondelaverdad.co, quién es la líder del ELC o bien remitir la solicitud solamente al correo de la líder del ELC.

3. Acciones del colaborador que realizará la normalización.

Para iniciar la actividad se debe tener en cuenta que la persona deberá tomar decisiones, esto es clave. En el computador, se deben tener a disposición el formato de términos a normalizar y el Tesauro (como mínimo), el ELC recomienda abrir un navegador (de preferencia propia del colaborador) para realizar búsquedas de términos y fuentes en caso de no encontrar algunos términos a normalizar en el Tesauro. Con la información de la proveniencia de los términos a normalizar, realice la búsqueda en el directorio pertinente y allí busque (se sugiere una primera búsqueda manual) el término con el cual puede normalizarse.

- La recomendación que hace el ELC es usar el botón de inicio y buscar en el directorio adecuado, no obstante, puede realizar la búsqueda a través de la barra de búsqueda o bien de manera alfabética.
- El colaborador no debe buscar sinónimos, pero sí, se dará cuenta que los términos normalizados tienen unas asociaciones semánticas que logran agrupar términos provenientes del lenguaje natural, también observará relaciones jerárquicas entre los directorios y los términos normalizados.
- Los términos consignados en el Tesauro son términos validados previamente por diferentes investigadores de la CEV, así que los términos allí consignados son correctos.

- Se deben usar los comandos automáticos del teclado, toda vez encuentre el término normalizado adecuado, es decir, una vez que encuentre el término subráyelo (con el ratón del ordenador) y en su teclado aplique la fórmula Ctrl + c para copiar el término y en la columna Reemplazar por, realice el comando Ctrl + v, de esta manera se asegura que el término pase de la misma manera como se encuentra en el tesaurus.
- Si se encuentran términos a normalizar que no están en el Tesaurus, se deben completar las columnas Observaciones y Fuente para que el ELC realice las validaciones oportunas y de ser necesario integre los términos del Tesaurus. Esto representa una alerta, así que no olvide reportarla al ELC. Se debe hacerlo a través de las tablas de gestión de información preestablecidas para la herramienta del tesaurus (revisar el Protocolo del Tesaurus).
- Durante la toma de decisiones y ejecución de la actividad por parte del colaborador se presentarán dudas, no es de preocuparse, el ELC sugiere colores a manera de convenciones, el amarillo es útil para identificar aquellos términos cuya normalización no está del todo segura y el rojo para aquellos que no se logren normalizar, esto le permite tanto al colaborador como a los integrantes del ELC realizar controles rigurosos a dichos términos.
- La normalización no es rígida, es decir, dentro de un mismo documento se pueden normalizar términos que pertenezcan a diferentes directorios del Tesaurus, por ejemplo, términos como: “era estudiante cuando pasaron los hechos” podría pensarse en el directorio de Profesiones y oficios, pero el ELC ha validado el término Estudiantes como parte del directorio de Poblaciones. Pero a continuación se puede encontrar términos, esta vez sí, en relación con el directorio Profesiones y oficios como “fabricaba zapatos” que puede normalizarse como Zapatero (directorio Profesiones y oficios > Oficios legales).
- Existen términos que pueden dividirse y normalizarse de dos o más maneras, por ejemplo, “sindicalista de la CUT” el colaborador puede normalizar como Sindicalistas (directorios de Poblaciones) y como Central Unitaria de Trabajadores

de Colombia – CUT (Directorio de Organizaciones > Organizaciones sindicales). En estos casos, la recomendación es realizar ambas normalizaciones usando la | para separar ambos términos normalizados.

- El ELC ha validado los complementos como “Miembro de”, “Agente de” o el rango cuando es posible reconocer la organización a la que pertenecen las personas (que desempeñan un rol dentro de dicha organización) a las que hace referencia el término, por ejemplo:

Ejemplo término a normalizar	Complemento adecuado
Era militar	Miembro del Comando General de las Fuerzas Militares
Fue policía	Agente de la Policía Nacional de Colombia
Era de las FARC	Miembro de las Fuerzas Armadas Revolucionarias de Colombia
Comandante del ejército	Comandante del Comando General de las Fuerzas Militares

- Se deben usar las normalizaciones más específicas que sean posibles, de no serlo, se pueden usar los términos genéricos (1-2 nivel), de usar esta forma, se debe usar la convención de color amarillo, para que el ELC revise dicha normalización.
- Para términos como “null”, “nada”, “no sabe”, “sin especificar” o similares es necesario usar el término normalizado No se especifica ya que es el término validado con el que el ELC ha venido adelantando los trabajos de normalización desde su conformación.

- Los directorios que más se usan son: Profesiones y oficios, Organizaciones y Poblaciones. Esto no implica que sean los únicos o que no se usen dentro de un mismo documento.
- Una vez termine el trabajo, se recomienda enviarlo al ELC para realizar su validación, el ELC le brindará todo el apoyo que sea necesario.

Sobre las acciones correctas.

En este apartado se muestran algunas acciones que no son del todo correctas extraídas con base en la experiencia de la propia actividad que los miembros del ELC han realizado en 2020 y 2021. Teniendo en cuenta lo anterior, se proponen una serie de sugerencias y tips para optimizar las actividades del proceso de normalización.

- Un colaborador que requiera el Tesauro para realizar normalizaciones optará por buscar el término a normalizar en la barra de búsqueda de este como lo haría en la barra de búsqueda de cualquier navegador web, pero está no es una acción recomendable ya que la función de la propia barra no es igual a la de los motores de búsqueda. La barra de búsqueda del Tesauro está desarrollada para buscar términos específicos, para realizar la búsqueda de manera adecuada, los datos que se escriben deben ser ortográficos, en ocasiones la barra puede responder con términos que responden a una naturaleza diferente al de la proveniente de los términos a normalizar. Si se pasa por alto este detalle, las normalizaciones hechas podrían no ser del todo correctas, y, al ser términos normalizados son aceptados por las bases de datos de la CEV y, así mismo, ELC tardaría un poco más en realizar correcciones ya que al ser una actividad manual se pueden pasar por alto estos términos. Por otra parte, la barra de búsqueda puede no mostrar el término por algún error en la búsqueda y el colaborador asume que dicho término no se encuentra en el Tesauro y sugerirlo al ELC, que al momento de hacer la validación se dará cuenta que dicho término si se encontraba en la herramienta.

La sugerencia del ELC es realizar la búsqueda del término en los directorios del Tesauro, basado en la procedencia de los términos a normalizar. Iniciar desde la página de inicio del Tesauro, allí se ubicará el directorio adecuado y se comenzará a realizar la actividad de análisis y selección, tardará un poco más pero el resultado será el esperado.

- Al pasar la normalización del Tesauro al formato, escribir en vez de usar los comandos automáticos en el teclado no es recomendable, ya que se pueden producir errores mínimos que se pasan por alto, por tal motivo, los comandos antes mencionados (Ctrl+c y Ctrl + v) representan la mejor manera de trasladar el término normalizado a la tabla de trabajo y garantiza, de paso, que el término normalizado sea el mismo que está en el Tesauro.
- Trabajar en estructuras y formatos diferentes es una actividad que un equipo o colaborador puede hacer, y será recibido por el ELC, no obstante, a estos equipos misionales o colaboradores se les entregará el documento con modificaciones que se ajustan a la estructura propuesta por el ELC, no se puede saber hasta qué punto le sean útiles o no dichas modificaciones.
- No realizar las alertas de manera temprana al ELC, esta es una actividad poco recomendable, ya que, por un lado, los miembros del ELC vienen trabajando en la actualización del Tesauro y por otro, conocen más a fondo las dinámicas propias de la normalización, por ende, son las personas idóneas para realizar validaciones. No para realizar los trabajos de normalización desde cero, pero sí para aportar el apoyo adecuado a los colaboradores que requieren datos normalizados.

Las Actividades de Normalización Realizadas por el ELC, Periodo 2020 - 2021.

A continuación se presenta un panorama general de las actividades que el ELC ha desarrollado desde la conformación de este, es decir desde el ingreso de las profesionales Udiluz Monsalve Muñoz y Mónica Ramírez Cano en junio de 2020. Así mismo, se recogen las actividades que el ELC ha encarado a través del profesional Alexis Peña vinculado al ELC en febrero de 2021 y cuyo contrato terminará en noviembre del mismo año.

En primera instancia, entre junio y octubre de 2020 el ELC desambigüó y normalizó las preguntas abiertas sobre el diligenciamiento de las entrevistas en el Módulo de Captura para apoyar los procesos de análisis cuantitativo.

Seguidamente, en agosto de 2020 se establece el marco de referencia a partir del Tesauro y se inician los procesos de normalización manuales para los metadatos del Buscador.

Con motivo del ingreso del profesional Alexis Peña, en febrero de 2021, se realizaron dos jornadas de capacitación lideradas por Udiluz Monsalve cuyos pormenores pueden verse en la Actas N° 1, 2, 3, 4, 5, 6, 7 y 8 (Carpeta ACTAS FEBRERO - Capacitaciones). En estas jornadas se realizó una nueva desambiguación y normalización de las preguntas abiertas para el Módulo de Captura.

Por otro lado, desde marzo y hasta julio de 2021 el ELC realizó una actividad de apoyo al equipo misional de Desarrollo, actividad solicitada por el ingeniero Eduar Ramos Barragan para la validación del etiquetado para efectos de desambiguación, cuyo objetivo es apoyar los procesos de automatización del etiquetado de entidades. En esta actividad se pudo validar el etiquetado en 86 entrevistas cumpliendo la meta acordada por el ingeniero y la líder del equipo Andrea Córdoba.

De acuerdo con comunicación sostenida con el Ingeniero Oliver Estuardo Mazariegos Zetina en la cual, él solicitó la reasignación de categorías para las preguntas en el Módulo:

1. ¿Puso en conocimiento a alguna entidad o autoridad? 1.1.1 Estatal; ¿Puso en conocimiento a alguna entidad o autoridad? 1.1.2 Comunitario; ¿Puso en conocimiento a

alguna entidad o autoridad? 1.3.1 Internacional; "1 ¿Puso en conocimiento a alguna entidad o autoridad? 1.1.2 ¿Por qué accedió a esta/s autoridad/es o entidad/es?" y "1 ¿Puso en conocimiento a alguna entidad o autoridad? 1.1.3 ¿Cuál era su objetivo principal al acceder a esta vía?", actividad que se realizó en mayo del presente año, su rapidez se explica ya que la misma actividad había sido asignada al ELC en el año 2020, pero debido a un error de Desarrollo los cambios realizados por ELC no fueron incluidos, no obstante, el ELC mantiene un archivo que usa el espacio institucional de la nube, provisto por la multinacional Google en el cual recopila todas las actividades y sus distintos procesos.

Durante junio el ELC apoyo a Desarrollo para la normalización de una serie de términos que integrarían el Lago de Datos (*DataLake*), lo cual dió una primera impresión de las actividades que posteriormente serían solicitadas al ELC.

Seguidamente, durante el mes de agosto se realizó la validación y normalización de 8621 términos suministrados por el equipo misional de Desarrollo. Esta actividad se realizó como apoyo al traslado de datos que se hará de diferentes bases de datos al Lago de Datos . Esta actividad contó con el apoyo del integrante de Desarrollo John Fredy Rodriguez Moreno quien fue el nexo entre ambos equipos misionales. En el mismo periodo de tiempo se realizó una actividad de apoyo al equipo de Territorio, para la normalización de términos que no se encontraban en el Tesauro. Como productos de ambas actividades se pudo presentar una tabla de sugerencias que serán revisadas por Udiluz Monsalve para confirmar su integración a la herramienta en el mes de septiembre.

Las actividades de normalización que el ELC afrontó en septiembre fue un apoyo al equipo de Desarrollo basado en las 86 entrevistas mencionadas anteriormente, para cargar los archivos en las entrevistas dentro del Módulo de captura. En el mismo periodo de tiempo se planteó una normalización a las nuevas categorías presentes en el Módulo de Captura (ingresadas desde el 8-10 de febrero de 2021), actividad que se realizó tomando como base la experiencia de las capacitaciones del mes de febrero.

Por otro lado, en apoyo al Equipo de Analítica se realizó una actividad de normalización y recategorización de las fichas en las 12 variables asociadas a Impactos, la variable de

Iniciativas de no repetición, las 5 variables de Afrontamientos y las 5 variables de Contextos. Para la realización de esta actividad se creó un desarrollo en el Módulo de captura en la sección Gestión > Catálogos > Recategorizar opciones a cargo del Ing. Oliver Mazariegos.

Normalización del vocabulario











Reclasificación de respuestas

Seleccione el listado a reclasificar

Afrontamiento colectivo - dificultades

Mostrar únicamente los pendientes de reclasificar

Opciones para el listado: **Afrontamiento colectivo - dificultades**

#		Texto original		Nuevo texto reclasificado
1	 Corregir el texto original	Adaptación al exilio	 Reclasificar	Pendiente de reclasificar
2	 Corregir el texto original	Asociado a hechos victimizantes	 Reclasificar	Pendiente de reclasificar
3	 Corregir el texto original	Con las instituciones	 Reclasificar	Pendiente de reclasificar
4	 Corregir el texto original	Conflictos interétnicos	 Reclasificar	Pendiente de reclasificar
5	 Corregir el texto original	De seguridad	 Reclasificar	Pendiente de reclasificar

En este desarrollo se pueden observar dos botones “Corregir el texto original” cuya función es la normalización del término, esto es, realizar todo tipo de correcciones de tipo ortográfico, y el botón “Reclasificar” con el cual se realizan las recategorizaciones de las fichas que solicitó el Equipo de Analítica. El Tesauro fue la herramienta básica para realizar ambas actividades ya que es el lenguaje controlado que el ELC ha construido a lo largo de su existencia y que los investigadores de los demás equipos misionales han validado.

Finalmente, este mismo desarrollo se utilizó para normalizar el campo **Ocupación / Ocupación actual (víctima, persona entrevistada)**, en este campo se usó el botón “Corregir el texto original” para estandarizar las diferentes entradas con base en el Tesauro, más precisamente los directorios Profesiones y oficios y Poblaciones. El botón “Reclasificar” se usó para recategorizar los términos según 32 sectores económicos validados por el Equipo de Analítica. Esta actividad apoya algunos análisis, estadísticas y gráficos que representarán información propia de las víctimas y por ende vital para la Comisión de la Verdad.

Referencias bibliográficas.

Gil Leiva, I. (1997). *La automatización de la indización, propuesta teórico-metodológica: aplicación al área de biblioteconomía y documentación*. Murcia, España: Universidad de Murcia.

Gil Urdiciain, B. (1992). *Función de los lenguajes documentales en el tratamiento de la información en las organizaciones*. *Revista general de información y documentación*, 2(2), 195-200.

Iglesias-Maturana, M. T. (1992). *Análisis y recuperación de información documental*.

Iglesias Maturana, M. T. (1997). *Beneficios de la utilización de lenguajes controlados en el análisis y recuperación de información*. Recuperado de <http://rev-ib.unam.mx/ib/index.php/ib/article/view/3854/3407>

Jiménez, A. G., Mateos, D. R., & García, B. C. (2019). *Estudio sobre la indización/etiquetado y los lenguajes documentales en cinco diarios españoles*. *Scire: representación y organización del conocimiento* (ISSNe 2340-7042; ISSN 1135-3716), 25(1), 55-64.

Méndez, E., & Moreiro González, J. A. (1999). *Lenguaje natural e indización automatizada*. *Ciencias de la Información*, 30(3), 11-24.

Ramírez Polo, L. (2012). *Los lenguajes controlados y la documentación técnica: mejorando la traducibilidad*. *Tradumática*, (10), 0192-204.

Slype, G. V., Hípola, P., & Moya Anegón, F. (1991). *Los lenguajes de indización: concepción, construcción y utilización en los sistemas documentales* (p. 198). Fundación Germán Sánchez Ruipérez; Pirámide.