



Entramado
ISSN: 1900-3803
ISSN: 2539-0279
Universidad Libre de Cali

Generación semiautomática de red semántica para describir comunidades víctimas del conflicto armado colombiano *

Mejía, José Soto; Gómez Suta, Manuela; Zuluaga Ramírez, Carlos Mauricio; Chavarro Porras, Julio Cesar
Generación semiautomática de red semántica para describir comunidades víctimas del conflicto armado colombiano *

Entramado, vol. 16, núm. 1, 2020

Universidad Libre de Cali

Disponible en: <http://www.redalyc.org/articulo.oa?id=265464211019>

DOI: 10.18041/1900-3803/entramado.1.6114

Generación semiautomática de red semántica para describir comunidades víctimas del conflicto armado colombiano *

Semi-automatic generation of semantic network to describe communities' victims of the Colombian armed conflict

Geração semiautomática de uma rede semântica para descrever as comunidades vítimas do conflito armado colombiano

José Soto Mejía ** jomejia@utp.edu.co

Universidad Tecnológica de Pereira, Colombia

ORCID: <http://orcid.org/0000-0002-0205-6863>

Manuela Gómez Suta *** madegomez@utp.edu.co

Universidad Tecnológica de Pereira, Colombia

ORCID: <http://orcid.org/0000-0001-7108-2689>

Carlos Mauricio Zuluaga Ramírez **** cmzuluaga@utp.edu.co

Universidad Tecnológica de Pereira, Colombia

ORCID: <http://orcid.org/0000-0002-8981-7409>

Julio Cesar Chavarro Porras ***** jchavar@utp.edu.co

Universidad Tecnológica de Pereira, Colombia

ORCID: <http://orcid.org/0000-0002-8981-7409>

Entramado, vol. 16, núm. 1, 2020

Universidad Libre de Cali

Recepción: 07 Mayo 2019

Aprobación: 20 Diciembre 2019

DOI: 10.18041/1900-3803/entramado.1.6114

CC BY-NC-SA

RESUMEN: La situación de violencia en Colombia ha producido dinámicas que escapan al control del Estado y dificultan la construcción de la agenda pública de las instituciones estatales, ya que estas poseen limitaciones para describir el entorno social y político que influye en el territorio. Para suplir estas limitaciones, es pertinente generar redes semánticas que describan información sobre las comunidades víctimas del conflicto armado. Lo anterior permitirá generar un insumo que las instituciones y autoridades locales puedan emplear para reconocer pilares en la formación de políticas públicas coherentes a las necesidades de cada comunidad. En el presente artículo, se expone el proceso para la generación semiautomática de una red semántica a partir del tratamiento de datos textuales. Para tal fin, se han empleado herramientas de minería de textos y técnicas de análisis multivariado. La red semántica generada es una primera aproximación para la descripción de las características del caso de estudio de la comunidad de Arauca durante los años 2013-2018.

PALABRAS CLAVE: Red semántica, políticas públicas, minería de textos, técnicas de análisis multivariado.

ABSTRACT: The situation of violence in Colombia has produced dynamics that are beyond the control of the State and make it difficult for state institutions to build the public agenda. These institutions have limitations to describe the social and political environment that influence the territory. To overcome these limitations, it is pertinent to generate semantic networks that describe information about communities victims of the armed conflict and, thus, generate an input that local institutions and authorities can use to recognize pillars in the formation of public policies consistent with the needs of each community. In this document is presented the process for the semi-automatic

generation of a semantic network from the processing of textual data, using text mining tools and multivariate analysis techniques. The semantic network generated is a first approximation for the description of the characteristics of the Arauca community during the years 2013-2018, which was selected as a case study.

KEYWORDS: Semantic network, public policy, text mining, multivariate analysis techniques.

RESUMO: A situação de violência na Colômbia tem produzido dinâmicas que estão além do controle do Estado e dificultam a construção de uma agenda pública por parte das instituições estatais, já que estas têm limitações na descrição do ambiente social e político que influencia o território. Para superar essas limitações, é pertinente gerar redes semânticas que descrevam as informações sobre as comunidades vítimas do conflito armado. Isto permitirá gerar um input que as instituições e autoridades locais possam utilizar para reconhecer pilares na formação de políticas públicas que sejam consistentes com as necessidades de cada comunidade. Neste artigo, é apresentado o processo para a geração semiautomática de uma rede semântica baseada no processamento de dados textuais. Para isso, foram utilizadas ferramentas de mineração de texto e técnicas de análise multivariada. A rede semântica gerada é uma primeira abordagem para descrever as características do estudo de caso da comunidade Arauca durante os anos 2013-2018.

PALAVRAS-CHAVE: Rede semântica, políticas públicas, mineração de texto, técnicas de análise multivariada.

1. Introducción

La situación de violencia en Colombia posee un eje político que se encuentra ligado a la acción de las guerrillas revolucionarias y de las fuerzas que las enfrentan. Sin embargo, también existen otras dimensiones que interfieren en ello como son: 1) el cultivo y tráfico de droga, 2) el crecimiento de los grupos armados y 3) la desorganización social que favorece la violencia. Estos aspectos han producido dinámicas que escapan al control del Estado, debilitan sus estructuras y producen que su autoridad desaparezca en gran parte del territorio. En esta dirección, las instituciones y autoridades locales son las encargadas de edificar la agenda pública mediante: 1) la caracterización del escenario, y 2) la formación de una visión compartida y ética de gobernanza que incluya las diferentes facciones de la sociedad con el fin de 3) definir las estrategias y arreglos que se deben efectuar para coordinar y armonizar las políticas y los recursos hacia los fines comunes del desarrollo sostenible y la paz (Alzate y Romo, 2014).

Adicional a la desaparición de la autoridad del Estado, la naturaleza de las instituciones y autoridades locales está guiada por un modelo de gestión pública cuya premisa es satisfacer las necesidades de los ciudadanos a través de la efectividad en el diseño y la aplicación de políticas para la profesionalización de la administración de lo público (Christensen y Laegreid, 2005). En este orden de ideas, la edificación de la agenda pública queda enmarcada en las competencias de los funcionarios delegados y no propiamente en un espacio en el cual diversos actores sociales, vinculados o no al gobierno, evidencien sus posiciones, argumentos y necesidades frente a una misma situación (Alzate y Romo, 2014). Lo anterior obstaculiza el proceso de formación de la política pública ya que, como señala Kingdon (1984), durante este proceso es imperioso reconocer y analizar: 1) los elementos del entorno social y político que

influyen en la configuración de una política pública y 2) las preferencias, creencias e intereses de los actores involucrados.

Sin embargo, el Estado, consciente de su responsabilidad, ha buscado herramientas que le permitan caracterizar y describir la comunidad de interés a la luz de la formación de la agenda pública. A las instituciones estatales se les han ofrecido herramientas cualitativas y estadísticas que les ayuden a obtener información sobre la comunidad de interés y sus problemáticas. No obstante, estos insumos, aun siendo valiosos para la comprensión profunda del medio sociocultural, presentan una serie de falencias, a saber: los proyectos antropológicos, sociológicos y etnográficos cualitativos, que pueden llevar años, requieren masivas investigaciones detalladas sobre una región o un grupo y sus resultados no son del todo escalables porque los métodos utilizados son, a menudo, específicos del contexto (Carley, Bigrigg y Diallo, 2012); por otro lado, las investigaciones estadísticas parten del análisis de observaciones en pocos puntos o un solo punto en el tiempo en un intervalo limitado (Ilgen y Hulin, 2000).

Desde el marco nacional colombiano, no se identifican propuestas que posibiliten a las instituciones estatales caracterizar el entorno de la comunidad que será afectada por la política pública. En el escenario internacional, investigadores, analistas y encargados de formular estas políticas hacen uso de redes semánticas que facilitan la comprensión rápida sobre la disposición de la comunidad afectada y los recientes cambios en el entorno sociocultural (Carley *et al.*, 2012). Estas redes constituyen formalismos que capturan el conocimiento humano en una abstracción adecuada para el procesamiento por medio de entidades inteligentes como las computadoras (Brachman, 1977).

La red semántica es un conjunto de términos interrelacionados (Carley, *et al.*, 2012) donde se representan objetos individuales, categorías de objetos y relaciones; por ende, "permite la representación de cualquier cosa que puede establecerse en lenguaje natural" (Quillian, 1969, p. 460). En consecuencia, una red es un esquema adecuado para representar las complejas dinámicas del conflicto ya que estas son comúnmente descritas mediante relatos escritos.

La propuesta de emplear la red semántica se sustenta al reconocer que desde las investigaciones de carácter cualitativo se procesan y analizan textos de fuente abierta a través de los cuales etnógrafos, antropólogos y sociólogos comprender la realidad del entorno de interés (Carley, *et al.*, 2012). No obstante, el actual volumen de textos que se debe analizar para caracterizar una comunidad sugiere la necesidad de un nuevo enfoque automatizado (van Holt, Johnson, Brinkley, Carley y Caspersen, 2012). Así, el conocimiento puede esquematizarse en un formato analizable para las entidades inteligentes.

En este sentido, los autores del presente artículo han identificado que la red semántica es una representación oportuna para describir información concerniente a las comunidades que han sido víctimas del conflicto armado colombiano y las dinámicas que este conlleva. Asimismo, esta representación constituye un insumo para la formación de políticas

públicas y planes de desarrollo que las instituciones y autoridades locales elaboran en calidad de actores en la construcción de la paz.

Por lo anterior, el objetivo del presente artículo es documentar el proceso de la generación de una red semántica que describe la comunidad de Arauca, víctima del conflicto armado, durante el periodo 2013-2018. La red se generó siguiendo la metodología planteada por Uschold y King (1995) y el método de Aussenac-Gilles, Biebow y Szulman (2002). Considerando este marco metodológico, se ha propuesto emplear herramientas de minería de textos y técnicas de análisis multivariado para su construcción semiautomática. En especial, se ha realizado una propuesta para la identificación de los conceptos que pertenecen a la red y facilitan describir la comunidad de interés.

En este sentido, el estudio divulgado mediante el presente documento es innovador dado que 1) a nivel nacional no se han utilizado las bondades de la red semántica para describir las características de escenarios afectados por el conflicto; con el propósito de brindar un insumo para que las autoridades logren mayor sincronía entre las políticas que se formulan, se ejecutan y el impacto de estas; y 2) la metodología empleada no restringe la identificación de conceptos a experiencias humanas, modelos teóricos anteriores, diccionarios semánticos, bases de conocimiento o ponderación de términos. Lo anterior, representa una marcada diferencia de las propuestas actuales.

Es oportuno aclarar que el empleo de la red semántica es un complemento a los actuales enfoques utilizados para comprender el perfil sociocultural de las comunidades de interés, no un sustituto de dicho perfil. En este sentido, dicha red reconoce rápidamente patrones y asociaciones semánticas de la información textual, sin limitar las posibles relaciones que puedan surgir de los datos examinados; considera diferentes variables y las relaciones entre ellas, llenando así una brecha clave en la evaluación tradicional de sistemas sociales; proporciona orientación sobre los principales conceptos presentes en una serie de datos (van Holt, *et al.*, 2012); y brinda una visión del uso del lenguaje, que hacen los miembros de la sociedad, con el fin de construir un consenso (Carley, *et al.*, 2012).

El siguiente apartado expone el sustento teórico que justifica la ejecución del estudio documentado así como la naturaleza metodológica del mismo.

2. Marco teórico

Roth Deubel (2002) expone que una política pública existe siempre y cuando instituciones estatales asuman total o parcialmente la tarea de alcanzar objetivos estimados como deseables o necesarios dada una situación percibida como problemática. En esta dirección, "las acciones gubernamentales se desarrollan para y a través de actores sociales, es decir, son seres humanos quienes conciben, deciden e implementan las políticas públicas y cuyos destinatarios son las personas" (Dubnick y Bardes, 1983, p. 5).

De este modo, las políticas públicas están encaminadas a 1) determinar los objetivos colectivos que se consideran como deseables o necesarios; 2) examinar los medios y actividades, total o parcialmente, que una institución u organización gubernamental debe seguir para el cumplimiento de los objetivos establecidos; y, además, 3) analizar los resultados esperados e imprevistos de las acciones encaminadas (Quintana, 2014).

En este orden, es preciso señalar que la política pública alcanza sus objetivos a través de una secuencia lógica denominada ciclo de la política, la cual está conformada por las siguientes etapas:

- **Identificación y estructuración del problema:** Consiste en la construcción de la contrariedad social como problema que necesita una intervención pública, es decir, es el reconocimiento de la realidad problemática y su comprensión. Conlleva la lectura y la discusión que los diferentes actores realizan del problema desde su perspectiva (Kingdon, 1984). Esta primera etapa tiene como finalidad describir la situación a abordar; por ende, en esta se consulta a diferentes individuos para comprender cómo funciona el problema, cuáles son las personas involucradas y cuáles son las lógicas comportamentales que se presentan. La identificación y estructuración del problema debe ser asumida por las instituciones y centros de investigación con el fin de proporcionar información real y eficaz que posibilite caracterizar la problemática apropiadamente. Esto permitirá generar diversos análisis y discusiones argumentados antes de determinar cómo regular el problema.
- **Formulación de soluciones:** Permite seleccionar la alternativa más factible para la disminución de la tensión entre el ser (la situación actual) y el deber ser (la situación deseada). La esquematización de una solución se fundamenta en clarificar, mediante el primer paso del ciclo de la política, cuáles son las necesidades de los actores involucrados en el problema; de esta forma, en esta etapa se busca constituir el objetivo de la política y, subsecuentemente, las actividades, el presupuesto y las mediciones para alcanzar la meta propuesta (Quintana, 2014).
- **Implementación y evaluación de soluciones:** Consiste en transformar la propuesta teórica en hechos. Kingdon (1984) establece que la implementación debe ejecutarse paralelamente con la evaluación para desarrollar constantemente la redefinición de las características del problema. Así, se pueden adaptar las actividades a ejecutar para que sean soluciones a problemáticas reales de la población. En este sentido, uno de los criterios para evaluar los resultados de la política pública es la capacidad, por parte de las instituciones, de recoger información pertinente sobre el problema a tratar para, luego, ajustar la política en función de estos elementos.

Conforme a lo anterior, el concepto de política pública y las etapas que conlleva desarrollarla permiten visualizar que la identificación y estructuración del problema son cruciales para la correcta formulación de dicha política pública, del adecuado establecimiento de metas y de la determinación de las actividades a ejecutar. Por su parte, esto incide en la evaluación que haga la institución de su capacidad para describir el problema que se desea resolver. Por consiguiente, es relevante reconocer las herramientas empleadas por las instituciones para describir las regiones y sus respectivos problemas ya que, a través de estas, se recolecta y analiza la información con la cual se formulan las políticas públicas.

La mayoría de las instituciones nacionales (el Observatorio de Derechos Humanos y Derecho Internacional Humanitario -SNIDH-, el Departamento Nacional de Planeación -DNP-, la Misión para la Transformación del Campo y el Centro de Recursos para Análisis de Conflictos -CERAC-) emplean estudios etnográficos de carácter cualitativo. Otras instituciones (el Departamento Administrativo Nacional de Estadística -DANE-, el DNP y el CERAC) parten de investigaciones de naturaleza estadística. Por lo anterior, aun cuando los insumos de información son valiosos, estos requieren cantidades masivas de minuciosa y detallada investigación sobre una región o un grupo determinado. De este modo, a menudo los métodos utilizados son específicos para el contexto (Carley, *et al.*, 2012), conllevan años, sus resultados no pueden ser escalables, o brindan descripciones estáticas que surgen del análisis de observaciones en un solo punto en el tiempo o de pocas observaciones en un intervalo limitado (Ilgen y Hulin, 2000).

Otra metodología no explorada a nivel nacional colombiano, pero que se lleva a cabo en otros contextos internacionales, es la utilización de redes semánticas (Carley, *et al.*, 2012; van Holt, *et al.*, 2012). Estas se entienden desde el trabajo de Quillian (1969) como un mecanismo asociativo general para catalogar los significados de las palabras. Dicho mecanismo permite desarrollar un esquema de codificación formal y uniforme que refleja la estructura semántica del fenómeno a modelar para que este pueda ser manejable mediante procedimientos generales. En el caso particular de emplear las redes semánticas como instrumento de apoyo en la formación de políticas públicas, el fenómeno a modelar o representar es la comunidad que padecerá la ejecución de la agenda pública.

El esquema propuesto por Quillian (1969) representa, mediante nodos, las afirmaciones factibles del fenómeno a modelar y, por medio de enlaces (arcos), la información asociada. En esta dirección, para la modelación del conocimiento mediante una red semántica, se requiere de una lista de hechos y un índice de conexiones asociativas que relacionan estos hechos. Este formato de representación simple permite codificar mucha información fáctica y facilita la interpretación del fenómeno representado (Brachman, 1977).

No obstante, no es apropiado llamar red semántica a cualquier representación que entrelace nodos con arcos ya que estos son solo entidades descriptivas que no incluyen la importancia epistemológica de

la red semántica (Brachman, 1979). En este sentido, la semántica de la red radica en que emplea los campos de los términos presentes en ella para expresar conceptos (Kashyap y Borgida, 2003), objetos, atributos y relaciones del fenómeno que se modela (Brachman, 1977).

Una red posee características semánticas porque "intenta combinar en un único mecanismo la capacidad no solo de almacenar conocimiento factual sino también de modelar las conexiones asociativas que hacen que ciertos elementos de información sean accesibles desde otros" (Woods, 1975, p. 106). De esta manera, en una red semántica, los nodos y enlaces forman conceptos del fenómeno de interés, semejantes a los presentados en el diccionario, donde los significados se construyen a partir de secuencias de palabras definidas en otro lugar del mismo diccionario (Brachman, 1979). Por ende, el significado completo de un concepto se puede comprender a través de un rastreo exhaustivo de todos los enlaces y nodos unidos al nodo principal en la construcción de la definición.

La forma en que la definición de un concepto se deriva y se relaciona con todas las demás entidades es la clave para hacer un uso inteligente del conocimiento incorporado en la red (Brachman, 1977). En este sentido, las redes semánticas brindan la oportunidad de examinar un posible esquema de representación para las entidades intencionales, y conllevan la necesidad de caracterizar los enlaces que permitan reconocer las modificaciones intencionales y las aserciones contingentes (Kashyap y Borgida, 2003).

En conclusión, la red semántica es candidata para la descripción de una comunidad víctima del conflicto armado colombiano, que soportará la potencial ejecución de una política pública, porque esta representación del conocimiento usa el lenguaje natural de forma axiomatizada. De esta manera, los nodos y enlaces plantean la intensión del fenómeno a simbolizar. Esto permite reconocer las características de la comunidad que se desea estudiar. Dadas las ventajas de emplear redes semánticas para la descripción de un entorno social, diferentes estudios de carácter internacional se han focalizado en la construcción de estas. A continuación, se mencionan algunos de estos esfuerzos.

El modelado de conocimiento mediante redes semánticas ha sido usado para describir comunidades con marcadas dificultades políticas y sociales. Por ejemplo, en el 2012, van Holt et al. caracterizaron, a través de una red de información construida con fuentes textuales, el conflicto étnico de Sudán entre 2003 y 2010. De esta forma, los autores comprobaron que la situación de paz en este país estaba condicionada a la relación entre la posición geoespacial de los grupos étnicos y los recursos del medio ambiente. Igualmente, se halla el trabajo de Pfeffer y Carley (2012) quienes procesaron información proveniente de periódicos con el objetivo de construir redes que describieron y comprendieron las causas de diferentes acontecimientos de la Primavera Árabe.

Bajo esta misma línea, se encuentra la investigación de Carley *et al.* (2012) quien obtuvo como resultados una red de información textual que caracterizaba la dinámica de Afganistán durante 2010 y 2013. De esta manera, los autores brindaron una herramienta con la cual se podían

responder preguntas sobre quiénes eran los actores clave; cuáles eran los temas relevantes, los sentimientos, los recursos y las actividades de la comunidad; además permitió identificar qué papel desempeñaban los diversos actores. Asimismo, las redes semánticas han sido aprovechadas para reconocer la dinámica de grupos de interés. Un ejemplo de esto es el trabajo de Hutchins y Benham-Hutchins (2010) quienes realizaron un análisis multimodal de datos estructurados y no estructurados para entender las relaciones entre las personas que conformaban la red criminal de narcotráfico en parte de los Estados Unidos de América, así como los procesos de financiamiento, suministro de transporte y reclutamiento.

De este modo, los trabajos previamente comentados construyeron, respectivamente, un producto que facilitó comprender la dinámica de la región de interés. Estos representan evidencias empíricas de la red semántica como un adecuado mecanismo para caracterizar los elementos del entorno social y los actores que interactúan en la formación de políticas públicas. De esta manera, estos estudios brindaron un sustento clave para que los autores de la presente investigación hayan decidido generar una red semántica que describa una comunidad víctima de conflicto en Colombia.

La red semántica propuesta surge del empleo de datos textuales, ya que los textos contienen información que se puede extraer y codificar como una red que vincula actores, sentimientos, recursos, actividades y lugares (Carley *et al.*, 2012). Este enfoque es posible dada la gran cantidad de información digital disponible, en gran parte, de forma no estructurada o de texto. Igualmente, se justifica seleccionar los recursos textuales como fuente para la construcción de una red que permita describir a una comunidad de interés porque, durante la creación de etnografías, la mayoría de los investigadores evalúan los datos de texto de fuente abierta para proporcionar orientación sobre las principales características de una región (van Holt *et al.*, 2012).

Es importante recalcar que la eficiencia de la red semántica es dependiente de la metodología y los procedimientos que hayan sido empleados para su construcción. Gómez-Pérez, Fernández-López y Corcho (2004) exponen que la edificación de una red semántica a partir de datos textuales conlleva: seleccionar el corpus de trabajo, instaurar el vocabulario, identificar conceptos entre los términos relevantes, promulgar relaciones entre los términos y conceptos, y esquematizar la red en un lenguaje de preferencia legible por entidades computacionales inteligentes.

En este sentido, parte de los estudios antes mencionados plantean los siguientes procedimientos para la creación de una red semántica:

- Selección del vocabulario y clasificación de los términos en categorías previamente establecidas, considerando que estos procesos los realiza un experto (o varios de ellos) de forma manual (van Holt *et al.*, 2012). En algunas ocasiones, previamente a la formación del vocabulario, se ponderan los términos siguiendo una medida seleccionada por los investigadores; después, se filtran aquellos que posean un valor menor a un

límite deseado (Carley *et al.*, 2012). Asimismo, existen proyectos que usan un corpus etiquetado, es decir, textos que ya han sido indexados en categorías predefinidas. De esta manera, la selección del vocabulario está atada a la información etiquetada en los documentos (Pfeffer y Carley, 2012). Por otra parte, investigaciones como las de Hutchins y Benham-Hutchins (2010) y Zhu, Wang y He (2016) utilizan bases de datos que facilitan el reconocimiento de entidades nombradas dentro del corpus, sin necesidad de que los documentos estén etiquetados.

- Establecimiento de conceptos mediante la ponderación de los términos que pertenecen al vocabulario. En otras palabras, a partir de una métrica seleccionada, los investigadores jerarquizan los términos y aquellos con mayor ponderación son reconocidos como conceptos cuya semántica recoge otras expresiones presentes en los textos analizados (Carley *et al.*, 2012; Zhu *et al.*, 2016).
- Las relaciones entre términos y conceptos se dan mediante análisis de coocurrencia en el corpus (Pfeffer y Carley, 2012; van Holt *et al.*, 2012; Zhu *et al.*, 2016). En este sentido, se parte del supuesto de que los términos son mutuamente dependientes cuando su utilización conjunta es muy común.

Por su parte, cabe mencionar que estas propuestas metodológicas poseen falencias sobre los tiempos de generación. Por ejemplo, la formación manual del vocabulario retrasa la construcción de la red (Zhu *et al.*, 2016) y, por ende, aminora una de sus principales ventajas en comparación con otros enfoques de investigación que se caracterizan por consumir mucho tiempo. Igualmente, el etiquetado manual de texto previo al procesamiento de datos limita la escalabilidad de la red, ya que cada nuevo texto que se desee procesar debe estar indexado (Carley *et al.*, 2012). Por lo anterior, la construcción de la red se ve condicionada a la presencia de personal capacitado que estudie los documentos anticipadamente.

Al emplear bases de conocimiento o diccionarios para la identificación de términos y conceptos, se restringen los componentes por hallar porque estos dependen de la naturaleza y diversidad de la información que se encuentre en la base de datos; frente a esto, es recomendable adaptar las bases de conocimiento para que respondan a las necesidades del dominio particular que se está analizando (Deerwester, Dumais, Furnas y Landauer, 1990).

La calidad de los conceptos identificados mediante la ponderación de términos es dependiente del 1) preprocesamiento de texto y 2) de la métrica utilizada para jerarquizar la relevancia de las palabras (Astrakhantsev, Fedorenko y Turdakov, 2015). Igualmente, no existe consenso sobre si, dada una métrica, el término con mayor ponderación es más relevante que lo anterior, a partir de otra medida porque, aun cuando existen muchas investigaciones sobre técnicas para la ponderación de términos, es poco el consenso sobre cuál métrica es apropiada en relación

con el tipo de corpus, el largo del documento y el idioma que se analice (Echeverry-Correa, 2015).

Por otra parte, establecer las relaciones entre las unidades textuales de la red mediante medidas de coocurrencia no permite identificar conexiones subyacentes en el texto, es decir, aquellos vínculos implícitos; ya que solo se analiza la información textual del corpus. Además, los resultados del análisis de coocurrencias son dependientes del tamaño de la ventana de palabras (Astrakhantsev *et al.*, 2015). Otra falencia de emplear medidas de coocurrencias subyace en el desconocimiento del tipo de relaciones que se identifican. Por ende, la interpretación de la red semántica es ambigua y se pierde la virtud de axiomatización que posee (Zouaq, Gasevic y Hatala, 2012).

Dadas las debilidades de los actuales planteamientos para construir redes semánticas, se propone emplear herramientas de minería de textos y técnicas de análisis multivariado con el fin de superar parte de las falencias de las investigaciones mencionadas. En esta dirección, se propone la generación semiautomática de la red semántica toda vez que 1) se estudian las alternativas de escritura establecidas mediante el análisis de la matriz de correlación de los términos extraídos de los textos; 2) se selecciona el vocabulario a partir de una ponderación de los términos y la verificación manual; 3) se identifican los conceptos al mejorar la calidad de la "matriz término documento (MTD)" a través del análisis latente semántico y se ejecuta la técnica de análisis factorial, 4) se etiquetan manualmente los factores (conceptos) resultantes; 5) se hallan, por medio del análisis de componentes principales, el conjunto de documentos asociados a cada concepto etiquetado para procesarlos nuevamente e identificar los subconceptos que describen características del concepto previamente nombrado; y, por último, 6) se reconocen las relaciones entre término y conceptos mediante el estudio e interpretación de la composición de cada factor principal para, después, etiquetar dicha relación gracias al juicio de expertos.

En este sentido, la contribución metodológica de esta propuesta consiste en la constitución de conceptos y subconceptos a través de técnicas de análisis factorial y análisis de componentes principales para, bajo este esquema, identificar relaciones entre las unidades textuales de la red. De este modo, este desarrollo propuesto se diferencia de los mencionados previamente porque no restringe la identificación de conceptos a experiencias humanas, modelos teóricos anteriores, diccionarios semánticos, bases de conocimiento o ponderación de términos.

3. Metodología

Desde finales del siglo XX, se han propuesto diferentes enfoques para construir redes semánticas desde cero y reutilizar formalismos de representación del conocimiento con fin de crear nuevos esquemas o generar herramientas semiautomáticas que faciliten la adquisición del conocimiento.

Gómez-Pérez *et al.* (2004) realizan una revisión de metodologías y métodos para la construcción de esquemas de representación del conocimiento. De esta forma, reconocen que la metodología a seguir se define, dependiendo de los recursos con los cuales se cuente al iniciar la generación de la red. Dado que el objeto de estudio de la presente investigación no contaba con ningún tipo de representación de conocimiento que describiera comunidades víctimas del conflicto, el insumo inicial fue el conjunto de textos a procesar (también denominado corpus) en lenguaje natural. Así, la metodología seleccionada fue la propuesta por Uschold y King (1995).

La metodología seleccionada no exige que la extracción de información se realice con ayuda de una base de conocimiento, modelos previos o representaciones existentes; por el contrario, plantea el reconocimiento de conceptos y relaciones como una actividad. Dicho de otro modo, la conceptualización de la red semántica se debe diseñar siguiendo algún paradigma de recuperación de información (Uschold y King, 1995). La estrategia empleada durante la ejecución del presente proyecto es conocida como *middle-out* (Gómez-Pérez *et al.*, 2004), y consiste en identificar el núcleo de términos base para, luego, comenzar a especificarlos y generalizarlos según lo considere necesario el experto.

La metodología de Uschold y King (1995) propone una etapa de planeación, construcción y evaluación de la red generada. Durante la planeación, se establece el objetivo y la aplicación de la red. La construcción consiste en capturar los conceptos y relaciones que se esquematizan en la red. Por último, la evaluación hace referencia a validar la funcionalidad de la representación construida.

La selección del método a emplear, al igual que la metodología, depende de los recursos con que cuente el proyecto. En este sentido, se eligió el método propuesto por Aussenac-Gilles *et al.* (2002) quienes plantean una serie para determinar la conceptualización de la red semántica a partir del estudio de un corpus. Este método propone tres niveles, a saber: 1) nivel lingüístico, el cual se refiere a hallar términos y relaciones léxicas de textos mediante el análisis lingüístico; 2) nivel de normalización, en el cual se crean clústeres entre los términos, formando los conceptos y relaciones conceptuales. De esta manera, se parte de relaciones léxicas y términos para llegar a la generación de relaciones semánticas y conceptos. Por último, 3) en el nivel formal, los conceptos y relaciones establecidas se formalizan por medio de un esquema gráfico.

La constitución del "corpus" se refiere a establecer aquellos documentos que contienen información sobre la comunidad de interés. Por su parte, el estudio lingüístico es la selección de herramientas y técnicas lingüísticas adecuadas para formar el vocabulario a partir del cual se identifican los conceptos que se expresan en la red semántica. Durante la normalización, se genera el modelo conceptual. Por último, en la formalización, se representan con un esquema gráfico las relaciones entre los conceptos y términos identificados (La Figura 1 resume el método seleccionado).

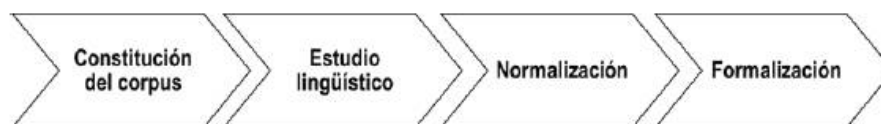


Figura 1

Método propuesto por Aussenac-Gilles *et al.* (2002)

Fuente: Adaptado de "Metodologías y métodos para la construcción de ontologías", de Gómez Pérez *et al.* (2004), *Ontological Engineering with examples from the areas of Knowledge Management e-Commerce and the Semantic Web*, p. 120.

Al construir representaciones de conocimiento como las redes semánticas, es usual que el cuello de botella sea la adquisición del conocimiento que describa el fenómeno a modelar, así como la conversión de estos datos a conceptos y relaciones relevantes (Gómez-Pérez *et al.*, 2004). Con el fin de reducir esta dificultad, diversas investigaciones (Carley *et al.*, 2012; van Holt *et al.*, 2012; Zhu *et al.*, 2016) han empleado herramientas de minería de textos (MT) que buscan minimizar el esfuerzo humano y el tiempo necesario para desarrollar redes semánticas. Estas herramientas brindan, con una precisión razonable, la posibilidad de identificar automática o semiautomáticamente diversas clases de entidades en documentos de texto, y permiten reconocer relaciones que se afirman entre ellas.

De esta forma, la MT busca brindar a las máquinas la capacidad, propia del ser humano, de distinguir y aplicar patrones lingüísticos a los datos no estructurados para comprender expresiones como jergas, variaciones de ortografía y significados contextuales (Kao y Poteet, 2007) que, sumado a la capacidad de las computadoras de procesar textos en grandes volúmenes y velocidad, facilita la adquisición de conocimiento necesario para la construcción de redes semánticas.

Conforme a lo anterior, la MT corresponde al "descubrimiento y la extracción de conocimiento interesante y no trivial del texto libre" (Kao y Poteet, 2007, p. 1). Así, esta realiza un análisis profundo de los datos no estructurados para hallar información y relaciones no evidentes (Carley *et al.*, 2012). Una de sus ventajas consiste en brindar herramientas para la administración de la alta dimensionalidad de los datos, considerando la relación que pueda darse entre los mismos. Además, la MT permite encontrar de forma eficiente patrones en archivos textuales en función de los contenidos (Fedorenko, Astrakhantsev, Turdakov y Solzhenitsyn, 2014).

Al emplear las herramientas de la MT y siguiendo el método de Aussenac-Gilles *et al.* (2002) para la generación semiautomática de la red semántica, se establecieron tres procesos y sus procedimientos asociados. A continuación, se describen cada uno de ellos.

Extracción de términos para la formación del vocabulario

Se enmarca en el proceso de la MT de extracción de información. En este punto, se transforman datos no estructurados en una base de datos

estructurada. Jacquemin y Bourigault (2012) exponen que el término es un dato textual mientras que el concepto es una arquitectura mental propia de un fenómeno.

En el tipo de extracción ejecutada no se restringió la selección de término a una lista de referencia. Los procedimientos de extracción de términos se enfocaron en localizar un conjunto específico de palabras que se hayan en documentos de interés (Lagutina, Mamedov, Lagutina, Paramonov, y Shchitov, 2016) para que dichas unidades léxicas proporcionasen insumos en los procesos de análisis, comprensión e interpretación del corpus (Jacquemin y Bourigault, 2012). Diversos estudios de extracción de términos, como los de Astrakhantsev *et al.* (2015) o Fedorenko *et al.* (2014), han expresado los siguientes procedimientos para desarrollar su recuperación, a saber:

- Establecimiento del corpus teniendo en cuenta el fenómeno que se desea modelar: En el caso documentado, el objeto a modelar estaba conformado por comunidades víctimas del conflicto armado colombiano; en especial, la comunidad de Arauca.
- Preprocesamiento de los datos textuales para transformar cada documento (del corpus) en un archivo plano dado que existen una serie de características del texto en lenguaje natural que pueden hacer que sea difícil procesarlo automáticamente (Manning y Schutza, 2000). En el estudio ejecutado se consideró tokenizar el corpus, normalizar los caracteres a minúscula, y eliminar la puntuación y los signos diacríticos.
- Colección de candidatos: Se refiere a obtener las unidades léxicas admisibles para ser términos (Astrakhantsev *et al.*, 2015). El estudio documentado consideró actividades como filtrar los caracteres que contuvieran números, determinar las alternativas de escritura, eliminar palabras de paradas y lematizar las unidades textuales resultantes. Para determinar las alternativas de escritura, se ponderaron las palabras siguiendo la métrica de frecuencia en el corpus (collection frequency -TC-) (Manning y Schutza, 2000). Después, se calculó la matriz de correlación de Pearson entre los términos. Aquellos que tenían una relación positiva perfecta fueron extraídos para ser analizados por los autores del presente manuscrito quienes, de forma manual, estudiaron que las alternativas de escritura identificadas tuvieran coherencia y sentido bajo el fenómeno que se modelaba.
- Cálculo de características: Se refiere a medir propiedades de los términos con el fin de ubicarlos en un espacio comparable (Astrakhantsev *et al.*, 2015; Fedorenko *et al.*, 2014). Para la generación de la red, se modelaron los términos bajo la hipótesis de la bolsa de palabras y se empleó la ponderación TD-IDF. Después, se esquematizaron los datos mediante la "matriz término documento (MTD)". En este sentido, se empleó el modelo de espacio vectorial (Manning y Schutza, 2000).
- Inferencia basada en características: Se realizó empleando el etiquetado manual del 25 % de los documentos empleados en el

corpus. En este procedimiento, las palabras etiquetadas fueron comparadas con aquellas extraídas y se calcularon medidas de precisión, reconocimiento y medida F (Lagutina *et al.*, 2016). De esta forma, se conformó el vocabulario con el que se reconocen unidades conceptuales.

Identificación y etiquetado de conceptos y subconceptos

Para manipular los datos se inicia con la MTD. No obstante, esta matriz es dispersa y poco informativa. Por lo tanto, se realiza una transformación mediante la técnica de análisis latente semántico, la cual permite recopilar todos los contextos dentro de los cuales aparecen los términos del vocabulario y reconocer factores comunes que representen conceptos subyacentes (Evangelopoulos, Zhang y Prybutok, 2012). Esta técnica asume que existe una estructura implícita en la coocurrencia de los términos y permite sobrellevar la suposición de independencia del modelo de espacio vectorial (Echeverry-Correa, 2015). Al realizar un análisis sobre la base de la muestra falible encontrada realmente, dicha técnica facilita la construcción de un sistema que prediga qué términos pertenecen a un concepto (Deerwester *et al.*, 1990).

En este punto, cabe acotar que se decidió realizar la transformación de la MTD, empleando la técnica de análisis latente semántico, para superar la falencia del modelo de espacio vectorial. Dicho modelo asume que no existe relación conceptual entre los términos (Echeverry-Correa, 2015). En la práctica, este supuesto del modelo de espacio vectorial no es cierto ya que en un documento, el texto es un orden secuencial de términos que guardan relación entre ellos con el fin de construir ideas. Por lo tanto, se presenta una relación conceptual entre los términos. Igualmente, no se empleó el modelo Latente Dirichlet Allocation (LDA) (Blei, NG y Jordan, 2003) porque este se enfoca en hallar tópicos subyacentes al corpus con el propósito de agrupar los documentos en relación con los tópicos descubiertos; es decir, el modelo LDA no permite obtener una versión más informativa de la MTD. Por ende, la aplicación del modelo LDA no hubiera suplido la necesidad de la investigación ejecutada.

Una vez planteada la matriz mejorada mediante el análisis latente semántico, se realizó un análisis factorial exploratorio con el fin de identificar los conceptos subyacentes a los términos ponderados en los diferentes documentos.

Se seleccionó esta técnica proveniente del análisis multivariado porque consiste en modelar la información relevante (representada en variables multivariantes, para el caso los términos del vocabulario) como proveniente de un número limitado de factores latentes que no son observables. No obstante, estos factores son mucho más interesantes para la comprensión teórica de las variables cuantificables (Uriel Jiménez y Aldás Manzano, 2005). De esta manera, el análisis factorial facilita la formación de conceptos cuando estos surgen de los factores latentes que reúnen algunas variables originales.

Dentro del análisis factorial, se siguió el criterio de la media aritmética para establecer el número de factores latentes. Asimismo, se empleó la rotación varimax que preserva la incorrelación entre los factores (conceptos) y favorece la interpretación de estos ya que permite visualizar aquellos rasgos comunes dados por las características de las variables asociadas (Evangelopoulos *et al.*, 2012).

El etiquetado del factor se realizó gracias a la interpretación del significado de aquellos términos que poseían mayor puntuación en el respectivo factor. Así, aunado al conocimiento de dominio, se nombró cada concepto. Durante el desarrollo de esta actividad, se consideró que la fundamentación del análisis factorial plantea que los factores resultantes son ortogonales entre sí. Por ende, se partió bajo la directriz de que los atributos, a través de los cuales se interpretaban y se nombraban los conceptos, debían posibilitar que los conceptos fueran diferenciables entre ellos.

Para hallar los subconceptos asociados a cada concepto etiquetado se ejecutó la técnica de análisis de componentes principales. En el marco de esta, se ubicó a cada individuo (documento) en el espacio factorial (de conceptos) y, después, se calculó una distancia coseno entre cada texto y concepto etiquetado. De esta forma, se identificaron aquellos documentos que apoyaban en mayor medida la creación de cada concepto y, por ende, contenían información sobre los subconceptos. Para cada concepto, se estableció un conjunto de documentos que fue procesado nuevamente bajo todas las actividades previamente mencionadas. Subsiguientemente, se identificaron los subconceptos mediante el análisis factorial y se etiquetaron manualmente.

Reconocimiento de relaciones entre las unidades conceptuales

Al analizar la composición de cada factor, se establecieron las relaciones entre conceptos y subconceptos. Conforme a los procedimientos anteriores, el enfoque metodológico ejecutado permitió identificar una relación entre unidades conceptuales, mas no su tipo. Por esta razón, se determinó la naturaleza de esta conexión a partir del conocimiento de dominio.

Por último, se representaron las unidades conceptuales y las relaciones entre ellas a través de un esquema gráfico no legible por las máquinas. Sin embargo, se seleccionó esta representación porque facilita al ser humano interpretar y comprender la información contenida en aquel.

Para realizar un ejercicio de verificación de la red, se solicitó a un experto responder una serie de preguntas usando solo el conocimiento que se hallaba en la red. De forma posterior, parte de los autores del presente artículo verificaron que las respuestas brindadas no presentaran información contradictoria a la esquematizada en la red. Además, el equipo se cercioró de que las contestaciones del experto no contradijeran la información de los documentos analizados. En este sentido, parte de los autores del manuscrito estudiaron cuidadosamente los textos. Este

enfoque de verificación se retoma del trabajo de Tapia-León, Rivera, Espinosa y Lujan-Mora (2018).

Dada la exposición del presente apartado, se puede colegir que la propuesta metodológica realizada para la generación semiautomática de una red semántica se diferencia de aquellos estudios descritos en el marco teórico, en los siguientes aspectos:

- Como paso previo a la formación del vocabulario, se examinan, por juicio de expertos, las alternativas de escritura mediante el análisis de la matriz de correlación de los términos extraídos. Dicho de otro modo, el enfoque propuesto no exige bases de conocimiento ni textos etiquetados previamente; por lo cual, la posibilidad de hallar información relevante no está restringida a modelos teóricos.
- La identificación de conceptos se lleva a cabo mejorando la calidad de la MTD mediante la técnica de análisis latente semántico y analizando la correlación de las variables observables (términos del vocabulario) en relación con las variables latentes (conceptos). De este modo, la identificación de conceptos no solo considera la participación de cada término en el corpus sino también la interacción entre las unidades léxicas.
- Igualmente, para identificar subconceptos, se ejecuta la técnica de análisis de componentes principales con el fin de reconocer aquellos documentos que soportan la formación del concepto. Después, dichos documentos son procesados y, mediante el análisis factorial, se reconocen los subconceptos.
- Se advierten las relaciones entre unidades conceptuales mediante el estudio de la composición de cada factor extraído. Por consiguiente, se considera la interrelación entre las variables y no la cercanía con la que aparecen en el corpus (aspectos examinados mediante el análisis de coocurrencias).

La propuesta metodológica presentada se ejecutó considerando como caso de interés la comunidad de Arauca durante los años 2013 y 2018. Este territorio ha sido objeto de múltiples interpretaciones (Gutiérrez, 2010). Allí, las características geográficas, económicas, sociales y culturales han producido dinámicas que hacen diferenciable el territorio. En parte, lo anterior ha sido la causa de que este departamento colombiano se haya transformado en una zona estratégica para el asentamiento de los actores armados ilegales (nacionales). Frente a esta situación, el Estado colombiano ha buscado fortalecer su presencia en el Departamento pero esta no se ha dado uniformemente en los diferentes municipios. En síntesis, Arauca es una comunidad donde la diversidad geográfica, económica, social y cultural ha producido dinámicas complejas que favorecen la presencia de conflictividades. Frente a estas dinámicas, el Estado debe responder a través de su accionar (representado, en parte, por las políticas públicas) (Gutiérrez, 2010). No obstante, las instituciones estatales necesitan reconocer estas dinámicas para que sus actuaciones sean eficaces.

En este sentido, se observa que la generación de una red semántica que describa la comunidad de Arauca es un insumo que soportaría el quehacer de las autoridades nacionales. De este modo, el periodo analizado (2013-2018) desea realizar una descripción de las características actuales de esta comunidad. Sin embargo, los autores del actual artículo son conscientes de que la ejecución de un análisis a profundidad requeriría estudiar una mayor ventana de tiempo.

En el siguiente apartado, se presentan los principales resultados de la ejecución metodológica planteada.

4. Resultados

Siguiendo la metodología de Uschold y King (1995) y la naturaleza del proyecto, se determinó que el objetivo de la red semántica debía describir comunidades víctimas del conflicto armado. De este modo, se tomó la comunidad de Arauca durante los años 2013 y 2018 como caso de interés. En esta dirección, la red se plantea como una herramienta de apoyo en la formulación de políticas públicas.

El corpus empleado para este trabajo constó de sesenta publicaciones periódicas, escritas en castellano y seleccionadas manualmente. Los documentos fueron extraídos mediante un buscador de internet, considerando que los criterios de selección consistían en que cada texto hubiera sido publicado entre el 2013 y el 2018, y que su contenido estuviera relacionado con la comunidad de Arauca.

Una vez capturado el corpus, este fue estudiado en la plataforma *KNIME Analytics Platform*, la cual brinda un módulo para la minería de textos y extensiones para hacer codificaciones en el lenguaje Python. Con estas codificaciones se realizó la implementación de las técnicas de análisis multivariado.

Desde el flujo de trabajo de *KNIME*, se realizó la extracción de términos. El tipo de palabras asociadas, como ya se ha mencionado, surgió de analizar la matriz de correlación entre los términos y el posterior juicio de expertos. A manera de ejemplo, se identificó una relación positiva perfecta entre "conflicto armado" y "conflicto interno". Más adelante, haciendo uso del juicio de expertos, se reconoció que dicha relación era correcta. Por tal motivo, se pudo establecer que estas palabras estaban asociadas y que, cada vez que surgiera en el corpus el segundo de los términos mencionados, se debía reemplazar por el primero para que las actividades subsecuentes (ponderación y aplicación de técnicas multivariadas) no fueran sesgadas por la múltiple aparición de palabras que semánticamente eran semejantes (Manning y Schutza, 2000).

A partir de la colección de candidatos y el cálculo de características, se estableció un conjunto de términos que se compararon con aquellos etiquetados manualmente en el 25 % de los textos. Las métricas de comparación se retomaron de Lagutina *et al.* (2016), quienes proponen emplear:

- Precisión (P): Es la proporción de términos recuperados que son relevantes, medidos por la relación entre el número de ítems recuperados relevantes y el número total de ítems recuperados.

Se expresa por:

$$P = \frac{|T_{rel} \cap T_{ext}|}{|T_{ext}|}$$

donde T_{rel} es el número de términos relevantes según lo señalado por los expertos y T_{ext} es el número de términos extraídos en referencia al procedimiento ejecutado.

- Recuperación (R): Es la proporción de términos relevantes recuperados, medidos por la relación entre el número de elementos recuperados relevantes y el número total de elementos relevantes en el corpus.

Se expresa por:

$$R = \frac{|T_{rel} \cap T_{ext}|}{|T_{rel}|}$$

Medida-F (F): Es una media armónica entre la precisión y la recuperación. Se supone que se desean términos claves con igual posibilidad de tener características de precisión y reconocimiento. Por lo anterior, se expresa mediante

$$F - measure = \frac{2PR}{P+R}$$

De esta manera, se obtuvo que, para $T_{ext} = 210$, $T_{rel} = 199$, $T_{rel} \cap T_{ext} = 186$ las métricas alcanzadas fueron:

$$P = \frac{186}{210} = 0.885 \quad [1]$$

$$R = \frac{186}{199} = 0.9346 \quad [2]$$

$$F = \frac{2PR}{P+R} = 0.90 \quad [3]$$

Así, dadas las características metodológicas descritas para la extracción de términos, se identificó que se alcanzó un 88,5 % de precisión (proporción de términos recuperados relevantes dado el juicio de

expertos), un 93,46 % de reconocimiento (es la razón entre ítems relevantes recuperados y el total de términos relevantes). Igualmente, la medida promedio entre la precisión y el reconocimiento fue del 90 %. Lo anterior demuestra que el proceso de extracción empleado permitió hallar términos relevantes para el dominio de interés (comunidad de Arauca-víctima del conflicto armado), lo cual garantiza datos de entrada coherentes en la construcción de conceptos para la red semántica buscada.

Subsiguientemente, los datos son esquematizados en una MTD (matriz término-documento) para la identificación y el etiquetado de conceptos. La posterior mejora de la calidad de esta matriz se lleva a cabo mediante la aplicación del análisis latente semántico. Para ejecutar esta técnica; primero, se extrajeron los valores propios de la MTD normalizada y se consideró la descomposición de valor singular (SVD- singular value decomposition) (Deerwester *et al.*, 1990); luego, se seleccionaron los k valores propios que eran mayor a la unidad; finalmente, se obtuvo la versión truncada de la MTD, es decir, a partir de las matrices resultantes después de ejecutar la SVD, se eligieron los k valores singulares que habían pasado el criterio de selección y sus respectivas columnas.

De esta manera, se obtuvo una matriz de las mismas dimensiones que la MTD, pero más informativa, ya que el análisis latente semántico recopila todos los contextos dentro de los cuales aparecen las palabras del vocabulario y establece factores (latentes) que representen información común subyacente.

En este orden de ideas, se aplicó la técnica de análisis factorial exploratorio para la cual se consideraron los siguientes planteamientos de Uriel Jiménez y Aldás Manzano (2005):

- Evaluar que el determinante de la matriz de correlación fuera igual o cercano a cero, con el objetivo de reconocer que las variables observables (términos del vocabulario) estuvieran altamente intercorrelacionadas entre sí y hubiera evidencia que soportara el desarrollo del modelo factorial. Durante la ejecución del proyecto, se halló que este determinante tenía valor cero, lo cual significó que algunas de las variables eran linealmente dependientes y era correcto continuar el análisis factorial.
- Establecer la matriz de componentes tipificados mediante el método de componentes principales para la extracción de factores. Se empleó el criterio de la media aritmética para la retención de factores. Así, se hallaron 28 factores (conceptos) del corpus seleccionado.
- Valorar el modelo factorial estimado al estudiar el valor absoluto de la diferencia entre los coeficientes de correlación observados y reproducidos. En este punto se identificó que el 6,76 % de los residuales no redundantes tenían valores absolutos mayores que 0,05. Esto indicó que la bondad del modelo no era discutible. Analizar la comunalidad de las variables originales en el modelo factorial. A través de este procedimiento, se reconoció que la varianza de todos los términos del vocabulario, producto de los

factores comunes estaba entre 0,97 y 0,99. Lo anterior es una buena señal de la bondad del modelo.

- Rotar la matriz de factores tipificados con el fin de analizar fácilmente las interrelaciones entre las variables. Como ya se ha expuesto, se aplicó la rotación *varimax*.

Como se mencionó anteriormente, se identificaron 28 conceptos. A continuación, se ejemplifica para el concepto (factor) número cuatro, el análisis a través del cual fue etiquetado. En la Tabla 1, se presentan los términos cuyas cargas son representativas en el factor de interés.

Tabla 1
Términos top del factor número cuatro.

No.	Términos	C3: Violencia contra la mujer araucana
1	Rangos altos	-0,907343946
2	CNMH	-0,884575397
3	Hipermasculinidad	-0,871164073
4	Compañeras sentimentales	-0,871164073
5	BVA	-0,810912082
6	Violencia género	-0,809474752

Fuente: Autores

Para realizar el análisis, se retomaron las palabras asociadas a cada término. Estas se exponen en la Tabla 2.

Tabla 2
Algunos términos y palabras asociadas.

Términos	Palabras asociadas
	superiores
	jerarquía de poder
rangos altos	posición de poder
	lugar subordinado frente al hombre
	relaciones de poder
cnmh (centro nacional de memoria histórica)	centro nacional de memoria histórica
	centro de memoria histórica
	llaneros son bravos
hipermasculinidad	guerrero grande fuerte y violento en extremo
	cohesión entre sus hombres
	Cultura sexista
	novias de miembros de la policía o el ejército
compañeras sentimentales	Compañeras de los muchachos
	Pocholeras
	relación con los soldados y policías
bva (bloque vencedores de Arauca)	bloque vencedores de Arauca
	bloque vencedores
	violencia contra las mujeres niñas y adolescentes
	discriminación contra la mujer
	violencia contra las araucanas
	violencia basada en género
violencia género	violencia que han vivido las mujeres
	daño a las mujeres
	violencia de género
	violencias contra ellas
	circunstancia traumática
	violencia ejercida contra ellas

Fuente: Autores

Asimismo, se identificó que los términos con alta ponderación en el factor cuatro poseían relación con la violencia contra la mujer. En la Tabla 1, los términos número uno y cuatro representan componentes en las formas de violencia contra la mujer araucana. Particularmente, los valores altos en "rangos altos" se pueden asociar con las relaciones desiguales que tienen hombres y mujeres en contextos como los grupos insurgentes, donde los altos rangos aprovechan su estatus de poder para entablar relaciones con milicianas y estas no tienen la posibilidad de negarse por miedo a contradecir a alguien superior en nivel jerárquico y sufrir las consecuencias que esto conlleva. Del mismo modo, el término "compañeras sentimentales" hace referencia a las compañeras de los miembros de la policía que han sufrido violencia sexual, física y feminicidios a mano de actores del conflicto (como los paramilitares, grupos guerrilleros, entre otros) por emprender relaciones sentimentales con miembros de la fuerza pública. Un caso específico de esto lo representan las pocholeras.

La "hipermasculinidad" hace relación a la cultura sexista que conlleva el conflicto armado. Conforme a esta cultura, los soldados debían representar fortaleza y violencia, atacando a personas inferiores y de menor fuerza física como las mujeres. El término número dos en la Tabla 1, expresa el Centro Nacional de Memoria Histórica (cnmh) quien es el encargado de documentar las formas de violencias que se dieron durante el conflicto en el período de tiempo analizado. En especial, esta entidad ha tenido como responsabilidad registrar la violencia sexual ejercida por el Bloque Vencedores de Arauca (bva). El término bva puede ser interpretado 1) como uno de los principales actores de violencia de género en el departamento de Arauca o 2) según su conexión con la función del cnmh, ya que este grupo paramilitar (bva) ha estado inactivo desde un tiempo mayor a la procedencia de los documentos analizados (desmovilizados en el 2011). En este sentido, se llega al término "violencia género", el cual se refiere al maltrato hacia la mujer por ser del género femenino. En consecuencia, se consideró apropiado nombrar el factor como el concepto "Violencia contra la mujer araucana".

El procedimiento arriba descrito para fijar la etiqueta del factor 4 se repitió para los otros 27 factores restantes. De esta manera, se llegó al listado que se presenta en la Tabla 3.

Tabla 3
Conceptos identificados y etiquetados.

C0: Resignación	C14: Cultura de violencia sexual
C1: Diagnostico	C15: Construcción paz: Necesidades
C2: Dinámica del conflicto	C16: Crecimiento país
C3: Violencia contra la mujer araucana	C17: Rol de la mujer en el posconflicto
C4: Cese fuego	C18: Rol de los movimientos sindicales
C5: Cultivos ilícitos	C19: Enfoque de género
C6: Violencia sexual	C20: Víctimas
C7: Circunscripción	C21: Negociación ELN y Gobierno
C8: Desplazamiento	C22: Caracterización de posconflicto
C9: Justicia especial para la paz	C23: Restitución de tierras
C10: Líderes y líderes sociales	C24: Garantías de participación
C11: Disidentes	C25: Desplazamiento al exterior
C12: Plebiscito	C26: Instancia especial
C13: Accionar del ELN	C27: Retos en el posconflicto

Fuente: Autores

Para identificar los subconceptos asociados (ver sección "Metodología", subsección "Identificación y etiquetado de conceptos y subconceptos") se calcularon las puntuaciones factoriales, es decir, los valores que toman cada uno de los documentos en los factores (conceptos) identificados (Uriel Jiménez y Aldás Manzano, 2005). Después, mediante la distancia coseno, se identificaron aquellos documentos que estaban más cerca de cada factor. Siguiendo el caso ejemplificado, la Tabla 4 expone los documentos más cercanos, en su valor factorial, al concepto "C3: violencia contra la mujer araucana".

Tabla 4
Documentos más cercanos al concepto C3

Documentos	C3: Violencia contra la mujer araucana
D15	-0,918395405
D2	-0,992744287
D21	-0,872454836
D35	-0,958959165
D5	-0,987098239
D55	-0,810293139
D8	-0,910293139
D9	-0,910293139

Fuente: Autores

Estos documentos fueron procesados nuevamente mediante las actividades previamente descritas para constituir el corpus de documentos originales (extracción de términos, identificación y etiquetado de subconceptos). De este modo, se identificaron y etiquetaron los subconceptos de: 1) violencia sexual, 2) desplazamiento, 3) desplazamiento al exterior, 4) feminicidios y 5) organizaciones de mujeres. En este sentido, se ratificó la existencia de la relación entre el concepto C3 y sus respectivos subconceptos. La naturaleza de dicha conexión fue establecida por los autores del trabajo a partir de su conocimiento del área.

Por último, se representaron gráficamente las unidades conceptuales identificadas, así como las relaciones entre ellas a través de la herramienta gratuita CmapsTools. En la Figura 2, se presenta la esquematización del concepto C3, donde "violencia sexual" es parte de "C3: Violencia contra la mujer araucana", Igualmente, "Desplazamiento", "Desplazamiento al exterior" y "Feminicidio" son parte de "C3: Violencia contra la mujer araucana".

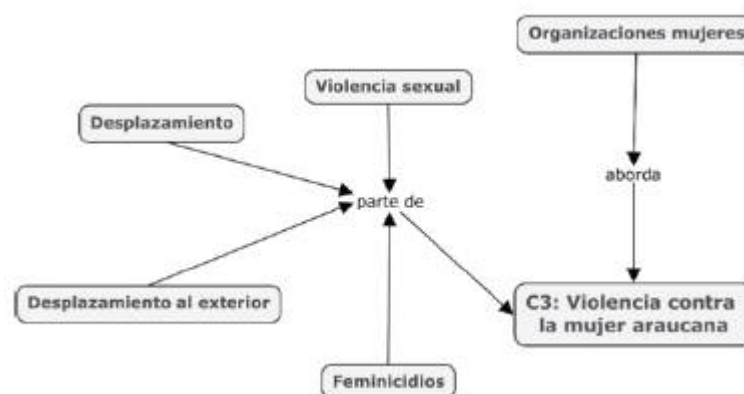


Figura 2

Red semántica que considera únicamente "C3: Violencia contra la mujer araucana".

Fuente: Autores

De esta forma, se reconocieron los subconceptos y relaciones conceptuales para cada uno de los 27 factores identificados y etiquetados. Con esto se obtuvo una red de 118 nodos, de los cuales 28 representaban

conceptos y los demás subconceptos. En los Anexos (ver Tablas A, B y C) se presenta cada concepto identificado y sus respectivos subconceptos.

Al identificar los subconceptos de "C0: Resignación", se reconoció la presencia de "violencia contra la mujer araucana" (ver Tabla A); una situación análoga sucedió al reconocer los subconceptos de "C27: Retos en el posconflicto" (ver Tabla C). En consecuencia, se determinaron nuevas características de "C3: Violencia contra la mujer araucana" al analizar la constitución de los subconceptos de otros factores. En este sentido, la esquematización del concepto C3 se amplía y se expone en la Figura 3. Allí, se puede observar que la violencia contra la mujer araucana *causa* resignación y *es un* reto en el posconflicto.

Igualmente, al aplicar la metodología ya descrita se identifica que las interpretaciones de los subconceptos reafirman hallazgos de otros conceptos. Lo anterior, se evidencia al analizar la Tabla B (ver "Anexos"). Allí, se puede observar que para "C9: Jurisdicción especial para la paz" se presenta el subconcepto "enfoque de género". A su vez, en la Tabla 7 (ver "Anexos"), se visualiza que existe el subconcepto "jurisdicción especial para la paz" para el concepto "CI9: Enfoque de género". Por lo anterior, la relación entre C19 y C9 se reafirma y se reconoce que dicha conexión es bidireccional.

El ejercicio de verificación (ver arriba sección "Metodología", subsección "Reconocimiento de relaciones entre las unidades conceptuales") de la red semántica permitió reconocer que la red semántica generada exponía correctamente la información contenida en el corpus. Además, la estructura diseñada no producía confusiones a la persona que la interpretaba.

Finalmente, en la sección "Conclusiones" se resaltan futuras líneas de investigación.

5. Conclusiones

Siguiendo la metodología expuesta y ejecutada, se generó una red semántica para la descripción de la comunidad de Arauca en Colombia, durante el periodo 2013-2018. En esa dirección, este estudio propone un enfoque metodológico innovador para la generación semiautomática de una red semántica. A continuación, se señalan algunas líneas de investigación que permitirán robustecer la metodología expuesta:

- Al aumentar el tamaño del corpus, se obtendrán resultados más certeros sobre los términos y relaciones extraídas. Además, se dará la posibilidad de hallar información poco frecuente en el dominio de interés.
- Evaluar diferentes ponderaciones de los términos permitirá progresar en la extracción no supervisada de unidades terminológicas (Astrakhantsev *et al.*, 2015).
- Integrar tesauros relacionados con el tema, facilitará el etiquetado de conceptos y subconceptos, y disminuirá la intervención manual.

- Emplear patrones lingüísticos como los propuestos por Zouaq *et al.* (2012) facilitará deducir el tipo de enlaces entre términos.
- Caracterizar la red semántica empleando herramientas de la teoría de redes, proporcionará la naturaleza y estructura semántica de los conceptos.
- Verificar la coherencia semántica de la red semántica mediante propuestas como la de Al Madi y Khan (2018), proporcionará más datos para retroalimentar la metodología propuesta.

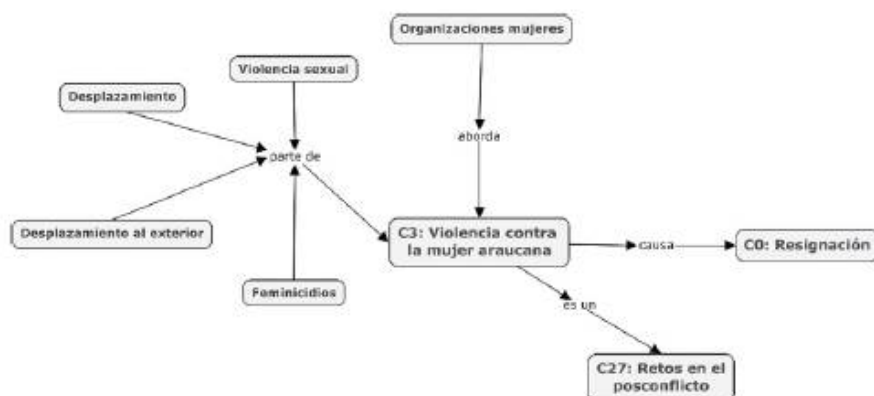


Figura 3

Red semántica considera únicamente "C3: Violencia contra la mujer araucana".

Fuente: Autores

Siguiendo las líneas de trabajo señaladas, futuros proyectos podrán generar una red semántica partiendo del enfoque metodológico aquí documentado. Luego, estos proyectos podrán transformar dicha red en un modelo ontológico que facilite el desarrollo de procesos automáticos y ejecutables por máquinas computacionales. Lo anterior, constituirá un insumo para que las instituciones estatales logren mayor sincronía entre las políticas públicas que formulan y ejecutan, y el impacto de estas.

Referencias bibliográficas

1. AL MADI, Naser y KHAN, Javed. Constructing Semantic Networks of Comprehension from Eye-Movement during Reading. En: Proceedings - 12th IEEE International Conference on Semantic Computing (12, 31-2, enero-febrero: Laguna Hills, California). ICSC. California: IEEE, 2018. p. 49-55. <https://doi.org/10.1109/ICSC.2018.00016>
2. ALZATE, Mary L. y ROMO, Gerardo. El enfoque de la gobernanza y su recepción en el marco gubernativo actual de las sociedades latinoamericanas. En: Opinião Pública. Diciembre, 2014. vol. 20, no. 3, p. 480-495. <https://doi.org/10.1590/1807-01912014203480>
3. ASTRAKHANTSEV, Nikita A., FEDORENKO, Denis G. y TURDAKOV, Denis Y. Methods for automatic term recognition in domain-specific text collections: A survey. En: Programming and Computer Software. Noviembre, 2015. vol. 41, no. 6, p. 336-349. <https://doi.org/10.1134/S036176881506002X>

4. AUSSENAC-GILLES, Nathalie; BIEBOW, Brigitte y Szulman, Sylvie. Revisiting Ontology Design: A Methodology Based on Corpus Analysis. En: EKAW 2000 (12, 2-6, octubre: Jean-les-Pins, Francia). Knowledge Engineering and Knowledge Management Methods, Models, and Tools. Berlin: Springer-Verlag Berlin Heidelberg, 2000. p. 172-188. https://doi.org/10.1007/3-540-39967-4_13
5. BLEI, David M.; NG, Andrew Y. y JORDAN, Michael I. Latente Dirichlet Allocation. En: The Journal of Machine Learning Research. Enero, 2003. vol. 3, p. 993-1022.
6. BRACHMAN, Ronald J. On the epistemological status of semantic networks. In: Associative Networks. Representation and Use of Knowledge by Computers. Nueva York: Academic Press, FINDLER, Nicholas V., 1979. p. 3-50. <https://doi.org/https://doi.org/10.1016/B978-0-12-256380-5.50007-4>
7. BRACHMAN, Ronald J. What's in a concept: structural foundations for semantic networks. In: International Journal of Man-Machine Studies. Marzo, 1977. vol. 9, no. 2, p. 127-152. [https://doi.org/10.1016/S0020-7373\(77\)80017-5](https://doi.org/10.1016/S0020-7373(77)80017-5)
8. CARLEY, Kathleen M., BIGRIGG, Michael W. y DIALLO, Boubacar. Data-to-model: A mixed initiative approach for rapid ethnographic assessment. In: Computational and Mathematical Organization Theory. Julio, 2012. vol. 18, no. 3, p. 300-327. <https://doi.org/10.1007/s10588-012-9125-y>
9. CHRISTENSEN, Tom y LAEGREID, Per. El estado fragmentado: los retos de combinar eficiencia, normas institucionales y democracia. En: Gestión y Política Pública. Julio- diciembre, 2005. vol. 14, no. 003, p. 557-598. <https://www.redalyc.org/pdf/133/13314306.pdf>
10. DEERWESTER, Scott; DUMAIS, Susan T.; FURNAS, George W.; LANDAUER, Thomas K. y HARSHMAN, Richard. Indexing by Latent Semantic Analysis. En: Journal of the American Society for Information Science. Septiembre, 1990. vol. 41, no. 6, p. 391-407. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41](https://doi.org/10.1002/(SICI)1097-4571(199009)41)
11. DUBNICK, Melvin J. y BARDES, Barbara A. Thinking about public policy, a problem-solving approach. Political Science Scholarship. Nueva York: John Wiley and Sons, 1983. 283 p. ISBN 0471864641 9780471864646.
12. ECHEVERRY-CORREA, Julian D. Contributions to speech analytics based on speech recognition and topic identification. Trabajo de grado PH.D. en Ingeniería de Sistemas Electrónicos. Madrid: Universidad Politécnica de Madrid. Escuela Técnica Superior de Ingenieros de Telecomunicación, 2015. 135 p. <https://doi.org/10.22517/9789587222678>
13. EVANGELOPOULOS, Nicholas; ZHANG, Xiaoni y PRYBUTOK, Victor R. Latent Semantic Analysis: five methodological recommendations. In: European Journal of Information Systems. Septiembre, 2012. vol. 21, no. 1, p. 70-86. <https://doi.org/10.1057/ejis.2010.61>
14. FEDORENKO, Denis G., ASTRAKHANTSEV, Nikita A. y TURDAKOV, Denis Y. Automatic recognition of domain-specific terms: An experimental evaluation. In: Proceedings of the Institute for System Programming. 2014. vol. 26, no. 4, p. 55-72. [https://doi.org/10.15514/I SPRAS-2014-26\(4\)-5](https://doi.org/10.15514/I SPRAS-2014-26(4)-5)

15. GÓMEZ-PÉREZ, Asunción, FERNÁNDEZ-LOPEZ, Mariano y CORCHO, Oscar. *Ontological Engineering: With examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Londres: Springer-Verlag London, 2004. 404 p. (Serie Advanced Information and Knowledge Processing). <https://doi.org/10.1007/b97353>
16. GUTIÉRREZ, Omar J. Arauca: espacio, conflicto e institucionalidad. En: *Análisis Político*. Mayo, 2010. vol. 23, no. 69, p. 3-34. <https://revistas.unal.edu.co/index.php/anpol/article/view/45778/47305>
17. HUTCHINS, Christopher E. y BENHAM-HUTCHINS, Marge. Hiding in plain sight: Criminal network analysis. En: *Computational and Mathematical Organization Theory*. Noviembre, 2010. vol. 16, p. 89-111. <https://doi.org/10.1007/s10588-009-9060-8>.
18. ILGEN, Daniel. R. y HULIN, Charles. L. *Computational Modeling of Behavior in Organizations: The Third Scientific Discipline*. Nueva York: American Psychological Association. 2000. 309 p. ISBN 978-1-55798-639-9.
19. JACQUEMIN, Christian y BOURIGAULT, Didier. Term Extraction and Automatic Indexing. In: *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press, MITKOV, Ruslan, 2012. <https://doi.org/10.1093/oxfordhb/9780199276349.013.0033>
20. KAO, Anne y POTEET, Stephen R. Overview. En: *Natural Language Processing and Text Mining*. Londres: Springer-Verlag London Limited, 2007. p. 1-7. <https://doi.org/10.1007/978-1-84628-754-1>
21. KASHYAP, Vipul y BORGIDA, Alex. Representing the UMLS® semantic network using OWL (or “what’s in a Semantic Web link?”). En: *2nd International Semantic Web Conference (2: 20-23, octubre: Sanibel Island, Florida). The Semantic Web - ISWC 2003*. Springer, Berlin, Heidelberg: ISWC, 2003, p. 1-16. https://doi.org/10.1007/978-3-540-39718-2_1.
22. KINGDON, John W. *Agendas, alternatives and public policy*. Boston: Little, Brown and Company, 1984. 240 p. ISBN 9780316493918.
23. LAGUTINA, Ksenia, MAMEDOV, Eldar, LAGUTINA, Nadezhda, PARAMONOV, Ilya y SHCHITOV, Ivan. Analysis of relation extraction methods for automatic generation of specialized thesauri: Prospect of hybrid methods. En: *19th Conference of Open Innovations Association (19: 7-11, noviembre: Jyväskylä, Finland). 2016 19th Conference of Open Innovations Association (FRUCT)*. Saint-Petersburg: FRUCT Oy, 2016. p. 138-144. <https://doi.org/10.23919/FRUCT.2016.7892193>
24. MANNING, Christopher y SCHUTZA, Hinrich. *Foundations of statistical natural language processing*. 2 ed. Cambridge: The MIT Press, 2000. 680 p. ISBN 9780262133609.
25. PFEFFER, Jürgen y Carley, Kathleen M. Rapid modeling and analyzing networks extracted from pre-structured news articles. En: *Computational and Mathematical Organization Theory*. Junio, 2012. vol. 18, no. 3, p. 280-299. <https://doi.org/10.1007/s10588-012-9122-1>.
26. QUILLIAN, M. Ross. The teachable language comprehender: a simulation program and theory of language. In: *Communications of the ACM*.

- Agosto, 1969. vol. 12, no. 8, p. 459-476. <https://doi.org/10.1145/363196.363214>
27. QUINTANA, Christian. Impacto social y económico en la implementación de políticas públicas. Trabajo de grado Especialista en Finanzas y Administración Pública. Bogotá D.C.: Universidad Militar Nueva Granada. Facultad de Ciencias Económicas, 2014. 34 p. http://repository.unimilitar.edu.co/bitstream/10654/11092/1/IMPACTOSOCIAL_Y_ECONÓMICO_EN_LA_IMPLEMENTACIÓN_OK.pdf
 28. ROTH DEUBEL, André-Noël. Políticas públicas. Formulación, implementación y evaluación. 10.ª edición. Bogotá D.C.: Ediciones Aurora, 2002. 134 p. ISBN 958913615X.
 29. TAPIA-LEÓN, Mariela; RIVERA, Abdon C.; ESPINOSA, Janneth. C. y LUJAN-MORA, Sergio. Representation of Latin American university syllabuses in a semantic network. En: 2nd International Conference on Information Systems and Computer Science. (2: 23-25, noviembre: Quito, Ecuador). 2017 International Conference on Information Systems and Computer Science (INCISCOS). Quito: CPS, 2018. p. 295-301. <https://doi.org/10.1109/INCISCOS.2017.28>
 30. URIEL JIMÉNEZ, Ezequiel y ALDÁS MANZANO, Joaquín. Análisis factorial. En: Análisis multivariante aplicado: Aplicaciones al marketing, investigación de mercados, dirección de empresas y turismo. España: Thomson-Paraninfo, 2005. p. 407-440.
 31. USCHOLD, Mike y KING, Martin. (1995). Towards a Methodology for Building Ontologies. Edinburgh: University of Edinburgh, 1995. AIAI-TR-183. <http://www.aiai.ed.ac.uk/publications/documents/1995/95-ont-ijcai95-ont-method.pdf>
 32. VAN HOLT, Tracy; JOHNSON, Jeffrey C.; BRINKLEY, James D.; CARLEY, Kathleen M. y CASPERSEN, Janna. Structure of ethnic violence in Sudan: A semi-automated network analysis of online news (2003- 2010). In: Computational and Mathematical Organization Theory. Septiembre, 2012. vol. 18, no. 3, p. 340-355. <https://doi.org/10.1007/s10588-012-9124-z>.
 33. WOODS, William A. WHAT'S IN A LINK: Foundations for Semantic Networks. En: Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence. California: Morgan Kaufmann Publishers, Inc., COLLINS, Allan y SMITH, Edward E., 1988. p. 102-125. <https://doi.org/10.1016/B978-1-4832-1446-7.50014-5>
 34. ZHU, Bing, WANG, Yu y HE, Chenglong. Global Social Event Extraction and Analysis by Processing Online News. In: 2016 International Conference on Information System and Artificial Intelligence. (24-26, junio: Hong Kong, China). ISAI 2016. Hong Kong: CPS, 2016. p. 73-76. <https://doi.org/10.1109/ISAI.2016.0024>.
 35. ZOUAQ, Amal, GASEVIC, Dragan y HATALA, Marek. Linguistic patterns for information extraction in OntoCmaps. En: 3rd Workshop on Ontology Patterns (3: 12, noviembre: Boston, Estados Unidos). Proceedings of the 3rd Canadian Semantic Web Symposium. Aachen: Rheinisch-Westfaelische Technische Hochschule Aachen * Lehrstuhl Informatik V, 2012. p. 61-72.

Anexos

En las Tablas A, B y C se representan los conceptos identificados y etiquetados, así como los subconceptos asociados.

Tabla A
Conceptos y subconceptos 1.

Concepto	Subconcepto	Concepto	Subconcepto
C0: Resignación	Violencia contra la mujer araucana	C4: Cese fuego	Saravena: Corazón del conflicto
	Desplazamiento al exterior		Zona frontera
	Accionar ELN		Características del cese al fuego
	Escepticismo en el posconflicto		Enfoque de género: Respeto y no repetición
	Cese fuego ELN y Gobierno	C5: Cultivos ilícitos	Desplazamiento
	Conflicto armado		Interrogantes y escepticismo: Narcotráfico
	Interrogantes en el posconflicto		Cultura
	Circunscripción		Retos
	Encuentro regional de paz	C6: Violencia sexual	Enfoque de género
	Deficientes reparaciones		Maltrato físico
	Crecimiento país		Categorías de violencia sexual
	Labor organizaciones ayuda: Desplazamiento		Cese fuego
	Atentados ELN		Inversión social
	Contenido regional en negociaciones con ELN		Congelamiento diálogos
	Conflicto armado		Diferencia zona rural y urbana
	Desplazamiento al exterior		Homicidios selectivos
	Dinámica ELN y FARC		Desmovilizados
	Modelo social		Violencia de género
	Desplazamiento		Diagnostico
	ELN principal actor del conflicto		Actores conflicto
C1: Diagnostico		C7: Circunscripción	Congelamiento diálogos
			Retos e incertidumbre
C2: Dinámica del conflicto			Cese fuego
			Justicia especial para la paz
C3: Violencia contra la mujer araucana	Violencia sexual		Operativos
	Desplazamiento		Organizaciones de trabajo
	Organizaciones de mujeres		Enfoque de género
	Desplazamiento al exterior		
	Feminicidios		

Fuente: Autores.

Tabla B
Conceptos y subconceptos 2.

Concepto	Subconcepto	Concepto	Subconcepto
C8: Desplazamiento	Desplazamiento al exterior	C15: Construcción paz: Necesidades	Frente Domingo Lain
	Causas desplazamiento		Características Arauca
	Congelamiento diálogos		Plebiscito
	Nuevos actores de desplazamiento		Zona frontera
	Desplazados venezolanos		Protección
C9: Jurisdicción especial para la paz	Enfoque de género		Rol Iglesia
	Debilidad institucional		Modelo social
	Seguridad		Revictimizada
	Construcción paz		Instancia especial
	Turismo		Enfoque de género
C10: Líderes y líderes sociales	Víctimas	C16: Crecimiento país	Proceso paz frenado
	Posconflicto		Reparaciones simbólicas
	Saneamiento		Reconciliación
	ELN		Políticas públicas
	Reconocimiento		Organizaciones mujeres
	Violencia sexual		Revictimizada
	Encuentro regional de paz	C17: Rol de la mujer en el posconflicto	Deporte
	Instancia Especial		Garantías de participación
C11: Disidentes	Movilizaciones sociales		Defensoras
	Elenos: redes de apoyo		Políticas públicas
	Retos		Líderes sociales
C12: Plebiscito	Conflicto armado		Movilizaciones sociales
	Concierto delinquir - extorsiones		Revictimizadas
	Desigualdad		Desmovilización
	Enfoque de género		Violencia intrafamiliar
	Reconocimiento		Negociación
	Negociación		Crecimiento país
C13: Accionar del ELN	Proceso paz frenado	C18: Rol de los movimientos sindicales	Desplazamiento
	Reconciliación		Diagnóstico de necesidades
	Extorsiones		Cultura violencia de género
	Líderes y líderes sociales		Crecimiento de país
	Narcotráfico		Rol Iglesia
C14: Cultura de violencia sexual	Congelamiento diálogos		Inversión social
	Organizaciones mujeres		Reconciliación
	Atentados		Garantías de participación
	Modelo social		Revictimización
	Inversión social		
	Novedad		
	Femicidios		

Fuente: Autores

Tabla C
Conceptos y subconceptos 3.

Concepto	Subconcepto	Concepto	Subconcepto
C19: Enfoque de género	Cultura de violencia sexual	C25: Desplazamiento al exterior	Revictimización
	Justicia especial paz		FARC
	Ideología		Trabajadora sexual
	Protección		Ideología
	Trabajo		Modelo social
C20: Víctimas	Petróleo - contrabando: ELN	C26: Instancia especial	Circunscripción
	Transporte: Zona rural		Defensores
	Necesidades de víctimas		Características de Arauca
	Persecución: paramilitares		Congelamiento diálogos
	Reconciliación		Desplazamiento
C21: Negociación ELN y Gobierno	Garantías de participación		Desmovilización
	Paz		Plebiscito
	Desescalamiento conflicto armado		Reparación simbólica
	Congelamiento diálogos		Educación
	Contenido regional		Enfoque de género
C22: Caracterización del Posconflicto	Incertidumbre	C27: Retos en el posconflicto	Turismo
	Paramilitares		Atentados ELN
	Plebiscito: Zona rural		Cultura de violencia de género
	Instancia especial		Cese fuego
	Cese al fuego		Accionar ELN
C23: Restitución de tierras	Revictimización		Desplazamiento
	Reconciliación		Retos del Posconflicto
	Protección derechos		Enfoque de género
	Conflicto armado		Líderes y líderes sociales
	Discriminación		Derecho libertad
C24: Garantías de participación	Revictimización		Turismo
	Comunidades étnicas		Revictimización
	Posconflicto		Educación
	Conflicto		Atentados
	Instituciones fuerza pública		Violencia contra la mujer araucana

Fuente: Autores

Notas

* Artículo de investigación científica y tecnológica, producto del proyecto: Generación de redes de información textual que describan factores socio-culturales de comunidades víctimas del conflicto armado como insumo para la formación de políticas públicas y planes de desarrollo. Este es un artículo Open Access bajo la licencia BY-NC-SA (<http://creativecommons.org/licenses/by-nc-sa/4.0/>)

Cómo citar este artículo: SOTO MEJÍA, José; GÓMEZ SUTA, Manuela; ZULUAGA RAMÍREZ, Carlos Mauricio y CHAVARRO PORRAS, Julio Cesar. Generación semiautomática de red semántica para describir comunidades víctimas del conflicto armado colombiano. En: Entramado. Enero - Junio, 2020 vol. 16, no. 1, p. 252-271 <https://dx.doi.org/10.18041/1900-3803/entramado.1.6114>

Conflicto de intereses Los autores declaran no tener ningún conflicto de intereses.