
Modelo y algoritmo para identificación y clasificación de afectaciones en un corpus de testimonios sobre desaparición forzada

Model and algorithm for identifying and classifying affectations in a corpus of testimonies on enforced disappearance

Ana-María TANGARIFE-PATIÑO (1), Sandra-Patricia ARENAS-GRISALES (1),
José-David RUIZ ÁLVAREZ (2), Fabián BAENA-HENAO (1), Natalia MUÑOZ-OSORIO (1),
Tatiana TIRADO-TAMAYO (1), Brayan-Alexánder MUÑOZ-BARRERA (2)

(1) Escuela Interamericana de Bibliotecología, Universidad de Antioquia. ana.tangarife@udea.edu.co, sandra.arenas@udea.edu.co, natalia.munoz@udea.edu.co, tatiana.tirado10367@udea.edu.co (2) Instituto de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia. josed.ruiz@udea.edu.co, balexander.munoz@udea.edu.co

Resumen

Se presenta, en este artículo, la metodología para la construcción de un modelo y un algoritmo para la identificación y clasificación de afectaciones (económicas, físicas, políticas, psicológicas y socioculturales) en un corpus de testimonios de familiares de víctimas de desaparición forzada. Se describe, inicialmente, el contexto y las características del fenómeno de desaparición forzada en Colombia y se enuncian las preguntas que guían la investigación en torno al procesamiento y la modelación de la información contenida en relación con el volumen de datos, con los modelos conceptuales para la clasificación y con las métricas para medir la eficiencia del algoritmo implementado. Luego, se detalla la metodología y la base conceptual sobre la que se inscribe el modelo, para discutir, posteriormente, las métricas y los criterios de valoración de estas, para comparar la eficacia entre la clasificación automática y la manual. Por último, se ofrecen reflexiones y líneas de trabajo futuro sobre el uso de técnicas de procesamiento de lenguaje natural en textos de memoria histórica.

Palabras clave: Algoritmos. Corpus. Desaparición forzada. Lingüística computacional. Procesamiento de lenguaje natural. Testimonios. Colombia.

1. Introducción

La clasificación y extracción de información en corpus de dominios especializados es una tarea de gran relevancia en el procesamiento de lenguaje natural. Distintas aproximaciones han propuesto métodos y técnicas para reconocimiento automático en análisis de contenido (Eriksson, 2007), (Aguado y otros, 2002), (Buendía, 2010). Se reconoce un amplio avance en muchas lenguas en donde se recogen aspectos semánticos del lenguaje en corpus como Wordnet.

Este artículo plantea una reflexión en un dominio concreto para atender a necesidades de clasificación y representación de información en el

Abstract

This article presents the methodology for the construction of a model and an algorithm for the identification and classification of affectations (economic, physical, political, psychological and sociocultural) in a corpus of testimonies of relatives of victims of enforced disappearance. Initially, the context and characteristics of the phenomenon of enforced disappearance in Colombia are described and the research questions guiding the processing and modeling of the information contained in relation to the volume of data, the conceptual models for the classification and the metrics to measure the efficiency of the implemented algorithm are stated. The methodology and the conceptual basis on which the model is based are then detailed, followed by a discussion of the metrics and their evaluation criteria to compare the efficiency between automatic and manual classification. Finally, reflections and lines of future work on the use of natural language processing techniques in historical memory texts are offered.

Keywords: Algorithms. Enforced disappearance. Computational linguistics. Natural language processing. Testimonies. Colombia.

campo específico de las afectaciones de familiares de víctimas de un fenómeno violento.

La desaparición forzada es una de las formas de violencia más recurrentes en Colombia. El Observatorio de Memoria y Conflicto, del Centro Nacional de Memoria Histórica (CNMH), en agosto de 2021 tenía información de 80 674 personas desaparecidas (Observatorio de Memoria y Conflicto, CNMH, 2021). Por su parte, la Unidad de Búsqueda de Personas dadas por Desaparecidas⁽²⁾ tiene datos sobre más de 120 000 personas desaparecidas (Ortiz Fonnegra, 2019).⁽³⁾ Para Gonzalo Sánchez, exdirector del CNMH, la desaparición forzada es un proceso inverso al re-

velado de una fotografía: lentamente, va borrando la imagen de la víctima hasta hacerla invisible (citado en CNMH, 2016, p. 14). El autor del crimen intenta eliminar todo rastro, todo registro, se esfuerza por ocultar su existencia, establece un manto de duda e incertidumbre sobre la víctima y su familia.

La lucha de familiares de víctimas y organizaciones de derechos humanos ha logrado visibilizar el fenómeno y ha presionado para la creación de leyes que lo tipifiquen como delito.⁽⁴⁾ Desde los años setenta, han formado comunidades afectivas y organizaciones sociales de víctimas para denunciar, visibilizar y exigir justicia. La agregación en comunidades y organizaciones es necesaria, por las características del crimen: impunidad de los responsables, riesgo de vida y estigmatización de miembros de la familia, e invisibilidad de las víctimas.

La desaparición forzada es definida por la “Convención internacional para la protección de todas las personas contra las desapariciones forzadas” (Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, 2006), en su artículo 2, como:

El arresto, la detención, el secuestro o cualquier otra forma de privación de libertad que sean obra de agentes del Estado o por personas o grupos de personas que actúan con la autorización, el apoyo o la aquiescencia del Estado, seguida de la negativa a reconocer dicha privación de libertad o del ocultamiento de la suerte o el paradero de la persona desaparecida, sustrayendo a la protección de la ley.

Además, es considerada por el “Estatuto de Roma” como crimen de lesa humanidad, cuando cumple con características como ser generalizada y sistemática, es decir, cuando deja multiplicidad de víctimas, y cuando se ha llevado a cabo de manera frecuente y reiterada (Corte Penal Internacional, 1998).

La desaparición forzada es una agresión que implica múltiples vulneraciones a los derechos humanos de víctimas directas, sus familiares y comunidades de las cuales hacen parte. Entre estas vulneraciones se evidencian principalmente la violación al derecho a la libertad, a la vida, a la seguridad, a la integridad, entre otras, y en algunos casos se suman a este hecho otros delitos, como tortura y tratos crueles, inhumanos y degradantes. Por esto, y por ser un hecho que se extiende en el tiempo, es uno de los crímenes más atroces, pues genera un sufrimiento que solo puede cesar en el momento en que se obtiene información sobre las circunstancias en las que sucedió y el paradero de la persona desaparecida.

Según el CNMH (2016), al interior de las familias, el daño sufrido es diverso, dependiendo del lugar

de la víctima en la familia y la comunidad. Si bien todos estos comparten la vulnerabilidad ante la violencia, algunas personas son aún más vulnerables, debido a factores como la pobreza, el desempleo o el analfabetismo, lo cual agrava las condiciones de inseguridad y la falta de acceso a instituciones de apoyo y protección.

Ante un delito que pretende borrar la existencia de un ser humano, un recurso para traer de vuelta al desaparecido es la palabra, el testimonio, la denuncia. Estos relatos permiten una aproximación a la experiencia y a las afectaciones que sufren los familiares y las comunidades que lo padecen.

Uno de los problemas que enfrentan hoy los investigadores que analizan este fenómeno en el marco del conflicto armado es cómo procesar ese volumen de información derivada de testimonios, entrevistas, escritos, videos, entre otros. En ese sentido, el objeto de estudio de esta investigación es el procesamiento, la modelación y la comunicación-difusión de ese cúmulo de testimonios sobre afectaciones en familiares de víctimas de desaparición forzada.

La investigación surge de preguntas relacionadas con el fenómeno y con la disposición de información sobre memoria; así como sobre el procesamiento y la modelación de esta información, de manera que puedan ofrecerse formas de comprensión y difusión desde los análisis textuales. Así pues, se busca responder a las preguntas: ¿cómo pueden ser analizados grandes volúmenes de información en relación con los testimonios de la desaparición forzada en Colombia? ¿Es posible crear herramientas automáticas para facilitar la labor de análisis y clasificación de esos testimonios? ¿Cuáles serían las métricas para evaluar la eficacia de un algoritmo de clasificación en comparación con lo que hace una máquina y lo que haría un analista?

En esta investigación se conforma un corpus de testimonios sobre afectaciones de los familiares de las víctimas de desaparición forzada en Colombia, con el objetivo de reconocer patrones para construir una base conceptual que permita el diseño de un algoritmo orientado a la clasificación automática, a partir de la identificación de afectaciones de la desaparición forzada mencionadas en testimonios. Si bien existen otros análisis de esta información concreta en términos de recolección de hechos, el aporte de este trabajo radica en el diseño de una herramienta para realizar análisis automático sobre un aspecto específico de estos testimonios abriendo la posibilidad de facilitar el análisis de amplios volúmenes de información sobre afectaciones a víctimas de violencia.

Este artículo presenta el proceso de construcción de dicho algoritmo y las métricas para evaluar su eficacia, haciendo hincapié en el ejercicio conceptual para la construcción de nodos de clasificación. Se presentan también los resultados que genera el algoritmo comparando sus resultados en relación con el análisis manual.

Para ello, inicialmente, se identifica el campo del procesamiento de lenguaje natural como punto de partida para analizar el corpus conformado. A continuación, se describe la metodología para analizar los testimonios y la construcción de un diccionario de afectaciones, que sirvió como base conceptual para el diseño de un modelo y de un algoritmo para la identificación y clasificación automática de las afectaciones. Después, se presenta el desarrollo del algoritmo y las métricas implementadas para medir la eficacia del mismo. Finalmente, se enuncian algunas reflexiones sobre las limitaciones, posibilidades y futuros desarrollos en la implementación del análisis de datos lingüísticos en corpus de memoria.

2. Procesamiento de lenguaje natural y analítica de datos

El corpus documental responde a un tipo de información que da cuenta de relatos y testimonios de familiares de víctimas en los cuales se aplicaron técnicas de procesamiento de lenguaje natural para la extracción de información y la relación existente entre distintos elementos.

Para ofrecer una plataforma de consulta sobre un corpus de memoria es necesario plantearse, por un lado la estructuración y anotación de los textos para permitir la consulta estructurada y, por otro lado, el desarrollo de herramientas de análisis conceptual que específicamente permitan analizar aspectos de los textos, y por ello nos centramos concretamente en las afectaciones.

El *procesamiento de lenguaje natural* vincula saberes de la lingüística, la analítica de datos y la informática, con el objeto de generar modelos computacionales que reproduzcan uno o más aspectos del lenguaje natural. Se plantea, igualmente, comprobar los modelos lingüísticos y las teorías, diseñando algoritmos y sistemas que puedan ser evaluados y comprendidos en máquinas computacionales.

La *lingüística computacional* está focalizada en el conjunto de procedimientos o métodos para el estudio del lenguaje, con miras a facilitar su exploración, de modo que pueda contribuir en análisis de textos leídos por máquina, a fin de leer, buscar y manipular información en donde la cantidad es relevante (McEnery & Hardie, 2012). Su campo de desarrollo se encuentra igualmente entre la

oralidad y la textualidad, para lo cual requiere disponer de recursos lingüísticos, entre los que se cuentan corpus, diccionarios y gramáticas.

Este proyecto de investigación emplea tecnologías del lenguaje para la construcción de un corpus cuyo análisis sirvió como muestra para la construcción conceptual y la posterior modelación.

Los *corpus* son definidos como un conjunto estructurado de textos que constituyen una muestra lo más realista posible del uso lingüístico. Debe tener un diseño coherente, introducción de marcas en los textos que definan su estructura según estándares comúnmente aceptados, y documentación completa que permita conocer la procedencia y las características de los materiales.

El concepto *corpus de memoria* hace referencia al conjunto de recursos de información relacionados con violaciones a derechos humanos y a procesos de defensa y lucha por la verdad y la justicia en el marco de conflictos armados. Son documentos y objetos de información (tanto digitales como analógicos) que son recopilados o producidos por instituciones u organizaciones de investigación, atención o defensa de los derechos humanos. Para este caso, son recursos dispuestos en la web y enriquecidos con técnicas de procesamiento de lenguaje natural para recuperar contenido significativo de modo que puedan relacionarse elementos como actores, entidades, impactos, modalidades de victimización, entre otros. Estos recursos tienen diversas características formales y de contenido pues van desde documentos relacionados con la denuncia y el registro de hechos hasta testimonios textuales, audiovisuales, y producciones escritas a partir de talleres o ejercicios que son realizados con personas que son víctimas o familiares de víctimas del conflicto.

Como evidencia lingüística, el corpus recoge muestras del uso del lenguaje natural y permite realizar diversos tipos de estudios de su comportamiento. Las colecciones de datos deben conformarse a partir de textos producidos en situaciones reales de comunicación, sean orales o escritas, lo cual dará un carácter de fiabilidad, en tanto esas situaciones sean una muestra representativa de la lengua. “La lingüística de corpus es importante porque permite la construcción de elementos de estudio que luego pueden servir para la comprobación de teorías lingüísticas” (Duque, 2009, p. 28).

Los *corpus textuales* están conformados por colecciones que pueden ser de tipo literario, periodístico, científico. Recopilan información lingüística para el procesamiento de grandes cantidades textuales, que son utilizadas en distintos recursos y aplicaciones. A partir de los corpus, se pueden

obtener conclusiones relacionadas con un escritor, una época, una variedad lingüística, cambios lingüísticos, diferencias en la adquisición y el uso de la lengua según un grupo social, un género, un tema, etc. (McEnery & Hardie, 2012). También pueden conformarse, como en el caso de esta investigación, a partir de asuntos más específicos relacionados con las afectaciones que se expresan en testimonios a través del discurso.

Por su parte, la *analítica de datos* hace referencia al proceso de tomar un conjunto de datos agrupados bajo unos criterios para el estudio de un proceso o fenómeno natural, y el posterior procesamiento para identificar patrones, recurrencias o características que terminen dando una explicación descriptiva y predictiva del proceso o fenómeno estudiado. Típicamente, este término se utiliza cuando el conjunto de datos es muy grande y se necesita desarrollar herramientas para hacer un análisis asistido por sistemas informáticos. En este último caso también se utiliza de forma más precisa el término en inglés *big data* o *minería de datos*.

3. Categorías para analizar el dominio

El análisis de testimonios se plantea a partir de algunas categorías para encontrar otras dimensiones del fenómeno, más allá de datos relacionados con los hechos. Por esa razón se decide abordar los testimonios desde las afectaciones.

Para comprender la dimensión de la afectación producida por la desaparición forzada es necesario visibilizar la vulnerabilidad a la que sus familiares se ven expuestos y que se manifiesta en los relatos sobre la manera como el evento transformó sus vidas.

Para efectos de esta investigación, se entiende por *afectaciones* las consecuencias o efectos de la desaparición forzada que alteran, perjudican, cambian de manera negativa y abrupta el curso de vida de individuos, familias y comunidades. Se agruparon las afectaciones en: económicas, físicas, políticas, psicológicas y socioculturales. Esta definición de afectaciones se basó en trabajos sobre el tema (Castrillón Rivera *et al.*, 2007; Charry Lozano, 2016; CNMH, 2014, 2016; Durán Pérez *et al.*, 2004; Rebolledo y Rondón, 2010). A continuación, se describen cada una de ellas.

Afectaciones económicas: derivadas del deterioro de condiciones económicas y materiales en las cuales se sustenta la calidad de vida y el desarrollo de individuos, familias y comunidades. Estas condiciones pueden estar relacionadas con el hecho de que la víctima directa de la desaparición es quien proveía el sustento familiar, o con la pérdida de propiedades como la vivienda, tierras,

fuentes de ingresos como negocios propios o el empleo. Sobre estas afectaciones, es pertinente mencionar que pueden ser producto de hechos asociados a la desaparición forzada en un contexto específico de violencia, que dé cuenta de intereses de actores armados, y que se caracterizan por procesos de desplazamiento forzado, donde familias y comunidades se ven obligadas a abandonar sus tierras y bienes por amenazas, o por temor a ser objeto de nuevas agresiones.

Afectaciones físicas: se relacionan con el deterioro de las condiciones de salud física y se pueden expresar en la aparición o agravamiento de enfermedades. Si bien las víctimas manifiestan un deterioro de la salud física en general y un decaimiento asociado con la ocurrencia del hecho victimizante, se puede hablar de algunas enfermedades comunes para los casos específicos de desaparición forzada, como alteraciones del sueño, cáncer, enfermedades cardiovasculares, desnutrición, dolores, entre otras (CNMH, 2016).

Afectaciones políticas: son aquellas que debilitan, desestructuran o eliminan identidades políticas, expresiones de movimientos o procesos políticos que se identifican como diferentes o contrarios a los intereses del perpetrador. Pueden estar relacionadas con prácticas dirigidas a desincentivar la participación político-electoral, a acallar la oposición política, a imponer el silenciamiento de voces disidentes y obstaculizar la organización comunitaria.

Afectaciones psicológicas: efectos de la desaparición forzada relacionados con el deterioro de la salud psíquica y emocional, que impiden el desenvolvimiento adecuado de los individuos en los distintos ámbitos de la vida. Estas afectaciones pueden expresarse en el desencadenamiento de trastornos mentales o en la aparición de síntomas, emociones, comportamientos y sentimientos que generan sufrimiento. En el campo de las emociones, son frecuentes las expresiones como angustia, culpa, impotencia, incertidumbre, miedo, ira, y la persistencia de un duelo alterado que no se concreta por la inexistencia de un cuerpo o la esperanza de encontrarlo con vida.

Afectaciones socioculturales: hacen referencia a los efectos de las desapariciones sobre el tejido social, el sistema de creencias y tradiciones de las comunidades, y los elementos que se desprenden de estos, como las formas de organización, la cosmovisión, los rituales y, en general, todo tipo de acuerdos tradicionales construidos sobre la base de la identidad cultural. Entre las afectaciones socioculturales más frecuentes en los testimonios se encuentra la estigmatización, entendida como la valoración que se hace de una

persona o colectivo con la que se busca identificar y resaltar su condición de diferente no aceptado, “a partir de un rasgo desacreditador que se contrapone al estereotipo acerca de cómo deben funcionar los hechos de la realidad” (Castrillón Rivera *et al.*, 2012, p. 45). Para el caso de víctimas de desaparición forzada, esta afectación se origina en discursos que circulan socialmente, que las asocian con conductas y comportamientos reprochables, o que las acusan de hacer parte de grupos armados, o tener algún tipo de vínculos, con lo que se busca justificar su victimización. También es frecuente la referencia al *descrédito*, es decir, el no reconocimiento del daño producido, asociado a la indiferencia, la estigmatización y falta de atención del Estado.

La memoria es construida a partir de los relatos y por tanto está en el mundo del lenguaje. Estos testimonios han sido generados en espacios de conversación dentro de investigaciones, atención psicosocial, instancias judiciales, entre otros escenarios. Son relatos que cuentan, además de las circunstancias en las que se dan los hechos para aportar en el proceso de búsqueda, la manera como éstos han marcado las vidas de familiares y comunidades, porque cuando se da la desaparición hablamos de varios niveles en los que se están afectando las vidas de las personas, que van desde lo emocional con relación a sentimientos de vacío, angustia, frustración y rabia, pasando por el cuerpo y por lo material, pues muchas familias, luego de la desaparición de un ser querido, se ven enfrentadas a cambiar su vida, a desplazarse, a vivir con miedo y temor del señalamiento.

La decisión de abordar el asunto de la clasificación de textos desde las afectaciones está relacionada con el interés de construir un léxico sobre emociones presentes en testimonios de víctimas de conflicto armado que sirva como herramienta para tareas de clasificación y análisis por parte de investigadores y como aporte a formas de difusión de la memoria. Y el trabajo en un corpus pequeño con una sola modalidad de victimización brinda las pistas para la construcción de corpus más amplios sobre violación de los derechos humanos en el marco del conflicto armado.

Aquí entran en consideración elementos para pensar el campo del trabajo con afectaciones, específicamente emocionales. El análisis de sentimientos, según Liu (2017), no es solo un problema de clasificación y por tanto estudiar la subjetividad presente en textos puede ayudar a reconocer variaciones y construir herramientas más sofisticadas que diferencian aspectos como el afecto, la emoción, el humor. El tratamiento de los sentimientos/emociones en el lenguaje in-

cluye la preocupación por la subjetividad, los sentimientos y las creencias (Lui, 2015). Este autor también diferencia la orientación, intensidad y calificación del sentimiento y aporta una relación de las emociones básicas planteadas desde distintas teorías.

El léxico o diccionario emocional puede luego ser usado en sistemas automáticos de clasificación con el propósito de ampliar las herramientas léxicas para el español en un dominio específico: la comprensión de las emociones representadas en los discursos o testimonios de víctimas de conflicto armado.

Joshi *et al.* (2017) definen el *Sentiment resource* (lexicon) como un repositorio de unidades textuales marcadas con etiquetas que representan un sentimiento y que está compuesto por una unidad textual y etiquetas correspondientes.

4. Método

El método implementado en esta investigación es mixto. Toma elementos de la lingüística de corpus y del procesamiento de lenguaje natural, para analizar un conjunto de testimonios y construir una base conceptual, con el objetivo de proponer un modelo de clasificación automática. Se usan también métodos cuantitativos, para establecer las métricas de evaluación del algoritmo.

Para el desarrollo de esta investigación, se hizo una descripción de las afectaciones de los familiares, enunciadas en testimonios como consecuencia de la desaparición forzada. Los 391 testimonios del corpus fueron recopilados de informes de memoria o investigaciones académicas realizadas por organizaciones oficiales, investigadores y asociaciones de víctimas. Se encuentran recursos de prensa, informes de instituciones jurídicas de defensa de derechos humanos, colectivos y organizaciones sociales y de víctimas, así como información producida en procesos académicos e investigativos.

El principal criterio de selección de testimonios para el corpus consistió en que el texto diera cuenta de alguna de las afectaciones definidas. Generalmente, los testimonios refieren información sobre los hechos y las circunstancias de desaparición de personas o de aspectos relacionados con el proceso de búsqueda de las víctimas. Por tal motivo, se tuvieron en cuenta sólo aquellos fragmentos que hacían referencia a algún tipo de afectación, considerando, además, elementos que aportaran contexto dentro del mismo fragmento. En este punto interesaba especialmente reconocer las distintas formas en las que en el texto se manifiestan afectaciones, en

tal sentido se reconocía tanto la palabra que explícitamente aparecía como algunas figuras retóricas (metáfora, ironía, hipérbaton, eufemismo) para referir afectaciones o daños.

Por otra parte, como criterio de extensión mínima de los fragmentos, se tomó como base que el mismo se constituyera en una expresión cuyo conjunto de elementos diera cuenta del sentido en el marco de las afectaciones. De esta manera, la forma más simple de un fragmento podía corresponder a una única oración de tipo declarativo, desiderativo, dubitativo o interrogativo siempre y cuando representaran situaciones y emociones relacionadas con las afectaciones definidas. Se estableció este criterio debido a las dificultades para acceder a testimonios referidos al delito de desaparición forzada y sus afectaciones, ya fuera por causas de no accesibilidad por protección de la información o porque los testimonios disponibles, como ya se mencionó, referían principalmente a hechos y circunstancias de la desaparición de las personas y aspectos relacionados con su búsqueda. Es así como la variación de la cantidad de palabras de los testimonios del corpus se encuentra en un rango de 10 a 4981 y en total el corpus se compone de 171 364 palabras.

Una vez recopilado el corpus, se realiza la etapa de modelación, donde se hace anotación manual, que consiste en la lectura de testimonios y la selección de los apartados o fragmentos y la anotación de las afectaciones identificadas en dichos testimonios, con ayuda del *Software Nvivo*, herramienta para análisis cualitativo que permite codificar texto y explorar los datos para visualización y descubrimiento de temas y asociación entre datos. Este proceso fue realizado por cuatro integrantes del equipo, a partir del análisis de contenido, se procedió a un etiquetado exploratorio, sobre el 30% del corpus equivalente a 116 testimonios. Previa creación de los nodos y subnodos en Nvivo (codificación deductiva), cada etiquetadora debía realizar el proceso de anotación a 29 testimonios. Adicionalmente, y atendiendo a la técnica de codificación inductiva, también se podían crear nodos y subnodos según se complementaran nuevos niveles de la categorización. Con estas condiciones establecidas se procedió a reclasificar conforme a la nueva distribución de nodos y al ajuste de las anotaciones de párrafos, los cuáles se fragmentaron a partir de la extracción de las oraciones de interés. Estas oraciones o expresiones sirvieron para conformar el diccionario de afectaciones al igual que los nodos y subnodos bajo los cuales fueron clasificadas.

En cuanto a la distribución del corpus, este 30% fue destinado para la etapa de entrenamiento del algoritmo y un 5% (20 testimonios) para testeo. La distribución se realizó de manera aleatoria

bajo el enfoque *Train-Test Split* (Brownie, 2020). De los testimonios de entrenamiento, 20 fueron analizados manualmente de forma paralela al algoritmo para contrastar resultados y medir la eficacia de la clasificación. Finalmente, una vez probado el algoritmo y sus métricas este fue aplicado a todo el corpus.

Por su parte, la elaboración del diccionario se desarrolló en función del trabajo conceptual y de los resultados de anotación. Para el trabajo conceptual se tomaron como principales referentes al CNMH (2016 y 2014) y a Pérez Sales, P., et.al. (1998). A partir de la elaboración de fichas analíticas se descubrieron y caracterizaron a través de definiciones, términos asociados y, en ocasiones, algunas expresiones, las categorías de *daños*, *impactos* y *afectaciones*.

El diccionario fue usado posteriormente en las fases de entrenamiento y testeo en la construcción del algoritmo, el cual fue desarrollado en Python 3.

Afectaciones	Sinónimos-palabras	Expresiones
Económicas	Desempleo, Desplazamiento, Escasez, Necesidad, Pérdida, Pobreza...	"La pobreza aumentó", "No pude volver a trabajar", "Nos tuvimos que desplazar"...
Físicas	Achaque, Afección, Decaimiento, Dolencia, Enfermedad, Insomnio, Malestar, Maluquera...	"Comencé a enfermarme", "He sufrido muchas enfermedades", "Murió de pena moral"...
Políticas	Clandestinidad, Desconfianza, Impunidad, Inseguridad, Persecución, Represión, Señalamientos...	"Desaparecidos de la Unión Patriótica", "La organización se debilitó", "Se generaron desconfianzas entre las personas"...
Psicológicas	Ansiedad, Sufrimiento, Zozobra...	"Casi no me puedo acostumbrar a ver las fotografías", "El dolor nunca pasa", "Ella se decayó mucho"...
Socioculturales	Indiferencia, No reconocimiento, Revictimización	"Allá nadie ve nada", "No recibí ninguna ayuda", "Por algo les pasó lo que les pasó"...

Tabla 1. Niveles de la estructura del diccionario

La *estructura del diccionario* se presenta en tres niveles. En el primero se encuentran las afectaciones según el aspecto de la vida en el que inciden de manera directa. Luego, en el segundo nivel, se ubican sinónimos que representan o están asociadas a cada una de las afectaciones. Y, finalmente, en el tercer nivel, están las expresiones (Tabla 1).

5. Modelo, algoritmo y métricas

Para evaluar la clasificación de datos se emplean distintas métricas las cuales pueden ser: métricas para evaluar la capacidad de generalización del clasificador entrenado, métricas de evaluación que determinan cuál es el mejor clasificador de distintos modelos entrenados; y las métricas de evaluación que permiten elegir una solución óptima para clasificar (Hossin & Sulaiman, 2015).

En este trabajo se define como métrica de evaluación la eficacia para reconocer términos y expresiones relacionados con las afectaciones en comparación con la selección y clasificación que se haría manualmente.

A partir del diccionario construido con las palabras y expresiones clasificadas según la afectación, se construyó un modelo para la identificación de las afectaciones en un testimonio. El *modelo* hace referencia a la construcción conceptual por medio de la cual se da cuenta de las afectaciones como fenómeno ligado a las desapariciones forzadas y que se evidencia en las palabras utilizadas en los testimonios.

La identificación de afectaciones en un testimonio de desaparición forzada normalmente implica la lectura manual por parte de un experto en el área, capaz de identificarlas. Este trabajo, si bien es completamente realizable para un número limitado de testimonios, cuando el conjunto de testimonios aumenta, se convierte en una labor desafiante. Por tanto, se vuelve importante desarrollar herramientas de lectura e identificación automática de afectaciones. El *algoritmo* se constituye, entonces, en la serie de pasos automáticos mediante los cuales es posible identificar afectaciones que aparecen en uno o más testimonios.

Si bien los tipos de afectaciones han sido ya especificadas en la sección anterior a nivel textual, se escoge una unidad de medida que indique la presencia u ocurrencia de una afectación en un testimonio. Para este proceso, se optó por dos unidades de análisis, para combinarlas en el modelo.

- *Unidad 1: la palabra.* La aparición de una palabra en un testimonio marca la ocurrencia de una o más afectaciones. Es decir, se puede atribuir, a una palabra, una o varias afectaciones, y el hecho de encontrar esta palabra en un testimonio sería indicador suficiente de que en el testimonio acaece(n) la(s) afectación(es) relacionada(s). Es, por tanto, indispensable construir un conjunto de palabras asociadas a cada una de las afectaciones, o en otros términos, bajo esta definición de unidad, una afectación vendría determinada por

un conjunto de palabras que marcan su existencia en un texto.

- *Unidad 2: la expresión.* Se reconoce que no solo en palabras está el contenido de las afectaciones, sino también en expresiones completas, frases que determinan una o varias afectaciones. Estas expresiones se tratan de forma rígida, es decir, se establece que dichas expresiones ocurren en el lenguaje bajo una única forma y, por tanto, para identificar la afectación o afectaciones asociadas, la expresión debe aparecer íntegramente en el texto.

Utilizamos ambas unidades para contar la cantidad de veces que ocurre cada afectación en un testimonio. Si aparece una palabra o una expresión de una afectación determinada, se cuenta por una ocurrencia. Con este método, se suma la cantidad de veces que aparece cada afectación y se establecen así varias métricas para categorizar los testimonios.

Todo el texto es procesado en Python 3 con la ayuda del paquete NLTK (Natural Language Toolkit), para hacer un preprocesamiento que incluye quitar las palabras irrelevantes, símbolos y signos de puntuación, así como la lematización.

Por otro lado, las *métricas* son la medida que se utiliza para poder categorizar cada testimonio de acuerdo con las afectaciones identificadas. Con el conteo de ocurrencias de las afectaciones por testimonio, se definieron dos métricas: la primera, para asignar al testimonio una única afectación como la de mayor ocurrencia. Si dos afectaciones ocurren la misma cantidad de veces, entonces se decide aleatoriamente entre ambas. Y la segunda, en la que se dividen las ocurrencias de cada afectación por el total de ocurrencias de todas las afectaciones, para obtener porcentajes de afectación por cada testimonio. Estas dos métricas permiten estudios diferenciales y un nivel de detalle diferente para entender el alcance y la precisión del algoritmo a la hora de encontrar afectaciones en testimonios.

Para medir la eficacia del algoritmo se hizo el testeó, separando un conjunto de testimonios que no habían sido etiquetados anteriormente y a los cuales se les identificaron las afectaciones manualmente, y se comparó con la identificación realizada por el algoritmo.

Utilizando la métrica 1, se pudo establecer que el algoritmo identifica la misma afectación predominante que la lectura manual en $56,25 \pm 28,57$ %; y si se agrupan en categorías de afectaciones, la eficiencia sube al $78,57 \pm 21,43$ % de las veces. Es importante precisar dos aspectos intrínsecos

al problema, en primer lugar la limitación estadística en *términos* del número de testimonios, la cantidad de palabras de los mismos y la población de afectaciones en función de los testimonios. Por lo tanto, se pueden esperar desviaciones estándar altas debido a la estadística del corpus y adicionalmente se entiende también que al agrupar los testimonios en categorías de afectaciones mejora la identificación puesto que se pueblan mejor cada categoría con la estadística limitada, a la vez que se pueblan también más ricamente los diccionarios por categoría de afectación. Finalmente, entiéndase las incertidumbres en este trabajo como intervalos de confianza máximos y no como mínimos dado que preferimos establecer un intervalo mayor pero con una mayor probabilidad de contener los valores verdaderos.

Ahora bien, con la métrica 2 se define una distancia promedio entre las clasificaciones para un mismo testimonio, de la siguiente forma:

$$D_T = \sum_{i=1}^n |A_i^M - A_i^A|$$

donde D_T es la distancia entre las clasificaciones A_i^M , el porcentaje de ocurrencia de la afectación i atribuido por la clasificación A_i^A , por la clasificación hecha por el algoritmo. Finalmente, n corresponde al total de afectaciones.

Mientras más pequeña sea esta distancia, quiere decir que ambas clasificaciones son más similares o se diferencian menos. La distancia media encontrada es 0,076. Dado que la definición de esta métrica se hace en función de las características del problema, resulta tremendamente difícil definir un punto de referencia para esta distancia sin incurrir en sesgos del problema mismo, por tanto los valores deben relativizarse al contexto de esta investigación.

Adicionalmente, se define una distancia entre todos los testimonios para cada afectación, con el fin de medir cómo el algoritmo puede reconocer cada una de las afectaciones. Esta *distancia por afectación* se define de la siguiente forma:

$$D_A = \frac{1}{m} \sum_{i=1}^m |A_i^M - A_i^A|$$

Aquí, la A_i^A es el promedio de porcentajes de todos los testimonios manuales para la afectación i , y A_i^M , el promedio de todos los testimonios del algoritmo para la afectación i . Además, m es el número total de testimonios.

La diferencia principal entre estas dos distancias es que la primera mide la similitud entre la clasificación manual y la del algoritmo en un solo testimonio, dando así una medida de la cercanía de

ambas clasificaciones teniendo en cuenta todas las afectaciones. La segunda distancia establece una medida de similitud entre las clasificaciones teniendo en cuenta todos los testimonios, para tratar de disminuir el sesgo que podríamos tener por la variabilidad de los elementos del corpus.

Afectación	$D_A \pm 0,066$
Económicas	0,092
Físicas	0,039
Políticas	0,004
Psicológicas	0,128
Angustia	0,118
Culpa	0,009
Impotencia	0,131
Incertidumbre	0,094
Miedo	0,023
Ira	0,003
Socioculturales	0,041
Estigmatización	0,028
Descrédito	0,009

Tabla II. Distancia por afectación

En la Tabla II se muestran las distancias por afectación de acuerdo con lo definido. Mientras más pequeña la distancia, quiere decir que el algoritmo identifica la afectación en cuestión de forma más similar al procedimiento manual.

Este proceso de validación demuestra que el algoritmo y el modelo desarrollado son aceptables para clasificar testimonios en categorías de afectaciones, y moderadamente aceptable para identificar afectaciones particulares.

Utilizando la métrica 1, se puede concluir que, basados en la clasificación de una única afectación por testimonio, el algoritmo es eficaz y utilizable para grandes bases de datos, que pueden dar una confianza moderada en su clasificación. Por otro lado, utilizando la métrica 2, se puede ver que la diferencia en las clasificaciones es baja y se espera que el espectro de afectaciones identificadas para un testimonio no se aparte de forma significativa de la clasificación manual.

Por otra parte, este mismo conjunto de testimonios utilizados para la validación sirvieron para medir una incertidumbre sobre las asignaciones de afectaciones encontradas por el algoritmo. Para esto, se definió un valor de *identificación correcto por afectación*, denotado por ϵ_i^{ID} , como el número de veces en que fue identificada correc-

tamente la afectación, dividido por el total de veces que ocurre dicha afectación en relación con la clasificación manual. Y también se definió un valor de *identificación erróneo por afectación*, denotado por ϵ_i^{ER} , como la cantidad de veces que se asignó la afectación, pero que la afectación identificada manualmente era distinta, y dividido por el total de veces que ocurren las demás afectaciones.

Afectación	ϵ_i^{ID}	ϵ_i^{ER}	σ_i
Económicas	0,308	0,022	0,015
Físicas	0,286	0,005	0,004
Políticas	0,000	0,000	-
Psicológicas	0,659	0,368	0,126
Angustia	0,927	0,167	0,012
Culpa	0,000	0,000	-
Impotencia	0,333	0,067	0,045
Incertidumbre	0,158	0,028	0,023
Miedo	0,000	0,000	-
Ira	0,000	0,000	-
Socioculturales	0,200	0,005	0,004
Estigmatización	0,000	0,005	-
Descrédito	0,000	0,000	-
Promedio	0,410	0,075	0,03

Tabla III. Medición de eficiencia e incertidumbre por afectación

Se definió la *incertidumbre del algoritmo por afectación* como

$$[\sigma_i = \epsilon] _i^{ER} [(1 - \epsilon) _i^{ID}]$$

Dado que se requerirían demasiados testimonios para poder determinar correctamente estas cantidades y debido a que la estadística es limitada, se define una incertidumbre promedio entre las que se pudo medir y se le asigna por igual a todas las afectaciones. Adicionalmente, en los casos en que no se logra medir una eficacia debido a las limitaciones estadísticas, no se considera posible medir una incertidumbre asociada. Si bien es difícil establecer qué tan justa es esta estimación al tomar el promedio, se encuentra en un punto medio, que se puede esperar que no subestime o sobreestime por mucho los valores reales de la incertidumbre. Los resultados para las eficiencias e incertidumbre discutidas están en la Tabla III. En este caso, y de acuerdo con las decisiones explicadas en la metodología descrita, no aventuramos valores aceptables para la incertidumbre, dadas las limitaciones estadísticas del problema.

Una vez definidas las métricas y la eficacia, se aplica extensivamente el algoritmo sobre el conjunto completo de los 391 testimonios, de los cuales se muestran los valores sintetizados en las Figuras 1 y 2.

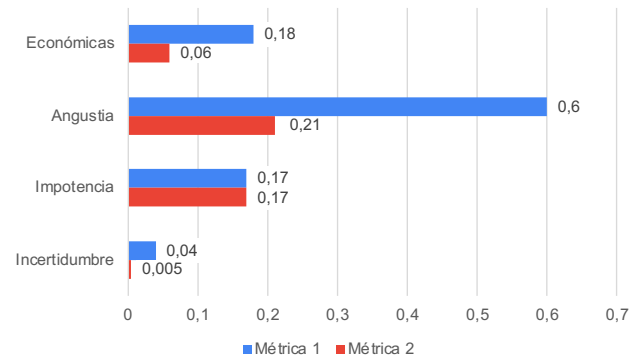


Figura 1. Comparativo de afectaciones utilizando las métricas 1 y 2

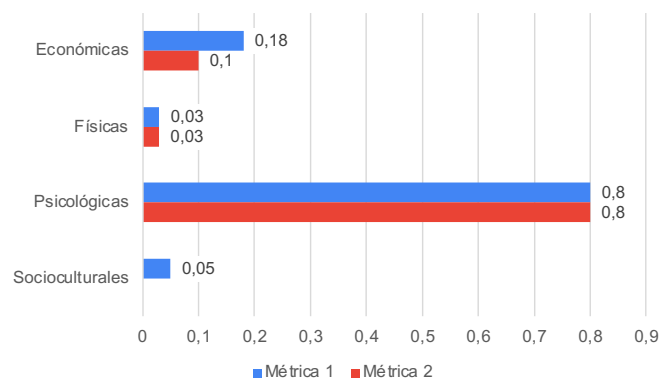


Figura 2. Categorías y su respectivo error en el total del corpus

De acuerdo con los resultados de nuestro modelo y algoritmo, y según se observa en las Figuras 1 y 2, las afectaciones psicológicas son fuertemente dominantes, seguidas por las afectaciones de orden económico.

En las Figuras 1 y 2 podemos ver que las afectaciones psicológicas y económicas tienden a ocurrir de manera simultánea en los testimonios; en otras palabras, típicamente los testimonios que reflejan afectaciones psicológicas también reportan afectaciones económicas. De las Figuras 1 y 2 podemos inferir que, en una muestra de testimonios, lo más probable sería encontrar afectaciones psicológicas y económicas.

Adicionalmente, de la Figura 1 identificamos que la mayoría de testimonios expresan angustia e impotencia ante las desapariciones forzadas. Estas dos afectaciones aparecen correlacionadas,

dándose de forma simultánea con facilidad en los testimonios analizados. Además notamos cómo, dentro de las afectaciones psicológicas, la angustia y la impotencia son dominantes sobre el resto de afectaciones encontradas.

6. Plataforma de divulgación

Con el fin de presentar los resultados y con la intención de visibilizar las afectaciones presentes en los testimonios se diseñó un sitio web⁽⁶⁾ que recoge los procedimientos ejecutados en el proyecto, con el fin de disponer el algoritmo en su fase demo para potenciales usuarios, como investigadores en áreas de memoria y desaparición forzada, y con intereses en procesamiento de lenguaje en corpus de testimonios de víctimas, organizaciones defensoras de derechos humanos, víctimas relacionadas con desaparición forzada, y visitantes en general.

Se consideró además la posibilidad de validar testimonios, tomando como base los diccionarios y el algoritmo referenciados anteriormente. Así mismo, es una alternativa que permite aportar nuevos testimonios al proyecto, que puedan ser integrados al corpus para mejorar la precisión de los resultados del algoritmo.

7. Discusión y conclusión

Aquí se plantean algunas cuestiones en relación con el método y con la evaluación de los resultados obtenidos con el uso del algoritmo, en contraste de lo que sería una lectura manual, y con la perspectiva de ofrecer herramientas que les permitan a investigadores sobre memoria del conflicto reconocer afectaciones y determinar una clasificación de textos a partir de ellas, tanto para el fenómeno de la desaparición forzada como para otros relacionados. Estas reflexiones están planteadas en la mención de algunas limitaciones intrínsecas al trabajo con analítica de datos, la valoración de resultados del modelo para investigación sobre memoria y algunas líneas de trabajo futuro en la lingüística computacional aplicada a corpus de memoria.

En cuanto a las limitaciones, se encuentra la necesidad de incluir más volumen de testimonios, que si bien es un asunto intrínseco al análisis de datos, contar con más testimonios permitiría evidenciar nuevas categorías en los ámbitos individual, familiar o colectivo, o en relación con actores o formas de victimización.

En cualquier caso, dado que esta investigación implicaba una conceptualización no solo del fenómeno, sino también de las afectaciones mismas, y de que los datos obtenidos fueron cuida-

dosamente seleccionados, los resultados comparativos entre la clasificación manual y la clasificación automática son bastante aceptables.

También se hace necesario aplicar el algoritmo a nuevos textos para corroborar su eficacia con corpus de otras formas de victimización y hacer una validación. Si bien implementamos técnicas para evitar los sesgos, sería importante usar otros datos que previamente no hayan sido tratados.

La modelación realizada con el algoritmo permite identificar porcentualmente las diferentes afectaciones. En las publicaciones analizadas para esta investigación estaban presentes todas las afectaciones, eran descritas y analizadas. Sin embargo, el modelo de análisis nos permite identificar la prevalencia y recurrencia de afectaciones psicológicas muy por encima de otras de carácter físico, económico, político, social, y dentro de aquellas, la angustia y la impotencia como dos emociones preponderantes.

Herramientas como estas pueden aportar a la comprensión de afectaciones generadas por desaparición forzada sobre individuos, familias y comunidades, en la perspectiva de avanzar en la política pública relacionada con los derechos de las víctimas, especialmente con el carácter reparador y el principio de integralidad de las acciones de resarcimiento que debe garantizar el Estado.

Sobre las perspectivas de estudios futuros acerca de la desaparición forzada u otros hechos victimizantes, puede ser de interés el análisis de los efectos diferenciales en la población, acudiendo a la posibilidad de ampliar el corpus con testimonios de sectores específicos, como mujeres, hombres, jóvenes, niños y niñas, personas que pertenecen a organizaciones comunitarias o movimientos sociales.

En cuanto a la aplicación de modelos de analítica de datos y lingüística computacional en corpus de memoria, se destacan la necesidad de construir más diccionarios y desarrollar conceptualmente un campo que permita hacer mejores clasificaciones, por ejemplo, ampliando o bien la relación de las afectaciones con otros elementos presentes en los textos o bien la variedad de corpus, para explorar cómo se comportan estas mismas afectaciones expresadas en relación con otras victimizaciones o con testimonios recogidos bajo otro enfoque.

Otro elemento para explorar es el análisis automático de textos utilizando herramientas de procesamiento de lenguaje natural. Esto permitiría entrenar corpus más grandes y a partir de la construcción de diccionarios más automáticos, que posibiliten cruzar afectaciones con información geográfica, nombres de entidades y actores,

de manera que puedan hacerse otros análisis mediante categorización simultánea en múltiples niveles. Esto, nuevamente, requiere datos en cantidad, porque la recopilación y la validación de la veracidad y la calidad de los datos debe ser garantizada por medio de esquemas o guías estandarizadas.

Por último, un campo que resulta de gran interés es el procesamiento de audios, puesto que gran cantidad de testimonios son recogidos en estos formatos, y la lectura automática y clasificación a partir de estos esquemas puede aportar sin duda en tareas de análisis por parte de investigadores sobre memoria histórica.

Notas

- (1) Este artículo es producto de la investigación "Bases para un modelo de análisis textual y etiquetado semántico de testimonios de víctimas y sobrevivientes del conflicto armado colombiano", financiada con recursos del Fondo Primer Proyecto. Vicerrectoría de Investigación. Comité para el Desarrollo de la Investigación de la Universidad de Antioquia. Código proyecto 2018-23597.
- (2) En el acuerdo firmado entre el Estado colombiano y las Fuerzas Armadas Revolucionarias de Colombia en 2016, se creó la Unidad de Búsqueda de Personas dadas por Desaparecidas, que se inserta en el Sistema Integral de Verdad, Justicia, Reparación y No Repetición, junto con la Comisión de la Verdad y la Justicia Especial de Paz. La Unidad de Búsqueda es una instancia que tiene fines humanitarios y extralegales, orientada exclusivamente a la búsqueda de personas desaparecidas, secuestradas y reclutadas por grupos armados ilegales en el contexto del conflicto armado colombiano (Comisión de la Verdad, Justicia Especial para la Paz y Unidad de Búsqueda de Personas Dadas por Desaparecidas, s. f.).
- (3) Si bien la Unidad de Búsqueda de Personas dadas por Desaparecidas recoge las cifras suministradas por el Observatorio de Memoria y Conflicto, las cifras aumentan, porque en sus registros incluyen a personas reclutadas, secuestradas y excombatientes dados por desaparecidos en el marco del conflicto armado.
- (4) En Colombia, la Ley 589 de 2000 (Colombia, Congreso de la República, 2020) atribuye las responsabilidades a particulares e integrantes de grupos armados y a servidores públicos o particulares que cometan el delito bajo la determinación o con la aquiescencia de estos.
- (5) Disponible para consulta en: <https://entrelines.com.co/>

Referencias

Aguado de Cea, Guadalupe; Álvarez de Mon; Rego, Inmaculada; Pareja Lora, Antonio (2002). Primeras aproximaciones a la anotación lingüístico-ontológica de documentos de la Web Semántica: OntoTag. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*. 17, 37–49

Browniee, Jason (2020). Train-Test Split for Evaluating Machine Learning Algorithms. // *Python Machine Learning*

Buendía, Miriam (2010). Anotación semántica en el dominio especializado de la Meteorología. // Caballero, Rosario; Pinar Sanz, María Jesús (Ed.). *Modos y formas de la comunicación humana (923-934)*. Cuenca: Ediciones de la Universidad de Castilla-La Mancha / AESLA.

Castrillón Rivera, D. G.; Liscano Pinzón, L. M.; Suárez Hueras, A. M. (2012). Contraste de las formas de estigmatización frente a la desaparición forzada: entre el discurso de las víctimas y la opinión de la sociedad civil. [Tesis de grado, Pontificia Universidad Javeriana]. Repositorio de la Pontificia Universidad Javeriana. <https://repository.javeriana.edu.co/handle/10554/7936>

Centro Nacional de Memoria Histórica (CNMH). (2014). *Desaparición forzada. Tomo III: Entre la incertidumbre y el dolor: impactos psicosociales de la desaparición forzada*. Imprenta Nacional.

Centro Nacional de Memoria Histórica (CNMH). (2016). *Hasta encontrarlos: el drama de la desaparición forzada en Colombia*. CNMH.

Charry Lozano, Liliana (2016). Impactos psicológicos y psicosociales en víctimas sobrevivientes de masacre selectiva en el marco del conflicto suroccidente colombiano en el año 2011. *Colombia Forense*, 3:2, 53-62. <https://doi.org/10.16925/cf.v3i2.1756>

Colombia, Congreso de la República (2000). Ley 589, por medio de la cual se tipifica el genocidio, la desaparición forzada, el desplazamiento forzado y la tortura; y se dictan otras disposiciones (2000, 6 de julio).

Comisión de la Verdad; Justicia Especial para la Paz; Unidad de Búsqueda de Personas Dadas por Desaparecidas (s. f.). *Sistema Integral de Verdad, Justicia, Reparación y No Repetición*. <https://www.jep.gov.co/Infografas/SIVJRNRES.pdf>

Corte Penal Internacional (1998). *Estatuto de Roma de la Corte Penal Internacional*. [https://www.un.org/spanish/law/icc/statute/spanish/rome_statute\(s\).pdf](https://www.un.org/spanish/law/icc/statute/spanish/rome_statute(s).pdf)

Duque, Erika Teresa (2009). Metodología para la extracción de metadatos semánticos de textos en español utilizando procesamiento de lenguaje natural: subaplicación para la identificación de contextos espaciales y temporales en textos que describan interacciones entre actores. Trabajo de grado, Universidad EAFIT, Medellín. https://repository.eafit.edu.co/bitstream/handle/10784/1261/erika_duque_2009.pdf

Durán Pérez, Teresa; Baci Herzfeld, Roberta; Pérez Sales, Pau (2004). Muerte y desaparición forzada en la Araucanía: una aproximación étnica. Universidad Católica de Temuco. <http://www.pauperez.cat/wp-content/uploads/2017/011/perez-sales-muerte-y-desaparicion-forzada.pdf>

Eriksson, Henrik (2007). An Annotation Tool for Semantic Documents. // Franconi E., Kifer M., May W. (eds). *The Semantic Web: Research and Applications. ESWC 2007. Lecture Notes in Computer Science*, vol 4519. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72667-8_54

Hossin, M; Sulaiman, M. N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*. 5: 2, 01-11.

Joshi, Aditya; Bhattacharyya, Pushpak; Ahire, Sagar. (2017). Sentiment Resources: Lexicons and Datasets. // *A practical guide to sentiment analysis*. Springer.

Liu, Bing. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Estados Unidos: Cambridge University Press

Liu, Bing (2017). *Many Facets of Sentiment Analysis. // A practical guide to sentiment analysis*. Springer.

McEnery, Tony; Hardie, Andrew. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Observatorio de Memoria y Conflicto, Centro Nacional de Memoria Histórica (CNMH) (2021). *El conflicto en cifras*. <http://micrositios.centrodememoriahistorica.gov.co/observatorio/portal-de-datos/el-conflicto-en-cifras/desaparicion-forzada/>

Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos (2006). Convención Internacional para la protección de todas las personas contra las desapariciones forzadas. <https://www.ohchr.org/sp/professionalinterest/pages/conventionced.aspx>

Ortiz Fonnegra, María Isabel (2019, mayo 29). "Hay que encontrar a más de 120.000 desaparecidos por el conflicto". El Tiempo. <https://bit.ly/32rRAYC>

Perez Sales, Pau; Bacic, Roberta; Durán, Teresa (1998). Muerte y Desaparición Forzada en la Araucanía. Una perspectiva étnica. Santiago de Chile: Editorial LOM.

Rebolledo, Olga; Rondón, Lina. (2010). Reflexiones y aproximaciones al trabajo psicosocial con víctimas individuales y colectivas en el marco del proceso de reparación. Revista de Estudios Sociales. 36, 40-50. <https://journals.openedition.org/revestudsoc/13259>.

Enviado: 2022-03-03. Segunda versión: 2022-10-27.
Aceptado: 2022-10-27.
