# **DICCIONARIO**

## ¿Qué es?

El diccionario colaborativo es una herramienta que permite a la ciudadanía explorar unos campos semánticos y una serie de términos, definidos y perfilados a partir del proceso misional de la Comisión de la Verdad. Se puede navegar sobre estos conceptos, explorando las entrevistas y los testimonios que configuran el contexto que les da cuerpo. Dicha integración facilita su comprensión y apropiación dentro de un público no especializado en temas del conflicto armado.

Un campo semántico es un conjunto de términos relacionados entre sí por su significado; esto quiere decir que los términos asociados a un campo semántico comparten elementos y/o características semánticas entre sí.

# ¿Cuál es su historia?

El diccionario es una herramienta derivada de los procesos de normalización del lenguaje llevados a cabo en la Comisión de la Verdad, con herramientas como el tesauro y la guía de etiquetas. Durante el proceso de etiquetado de las entrevistas realizadas por la institución, se vio la necesidad de gestionar un diccionario de etiquetas que sirviera de soporte conceptual para el procesamiento de la información; de esta manera, el equipo de lenguaje controlado, de la mano con los investigadores, comenzaron a organizar una serie de definiciones que servían como punto de referencia para aplicar los códigos y/o etiquetas respectivas. Es en este punto donde el equipo de lenguaje controlado ve una riqueza importante con toda la información que comienza a generarse.

A partir de este proceso colectivo, el equipo de lenguaje controlado propone la construcción de un diccionario digital, especializado en temas del conflicto, que apoye la navegación por el repositorio, las herramientas y el contenido legado por la Comisión de la Verdad; que apoye la lectura del Informe Final por parte de la ciudadanía; y que apoye la labor docente y los procesos académicos en escuelas y universidades.

# **Funcionalidad**

## **Usuarios**

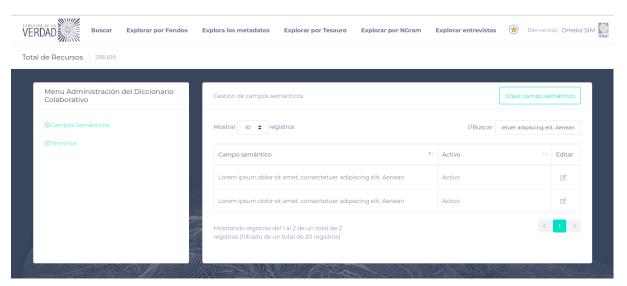
El diccionario colaborativo está pensado para dos clases de usuario:

- Funcionario-Administrador: usuarios encargados de definir e ingresar los campos semánticos y los términos, y subir los respectivos fragmentos de entrevistas, audios e imágenes asociados. Definen el contenido que va a ser mostrado a la ciudadanía. Requiere de autenticación.
- **Ciudadanos:** personas del común que pueden navegar por el diccionario a partir de índices (es decir, visualizan la información administrada por el funcionario); no requieren autenticación para la exploración.

# Tareas que el sistema proporciona

### 1) Gestión de campos semánticos y términos.

Espacio para la administración y gestión de los campos semánticos y los términos que son dispuestos para la consulta. A partir de allí se ingresan, editan, activan y/o desactivan los campos semánticos y los términos; y se cargan los fragmentos de entrevistas, los audios, las imágenes y los metadatos que describen y dan contexto a cada uno de estos elementos.



Administración campos semánticos y términos

#### 2) Navegación por los campos semánticos y términos.

Funcionalidad a través de la cual los ciudadanos navegan por el corpus semántico definido y perfilado por los funcionarios-administradores. La consulta puede hacerse por campo semántico o por términos, a través del índice de letras. Si selecciona por campo semántico, se podrán visualizar todos los términos asociados; y si selecciona por término, aparecerán los términos asociados a esa letra particular. Adicional a esta navegación, los usuarios pueden reproducir los audios y visualizar los fragmentos de entrevista integrados a cada campo y a cada término.



Página principal para la navegación por el diccionario

# Requerimientos

# Limitaciones y alcances

#### Limitaciones y posibilidades de mejora

- Una mejora es la relación de los términos con las entrevistas; es decir, la información, en este caso las entrevistas, tiene unos niveles de seguridad que definen las posibilidades de acceso libre o de acceso restringido. En este sentido, es importante que el aplicativo pueda decantar la información por niveles de seguridad, y que la carga de fragmentos de entrevistas no se haga de manera manual, sino que pueda ser correlacionada de manera automática. En el momento, el aplicativo permite visualizar términos y asociarlos con fragmentos de entrevistas los cuales han sido cargados, previa y manualmente, por un administrador conservando el anonimato de las personas (no se cargan en voz propia, sino en voz de los funcionarios).
- Con esta mejora, el administrador no tendría que subir ningún audio, no tendría que subir ningún fragmento de la entrevista, sino que, a partir del lago de datos, se automatiza la carga de esta información. Sin embargo, se deben precisar dos cosas importantes:
  - Debe existir unos niveles de acceso a la información, previamente definidos y catalogados por la entidad (o, en su defecto, un comité).
  - El funcionario debe estar monitoreando, constantemente, la información que va a publicarse, para tener una mayor seguridad de lo que va a mostrarse a la ciudadanía.

Es importante que al implementar la funcionalidad de relacionar un término con una entrevista, no se le pegue directamente a la base de datos porque, al ser tantos recursos, la consulta se vuelve compleja y afecta el rendimiento general de la herramienta. En este sentido, se recomienda realizar unos índices sobre ElasticSearch, y sobre ello extraer las consultas.

Al momento de relacionar un término con una entrevista, si no se estructuran los índices con ElasticSearch, el diccionario puede colapsar dado que no se darían las condiciones de rendimiento para responder a las necesidades de concurrencia. Es decir, si no se implementa ElasticSearch, la petición va a ser bastante lenta.

•	Una mejora en términos de visualización y facilidades para la ciudadanía, tiene que ver con la posibilidad de resaltar el texto de la entrevista mientras el audio, que recoge el testimonio, se va reproduciendo.

# Entornos y metodologías

### Desarrollo

El diccionario colaborativo estuvo ceñido a lineamientos definidos en el metabuscador desarrollado en la Comisión de la Verdad. En este sentido, hereda la arquitectura y las tecnologías usadas para este desarrollo. Tenemos entonces los siguientes entornos:

### Tecnologías y arquitectura

#### Para el FRONTEND:

ReactJS; versión 16.13.1

https://es.reactjs.org/

Material-UI; versión 4.11.4
 https://v4.mui.com/getting-started/usage/

#### Para el BACKEND:

MongoDB; versión 5.0.3

https://docs.mongodb.com/

NestJS; versión 7.1.5

https://docs.nestjs.com/

- Comunicación entre el frontend y el backend, se realiza a nivel de servicios REST (arquitectura web basada en el protocolo HTTP para mejorar las comunicaciones cliente-servidor). Las peticiones se hacen por GET, POST y PUT; en el caso del diccionario, no se usa DELETE.
- Arquitectura orientada a servicios: los desarrolladores crean los servicios que luego son consumidos por la aplicación; esto con el fin de reutilizar servicios para los distintos componentes que se van creando. Adicionalmente, permite una identidad misional para todos los desarrollos de la Comisión de la Verdad. Este desarrollo hereda la arquitectura del metabuscador.

 Los archivos en los que se desarrolla el código de los formularios, se encuentran en el frontend en:

sim-frontend/src/app/pages/dictionary

Copied!

Si, por ejemplo, se quiere añadir un campo semántico o modificar alguna variable se debe buscar, dentro del directorio del backend, el archivo semantic.controller.ts, con el fin de agregar el campo primero a la estructura y luego modificar el frontend. Sin embargo, si la persona no es especialista no se recomienda modificar la estructura del backend, ya que allí se encuentra toda la configuración del núcleo y las conexiones.

### Metodologías

 Modelo atómico: bajo esta metodología se definen: páginas, organismos, moléculas y átomos, que luego pueden ser reutilizados en otros desarrollos. Todos los componentes del metabuscador desarrollado por la Comisión de la Verdad, usan esta misma estrategia con el fin de tener una estructura homogénea.



Modelo atómico. Fuente: https://www.uifrommars.com/atomic-design-ventajas/

#### • Programación Sprint Scrum:

Historias de usuario

Definición de tableros y actividades (sprints)

Planeación y seguimientos

# **Buenas prácticas**

**Programación Scrum:** se recomienda trabajar bajo la metodología de programación Sprint Scrum, dado que es una metodología que permite la

articulación fácil y rápida de modelos simples y funcionales, los cuales pueden ir evolucionando con el feedback y las necesidades del cliente.

Bajo esta metodología estaremos en función de un desarrollo iterativo o incremental, el cual consiste en dividir el trabajo en pequeñas partes o bloques temporales (sprint). Al final de cada etapa se entrega una funcionalidad completa. Para estructurar la evolución se recomienda crear el Mínimo Producto Viable (Minimum Viable Product, MVP): producto con suficientes características para satisfacer a los clientes y proporcionar retroalimentación para el desarrollo futuro.

**MVP:** herramienta que permite aprender mientras se desarrolla. Gracias a la iteración, el producto evoluciona y se reduce el tiempo para la validación de nuevas ideas.

**Estandarizar código:** definir unas reglas de trabajo; es decir, estandarizar la manera en que se van a crear y llamar las funciones, los métodos, las variables, los atributos, etc. La normalización del código es fundamental para el mantenimiento óptimo del desarrollo.

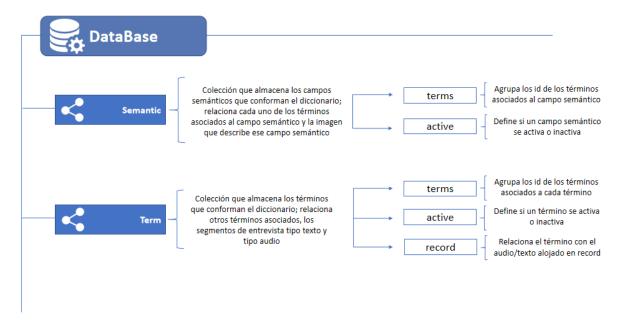
**Comentar el código:** es una buena práctica comentar el código a fin de que se convierta en un texto legible, autoexplicativo y facilite las modificaciones y el mantenimiento.

**Interacción con usuarios:** es una buena práctica tener entrevistas previas con los usuarios del sistema; dichos encuentros permiten entender los requerimientos puntuales y las funcionalidades concretas sobre las que debe basarse el desarrollo. Además, se recomienda encontrar un lenguaje común entre lo técnico del sistema y el requerimiento del usuario, a fin de no generar reprocesos y lograr comunicar fácilmente las funcionalidades del sistema.

# **Colecciones DB**

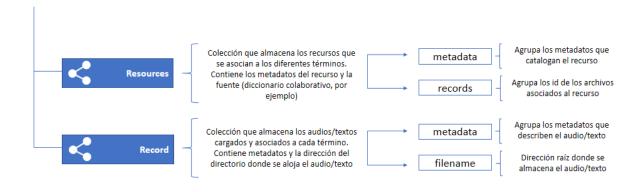
# **Arquitectura DataBase**

#### Colecciones usadas en MongoDB



Arquitectura DB parte I

Las colecciones Semantic y Term, son las colecciones que almacenan toda la gestión de los campos semánticos y los términos, las relaciones existentes entre ellos y las referencias a los audios y/o recursos asociados a cada término.



Arquitectura DB parte II

Las colecciones Resources y Record, son las colecciones que almacenan y gestionan los recursos asociados a los términos (cada término tiene asociado un segmento de entrevista tipo texto o tipo audio, con el fin de ilustrar su contexto).

# **Desarrolladores**

El desarrollo del diccionario colaborativo fue una construcción conjunta con el equipo de lenguaje controlado del Sistema de Información Misional de la Comisión de la Verdad. De hecho, el diccionario fue un proyecto colectivo formulado desde el equipo que ha administrado el tesauro de la Comisión. Según los desarrolladores, el diseño y el mockup fueron entregados, directamente, por el equipo de lenguaje controlado.

# ¿Por qué se ha logrado una buena herramienta?

El diccionario es una buena herramienta porque le permite a la ciudadanía conocer que, más allá de hechos victimizantes como el homicidio y el desplazamiento -sin restarles importancia-, hay toda una serie de términos que definen y nombran el conflicto armado colombiano; en otras palabras, es una herramienta que permite identificar y entender, de manera simple, factores, causas, consecuencias y todos aquellos elementos que configuran y constituyen el conflicto armado en Colombia. Y lo hace de manera intencionada; es decir, un término o un campo semántico, que puede no decir mucho de manera aislada, dentro de la herramienta se ubica en un contexto real que lo ilustra y ejemplifica. Es por ello que los términos y los campos semánticos que componen el diccionario son, selectivamente, definidos y catalogados.

De acuerdo con lo anterior, el diccionario permite salirse de la comodidad, de la zona de confort, para pensarse el conflicto armado desde la profundidad de sus procesos, desde definiciones que, de otro modo, no hubieran podido conocerse. Así lo relatan los mismos ingenieros:

A nivel personal, el crecimiento personal, el hecho de no encapsularse solo en el mundo del desarrollo y en tu casa, sino conocer que hay muchas personas que han sido afectadas por muchas situaciones y han tenido que dejar su familia; o casos reales, complicado... Y se sigue llenando uno de información. Entonces, de mi parte, me gustaría que esto nos ayudara a conocer que hay muchos términos que afectan a muchas personas a nivel Colombia, y que han pasado por muchos hechos victimizantes, y que uno solo conoce desplazamiento forzado... y hay muchos términos que describen las maneras en que han sido afectadas las personas. Esto nos va a ayudar a tener una visión más clara y más humana de lo que pasa en Colombia; y no solo el término, qué es eso, sino también qué historia hay detrás de ese término...

## Aprendizajes y experiencias

La interacción con el equipo de lenguaje controlado dejó una experiencia valiosa, la cual, según el grupo de ingenieros desarrolladores, adquirió un valor importante en tanto se asemeja a situaciones de la vida cotidiana, por ejemplo, la visita a un familiar o amigo. En efecto, según el equipo desarrollador, levantar los requerimientos dentro de un proyecto ya definido es como ir a la casa de un familiar: implica, primero, comportarse bien y, segundo, ser receptivo y comunicativo, actitudes que resultan en una experiencia memorable.

Esta experiencia se da de esta manera ya que, según el equipo de desarrollo, el cliente, en este caso los profesionales del equipo de lenguaje controlado, conocía muy bien el proceso y las necesidades, de allí que fuera sencillo levantar los requerimientos y coordinar un trabajo de co-creación (se estructuraba sobre sobre las ideas de ellos). Este punto es fundamental para resaltar que la labor de los ingenieros no se circunscribe solamente a "echar código", sino que está cargada de toda una labor de escucha y comprensión.

Ahora bien, esta labor y este trabajo colaborativo es posible gracias a la construcción de un lenguaje común entre profesionales de disciplinas muy diversas; un lenguaje común que implica bajar tecnicismos a un nivel coloquial que, sin restar rigurosidad al ejercicio, permita tender puentes de entendimiento mutuo. Por ejemplo, qué es un campo semántico o qué es un formato de audio, son conceptos que emergen del proceso conjunto, los cuales, al formularse en tonos cotidianos, generan un aprendizaje colectivo que se nutre, más que limitarse.

Me decían "mira el diseño como está...", y les hacía preguntas no tan técnicas, sino en palabras simples: "te gustaría que mostrara el audio en qué formato", "qué es formato", "un MP3, WAP...", etc.; entonces, esa relación es un ejemplo muy claro... es una persona que va a la casa y que necesita conocer esa estructura... Adicionalmente, intento no entorpecer el proceso de ellos, imagínate si yo en lugar de poner términos le pongo definiciones, que parecen ser sinónimos, pero en la orientación lingüística que ellos le dan, no lo son...

Adicional al proceso de escucha, otra experiencia importante tiene que ver con la necesidad de aterrizar a los usuarios en ciertas precisiones:

Por ejemplo, un requerimiento fue cuando hablábamos de extraer el audio de la entrevista... yo les decía que es posible, pero que requiere un tiempo. Si se aborda ese requerimiento, no se va a poder tener un producto funcional en pocos meses, porque es un desarrollo que implica una inversión importante. Que sí se puede hacer, pero ya son mejoras... tenemos que tener una base para poder iniciar.

Porque qué es lo que sucede, se puede entrar en reproceso... se puede hacer un proyecto muy grande, muy chévere, pero por el tiempo no alcanzamos... se quedó en un 40%, y para el ciudadano, no hicieron nada; entonces, es mejor decir: ya está este producto, ya se puede navegar, y en un futuro van a haber estas mejoras...

# **Usuarios**

El diccionario es una herramienta dispuesta en el archivo del esclarecimiento de la Comisión de la Verdad. El equipo de lenguaje controlado ha tenido un rol fundamental en su construcción, siendo co-creadores y líderes en el diseño de la metodología y de los mockups que definieron y perfilaron la herramienta. Adicionalmente, han sido las personas encargadas de crear el contenido que nutre la herramienta.

# Experiencias: aprendizajes, retos y oportunidades

#### Retos y oportunidades

El diccionario es una herramienta pensada para facilitar procesos que van más allá de la Comisión de la Verdad; y ese es su reto más grande: generar posibilidades de diálogo y construcción colectiva con la ciudadanía. En este sentido, y según el equipo de lenguaje controlado, la herramienta ha sido pensada para que apoye la lectura del Informe Final, para que apoye procesos pedagógicos en escuelas y universidades, y para que sirva de guía en la navegación por el repositorio legado por la Comisión de la Verdad.

De acuerdo con este reto fundamental, el diccionario se ha centrado, por un lado, en reunir las definiciones construidas durante el proceso de investigación; y, por otro lado, en darles un contexto para ampliarlas y aterrizarlas a través de casos ilustrativos como una entrevista o un recurso.

#### Diálogos interdisciplinares

El diseño y desarrollo de la herramienta fue un proceso colectivo donde diferentes experticias aportaron esfuerzos que consolidaron un trabajo importante para la concreción del proyecto. En este sentido, y según el equipo de lenguaje controlado, un primer momento estuvo guiado por el equipo de diseñadores, quienes construyeron los mockups; un segundo momento estuvo a cargo de los ingenieros, quienes implementaron la herramienta a partir de los diseños establecidos y las necesidades específicas del proyecto (para qué va a servir). De acuerdo con el equipo de lenguaje controlado, los ingenieros, más allá de entender un requerimiento, buscaron comprender la usabilidad, los usuarios y la finalidad del

proyecto; de esta manera, las reuniones interdisciplinares se vuelven clave para lograr los objetivos.

Esas primeras reuniones no fueron para hablar como tal de la herramienta, sino que en principio tocó hacer una discusión teórica del proceso que queríamos hacer. Tocó explicar a los compañeros por qué queríamos hacer un diccionario y cómo nos diferenciábamos del tesauro... y a partir de esa información, que yo llamaría teórica, fue que los compañeros dijeron: ahora sí entendemos, sí es necesario hacer un diccionario, vamos a esforzarnos por desarrollar esta herramienta...

Estas reuniones teóricas, según el equipo, llevaron a que los ingenieros se apropiaran conceptualmente de la herramienta logrando un resultado importante dentro del proyecto; son ingenieros que pueden hablar de campos semánticos y de términos, aunque no sea su fuerte disciplinar. Ahora bien, para lograr este lenguaje común fue necesario que ambos equipos (lenguaje controlado e ingenieros) usaran ejemplos que aterrizaran los tecnicismos a lenguajes cotidianos que permitieran comprender de la mejor manera el proceso:

Nos decían: por qué hablan de términos y no de palabras. Siempre recuerdo que se les daba el ejemplo: el término siempre está compuesto de varias palabras, por eso en el diccionario se habla de términos y no de palabras (por ejemplo, un término es si se habla de falsos positivos). De la misma manera, ellos tienen unos términos en inglés, y siempre los traducen y nos explican. Para que comprendieran qué es un campo semántico, tocó dar muchos ejemplos: que un campo semántico está conformado por varios términos... y queríamos que eso se reflejara en el diccionario, que son como temas que abarcan diferentes términos. Ellos con esa explicación, muy sencilla, se apropiaron conceptualmente de todo...

#### Tesauro y diccionario ¿cómo diferenciarlos?

Una de las dificultades a las que se enfrentó el equipo de lenguaje controlado fue transmitir las diferencias entre un tesauro y un diccionario; diferencias que no son fácilmente palpables si no se generan diálogos interdisciplinares que permitan comprenderlas. En este sentido, el proyecto fue una apuesta que implicó sentar unas diferencias y unos horizontes claros que permitieran identificar la propuesta social instaurada en la base de la idea de un diccionario. De acuerdo con esto, y según el equipo, el tesauro fue pensado para organizar la información misional (información que surge de la entidad para catalogar, para etiquetar y para normalizar); mientras que, por su parte, el diccionario fue pensado para la gente, para que la ciudadanía pudiera ubicarse conceptualmente en el trabajo realizado por la Comisión de la Verdad.

Esta diferencia da entrada a lo que es el diccionario. Según el equipo de lenguaje controlado, el diccionario es una herramienta pedagógica especializada para la comprensión del conflicto por parte de la ciudadanía; y es especializada porque son

los investigadores de la Comisión de la Verdad los que ayudaron en la construcción de las definiciones y en la revisión de los términos. En este sentido, el diccionario tiene dos vías: especializado, porque tiene el enfoque investigativo; y pedagógico, porque busca que la ciudadanía pueda comprender el conflicto armado colombiano a partir de una serie de términos y de campos semánticos.

#### Apuestas y aprendizajes

Según el equipo de lenguaje controlado, llegar al diccionario implicó identificar qué otros diccionarios especializados existen en el país; sin embargo, y según relata el equipo, sobre el conflicto colombiano los diccionarios están más centrados en aspectos jurídicos y hechos victimizantes. En este sentido, el diccionario legado por la Comisión de la Verdad se convierte en el primer intento que hay, a nivel nacional, de un diccionario con una terminología distinta que narra desde otras perspectivas el conflicto.

Uno dirá: gestor cultural qué tiene que ver con violencia. Sin embargo, cuando se miran las definiciones de un diccionario de estos y se mira gestor cultural, se puede ver esa definición aterrizada a cómo los gestores culturales han tenido que transformar su quehacer en medio de contextos de conflicto...

Una de las apuestas, entonces, tiene que ver con relacionar palabras y términos que, en apariencia, no tienen nada que ver con el conflicto armado pero, al mirarlas en contexto, dicen mucho de las transformaciones, dificultades y procesos que ha traído consigo el conflicto colombiano.

#### Aportes a la ciudadanía

Uno de los aportes que el equipo señala como importante es la posibilidad de ofrecer un diccionario especializado: un diccionario construido con base en las experiencias de investigación de la Comisión de la Verdad; hecho que implica que haya una terminología que no se encuentra en otros lugares o en otras herramientas. Este punto hace que el diccionario sea una herramienta de referencia para la ciudadanía: facilita la lectura del Informe Final y la exploración del archivo del esclarecimiento de la Comisión de la Verdad. Adicionalmente, y en consecuencia, el aporte más grande es la posibilidad de comprender el conflicto a partir de miradas tal vez inéditas.

Nosotros asumimos que, como hemos vivido en el conflicto, entendemos por qué se da el conflicto, y no es tan así... la mayoría de nosotros no lo entendemos. La palabra es como la unidad mínima de significado, y si nosotros tenemos significados claros sobre la vida, sobre nuestra experiencia, seguramente vamos a comprender mucho mejor las cosas; no sé si tan poéticamente como lo diría un lingüista o un

semántico, pero sé que la palabra es la que a nosotros nos empieza a formar el pensamiento, entender lo que la palabra nos quiere decir... y en ese sentido, los diccionarios cobran todo el sentido...

El diccionario es la posibilidad que tienen los investigadores de exponer el trabajo que hicieron durante todo el mandato de la Comisión de la Verdad; y es la posibilidad que tiene la ciudadanía para conocer esos puntos de vista que dan cuenta de ese otro conflicto que muchas veces pasa inadvertido (por ejemplo, la posibilidad de entender otras perspectivas en el tema del narcotráfico). En otras palabras, y según el equipo de lenguaje controlado, el diccionario brinda la posibilidad de salir de esos lugares comunes que han definido una sola mirada para entender el conflicto; es la posibilidad de que muchas situaciones se nombren, se conozcan y se reconozcan como parte del conflicto.

Este último punto da entrada a otro de los aportes resaltados por el equipo de lenguaje controlado: más allá de ser un diccionario normativo (cómo debería ser nombrado el homicidio, por ejemplo), la herramienta permite que las personas entiendan cómo la Comisión de la Verdad entendió el conflicto, explicando, en un lenguaje muy cotidiano, cómo se pensaron ciertas categorías de análisis y cómo se llegó a ciertas ideas. Es por ello que, resalta el equipo, el diccionario fue una construcción colaborativa, a partir de un equipo interdisciplinario.

Queremos salirnos un poco de eso, y por eso metimos todo lo de los casos ilustrativos, porque queríamos aterrizar en el discurso de las víctimas, de los actores, de los responsables, cómo se dan las cosas que están ahí plasmadas... eso es importante resaltarlo porque si la gente está esperando que la Comisión entregue un diccionario normativo, eso nunca va a pasar; esto no es un diccionario de la lengua, es un diccionario de un contexto que tenemos que vivir los colombianos, y que una entidad que investigó sobre eso quiere ofrecerle a la ciudadanía...

#### **Expectativas**

La ventaja de que el diccionario sea digital es que puede seguir creciendo con nuevos términos. Una de las expectativas del equipo es que el diccionario tenga muchas más entradas, resultado de otras investigaciones; asimismo, que pueda generar espacios de retroalimentación que complejicen los términos y campos semánticos que ya se han definido. En últimas, se quiere que la herramienta no deje de crecer, no quede estática, y realmente genere posibilidades de diálogo con comunidades, universidades e instituciones públicas y privadas. En este sentido, el diccionario debería tener la posibilidad de generar comentarios para refutar y/o reforzar las discusiones que se proponen; es decir, que la herramienta tenga un componente interactivo que permita el diálogo ciudadano.