

Prueba para certificado turniting a entregar a convocatoria Fodein

by MICHAEL STEVEN ALFARO DIAZ

Submission date: 29-Jul-2019 05:18PM (UTC-0500)

Submission ID: 1156051440

File name:

7544_MICHAEL_STEVEN_ALFARO_DIAZ_Prueba_para_certificado_turniting_a_entregar_a_convocatoria_Fodein_274562_1738794016.docx
(348.3K)

Word count: 5405

Character count: 30020



1
DECIMOQUINTA CONVOCATORIA PARA EL FOMENTO DE LA INVESTIGACIÓN Y LA INNOVACIÓN 2020

Título del proyecto	
Análisis de la contaminación del aire por material particulado usando minería de texto y análisis de series temporales en Colombia	
Campo de acción	Transdisciplinariedad - Aporte al PIM
Ambiental	La investigación a desarrollar aporta a las líneas de acción del Plan Integral Multicampus 3: Proyección social e investigación pertinentes y 5: Personas que transforman sociedad. Con relación a la línea de acción del PIM número 3, la propuesta establece resultados con impacto a la sociedad, en el marco de la Salud Ambiental, prioridad de la Facultad de Ingeniería Ambiental y que no sólo se enmarcan en el ámbito de ciudad capital sino que analizan el comportamiento de información de otro territorio del país, lo que involucra el campo de acción de la USTA Colombia; además se presenta un enfoque de análisis de política internacional. A través de la generación de conocimiento y formación de estudiantes es posible el fomento a los procesos de investigación en pregrado, contribuyendo no sólo con la producción científica de la USTA, sino a mejorar la clasificación del grupo de investigación INAM –USTA. De otro lado, la generación de los productos propuestos recogen la formación curricular establecida en el programa de ingeniería ambiental y se armonizan a competencias de análisis de grandes volúmenes de datos, en línea con las tecnologías de la información y la comunicación, dando paso a una formación humana integral, con el análisis de visión de comunidad en el marco de los objetivos de desarrollo sostenible y cuyo análisis de resultados orientarán un análisis de contenido y responsabilidad social. Todo esto en el marco de la línea de acción “Personas que transforman sociedad”. Además este proyecto soporta el desarrollo de acciones que incluyen el éxito académico, graduación oportuna y fortalecimiento de competencias ciudadana con impacto social. Este trabajo tiene un enfoque innovador en el país respondiendo a los retos del desarrollo sostenible y de la política de Salud Ambiental.
Articulación con funciones sustantivas y el sector social y productivo	
Los productos y resultados de la ejecución de la propuesta de investigación, establecen una orientación de transformación social al analizar la información divulgada a la comunidad acerca de un factor de deterioro ambiental como los es la contaminación del aire. En este estudio se observa a la comunidad integrada por el sector académico, productivo e institucional, quienes hacen parte de la comunidad y a quien se dirige la información comunicada por los medios de prensa; es la comunidad que en el marco de la salud ambiental se ve afectada por unas condiciones de calidad del aire. Desde la función sustantiva investigación se busca generar productos que orienten una transformación social enfocada a soportar el alcance de los Objetivos de Desarrollo Sostenible (ODS), en especial el ODS 11 comunidades y ciudades sostenibles. Por otra parte, a través de la dirección de trabajos de grado con la participación de auxiliares de investigación, la función docencia soporta el que hacer del futuro egresado de ingeniería ambiental y su impacto en la sociedad con una visión holística y enfocada de visión ambiental a la comunidad, se abarca en la generación de conocimiento para el desarrollo de la línea	



de profundización de salud ambiental de la Facultad de Ingeniería Ambiental; además el conocimiento generado se divulgará mediante la participación en evento académico y artículo de investigación.

Grupo de investigación	Línea de investigación en la que se inscribe el proyecto
INAM USTA	Salud Ambiental

Nombre del Investigador principal	Enlace CvLAC	Enlace ORCID	Enlace Google Académico
Nidia Isabel Molina Gómez	http://scienti.colciencias.gov.co:8081/cvlab/viualizador/generarCurriculoCv.do?cod_rh=0001403008	https://orcid.org/0000-0003-4485-262X	https://scholar.google.es/citations?user=Y4UC0goAAAAJ&hl=es
División	Facultad	Programa	Grupo de investigación
Ingenierías	Ingeniería Ambiental	Ingeniería Ambiental	INAM-USTA
Nombre del Co-investigador	Enlace CvLAC	Enlace ORCID	Enlace Google Académico
Claudia Fernanda Navarrete López	http://scienti.colciencias.gov.co:8081/cvlab/viualizador/generarCurriculoCv.do?cod_rh=0000652920	https://orcid.org/0000-0001-5971-2369	https://scholar.google.com/citations?user=4rHlhgoAAAAJ&hl=es
División	Facultad	Programa	Grupo de investigación
Ingenierías	Ingeniería Ambiental	Ingeniería Ambiental	INAM-USTA

Resumen de la propuesta	Palabras clave
Cada año se registra la pérdida de vidas a causa de la contaminación del aire, y sobre el particular las diferentes entidades han desarrollado acciones que implican desde la formulación de la regulación hasta la implementación de los planes de descontaminación y de salud. En Colombia esta cifra corresponde con cerca de 8000 muertes anuales que se relacionan con enfermedades del sistema respiratorio y cardiovascular, influenciadas por la exposición a contaminantes atmosféricos. No obstante, la información relacionada no se encuentra al alcance de toda la comunidad; a pesar que la información pueda concentrarse en informes técnicos o a nivel institucional,	ODS, artículos de prensa, material particulado, salud ambiental, estimación de patrones



su alcance a toda la comunidad es limitado. Se encuentra además que en estaciones de monitoreo de Medellín en jurisdicción del Área Metropolitana del Valle de Aburrá (AMVA) y en Bogotá, en jurisdicción de la Secretaría Distrital de Ambiente (SDA), se han presentado las mayores concentraciones de material particulado en el país, en diferentes periodos de tiempos en los últimos diez años. En este orden de ideas, se observa la necesidad de identificar el grado de divulgación a la comunidad a través de medios masivos de comunicación, como los artículos de prensa y su relación con el comportamiento del contaminante atmosférico material particulado en su fracción PM_{10} y $PM_{2.5}$. Para dicho propósito es necesario aplicar herramientas de minería de texto, análisis de series temporales y de aprendizaje automático que permitan identificar la frecuencia de términos, agrupación temática, la posible relación entre la información publicada y el comportamiento de los contaminantes del aire registrados por las estaciones que monitorean la calidad de aire en jurisdicción del AMVA y de la SDA con mayor incumplimiento a la norma registrado en los últimos diez años. Se espera que este proyecto identifique los aspectos requeridos en materia de divulgación a la comunidad, así como la información más frecuentemente divulgada como soporte a la implementación de medidas que apoyen en la prevención y mitigación de los efectos de exposición al material particulado y contribuyan al logro del ODS 11 en su componente de calidad del aire.

Problema de investigación

La calidad del aire es un indicador de interés comunitario y requiere del desarrollo de acciones de prevención y mitigación como bien necesario para la vida humana. El desarrollo de acciones para su mejoramiento es una tarea que, a pesar de recaer en la dirección de las autoridades ambientales, con la implementación del sector privado e institucional, requiere de la comunidad como apoyo en la implementación de las decisiones de política. La contaminación del aire por material particulado (PM_{10} y $PM_{2.5}$) es un tema de interés público al afectar las condiciones de salud de la población; ha cobrado cada año cerca de 7 millones de vidas humanas, dada la exposición a partículas finas contenidas en el aire contaminado. Tan sólo en 2016 cobró 4,2 millones de muertes por exposición a aire ambiente y 3,8 millones de muertes por exposición interna (OMS, 2018).

Según la Organización Mundial de la Salud (WHO, 2014), la contaminación del aire es un factor de riesgo crítico, se estima que causa el 43% de muertes por neumopatía obstructiva crónica, el 29% por cáncer de pulmón, 25% por accidentes cerebrovasculares y 24% de muertes por cardiopatías. Dada su importancia, se ha incluido en el Objetivo de Desarrollo Sostenible (ODS) 11, denominado ciudades y comunidades sostenibles, fijando como meta a nivel internacional con alcance local y temporal a 2030, la reducción del impacto ambiental negativo per cápita de las ciudades. Sobre el particular, Colombia ha



suscrito diferentes regulaciones, políticas y acuerdos; se incluye el documento CONPES 3918 de 2018 que fija la hoja de ruta en la implementación de los ODS a 2030 y una de las herramientas para constatar su cumplimiento será la información que se reporte a través de las redes de monitoreo de calidad del aire en jurisdicción de las entidades territoriales. Como línea base se encuentra que, de acuerdo con el informe generado por el Instituto de Hidrología Meteorología y Estudios Ambientales (IDEAM), en el periodo 2011 a 2015, las estaciones de monitoreo ubicadas en Centro y Éxito en la jurisdicción del Área Metropolitana del Valle de Aburrá (AMVA) y Carvajal-Sevillana y Kennedy en Bogotá (jurisdicción SDA) fueron las estaciones que presentaron la mayor concentración de material particulado en el país. Resultados que se contrastan con el comportamiento registrado en 2017. Se destacan además, las alertas y estado de emergencia establecidas en virtud del comportamiento de calidad del aire en Medellín y Bogotá en el primer trimestre de 2019.

Es del caso mencionar que en la implementación y alcance de las metas de los ODS, interactúan diversos grupos de interés. De una parte se encuentran, los medios de comunicación, quienes transmiten información a la comunidad acerca de hechos de interés, que orientan tanto posiciones, acciones y conceptos por parte de la sociedad lectora o receptora de la información comunicada. Por otra parte, se encuentran las entidades de gobierno, que para el ámbito de estudio, corresponde a las autoridades ambientales, quienes se encargan entre otros aspectos, de monitorear los factores de deterioro ambiental; como se ha mencionado cuentan con sistemas de monitoreo para el registro del comportamiento de los componentes ambientales y la generación de la información requerida como soporte para la toma de decisiones. Adicionalmente, se encuentra un tercer actor, la comunidad, sobre quien recaen tanto las acciones desarrolladas por las autoridades ambientales, así como la información que divulgan los medios de comunicación. Se incluyen como otros grupos de interés a los gremios, la industria y la academia, quienes en últimas también hacen parte de la comunidad.

Una comunidad apropiadamente informada es capaz de tomar decisiones reduciendo su vulnerabilidad frente a diversos acontecimientos. De allí la importancia de los medios de comunicación frente a hechos que incluyen a la dimensión ambiental. A pesar que existen diversos estudios de carácter científico que han estudiado el comportamiento de la calidad del aire y su influencia en la salud de la población, no siempre estos documentos se encuentran a su alcance. Se destaca que además de los medios de comunicación de carácter científico que analiza comportamientos y estados de la contaminación del aire, existen otras fuentes de información de nivel especializado, medios que reportan información ambiental y los medios de comunicación masivos, que divulgan información a la comunidad acerca de los principales hechos del país. El alcance de la información por parte de la comunidad se va limitando en virtud de su interés, formación y acceso a dichos medios. En este orden de ideas podría decirse que los medios de comunicación de carácter masivo presentan un mayor alcance en materia de divulgación de los hechos de interés para la comunidad.

Considerando que la contaminación del aire por material particulado es un factor de deterioro no sólo ambiental sino de la salud de la población, que atañe directamente a la comunidad, el presente proyecto pretende resolver las siguiente pregunta de investigación: ¿Cuál es el grado de relación entre el comportamiento de la calidad del aire y los patrones de información presentados por los medios de comunicación en prensa?

Justificación

SEDE PRINCIPAL BOGOTÁ - PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9.ª n.º 51-11 / contactenos@usantotomas.edu.co
www.usta.edu.co

DIVISIÓN DE EDUCACIÓN ABIERTA Y A DISTANCIA
PBX: (571) 595 00 00 ext. 2044 / Carrera 10.ª n.º 72-50 / admisiones@ustadistancia.edu.co
www.ustadistancia.edu.co



31 En materia de salud ambiental, la calidad del aire es un determinante que puede influir en la presencia o ausencia de enfermedad en la población, se incluyen además de este componente ambiental la biología humana, los sistemas de atención médica y los estilos de vida de la población expuesta a los contaminantes atmosféricos. Las acciones sobre cada uno de los determinantes, pueden orientar una mejora en la calidad de vida y bienestar del individuo expuesto.

La contaminación del aire y la exposición a material particulado es un evento que ha sido estudiado por décadas, así como sus efectos en la salud; estableciendo desde el ámbito internacional y nacional las regulaciones necesarias para el sector privado en materia de emisión de contaminantes y los estándares de calidad del aire a cumplir en el territorio distrital, nacional e internacional. Se suman los planes de descontaminación con las medidas de prevención, mitigación, los sistemas de alerta temprana y los sistemas de monitoreo de la calidad del aire. A nivel nacional, el IDEAM indicó que en virtud del análisis de información generada por los sistemas de vigilancia a la calidad del aire en Colombia para el 2017, algunas de las estaciones de monitoreo ubicadas en Bogotá y el Área Metropolitana del Valle de Aburrá (AMVA) son las que han superado los niveles permisibles de PM₁₀ y PM_{2.5}, presentando los niveles más altos de estos contaminantes. Además, de acuerdo con el (IDEAM, 2016) para el periodo 2011 a 2015 se observaron para PM₁₀ y PM_{2.5} las mayores concentraciones en algunas estaciones de monitoreo correspondientes al AMVA, Bogotá y a la Zona Minera del Departamento del Cesar; en Bogotá se presentaron incluso excedencias a la norma de PM_{2.5}. Este mismo estudio realizó un análisis de tendencia y relación de atención por infección respiratoria aguda y el comportamiento del material particulado, encontrando coincidencias en el comportamiento de ambas variables con algunos casos que se justifican principalmente en la influencia de variables meteorológicas.

La población es un actor clave frente a la implementación de las medidas de descontaminación del aire en una ciudad, y fundamental en la implementación de medidas que inciden sobre los determinantes de la salud ambiental: estilos de vida y ambiente. De allí la importancia de involucrar a la comunidad en la implementación de los planes de salud pública, planes de descontaminación del aire, campañas y acciones que orienten tanto prevención como mitigación de eventos que puedan afectar su salud. No obstante el trabajo de las autoridades, existe aún la necesidad de hacer masiva la comunicación de la situación ambiental y de salud a la comunidad, a tal punto de orientar patrones de comportamiento, así como estilos de vida cada vez más saludables. Dando paso a que la medicina no sea una medicina de tratamiento de la enfermedad, sino una medicina orientada específicamente a promover la salud de la población.

11 A pesar de los diferentes medios de comunicación utilizados por las entidades para divulgar sus planes, medios virtuales a través de sus páginas web, escritos con informes técnicos, verbales con entrevistas y desarrollo de eventos de divulgación; los medios de comunicación escritos en prensa son los más preferidos al momento de aclarar información específica de una situación. Si bien es cierto, la comunicación a través de los medios sociales son uno de los medios más utilizados por la comunidad, al momento de buscar información más profunda, son los medios de prensa escrita los más consultados por la comunidad en sus diferentes grupos etarios. Es necesario establecer tendencias o patrones que brinden información útil y que permitan generar nuevo conocimiento de la problemática de contaminación.



28 Por lo anterior y, considerando que los grupos de interés involucrados en la implementación del ODS 11 no sólo corresponden con las autoridades ambientales, con el sector privado y gremios, sino que la comunidad es en un actor sobre quien recae el desarrollo de las acciones, es indispensable realizar un análisis de minería de texto aplicable a los medios de comunicación escrita de alta difusión, identificando los aspectos centrales de la difusión, términos frecuentes, tendencias en el comportamiento de la información publicada, asociación de términos y análisis de correlación de términos. Esta información permitirá identificar el grado de divulgación a la comunidad en el ámbito nacional, en materia de calidad del aire, además de identificar las ciudades o zonas con mayor relevancia en los medios de comunicación. En contraste, es necesario además realizar un análisis de series temporales de comportamiento de los contaminantes PM₁₀ y PM_{2.5} registrados por las estaciones de monitoreo Carvajal-Sevillana, Kennedy en Bogotá (jurisdicción SDA) y Tránsito Centro y Éxito en Medellín de las jurisdicción del AMVA, al ser las estaciones que de manera histórica han evidenciado excedencias a la norma de calidad del aire y las mayores concentraciones de los contaminantes anteriormente citados. Un análisis comparativo de la información divulgada en medios de comunicación y de comportamiento del material particulado para los últimos 10 años (2009 a 2019) permitirá establecer si existe coherencia entre el comportamiento de la calidad del aire en función del contaminante material particulado y lo informado en los medios de comunicación, la frecuencia de divulgación de la temática y su relación con el comportamiento en las series analizadas. Este análisis establecerá un soporte técnico, en la medida en que es necesario cambiar la estrategia tangencial de inclusión de la comunidad, orientando patrones de comportamiento desde los medios de comunicación. Además como se ha indicado el ambiente es un determinante en la salud de la población y las acciones en sus patrones de comportamiento podrían orientar una influencia no sólo en su salud, sino en las dimensiones económica, social, institucional y ambiental del desarrollo sostenible, que en suma hacen sinergia para el logro de las metas ODS.

Objetivo general

40 39 Aplicar herramientas de minería de texto y de análisis de series temporales al comportamiento de material particulado para determinar el grado de relación entre la información publicada en medios de comunicación y el comportamiento de la calidad del aire en estaciones de monitoreo de calidad del aire del Área Metropolitana del Valle de Aburrá en Medellín y de la Secretaría Distrital de Ambiente en Bogotá.

Objetivos específicos

- 13 1. Identificar el grado de socialización de los problemas de contaminación del aire en el país con los niveles de contaminación, mediante la aplicación de herramientas de minería de texto a artículos de prensa publicados en medios digitales para el periodo 2009 a 2019
2. Realizar un análisis de series temporales para material particulado de la información reportada para estaciones de monitoreo en jurisdicción del AMVA y Bogotá en el periodo 2009-2019
3. Realizar un análisis de correlación de la información publicada en medios de prensa, publicados en medios digitales, con la información registrada por el sistema de monitoreo de calidad del aire en jurisdicción del AMVA y Bogotá para el periodo 2009 - 2019

Estado del arte y marco conceptual

SEDE PRINCIPAL BOGOTÁ - PBX: (571) 587 87 97 Línea gratuita nacional: 01 8000 111 180
Carrera 9.ª n.º 51-11 / contactenos@usantotomas.edu.co

www.usta.edu.co

DIVISIÓN DE EDUCACIÓN ABIERTA Y A DISTANCIA

PBX: (571) 595 00 00 ext. 2044 / Carrera 10.ª n.º 72-50 / admisiones@ustadistancia.edu.co

www.ustadistancia.edu.co





Un análisis de textos corresponde con el procesamiento de documentos no estructurados con el fin de identificar aspectos relevantes en los mismos, y generar información para la toma de decisiones desde el agrupamiento, reconocimiento de patrones, información frecuente, asociación de términos, análisis de sentimientos, entre otros. Esta herramienta se ha analizado con el manejo de grandes volúmenes de datos en el sector financiero, comercial (Guerreiro, Rita, & Trigueiros, 2016), técnica y social, al desarrollar análisis de sentimientos en la comunidad acerca de un tema en especial; por ejemplo (Belkahla Driss, Mellouli, & Trabelsi, 2019), desarrollaron un estudio en el que hicieron un análisis de sentimientos en virtud de las actividades gubernamentales mediante la minería de texto aplicada a medios sociales. La minería de texto ha sido aplicada además en redes sociales, en el ámbito científico para establecer el grado de relación entre colaboraciones científicas, tendencias y el producto de ello (Li, Han, & Lu, 2019). Se incluyen estudios como el análisis de métricas en ahorro energético (García et al., 2018), de información sanitaria de pacientes (Pastor-Duran, 2017), de eventos en salud los relacionados a exposición a sustancias químicas (Larsson et al., 2017). Al respecto de contaminantes atmosféricos se realizó un estudio en el que se analizaron datos de noticias con temáticas específicas de exposición personal a humo de tabaco en China (Q. Liu et al., 2019), se aplicó el método de distancia latente (LDA) para la definición de grupos de términos, concluyendo en el grado de relevancia de las temáticas de humo de tercera mano y las enfermedades respiratorias, restricciones y controles al contaminante. En estos estudios se realizó el análisis de sentimientos, frecuencia de palabras y agrupamiento temático. Otros estudios analizaron la información publicada por los diarios analizados con eventos específicos en materia del recurso agua y otro desarrollado por (Chen, Kan, Li, Liang, & Hong, 2018) incluyó un análisis de comportamiento de la calidad del aire en China. Estos estudios han permitido la identificación de ejes centrales de información, fechas, definición de entidades u organismos relevantes descritos en la información proporcionada a través de los medios masivos de comunicación. A nivel nacional no se encuentra un estudio que involucre el análisis de minería de texto aplicado a artículos de prensa, con el enfoque de los términos relacionados a la calidad del aire y salud ambiental.

Por su parte, el análisis de series temporales consiste en un análisis de tendencias en comportamiento de series de información, que permite conocer su comportamiento en un periodo de tiempo determinado (Grechka & Heigl, 2017). Este análisis se ha utilizado, además, con fines de pronóstico de eventos próximos. En contraste con la aplicación de herramientas de minería de texto a temas específicos como la contaminación del aire, en donde ha sido reducido el desarrollo de productos de carácter científico; la aplicación de series temporales ha tomado aspectos técnicos más específicos, ejemplo de ello son los estudios desarrollados por (Souza et al., 2014) y (Royé, Zarrabeitia, Riancho, & Santurtún, 2019); el primero, hizo uso de las series temporales para el análisis de la influencia de la contaminación del aire sobre hospitalizaciones por infecciones respiratorias, mientras que el estudio desarrollado por (Royé et al., 2019), aplicó las series para el análisis de accidentes cerebro vascular isquémicos y factores ambientales. En Colombia el análisis de series temporales se aplicó al estudio de asociación entre la contaminación del aire y suicidios, para la determinación de la influencia de la cobertura vegetal en la contaminación del aire (Zafra, Ángel, & Torres, 2017), así como al análisis de los efectos de la contaminación del aire en la morbilidad respiratoria y cardiovascular en 4 ciudades de Colombia (Fernández-Niño et al., 2018). Además en Bogotá se realizó un estudio que incluyó el análisis de los contaminantes atmosféricos en el periodo 1997 a 2007, concluyendo en que los contaminantes gaseosos no superan la norma, al contrario del PM (Gaitán, Cancino, & Behrentz, 2007).

Las metodologías aplicadas y métricas de evaluación corresponden con el coeficiente de auto correlación, así como el índice de NASH.



Metodología

Este proyecto comprende la realización de las siguientes actividades:

1. **Recopilación de información.** La recopilación de información tendrá dos enfoques principales: por una parte la información publicada en artículos de prensa en medio digital para el periodo 2009 a 2019 y de otra parte, el registro de información del contaminante material particulado en sus fracciones de PM_{10} y $PM_{2.5}$ en el mismo periodo de tiempo, monitoreado por las autoridades ambientales del AMVA y de Bogotá. Teniendo claros los enfoques se realizará, para el caso de la información de los medios digitales, la recopilación de información relacionada con textos informativos publicados en artículos de prensa en medios digitales tipo web, en el ámbito nacional. La recopilación de información de artículos de prensa tendrá como claves de búsqueda central: <contaminación del aire, calidad del aire, PM_{10} , $PM_{2.5}$ y enfermedad respiratoria>. Adicionalmente, se realizará la recuperación de información de comportamiento diario de los contaminantes atmosféricos PM_{10} y $PM_{2.5}$ en la ventana de tiempo 2009 a 2019, mediante solicitud a las autoridades AMVA y Secretaría Distrital de Ambiente, además de variables meteorológicas precipitación, velocidad y dirección de viento, temperatura que puedan involucrarse en el análisis de comportamiento de las series temporales. De manera transversal a la recopilación de la información de artículos de prensa y comportamiento de contaminantes, se hará recopilación de las principales normas y políticas en el ámbito nacional y de Bogotá relacionadas con el componente atmosférico para el periodo (2009-2019)
2. **Pre-procesamiento de la información.** Atendiendo a los tipos de información recopilada: a) información divulgada por medios de comunicación en la web e b) información cuantitativa de contaminantes criterio PM_{10} y $PM_{2.5}$, se desarrollarán las siguientes actividades:
 - a. Para el caso de la información de medios de comunicación se aplicarán técnicas de pre-procesamiento de texto que incluyen la limpieza del corpus documental, tokenización y lematización, se realizará identificación de los términos más frecuentes y su relación con el periodo de análisis.
 - b. Para el caso de la información de contaminantes criterio se realizará la depuración y organización de la base de datos de acuerdo con la periodicidad en cada caso y por contaminante.
3. **Análisis de comportamiento de la información.** Una vez organizadas las bases de datos se desarrollará la estadística descriptiva correspondiente al análisis de frecuencias, máximos, mínimos, desviación estándar, entre otras medidas.
 - a. Para el caso de la información de los artículos de prensa, se realizará un análisis de textos no estructurados, se identificarán etiquetas de información con base en la aplicación de herramientas de *machine learning* no supervisadas y se realizará análisis semántico de los textos. Se identificarán en cada caso: la fuente de la cual se extrajo cada artículo de prensa, la extensión de los artículos de prensa para identificar la importancia de la temática en materia de calidad del aire para los medios analizados y en la generalidad de la información recopilada. Identificación de zonas del país en los que se presentó la mayor información y las fuentes de contaminación relacionadas y asociación de términos. Identificación de actores, organizaciones más frecuentes relacionadas con la información presentada. Identificación del tono de la información presentada, en virtud de las dimensiones del desarrollo sostenible: dimensión ambiental, si el tema solo se atañe a tema ambiental, o en la dimensión social, económica o institucional, siendo estas dimensiones las establecidas bajo el marco de naciones unidas. La agrupación de términos, según etiquetas, se realizará



definiendo un número limitado de tópicos de agrupación, que se establecerá considerando las palabras frecuentes, asociación de términos y extensión de la base de información no relacional recopilada.

- b. Para el caso de la información de los contaminantes PM_{10} y $PM_{2.5}$ se realizará el análisis de las series temporales; se identificarán las tendencias o anomalías y se realizará análisis de correlación con políticas y acciones que se hayan implementado en el periodo de análisis, determinando si posiblemente, según los hitos de política o regulación, se presentó algún grado de influencia en el comportamiento de las series temporales.
 - c. Para ambos casos se identificarán las coincidencias temporales acerca del comportamiento de la información, mediante la unión de las bases de información analizadas. Este procedimiento requiere de la aplicación de herramientas de tratamiento de datos pasando la información textual a datos numéricos, de tal manera que esta transformación permita llevar los datos a series temporales agregadas conforme la agregación establecida para los contaminantes atmosféricos.
4. **Evaluación de los resultados de agrupamiento de información textual y de implementación de las series temporales.** Una vez realizado el tratamiento de estadística descriptiva e inferencial y aplicación de herramientas de minería de texto, para el caso de la información de prensa y, de series temporales para el caso de los registros de calidad del aire, se aplicarán métricas de evaluación consistentes en el coeficiente cofenético para el caso de la información textual y estadísticos como el coeficiente de NASH, para el caso de las series temporales. La evaluación de imputación de datos en el caso de las series temporales estará dada por métricas como la exactitud en el pronóstico de los datos de imputación utilizados y el error cuadrático medio.

Para el desarrollo del proyecto se hará uso del software Python, y sus librerías para análisis de texto como NTKL (*natural language tool kit*) y de aprendizaje de máquina (*machine learning for language toolkit*).

Resultados esperados

La apuesta inter y trans-disciplinar contenida en la presente propuesta permite que a través de los resultados esperados se acojan los campos de acción: sociedad y ambiente, con la vinculación del análisis en la información que los medios de prensa escrita divulgan a la comunidad y la información técnica contenida en bases de información públicas de autoridad ambiental. Los resultados alcanzables son de resorte en materia ambiental y social. En este orden de ideas, en desarrollo del trabajo de investigación, se pretende alcanzar los siguientes resultados y productos:

1. Identificación de los principales patrones y retos en materia de divulgación del estado de la calidad del aire, acciones, medidas a través de los medios de comunicación escrita en prensa. Las metas establecidas en los ODs requieren de acciones sinérgicas desde el ciudadano de a pie, por lo que se espera identificar las tendencias en comunicación acerca de la calidad del aire, los efectos sobre ésta, así como los puntos de inflexión en los que sea posible divulgar a la comunidad para orientar dichos patrones en el desarrollo de acciones desde este grupo de interés y desde este ámbito hacia los diferentes eslabones y grupos implicados en la implementación de los ODS y beneficiarios de los mismos.
2. Identificación del comportamiento de la calidad del aire en dos zonas del país bajo un histórico de diez años, estableciendo su comportamiento y su relación con los principales hitos en el marco político y regulatorio nacional y local. El estudio de comportamiento de la calidad del aire y su análisis de



asociación a hitos regulatorios y/o políticos permite un conocimiento de su influencia en el comportamiento de la calidad del aire para el periodo de estudio. A partir de este análisis será posible identificar el comportamiento regulatorio y, su relación al comportamiento de contaminantes atmosféricos.

3. Identificación de los puntos en los cuales se encuentran las mayores o menores posibilidades de divulgación en prensa a la comunidad acerca de las condiciones de calidad del aire.

De acuerdo con los resultados establecidos se generarán los siguientes productos:

CARACTERÍSTICAS DEL MODELO NACIONAL DE COLCIENCIAS	PRODUCTO	POSIBLE REVISTA/EVENTO
Productos resultado de actividades de generación de nuevo conocimiento	Artículo de Investigación en revista indexada Q4 o Q3	Revista de Salud Ambiental, Journal of Environmental Health ó International Journal of Sustainable Development and Planning
Productos resultado de actividades de desarrollo tecnológico e innovación	Informe técnico	Informe técnico que dará cuenta de los aspectos técnicos desarrollados en el proyecto.
Productos resultado de actividades de apropiación social del conocimiento	Circulación de conocimiento especializado a través de la participación en evento científico nacional (1) e internacional (1) con la divulgación de conocimiento especializado.	XIII Congreso Internacional de Medio Ambiente (evento nacional)
Productos de actividades relacionadas con la formación de recurso humano para la CTEI	Dirección de dos (2) trabajos de grado de pregrado	Dos (2) trabajos de grado de pregrado: Pendientes por definir.

Cronograma

ACTIVIDADES	RESPONSABLE	FECHA		17 FEBRERO	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOSTO	SEPTIEMBRE	OCTUBRE	NOVIEMBRE
		Inicio	Fin	1 2 3 4	5 6 7 8 9	10 11 12	13 14 15 16	17 18 19 20	21 22 23 24	25 26 27 28 29	30 31	32 33 34 35	36 37 38 39 40 41
Recopilación de información de artículos en prensa	Isabel Molina, Claudia Navarrete	03/02	31/03										
Recopilación, solicitud y recepción de información de	Isabel Molina, Claudia Navarrete	03/02	30/04										

[illegible]

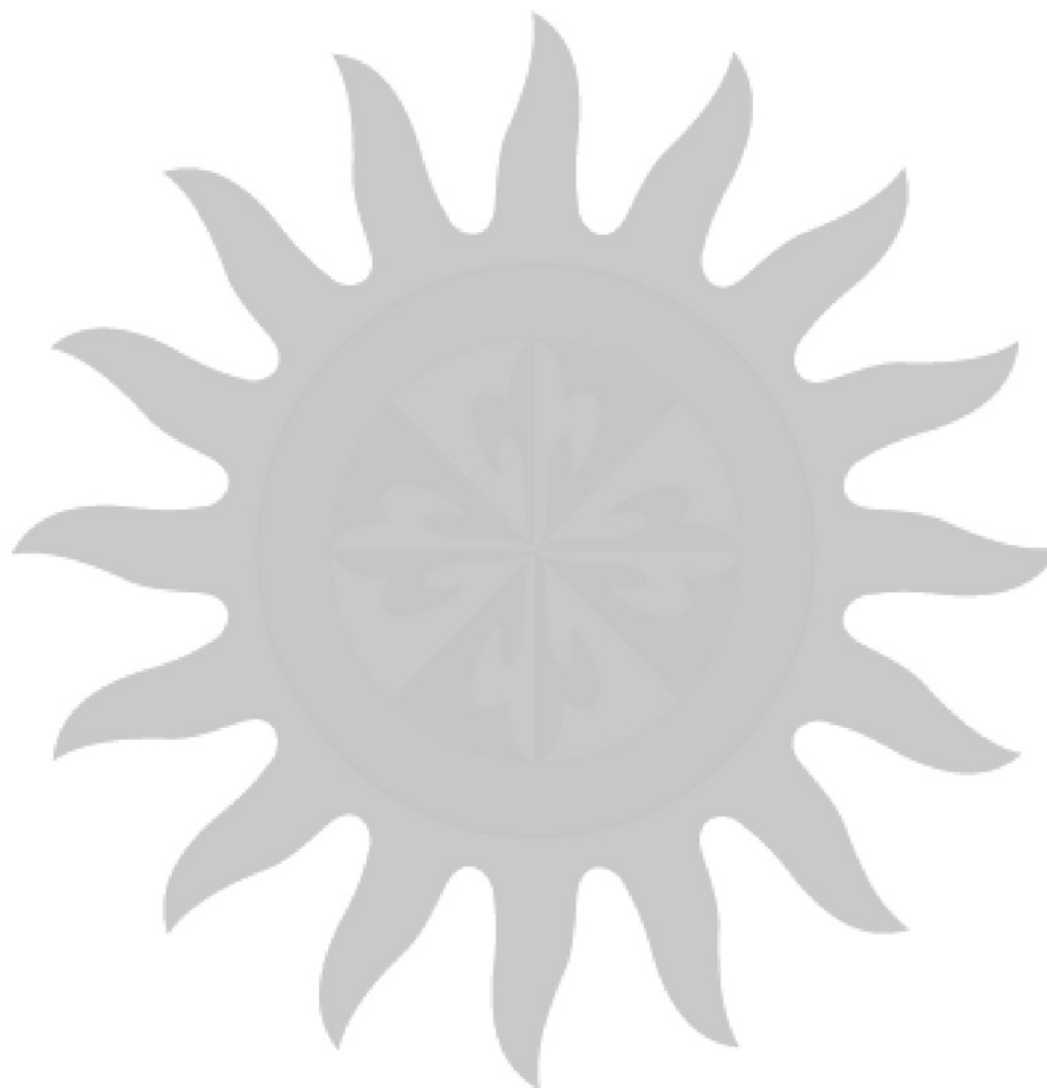


UNIVERSIDAD SANTO TOMÁS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

VIGILADA MINEDUCACIÓN - SNIES: 1704



Página 13 de 14



Nit. 860.012.357-6

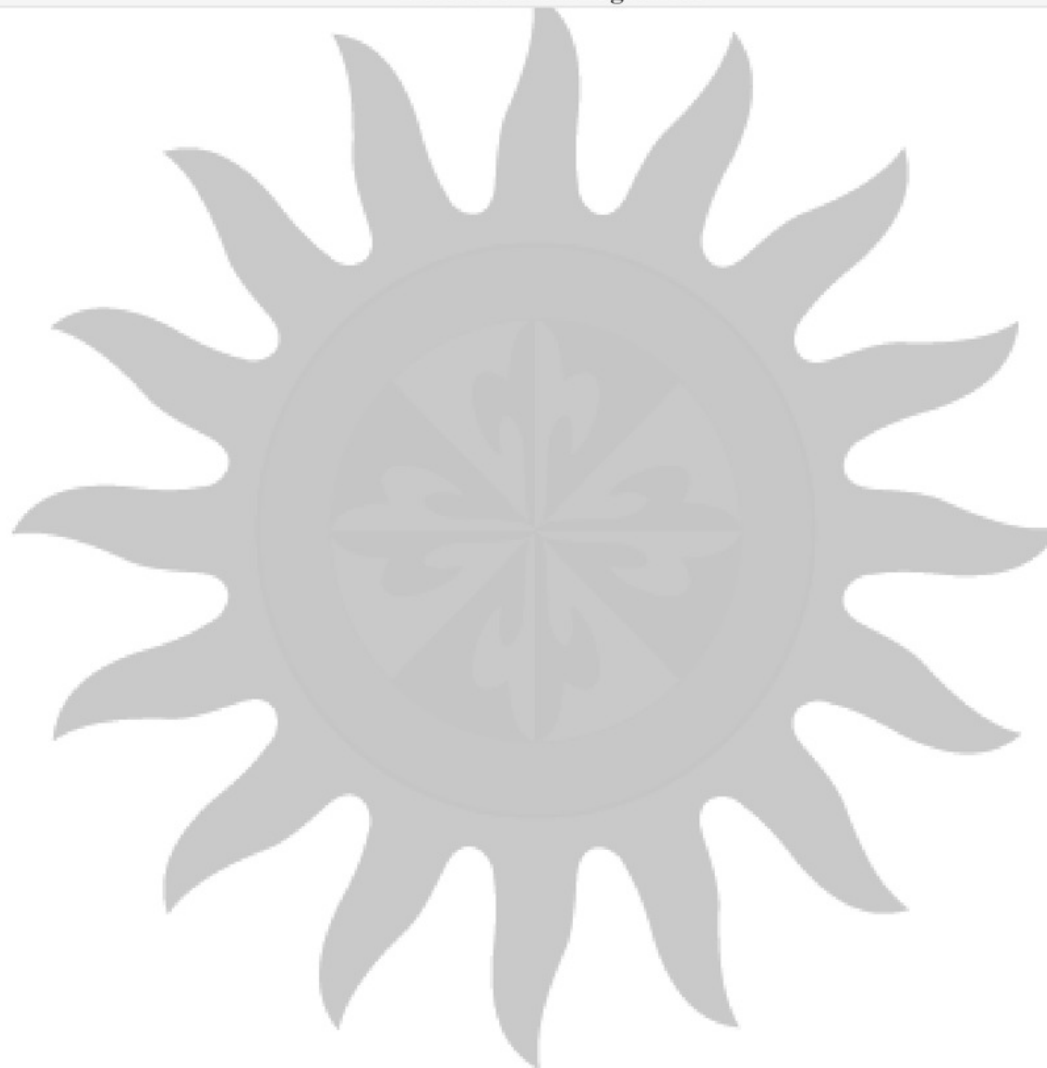
SEDE PRINCIPAL BOGOTÁ - PBX: (571) 587.87.97 Línea gratuita nacional: 01.8000.111.180
Carrera 9.ª n.º 51-11 / contactenos@usantotomas.edu.co
www.usta.edu.co

DIVISIÓN DE EDUCACIÓN ABIERTA Y A DISTANCIA
PBX: (571) 595.00.00 ext. 2044 / Carrera 10.ª n.º 72-50 / admisiones@ustadistancia.edu.co
www.ustadistancia.edu.co





Referencia bibliográficas



Nit. 860.012.357-6

SEDE PRINCIPAL BOGOTÁ - PBX: (571) 587.87.97 Línea gratuita nacional: 01.8000.111.180
Carrera 9.ª n.º 51-11 / contactenos@usantotomas.edu.co
www.usta.edu.co

DIVISIÓN DE EDUCACIÓN ABIERTA Y A DISTANCIA
PBX: (571) 595.00.00 ext. 2044 / Carrera 10.ª n.º 72-50 / admisiones@ustadistancia.edu.co
www.ustadistancia.edu.co



Prueba para certificado turniting a entregar a convocatoria Fodein

ORIGINALITY REPORT

15%	10%	2%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universidad Santo Tomas Student Paper	3%
2	sat.semarnat.gob.mx Internet Source	1%
3	www.postershop-espana.com Internet Source	1%
4	Submitted to INACAP Student Paper	1%
5	www.sedespa.gob.mx Internet Source	1%
6	www.usco.edu.co Internet Source	1%

7	Submitted to Universidad de Medellin Student Paper	<1 %
8	Submitted to Universidad Internacional de la Rioja Student Paper	<1 %
9	www.paot.org.mx Internet Source	<1 %
10	Submitted to Fundación Universitaria del Area Andina Student Paper	<1 %
11	www.cede.es Internet Source	<1 %
12	Submitted to Universidad Sergio Arboleda Student Paper	<1 %
13	www.ecoamerica.cl Internet Source	<1 %
14	www.asocars.org.co Internet Source	<1 %
15	Submitted to Fundación Universitaria CEIPA Student Paper	<1 %

16	www.inval.gov.co Internet Source	<1 %
17	www.lascondes.cl Internet Source	<1 %
18	scielo.org Internet Source	<1 %
19	www.proteccioncivilsonora.gob.mx Internet Source	<1 %
20	www.cepal.org Internet Source	<1 %
21	Submitted to Esumer Institucion Universitaria Student Paper	<1 %
22	www.lapaz.bo Internet Source	<1 %
23	www.elojodigital.com Internet Source	<1 %
24	Submitted to University of Edinburgh Student Paper	

<1 %

25

Submitted to Escuela De Ingenieria De Antiquia - Columbia

Student Paper

<1 %

26

cia.corantioquia.gov.co

Internet Source

<1 %

27

www.intergraph.com.ve

Internet Source

<1 %

28

www.hivos.nl

Internet Source

<1 %

29

imagourbis.unq.edu.ar

Internet Source

<1 %

30

pifrecv.utralca.cl

Internet Source

<1 %

31

Submitted to International School of the Hague

Student Paper

<1 %

32

grid2.cr.usgs.gov

Internet Source

<1 %

33	www.proventionconsortium.org Internet Source	<1 %
34	www.publicnow.com Internet Source	<1 %
35	agenda21ens.cicese.mx Internet Source	<1 %
36	Submitted to Universidad de Jaén Student Paper	<1 %
37	Submitted to Systems Link Student Paper	<1 %
38	argentina.emc.com Internet Source	<1 %
39	B. J. Pusey. "Discharge variability and the development of predictive models relating stream fish assemblage structure to habitat in northeastern Australia", Ecology Of Freshwater Fish, 6/2000 Publication	<1 %
40	Submitted to Universitat Politècnica de València	

<1 %

41 www.gestiopolis.com
Internet Source

<1 %

42 Submitted to UNIACC
Student Paper

<1 %

43 www.eluniversal.com
Internet Source

<1 %

44 www.televen.com
Internet Source

<1 %

45 www.rlc.fao.org
Internet Source

<1 %

46 www.ceh.cl
Internet Source

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off

Prueba para certificado turniting a entregar a convocatoria Fodein

GRADEMARK REPORT

FINAL GRADE

/100

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

