Tecnológico de Antioquia
Institución Universitaria
VIGILADA MINEDUCACIÓN

SELLO EDITORIAL TdeA

GIISTA
Grupo de Investigación en Ingeniería de Software del Tecnológico de Antioquia

Investigación e Innovación

en Ingeniería de Software

Volumen 4

Investigación e Innovación
en Ingeniería de Software
Volumen 4

Tecnológico de Antioquia
Institución Universitaria
VIGILADA MINEDUCACIÓN

SELLO EDITORIAL TdeA

Investigación e Innovación en Ingeniería de Software / volumen 4 Fabio Alberto Vargas Agudelo, Darío Enrique Soto Duran, Juan Camilo Giraldo Mejía, compiladores. – Medellín: Sello Editorial TdeA, 2020.

Correción de estilo, diseño, diagramación
Divegráficas S.A.S.
divegraficas@gmail.com

Derechos reservados del Tecnológico de Antioquia – Institución Universitaria

"Los artículos publicados en este libro incorporan contenidos derivados de procesos de investigación y reflexión académicas no representan necesariamente los criterios institucionales del Tecnológico de Antioquia. Los contenidos son responsabilidad exclusiva de sus autores. Obra protegida por el derecho de autor. Queda estrictamente prohibida su reproducción, comunicación, divulgación, copia, distribución, comercialización, transformación, puesta a disposición o transferencia en cualquier forma y por cualquier medio, sin la autorización previa, expresa y por escrito de su titular. El incumplimiento de la mencionada restricción podrá dar lugar a las acciones civiles y penales correspondientes. © 2020 Todos los derechos de autor reservados."

Hecho en Colombia

Investigación e Innovación en Ingeniería de Software
Tecnológico de Antioquia – Institución Universitaria

© Sello Editorial Tecnológico de Antioquia Institución Universitaria
Tecnológico de Antioquia
Calle 78b # 72A - 220
(+57 4) 444 3700
Medellín – Colombia

Dr. Leonardo García Botero
Rector

Dr. Fabio Alberto Vargas Agudelo
Director de Investigación – Coordinación del Sello Editorial

Dr. Dario Enrique Soto Duran
Decano de la Facultad de Ingeniería

Dr. Juan Camilo Giraldo Mejía
Lider de Grupo GIISTA

Dr. Leonardo García Botero
Rector

Dr. Fabio Alberto Vargas Agudelo
Director de Investigación – Coordinación del Sello Editorial

Dr. Dario Enrique Soto Duran
Decano de la Facultad de Ingeniería

Dr. Juan Camilo Giraldo Mejía
Lider de Grupo GIISTA

Autores:
Fabio Alberto Vargas Agudelo
Carlos Mario Zapata Jaramillo
Luis Eduardo Peláez Valencia
Miguel A. Cohuo Ávila
Alonso Toro Lazo
Juan Luis Arias Vargas
Daniel Eduardo Rodríguez Franco
Yulieth Gómez Osorno
Diego Alejandro Pérez Correa
Mauricio Escobar Tobón
Juan Carlos Valdés Quintero
Laura María Marín Arroyave
María Isabel Duque Cano
Claudia Elena Durango Vanegas
Valeria Carvajal Hurtado
Alicia Martínez Rebollar
Javier Ortiz Hernández
León Alberne Torres Restrepo
Hugo Estrada Esquivel
María Isabel Marín Morales
Jorge Molina
Rodrigo Alberto Méndez
Brahayan Estiven Rojas Golondrino
Jenifer Cuatindioy
Laura Martínez Sandoval
Christian Gustavo Arias Iragorri
Fredy Alonso Vidal Alegría
Víctor Daniel Gil vera
Catalina Quintero López
Sebastián Gómez Jaramillo
Juan David Aránzazu
Jorge Mario Cadavid
Darío Enrique Soto Duran
Adriana Xiomara Reyes Gamboa
Juan Camilo Giraldo Mejía
Julián Grisales Z.
Mauricio Amariles Camacho

ISBN impreso: 978-958-8628-64-6
ISBN digital: 978-958-8628-65-3

Investigación e Innovación en Ingeniería de Software / volumen 4 Fabio Alberto Vargas Agudelo, Darío Enrique Soto, Juan Camilo Giraldo, compiladores. – Medellín: Sello Editorial TdeA, 2020.

CONTENIDO

7 **PREFACIO**

9 **Capítulo 1**
INTELIGENCIA DE NEGOCIOS APLICADA A LOS DELITOS SEXUALES EN COLOMBIA 2010-2019

31 **Capítulo 2**
DISEÑO DE UN SISTEMA DE INFORMACIÓN QUE IMPLEMENTA EL MODELO AUTOMATIZADO PARA ASEGURAR LA CALIDAD DE LOS REQUERIMIENTOS EN PROYECTOS DE SOFTWARE

59 **Capítulo 3**
FORMALIZATION OF DOMAIN PROBLEMS DESCRIBED IN THE CAUSE-AND-EFFECT DIAGRAM IN THE CONTEXT OF THE SOFTWARE DEVELOPMENT PROCESS.

75 **Capítulo 4**
CARACTERIZACIÓN GEOMORFOLÓGICA DE LA CUENCA DEL RÍO NEGRO UTILIZANDO SISTEMAS DE INFORMACIÓN GEOGRÁFICA (SIG)

89 **Capítulo 5**
COVID-19: UN ANÁLISIS BASADO EN MINERÍA DE TEXTO

103 **Capítulo 6**
PROPUESTA DE UN MÉTODO DE EVALUACIÓN MULTICRITERIO COMO HERRAMIENTA PARA DETERMINAR ZONAS ÓPTIMAS DE UBICACIÓN DE RELLENOS SANITARIOS UTILIZANDO SISTEMAS DE INFORMACIÓN GEOGRÁFICA

141 **Capítulo 7**
DETECCIÓN AUTOMÁTICA DE EVENTOS GENERADORES DE ACCIDENTES DE TRÁNSITO

163 Capítulo 8
UN MÉTODO DE ETIQUETADO SEMIAUTOMÁTICO DE CORPUS LINGÜÍSTICOS

181 Capítulo 9
MODELO DE TRANSFORMACIÓN DIGITAL PARA LA CREACIÓN DE DESTINOS TURÍSTICOS INTELIGENTES CASO DE ESTUDIO: TÁMESIS

199 Capítulo 10
GESTIÓN DEL CONOCIMIENTO APLICADA AL PROCESO DE AUTOEVALUACIÓN

215 Capítulo 11
MONITOREO DE INDICADORES DE RENDIMIENTO EN CADENAS *BLOCKCHAIN ETHEREUM*

PREFACIO

Para las organizaciones es muy importante mantener interconectados los flujos de procesos, con el propósito de garantizar competitividad y ganar posicionamiento y calidad. Esto se alcanza con la incorporación de tecnologías que permiten la automatización a través de la gestión de grandes volúmenes de datos (*Big Data*), la inteligencia artificial articulada con electrónica y mecánica para soportar procedimientos industriales, la protección de los datos por medio de la seguridad de la información, y la ubicación y geoposicionamiento de puntos estratégicos con la utilización de sistemas de información geográfica. Todas estas tecnologías se presentan en la industria 4.0 con la integración de diferentes disciplinas, estrategias y métodos buscando fortalecer los mercados y, por ende, el crecimiento económico global.

El libro *Investigación e Innovación en Ingeniería de Software,* en su cuarto volumen, compila once artículos que relacionan las disciplinas y tópicos de la industria 4.0, con referencia a la especificación de requisitos de software, la inteligencia de negocios, la gestión del conocimiento, los sistemas de información geográfica, la seguridad de la información, la lingüística computacional y la inteligencia artificial.

Todos los temas que se encuentran en este volumen, y se presentan como capítulos, contienen un contexto general de la investigación desarrollada, antecedentes, revisión de literatura y resultados, contribuyen al desarrollo de la investigación y la innovación alrededor de Ingeniería de software, lo cual se ve reflejado en productos de nuevo conocimiento y en el fortalecimiento de los programas académicos de pregrado y posgrado, así como en los grupos, líneas y semilleros de investigación.

Muchos de los autores de los capítulos de este volumen son profesores adscritos a la Facultad de Ingeniería del Tecnológico de Antioquia, quienes también han participado en ediciones anteriores, gracias a los resultados obtenidos en su ejercicio investigativo. Otros autores pertenecen a instituciones de educación superior aliadas que contribuyen en este volumen con los resultados de sus proyectos de investigación.

Tenemos la certeza de que los resultados publicados en este nuevo volumen se convertirán en un soporte para seguir contribuyendo a la investigación e innovación en el área de ingeniería de software en las instituciones de educación superior del país y en el sector productivo.

Juan Camilo Giraldo Mejía
Fabio Alberto Vargas Agudelo
Darío Enrique Soto Durán
Compiladores

Capítulo I
Inteligencia de Negocios Aplicada a los Delitos Sexuales en Colombia 2010-2019

Rodrigo Alberto Méndez - rmendez@unimayor.edu.co
Brahayan Estiven Rojas Golondrino - estivenrojas@unimayor.edu.co
Jenifer Cuatindioy - jcuatindioy@unimayor.edu.co
Laura Martínez Sandoval - Lmartinez@unimayor.edu.co
Christian Gustavo Arias Iragorri - Charias98@unimayor.edu.co
Fredy Alonso Vidal Alegría - fvidal@unimayor.edu.co
Colegio Mayor del Cauca

I. INTRODUCCIÓN

La Organización Mundial de la Salud (OMS) define la violencia sexual "como todo acto sexual, la tentativa de consumarlo, los comentarios, insinuaciones sexuales no deseadas, o las acciones para comercializar o utilizar de cualquier otro modo la sexualidad de una persona, independientemente de la relación de ésta con la víctima y en cualquier ámbito, incluidos el hogar y el trabajo" [1].

Estudios internacionales y nacionales sobre el tema muestran que la violencia sexual puede tener lugar a lo largo de todo el ciclo vital, desde la infancia hasta la vejez, incluye a mujeres y hombres, ambos como víctimas y agresores. Aunque afecta a ambos sexos, son las niñas, las mujeres adolescentes y las mujeres adultas la población más prevalente en condición de víctima, esta violencia puede existir entre miembros de una misma familia, personas de confianza, entre conocidos y extraños. Ante este panorama, el delito sexual se ha convertido en una preocupación constante del Gobierno Nacional y de los ciudadanos. El presente artículo está centrado en el análisis de los registros de delitos sexuales en Colombia entre los años 2010-2019, según la base de datos de la Policía Nacional. Esta investigación nace con la finalidad de exponer las características comunes de los delitos sexuales en Colombia,

con un total de 176.616 registros a los cuales se les hace un análisis individual por variables, así como un análisis en grupo de variables para comprobar las relaciones que pueden existir. Por lo tanto, se pretende realizar el análisis de cada una de las variables con el fin de llegar a una conclusión con el total de los registros obtenidos. Se espera que los resultados de esta investigación sirvan como referencia para posteriores investigaciones y contribuir con la mejora de un bien común.

II. MATERIALES Y MÉTODOS

En 2019, el Gobierno colombiano definió su nueva estrategia contra la violencia y el abuso sexual infantil en su plan de desarrollo, para llevar su oferta institucional del ICBF y otras instituciones a las zonas más críticas con base en el análisis de las cifras de años anteriores. Según las cifras del Instituto Nacional de Medicina Legal y Ciencias Forenses, el 86 % del presunto delito sexual fue sobre menores de edad y el 73,8 %, sobre niñas y adolescentes. Casos que, en su mayoría, ocurren en casa. Asimismo, existen otras fuentes de información con resultados similares.

La Organización Panamericana de la Salud realizó un trabajo concienzudo que recolectó resultados de diversos estudios realizados en Latinoamérica donde se evidencia que las mujeres corren mayor riesgo de sufrir violencia sexual perpetrada por sus parejas, y encontró una prevalencia a lo largo de la vida de relaciones sexuales forzadas en una pareja, que varían entre el 5 y el 47 % [2]. Dicho trabajo recolectó distintas fuentes de información como [3], [4] y tras un arduo trabajo de homogenización de datos y catalogación, generó un conjunto de datos agregados que posteriormente fueron presentados de forma gráfica tipo *dashboard*.

En la Figura 1 se muestra el porcentaje de mujeres, entre 15 y 49 años, que le contaron a alguien o buscaron ayuda institucional a raíz de violencia por parte de la pareja entre los años 2004-2012.

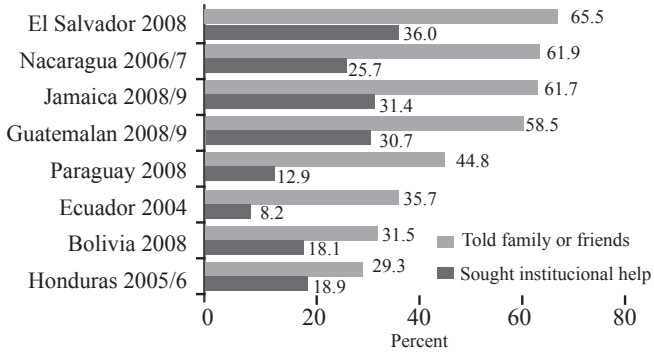


Fig. 1. Porcentaje de mujeres que le contaron a alguien o buscaron ayuda institucional (2004-2012)

Para este tipo de estudios es supremamente importante mostrar la información de una forma clara y condensada, que se puede generar a través de procesos de inteligencia de negocios, los cuales pueden ser aplicados a dominios de negocio diferentes al empresarial, como lo son el ámbito social y de investigación aplicada [5], [6].

El marco de trabajo utilizado para Inteligencia de Negocios aplicada a los delitos sexuales en Colombia utiliza la información pública de los datos de la Policía Nacional 2010-2019, que obedece al plan de transparencia y acceso a la información de la policía y sus informes institucionales de gestión y evaluación [7].

La metodología aplicada en el desarrollo de este proyecto es la creada por Ralph Kimball, basada en el Ciclo de Vida Dimensional del Negocio (*Business Dimensional Lifecycle*). Este proceso se divide en diferentes tareas que están agrupadas por fases, las cuales se especifican en la siguiente relación para el diseño del *datamart* [8]:

A. Planificación

- Precisar el alcance.
- Identificar las tareas.
- Programar las tareas.
- Planificar el uso de los recursos.
- Asignar la carga de trabajo a los recursos.
- Elaborar un documento final que representa un plan del proyecto [9].

En esta etapa de Planificación hay que precisar el alcance con la información obtenida, puesto que dicha base de datos contiene los delitos sexuales en Colombia, se planifica qué se va a hacer con dicha información, cuál va a ser el objeto de estudio, se determina qué se requiere para hacer todo el proceso y se organiza un

documento con toda la proyección de lo que se va a realizar, el cual se denomina anteproyecto.

B. Análisis de Requerimientos

- Identificar las necesidades del negocio.
- Identificación de fuentes de información.
- Tener claro los conceptos y terminología del dominio del proyecto.

1) Modelo relacional

En esta etapa de Análisis de Requerimientos según la base de datos adquirida, se detecta que la información abarca todo el territorio colombiano, desde el año 2010 hasta el primer semestre del año 2019.

En la Figura 2 se muestra el Modelo Relacional de la Base de Datos (*Staging*).

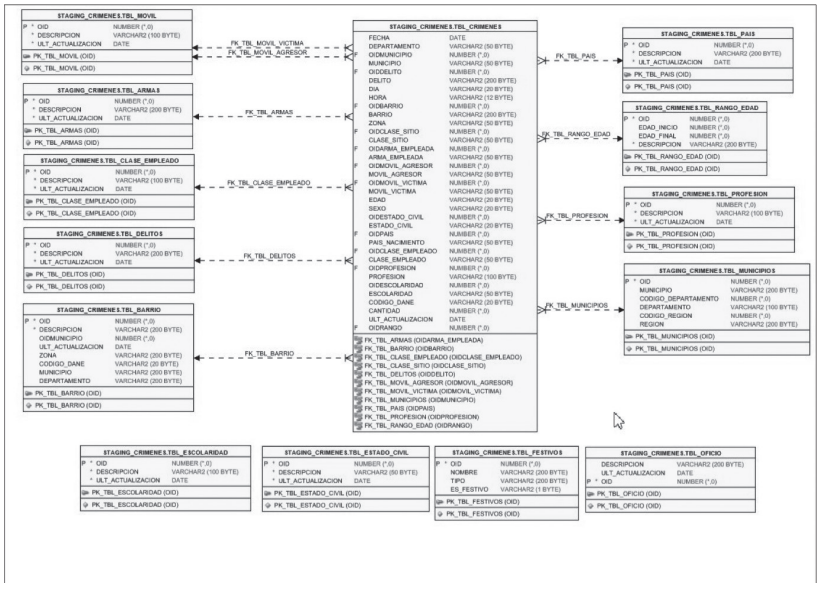


Fig. 2. Modelo Relacional Área de *Staging*

2) Modelado dimensional

Es esta etapa del modelado dimensional se determina el nivel de granularidad en horas y días, donde se toma como apoyo la elaboración de la matriz para determinar las dimensiones.

Según la información analizada, se toma la decisión de que las siguientes dimensiones son necesarias para poder obtener los resultados esperados de este trabajo:

Tiempo: El cual se define como día (también puede estar clasificado si es festivo o no, o si corresponde a alguna festividad especial, p. ej., Día de la Madre, Halloween, etc.) y hora, es sumamente importante para determinar en qué hora del día son más propensos los delitos.

Demografía: En esta dimensión se realizó un producto cartesiano de las tablas del modelo relacional de sexo, escolaridad, estado civil y edad (se organizó por rangos).

Oficio: En esta dimensión está almacenada la información de la víctima, que corresponde a si es estudiante, ama de casa, etc.

País Nacimiento: La base de datos corresponde solo al territorio colombiano, pero se encontró que muchas de las víctimas de estos delitos son extranjeros y que dicho delito se presentó dentro del país con personas que no son colombianos, esto también es vital, puesto que también se puede determinar qué tan susceptibles son estas personas en Colombia.

Delito: En esta dimensión están almacenados todos los delitos con su respectivo código y descripción.

Arma: Se crea con la finalidad de que guarde la información sobre el tipo de arma empleada en el delito, ejemplo: arma de fuego, cuchillo, cinta adhesiva, etc.

Ubicación: Corresponde al sitio en el país donde se cometió el delito, el cual registra la siguiente información: país, departamento, ciudad, zona, barrio, código DANE, región.

Tipo Sitio: Corresponde a la información donde se llevó a cabo de forma más concisa el delito, ejemplo: lote baldío, puente, casa abandonada, etc.

Transporte: Corresponde al medio de transporte en el cual se movilizaban el agresor y la víctima en el momento de la ejecución del delito, puede ser transporte público, moto, carro, etc.

En la Tabla 1 se muestra la Matriz de procesos/dimensiones resultantes.

Tabla 1. Matriz de procesos/dimensiones

Matriz de procesos/ dimensiones	TIEMPO	DEMOGRAFÍA	OFICIO	PAÍS NACIMIENTO	DELITO	ARMA	UBICACIÓN	TIPO SITIO	TRANSPORTE
Clase de arma	x	x				x			x
Número de víctimas	x		x				x		
Rango de edad	x	x					x		
Días	x							x	
Delitos	x				x	x	x		
Víctimas Extranjeras	x			x			x		

Posteriormente, se realiza el diseño del nuevo modelo de la bodega de datos que se muestra en la Figura 3: Modelo Relacional *Datamart* de Crímenes resultante.

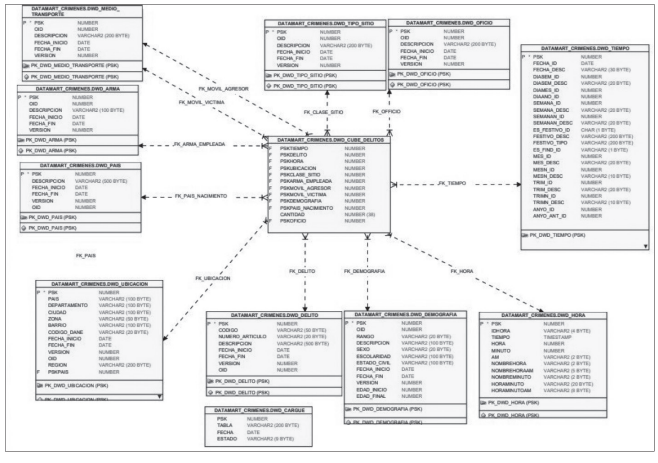


Fig. 3. Modelo Relacional Datamart de Crímenes

3) Diseño Físico

A continuación, se presenta la respectiva Tabla de Hechos (Tabla 2).

Tabla 2. Tabla de hechos

Nombre	Descripción	Regla de Agregación predeterminada
Cantidad	Numero de cada una de las dimensiones	Count

4) Diseño del sistema de Extracción, Transformación y Carga (ETL)

Para realizar la extracción de la información para la respectiva bodega se realiza *Job,* que consta de tres procesos:

El primer proceso corresponde a la Precarga, en él se identifica cuál es la dimensión que se va a cargar y se ordena ir al *log* a buscar cuál fue la última fecha en la que se cargó información para validar si la información que se está a punto de cargar es nueva o no.

El segundo proceso corresponde a la Carga, en el cual se debe subir la nueva información a la dimensión que corresponde.

El tercer proceso corresponde al de Postcarga, en el que, una vez almacenada la nueva información en la dimensión correspondiente, se procede en la tabla del *log* a hacer un registro con el nombre de la dimensión y la fecha del proceso del cargue actual.

En la Figura 4 se muestra el *Job* de Proceso de Cargue de la información.

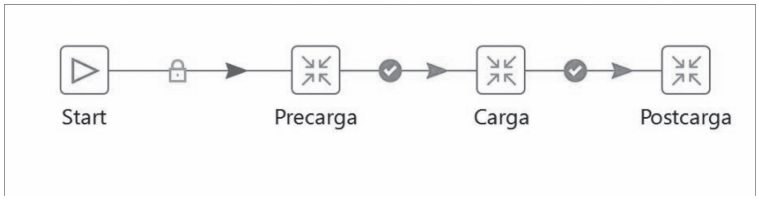


Fig. 4. Job de Proceso de Cargue de la Información

Mediante el desarrollo de esta etapa se garantiza al usuario acceso simple, sencillo y dinámico a la bodega por medio de la herramienta software para la generación de informes, que le permitan ver los perfiles conductuales de los agresores de delitos sexuales.

A continuación, se visualizan todos los procesos de cargue de las dimensiones: DIMENSIÓN ARMA, MEDIO DE TRANSPORTE, PAÍS, UBICACIÓN, DELITO, DEMOGRAFÍA, OFICIO, TIPO SITIO.

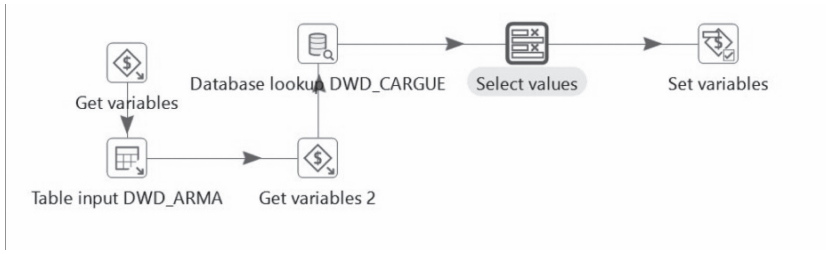


Fig. 5. Precarga de la Dimensión Arma

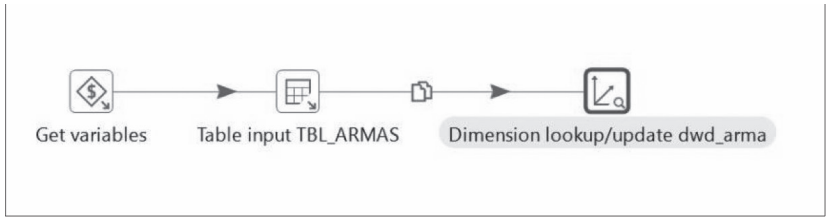


Fig. 6. Carga de la Dimensión Arma

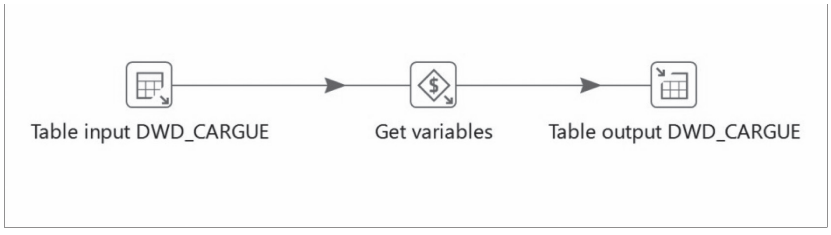


Fig. 7. Carga de la Dimensión Arma

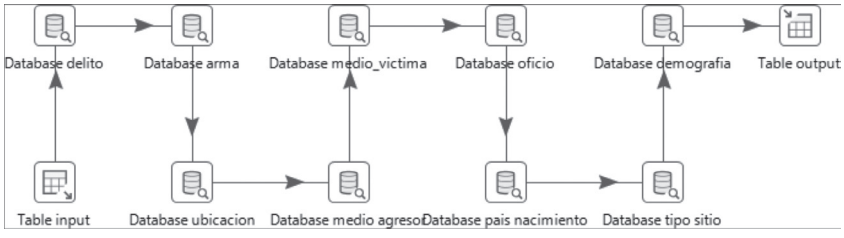


Fig. 8. Diseño del sistema de Extracción, Transformación y Carga (ETL)

III. RESULTADOS

Con los datos obtenidos después de todo el proceso antes mencionado en la metodología, se obtiene como resultado la bodega de datos con la información correspondiente a los delitos sexuales desde el 2010 hasta el primer semestre del año 2019, y se generan reportes de información de interés para la comunidad en general. El trabajo realizado es un primer acercamiento para generar un laboratorio de delitos sexuales en el departamento del Cauca. Cabe resaltar que, a continuación, se visualizan reportes con el ánimo de crear conciencia y prevención, puesto que todos estamos expuestos a ser víctimas de un delito o agresión.

En la Figura 9 se puede apreciar que la hora en la que los criminales cometen más delitos sexuales es la medianoche. Se puede observar que la medianoche es cuando se cometen más delitos sexuales (Figuras 9 y 10), pero en especial los dos delitos que más se cometen son actos sexuales y acceso carnal abusivo con menor de 14 años, es claro que los y las menores de 14 años son los más expuestos a este tipo de abusos según la información resultante.

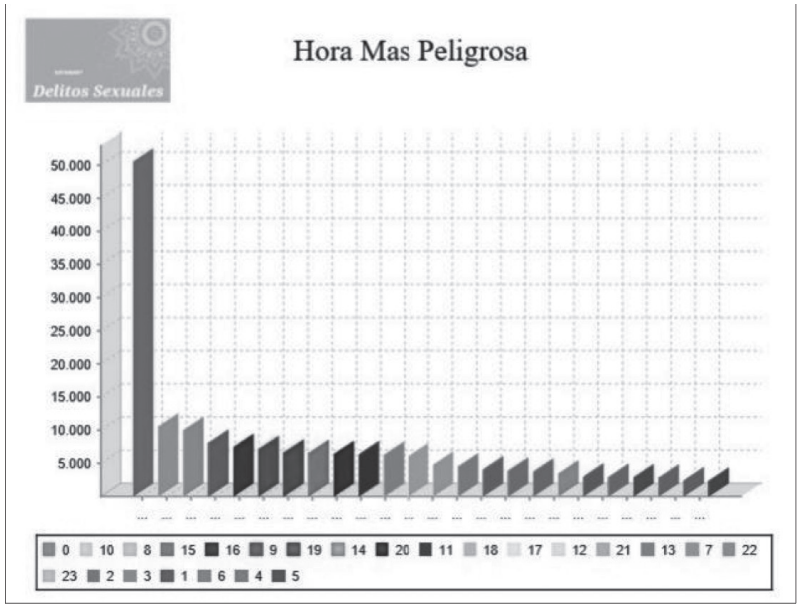


Fig. 9. Reporte Hora más peligrosa por Delito

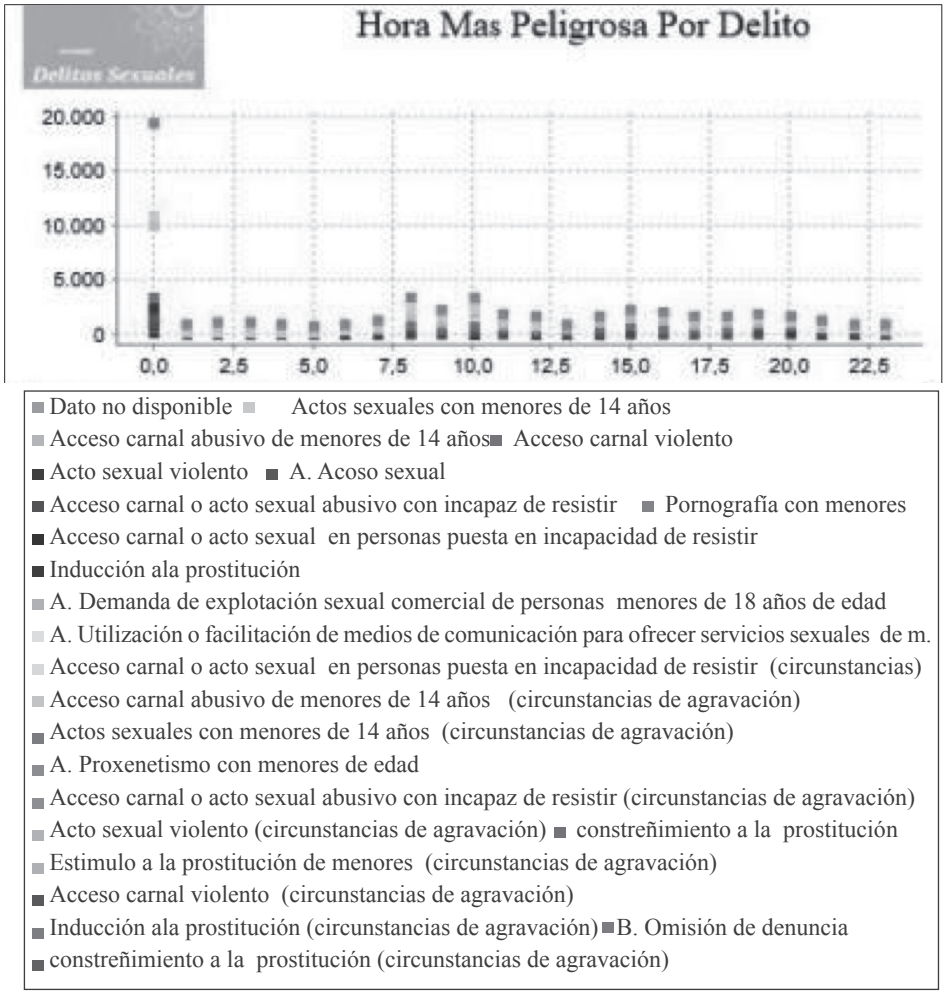


Fig. 10. Reporte Hora más peligrosa por Delito-Edad

En la Figura 11 se observa y afirma nuevamente que el rango de edad más afectado por los diferentes abusos y delitos está entre los 6 y 14 años, lo que reitera que los y las menores de 14 años son la mayoría de víctimas de estos abusos en Colombia, lo cual alerta a la comunidad en general a tener prevención y precaución con este tipo de población infantil.

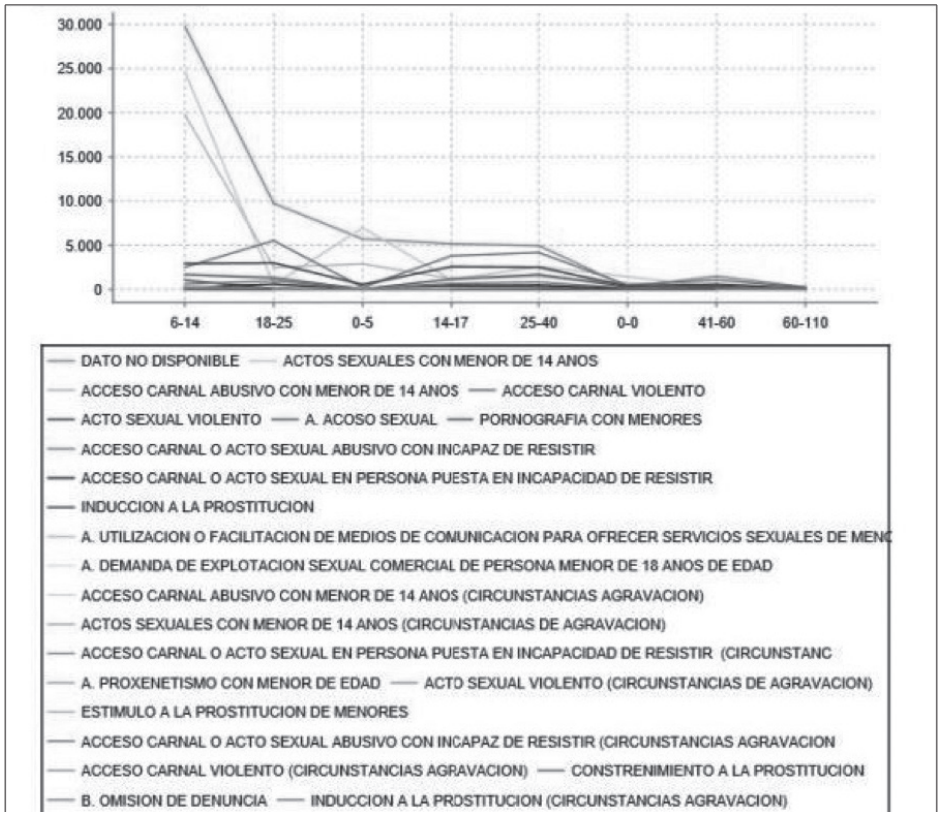


Fig. 11. Reporte de delitos sexuales por rango de edad

Después de analizar los datos correspondientes a 10 años, se observa claramente que las víctimas de delitos sexuales no se determinan por el tipo de sexo, hay víctimas tanto femeninas como masculinas, un total de 148.662 víctimas mujeres (85,69 % aproximadamente), un total de 23.569 víctimas hombres (13,58 % aproximadamente), y el resto corresponde a 383 víctimas sin clasificación de género (0,22 % aproximadamente) (ver Figura 12).

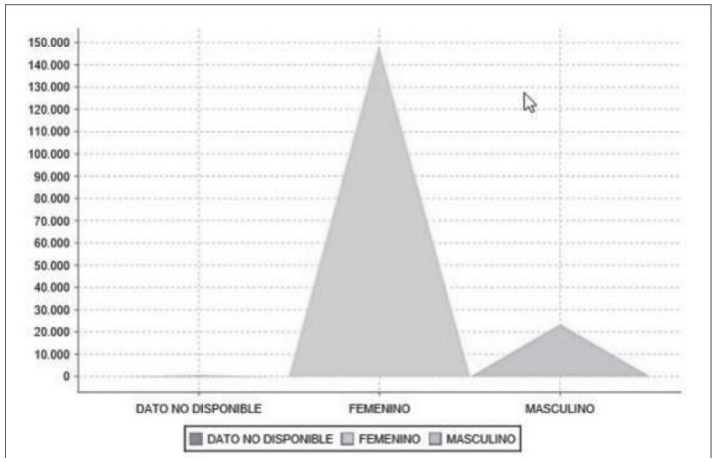


Fig. 12. Reporte de delitos sexuales por sexo biológico

Se puede valorar que la fecha en la que se presentan más delitos es el 1.° de enero con un porcentaje bastante alto con respecto a los demás días del año. Esto indica que luego de una festividad, como el año nuevo, los agresores aprovechan para atacar a sus víctimas (ver Figura 13).

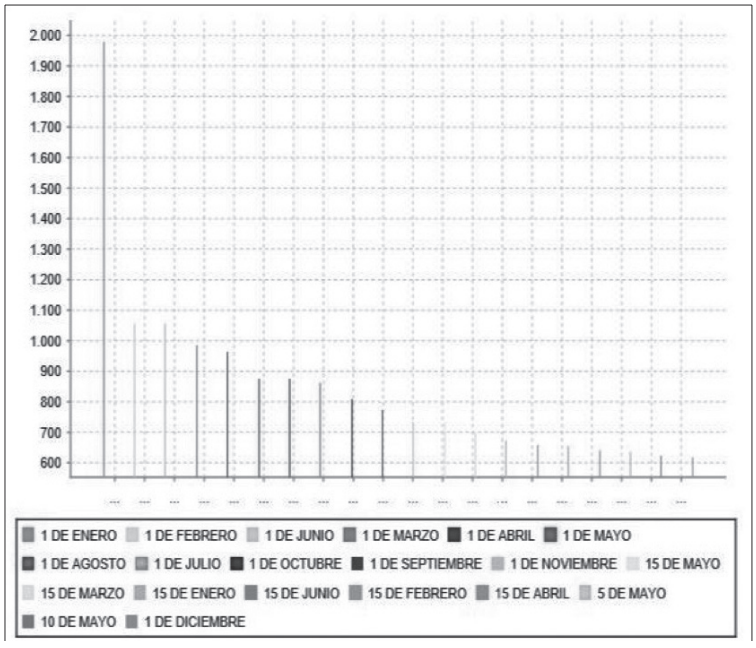


Fig. 13. Reporte de los 20 días más peligrosos del año

El reporte de delitos sexuales por municipios, por cada mil habitantes, se puede evaluar en la Figura 14. En el municipio de San Joaquín, en el departamento de Santander, se presenta el índice de los delitos más alto a nivel nacional, por ejemplo: de cada mil personas, 66 son víctimas del delito de acto sexual con menor de 14 años, asimismo, de cada mil personas 36 son víctimas de acceso carnal abusivo con menor de 14 años y, en tercer lugar, se puede observar que de cada mil habitantes del municipio 19 son víctimas de acto sexual violento. Es lamentable que de los dos primeros delitos que muestra este reporte las víctimas sean menores de 14 años.

Delitos Sexuales Por Cada Mil Habitantes Por Municipio

Departamento	Municipio	Delito	Cantidad por cada mil hab
SANTANDER	SAN JOAQUIN	ACTOS SEXUALES CON MENOR DE 14 ANOS	66,4882
SANTANDER	SAN JOAQUIN	ACCESO CARNAL ABUSIVO CON MENOR DE 14 ANOS	36,5908
SANTANDER	SAN JOAQUIN	ACTO SEXUAL VIOLENTO	19,1879
SANTANDER	SAN JOAQUIN	A. ACOSO SEXUAL	16,9567
SANTANDER	SAN JOAQUIN	ACCESO CARNAL VIOLENTO	15,618
BOLIVAR	ZAMBRANO	ACCESO CARNAL ABUSIVO CON MENOR DE 14 ANOS	14,5623
BOYACA	SANTANA	ACTOS SEXUALES CON MENOR DE 14 ANOS	13,6153
BOYACA	OICATA	ACTOS SEXUALES CON MENOR DE 14 ANOS	13,1159
ANTIOQUIA	GUARNE	ACCESO CARNAL ABUSIVO CON MENOR DE 14 ANOS	12,5905
HUILA	GUADALUPE	ACCESO CARNAL ABUSIVO CON MENOR DE 14 ANOS	11,6825
CHOCO	CERTEGUI	ACCESO CARNAL ABUSIVO CON MENOR DE 14 ANOS	10,1154
HUILA	GUADALUPE	ACTOS SEXUALES CON MENOR DE 14 ANOS	9,8938
CHOCO	RIO QUITO	ACCESO CARNAL VIOLENTO	8,8635
BOLIVAR	ZAMBRANO	ACTOS SEXUALES CON MENOR DE 14 ANOS	8,67
BOYACA	BETEITIVA	ACCESO CARNAL VIOLENTO	8,6602
MAGDALENA	SANTA BARBARA DE PINTO	ACCESO CARNAL ABUSIVO CON MENOR DE 14 ANOS	8,3215

Fig. 14. Reporte de los delitos sexuales por municipio por cada mil habitantes

En la Figura 15, correspondiente al reporte de delitos sexuales por departamento, y por cada mil habitantes, se pueden apreciar los departamentos en los cuales se ha presentado el índice más alto por violencia, en los tres primeros lugares se encuentran: Guaviare, Santander y Risaralda.

Delitos Sexuales

DELITOS POR DEPARTAMENTO POR CADA MIL HABITANTES

DEPARTAMENTO	DELITO	CANTIDAD	POBLACION	CANTIDADX1000 HAB
GUAVIARE	ACCESO CARNAL ABUSIVO CON MENOR DE 14 ANOS	195	82.767	2
SANTANDER	ACTOS SEXUALES CON MENOR DE 14 ANOS	3.313	2.184.837	2
RISARALDA	ACTOS SEXUALES CON MENOR DE 14 ANOS	1.740	943.401	2
AMAZONAS	ACTOS SEXUALES CON MENOR DE 14 ANOS	189	76.589	2
AMAZONAS	ACCESO CARNAL ABUSIVO CON MENOR DE 14 ANOS	119	76.589	2
CAQUETA	ACCESO CARNAL ABUSIVO CON MENOR DE 14 ANOS	637	401.849	2
QUINDIO	ACTOS SEXUALES CON MENOR DE 14 ANOS	833	539.904	2
HUILA	ACTOS SEXUALES CON MENOR DE 14 ANOS	1.746	1.100.386	2
CALDAS	ACCESO CARNAL ABUSIVO CON MENOR DE 14 ANOS	589	998.255	1
CAQUETA	ACTOS SEXUALES CON MENOR DE 14 ANOS	562	401.849	1
AMAZONAS	ACTO SEXUAL VIOLENTO	71	76.589	1
MAGDALENA	ACTOS SEXUALES CON MENOR DE 14 ANOS	797	1.341.746	1
ANTIOQUIA	ACCESO CARNAL ABUSIVO CON MENOR DE 14 ANOS	3.251	6.407.102	1

Fig. 15. Reporte de los delitos sexuales por departamento por cada mil habitantes

En la Figura 16 se muestra el reporte de delitos sexuales en los que se emplean armas: en el 45 % no se usan armas en la agresión o delito, en el 29 % no hay dato disponible o no se puede establecer si en realidad se usó o no un arma para cometer el delito, en el 16 % se empleó algún arma contundente, por ejemplo cualquier objeto que pueda ser usado como arma (botella, martillo, florero, etc.), y de ahí en adelante, con porcentajes menores, está el uso de arma de fuego, corto-punzantes, licor adulterado, cintas, escopolamina, entre otros. Es claro tener en cuenta que, según este reporte, para cometer el delito el agresor no necesita un arma.

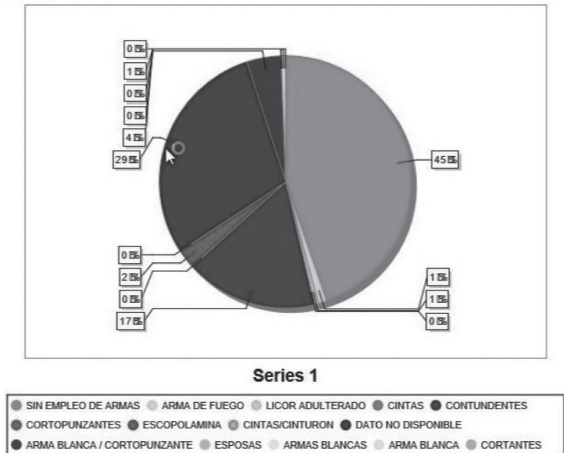


Fig. 16. Reporte de porcentaje de delitos sexuales por arma

En la Figura 17 se hace referencia al reporte del *ranking* de los tres delitos más relevantes por año y se puede apreciar que desde el 2010 hasta el primer semestre del año 2019 han persistido los mismos delitos: actos sexuales con menor de 14 años, acceso carnal violento y acceso carnal abusivo con menor de 14 años, solo para el año 2011 no se tiene información sobre cuáles fueron los delitos más cometidos, pero hay agresiones. Es importante ratificar que las agresiones contra menores de 14 años son las que más se presentan en el país a lo largo de estos 10 años.

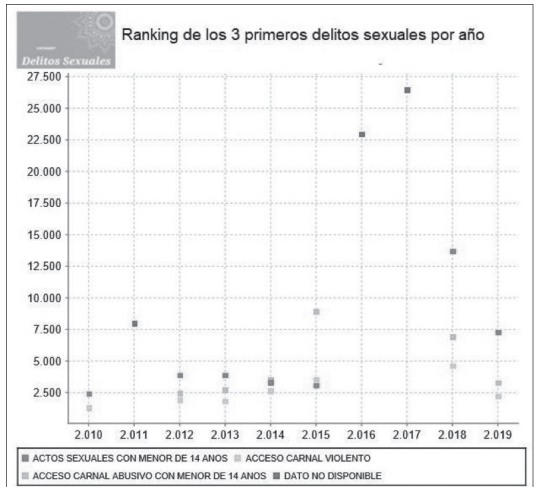
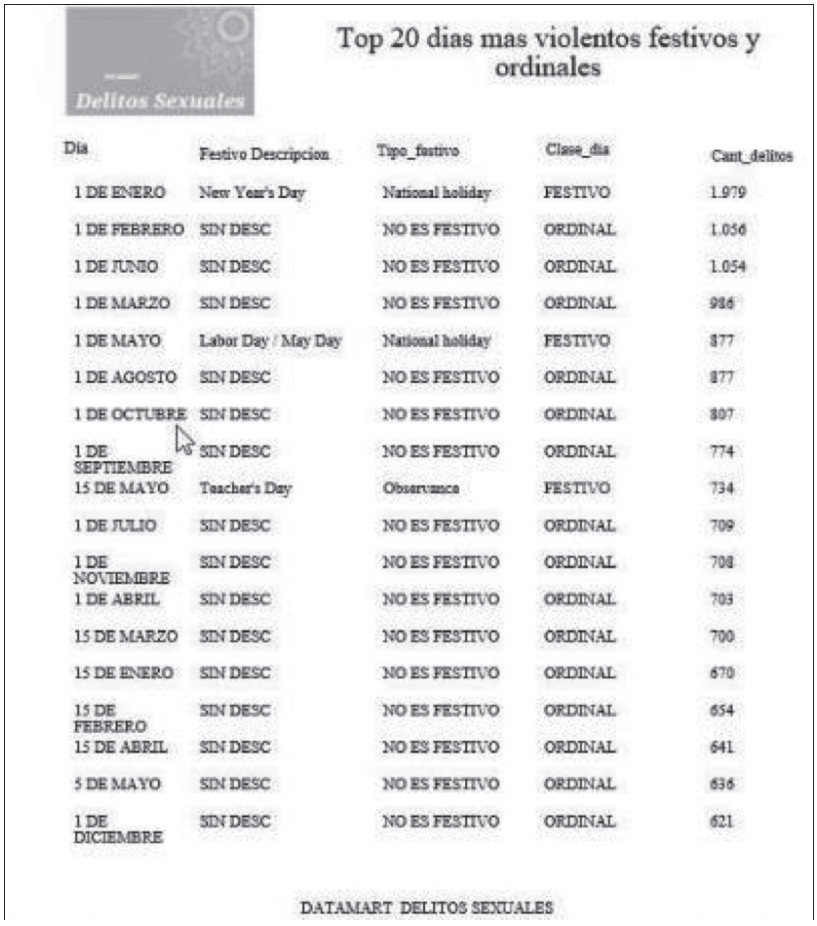


Fig. 17. Reporte del *ranking* de los primeros tres delitos sexuales por año

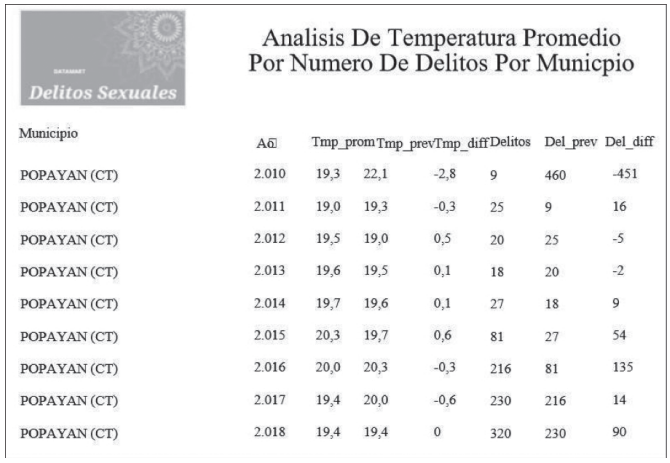
En la Figura 18 se presenta el reporte de los 20 días festivos y ordinales más violentos, se puede apreciar que el 1.° de enero, es decir, año nuevo, es el festivo con más índice de delitos sexuales, le siguen el 1.° de febrero, 1.° de marzo y 1.° junio como días ordinarios y continúa con el 1.° de mayo, que corresponde a la festividad del día del trabajo. Es importante ver la tendencia, según la cual hay una marcada línea de delitos el primer día del mes o a mitad de mes, sea festivo o no.



Dia	Festivo Descripcion	Tipo_festivo	Clase_dia	Cant_delitos
1 DE ENERO	New Year's Day	National holiday	FESTIVO	1.979
1 DE FEBRERO	SIN DESC	NO ES FESTIVO	ORDINAL	1.056
1 DE JUNIO	SIN DESC	NO ES FESTIVO	ORDINAL	1.054
1 DE MARZO	SIN DESC	NO ES FESTIVO	ORDINAL	986
1 DE MAYO	Labor Day / May Day	National holiday	FESTIVO	877
1 DE AGOSTO	SIN DESC	NO ES FESTIVO	ORDINAL	877
1 DE OCTUBRE	SIN DESC	NO ES FESTIVO	ORDINAL	807
1 DE SEPTIEMBRE	SIN DESC	NO ES FESTIVO	ORDINAL	774
15 DE MAYO	Teacher's Day	Observance	FESTIVO	734
1 DE JULIO	SIN DESC	NO ES FESTIVO	ORDINAL	709
1 DE NOVIEMBRE	SIN DESC	NO ES FESTIVO	ORDINAL	708
1 DE ABRIL	SIN DESC	NO ES FESTIVO	ORDINAL	703
15 DE MARZO	SIN DESC	NO ES FESTIVO	ORDINAL	700
15 DE ENERO	SIN DESC	NO ES FESTIVO	ORDINAL	670
15 DE FEBRERO	SIN DESC	NO ES FESTIVO	ORDINAL	654
15 DE ABRIL	SIN DESC	NO ES FESTIVO	ORDINAL	641
5 DE MAYO	SIN DESC	NO ES FESTIVO	ORDINAL	636
1 DE DICIEMBRE	SIN DESC	NO ES FESTIVO	ORDINAL	621

Fig. 18. Reporte Top 20 días festivos y ordinales más violentos

En la Figura 19 se aprecia el reporte de análisis de temperatura promedio en Popayán, y se puede evidenciar que la temperatura no define el aumento o la disminución de los delitos sexuales.



Delitos Sexuales

Analisis De Temperatura Promedio
Por Numero De Delitos Por Municpio

Municipio	Año	Tmp_prom	Tmp_prev	Tmp_diff	Delitos	Del_prev	Del_diff
POPAYAN (CT)	2.010	19,3	22,1	-2,8	9	460	-451
POPAYAN (CT)	2.011	19,0	19,3	-0,3	25	9	16
POPAYAN (CT)	2.012	19,5	19,0	0,5	20	25	-5
POPAYAN (CT)	2.013	19,6	19,5	0,1	18	20	-2
POPAYAN (CT)	2.014	19,7	19,6	0,1	27	18	9
POPAYAN (CT)	2.015	20,3	19,7	0,6	81	27	54
POPAYAN (CT)	2.016	20,0	20,3	-0,3	216	81	135
POPAYAN (CT)	2.017	19,4	20,0	-0,6	230	216	14
POPAYAN (CT)	2.018	19,4	19,4	0	320	230	90

Fig. 19. Reporte de análisis de temperatura promedio por número de delitos por municipio

En la Figura 20, entre los medios de transporte más utilizados por los abusadores, se puede determinar que en su mayoría el agresor no usa ningún medio de transporte para perpetrar sus delitos, esto permite inferir que el agresor puede estar en cualquier sitio, ya que no requiere de un medio de transporte para cometerlos.

Delitos Sexuales

RANKING DE LOS CINCO MEDIOS DE TRANSPORTE MAS UTILIZADOS POR LOS ABUSADORES SEXUALES

AÑO	PUESTO	MEDIO DE TRANSPORTE AGRESOR	CANTIDAD
2.010	1	A PIE	5.308
2.010	2	VEHICULO	71
2.010	3	CONDUCTOR MOTOCICLETA	70
2.010	4	CONDUCTOR TAXI	34
2.010	5	BICICLETA	19
2.011	1	A PIE	6.480
2.011	2	CONDUCTOR VEHICULO	56
2.011	3	CONDUCTOR MOTOCICLETA	53
2.011	4	CONDUCTOR TAXI	31
2.011	5	BICICLETA	13
2.012	1	A PIE	9.072
2.012	2	CONDUCTOR MOTOCICLETA	129
2.012	3	VEHICULO	69
2.012	4	CONDUCTOR TAXI	36
2.012	5	BICICLETA	12
2.013	1	A PIE	10.295

DATAMART DELITOS SEXUALES

Fig. 20. Reporte del *ranking* de los cinco medios de transporte más utilizados por los abusadores

En la Figura 21 se puede concluir que, indiscutiblemente, como se pudo observar en reportes anteriores, la hora en la que se presenta un gran número de delitos es la medianoche y el sitio preferido de los abusadores se presenta principalmente en vías públicas, le siguen casas de habitación, fincas, entre otros. Es importante mencionar que dentro de las viviendas y en zonas escolares y estatales también se presentaron delitos.

Tipo de sitio y hora preferido por el abusador

Delitos Sexuales

Hora	Tipo Stio	Cantidad
0	VIAS PUBLICAS	32.213
0	CASAS DE HABITACION	8.925
0	OTRO	2.019
0	FINCAS Y SIMILARES	1.847
0	DENTRO DE LA VIVIENDA	1.526
0	ESTABLECIMIENTO EDUCATIVO	584
0	COLEGIOS ESCUELAS	357
0	ENTIDAD PUBLICA / ESTATAL	344
0	EDIFICIO	262
0	HOTELES RESIDENCIAS Y SIMILARES.	259
0	APARTAMENTO	230
0	RED VIRTUAL O INTERNET	145
0	DATO NO DISPONIBLE	121
0	LUGAR DE TRABAJO	112
0	ESTABLECIMIENTO DE SALUD	91
0	OFICINAS	81

DATAMART DELITOS SEXUALES

Fig. 21. Reporte de sitio y hora preferida por el abusador

De acuerdo con la Figura 22, se puede deducir que los extranjeros que más han sido víctimas de delitos sexuales son los venezolanos, esto se debe a que en los últimos años se ha presentado un aumento de personas procedentes de este país.

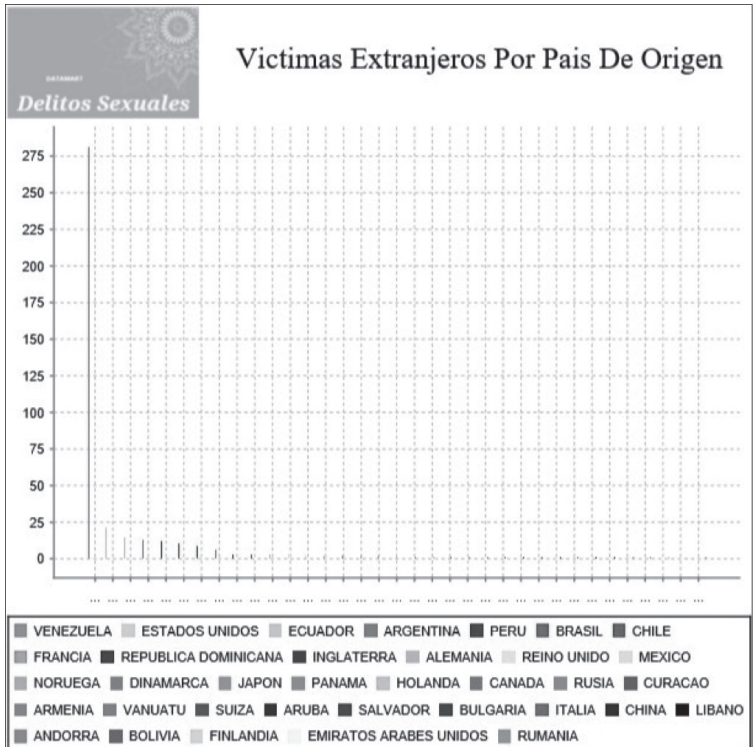


Fig. 22. Reporte de v*í*ctimas extranjeras por país de origen

IV. CONCLUSIONES

Después de realizar un análisis de datos entre el año 2010 y el primer semestre de 2019 se genera un resultado en el que no existe distinción de las víctimas de delitos sexuales por sexo, pero el porcentaje más alto se presenta en mujeres.

Al realizar un análisis de los delitos sexuales en relación con la hora en que ocurren, se observa que la medianoche es cuando se cometen más delitos sexuales, principalmente los actos sexuales con menor de 14 años y acceso carnal abusivo con menor de 14 años, y es claro que los menores de 14 años son los más expuestos a este tipo de abusos, según la información.

Es importante tener en cuenta los delitos por fecha: el 1.° de enero es el día en que se presentan más delitos, con un aproximado de 1950 delitos por agresión sexual, con respecto a los demás días del año. Esto indica que posterior a una festividad como año nuevo los agresores aprovechan para atacar a sus víctimas.

Es relevante destacar que, en el reporte de delitos sexuales por municipios, el índice de los delitos más alto a nivel nacional ocurre en el departamento de Santander, municipio de San Joaquín, y es también el más frecuente con víctimas menores de 14 años.

Después de analizar los datos reportados entre el año 2010 hasta el primer periodo del año 2019, han persistido los delitos de actos sexuales con menor de 14 años, acceso carnal violento y acceso carnal abusivo con menor de 14 años y se ratifica que las agresiones contra esta población son las que más se presentan en el país.

Con la información recolectada se puede proponer para municipios del Cauca la creación de un laboratorio para el análisis de resultados, lo cual favorecerá el diseño de estrategias y campañas de prevención en contra de los abusos en especial a las mujeres y a los menores de 14 años.

Los reportes generados dan una aproximación más a la realidad de la problemática que se presenta en Colombia en cuanto a la temática de delitos sexuales, puesto que todo este proceso nos permite concluir que los delitos más cometidos por los agresores son: actos sexuales con menor de 14 años, acceso carnal violento y acceso carnal abusivo con menor de 14 años, de donde se infiere que la población infantil es la más afectada.

Se recomienda hacer campañas de educación sexual orientadas a los niños entre los 7 y 14 años de edad. En las que se les explique que sus cuerpos solo les pertenecen a ellos y que tienen derecho a poner límites, además es importante decirles que hay adultos que podrían intentar lastimarlos u obligarles a hacer cosas molestas o que ellos no desean hacer, alertar que el adulto podría ser una persona conocida, amada o de confianza.

Es necesario generar conciencia en los niños que si alguna persona tiene algún comportamiento extraño como manosear, tomar fotografías desnudas u ofrecer regalos, sus padres o adultos responsables deben saberlo, es muy importante explicar este tipo de situaciones y tranquilizar para que ellos sepan que no son culpables de lo que haga el adulto. Se debe educar a los niños para que nunca acepten que personas desconocidas les lleven o acompañen a su casa u otros lugares, aun cuando los hayan conocido en sitios públicos y se noten amistosos o atractivos, todos estos aspectos generan prevención y protección para los niños en nuestro país.

V. REFERENCIAS

[1] Y. Jaramillo Giraldo y A. C. Córdoba Ospina, "Caracterización del abuso sexual en la ciudad de Medellín en los años 2011-2012 en edades comprendidas de 0 a 24 años de edad", Medellín, 2013.

[2] S. Bott, A. Guedes, A. P. Ruiz-Celis, J. A. Mendoza, Violence against women in latin america and the caribbean. A comparative analysis of population-based data, PAHO HQ Library Cataloguing-in-Publication Data, 2019.

[3] O. Pallitto, "The relationship between intimate partner violence and unintended pregnancy: analysis of a national sample from Colombia", International Family Planning Perspectives, vol. 30, n.º 4, pp. 165-238, 2004.

[4] M. T. Falcon, "Entre el silencio y la impunidad: violencia sexual en escenarios de conflicto", La ventana, vol. 5, n.º 41, 2015.

[5] A. Bustamante, "Soluciones de inteligencia de negocios en la práctica: apoyo a la toma de decisiones en proyectos educativos para población infantil vulnerable en el Caribe colombiano", UIS Ingenierías, vol. 10, n.º 2, 2011.

[6] Naciones Unidas - Oficina contra la Droga y el Delito, "Violencia, Crimen y tráfico ilegal de armas en Colombia". Dsisponible en: https://www.unodc.org/pdf/Colombia_Dec06_es.pdf

[7] Policía Nacional de Colombia, "Estadística Delictiva" [En línea]. Disponible en: https://www.policia. gov. co/grupo- informaci %C3% B3n- criminalidad/estadistica- delictiva

[8] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy, and B. Becker, The Data Warehouse Lifecycle Toolkit, John Wiley & Sons, 2008.

[9] G. R. Rivadera, "La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)", Cuadernos de la Facultad, n.° 5, Universidad Católica de Salta, 2010. [En línea]. Disponible en: https://www.ucasal.edu.ar/htm/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf

Capítulo 2
Diseño de un Sistema de Información que Implementa el Modelo Automatizado para Asegurar la Calidad de los Requerimientos en Proyectos de Software

Luis Eduardo Peláez Valencia - Luiseduardo.pelaez@gmail.com
Universidad Tecnológica de Pereira
Miguel A. Cohuo Ávila - Macohuo@itescam.edu.mx
Instituto Tecnológico Superior de Calkiní en el estado de Campeche (ITESCAM)
Alonso Toro Lazo - Alonso.toro@ucp.edu.co
Juan Luis Arias Vargas - Juan.arias@ucp.edu.co
Universidad Católica de Pereira
Daniel Eduardo Rodríguez Franco - Danieleduardorodriguez09@gmail.com
Tecnológico Nacional de México
Yulieth Gómez Osorno - Yulieth.gomez@ucp.edu.co
Universidad Católica de Pereira
Diego Alejandro Pérez Correa - Dalejandropc@gmail.com
Universidad Católica de Pereira

I. INTRODUCCIÓN

Abordar los proyectos de software para usuarios finales conlleva un ejercicio riguroso y responsable que debe asegurar el cumplimiento de las expectativas del cliente, además de las recomendaciones y especificaciones que hayan sido establecidas en las etapas iniciales de un determinado proyecto, de tal manera que la disposición y el cumplimiento de las actividades llevadas a cabo durante el proceso de desarrollo concluyan en atributos de calidad que posea el producto software resultante, de modo que pueda ser calificado como de calidad. Esto, normalmente configurado en la triple restricción de alcance o funcionalidad, tiempo y costo [1], variables propias del proceso de aseguramiento de la calidad del software.

La ingeniería del software establece una serie de criterios indispensables para lograr procesos y productos software de calidad mediante el desarrollo de actividades de análisis y diseño previos y de integración y verificación posteriores [2], distribuyéndolas en el ciclo de vida del desarrollo del software. En las etapas iniciales del

desarrollo, el análisis y diseño permiten comprender los requerimientos del cliente y plasmarlos fielmente en especificaciones concretas [3] realizada como parte del proyecto de investigación denominado "Procedimiento para especificar y validar requisitos de software en MIPYMES desarrolladoras de software, basado en estudios previos en la región". Lo anterior responde a la necesidad de indagar sobre los problemas de calidad del sector del software relacionado con el tratamiento que se les hace a los requisitos como un punto clave para lograr productos y procesos de calidad, con el ánimo de desarrollar propuestas que contribuyan a mejorarla, inicialmente en la industria que se representa desde la ciudad de Pereira. Para ello, se hace un recorrido por algunos trabajos que permiten dar cuenta de los logros que se han alcanzado en investigaciones similares y que tienen como objeto de estudio los requisitos, reconociendo, a partir de unos antecedentes regionales, nacionales e internacionales los orígenes de ese problema y las propuestas que se han desarrollado para ayudar a corregirlo. Finalmente, y como conclusión, se encuentra que han sido muchas las propuestas que de una u otra manera pretenden contribuir al mejoramiento de la calidad del software a partir de los requisitos, pero en ninguna de las revisadas se ha propuesto un método, proceso o lineamiento claramente definido para elaborar una completa, consistente, modificable y trazable, específicamente de requisitos de software en las micro, pequeñas y medianas empresas (MIPYMES, con lo cual los expertos de desarrollo cumplen con las expectativas de calidad.

De acuerdo con [4], estos requerimientos "expresan las necesidades y limitaciones impuestas a un producto de software que contribuyen a la solución de algún problema en el mundo real". En este sentido, la adopción de una disciplina como RQA (*Requirement Quality Assurance*) aporta

> [un] conjunto de instrumentos y buenas prácticas que proporcionan la seguridad de que los productos y procesos relacionados con la definición de un problema que se resolverá con software nuevo o modificado, cumplirán con lo especificado previamente, para generar la confianza necesaria que los lleve a ser considerados como de calidad [5].

Lo anterior, unido a un proceso de SQA (*Software Quality Assurance*),

> proporciona la seguridad de que los productos y procesos en el ciclo de vida del proyecto cumplan con sus requisitos especificados mediante la planificación, promulgando y llevando a cabo un conjunto de actividades de suficiente confianza, e indicando la calidad que radica en el software [6].

Esto permite adoptar un proceso de calidad en todas las etapas del ciclo de vida del software, que comprende desde las etapas iniciales de requerimientos hasta las pruebas finales del sistema, lo que promueve la utilización y adaptación de técnicas y buenas prácticas que ayudarán a la consecución de la calidad.

De esta manera, y para efectos del desarrollo del sistema de información que permita automatizar el "modelo para el aseguramiento de la calidad de los requerimientos en proyectos de software", llevado a cabo por la línea Ingeniería de Software del Grupo de Investigación entre Ciencia e Ingeniería, este artículo pretende aportar a la aplicación práctica de los conceptos RQA y SQA evidenciando algunos de los artefactos y técnicas de especificación, diseño y modelado propuestos por la ingeniería de software para el desarrollo de la aplicación denominada CHAMI, que implementa el modelo automatizado antes mencionado.

Por consiguiente, se detallan algunos procesos, técnicas y buenas prácticas de ingeniería de software implementadas, específicamente las que hacen parte de la etapa de diseño. Es importante aclarar que este artículo no se enfoca en el modelo para el aseguramiento de la calidad de los requerimientos, sino en el diseño del sistema de información que lo implementa; más concretamente en la modelación de los procesos.

El desarrollo y la especificación de requerimientos en requisitos es quizá la etapa más importante y determinante del proceso de desarrollo de software, pues permite garantizar, desde las primeras etapas del proceso, la calidad de las aplicaciones resultantes [7] [8].

Así, dentro de la etapa de requerimientos es indispensable aplicar técnicas y procedimientos que permitan la elaboración de un documento de especificación de requisitos de software (SRS, por sus siglas en inglés [9]) completo, consistente, modificable, trazable [10] y lo suficientemente detallado como para permitir, en primera instancia, una ayuda fundamental al criterio del analista y/o profesional encargado del proceso de requerimientos en su labor.

Por ello, el presente artículo se enfoca, principalmente, en explicar y documentar la fase de diseño de software, para la cual, a partir de la especificación de requisitos de software entregada, se elaboran diferentes diagramas estáticos y dinámicos de acuerdo con el estándar UML [11], para modelar la estructura y el comportamiento del sistema que se va a desarrollar, así como algunos prototipos y otros procesos que hacen parte del ciclo de vida del desarrollo del software seleccionado para el proyecto.

El artículo se estructura de la siguiente manera: en la sección 2 se presentan los materiales y métodos contemplados en el proyecto; en la sección 3, se describen los resultados y artefactos obtenidos de la ejecución de cada una de las fases llevadas a cabo en el desarrollo del sistema de información; la discusión de resultados se presenta en la sección 4, en la sección 5, las conclusiones y, por último, la bibliografía consultada.

II. MATERIALES Y MÉTODOS

La definición de las técnicas de diseño y modelado del sistema de información, que implementa el modelo automatizado para el aseguramiento de la calidad de los requerimientos en proyectos de software, se justifica a partir del ejercicio y la exploración de bibliografía y levantamiento de información que sustenten la validez y pertinencia del modelo, con base en el aseguramiento de la calidad de los requerimientos, así como la inclusión y utilización de técnicas de diseño y modelado. De esta forma, se proponen las fases para la ejecución del proyecto (Figura 1), considerando que la fase 6, a pesar de haber sido parte del proyecto, no se presenta en este documento:

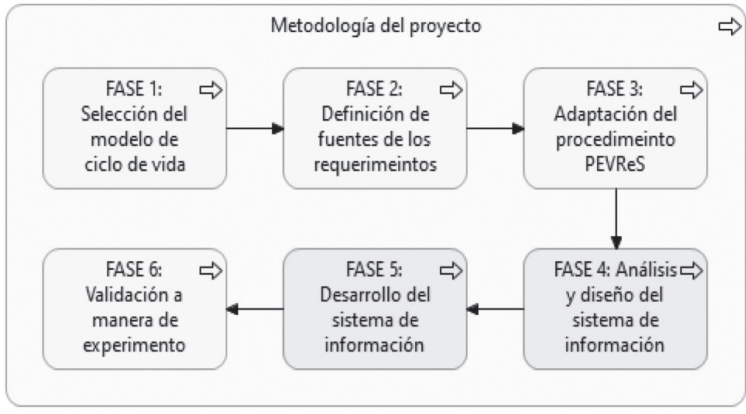


Fig. 1. Fases de la metodología de ejecución del proyecto

A. Fase 1: Selección del modelo de ciclo de vida del software (MCVS)

Dentro del ejercicio del diseño e implementación del modelo automatizado CHAMI, y con la intención de establecer un marco de trabajo que ayude a la estructura, planificación y control del desarrollo del producto software, se hace importante seguir un modelo de ciclo de vida del desarrollo de software que aborde cada uno de los aspectos antes considerados y que, a su vez, logre sintetizar y justificar la construcción basada en el diseño.

Actualmente, existe una amplia variedad de metodologías y modelos dispuestos a abordar desarrollos de software, sin embargo, tienen limitaciones inherentes al contexto de aplicación. Es decir que no existe una metodología que pueda ser aplicada perfectamente en todos los contextos de trabajo y todos los tamaños de proyecto [12].

En este sentido, de acuerdo con el estudio y la clasificación de modelos de ciclo de vida para el desarrollo de software (MCVDS), el rumbo de la investigación se encamina hacia la utilización del modelo de prototipos, el cual, a partir de la experiencia del equipo de desarrollo y la participación del usuario en la elaboración de prototipos, su evaluación y desarrollo, debe permitir la definición de especificaciones completas y acordes con el producto definitivo, lo que en consecuencia se ajusta al propósito y los objetivos del proyecto respecto a la calidad en el proceso de desarrollo y el producto software.

El modelo de prototipos inicia con la actividad de comunicación, continúa con la realización de un plan rápido y modelado o diseño rápido, para luego construir el prototipo y desarrollarlo [13], lo que permitirá una retroalimentación por parte del cliente/usuario que aclare las funcionalidades y requerimientos que debe poseer el sistema, lo cual tendrá sin dudas un impacto positivo al obtener mejores resultados en el producto final entregado, asegurando que cumpla con la mayor especificidad de las necesidades del cliente.

De acuerdo con la caracterización realizada, el modelo prototipo se centra en una representación de aquellos aspectos del software que serán visibles para el cliente y/o usuario final, de esta manera, el modelo permite realizar pruebas y modificaciones cuando es necesario. La clasificación y las etapas consideradas en el modelo se especifican en la Figura 2, las fases 1, 3, 4 y 5 se adaptan al esquema propuesto por el modelo para el aseguramiento de la calidad de los requerimientos en proyectos de software, en el desarrollo de este proyecto.

Sin embargo, por el enfoque de este documento, se presentarán solo algunos elementos correspondientes a las fases de modelado-diseño y construcción del prototipo, aunque todas las fases propuestas fueron llevadas a cabo.

Fig. 2. Modelo de ciclo de vida de prototipos. Adaptado de [13]

B. Fase 2: Recopilación y fuentes de los requerimientos

De acuerdo con el proceso de definición e identificación de los requerimientos como etapa inicial dentro del proceso de desarrollo de software, y con el fin de comprender y establecer las necesidades en función del sistema, es de vital importancia definir claramente los requerimientos del software, de tal manera que se tenga una especificación completa y consistente que sea, como lo indica [14], una base sólida para el proceso de desarrollo de software que, en consecuencia con el modelo desarrollado, tenga la calidad esperada para el diseño y la construcción del sistema de información CHAMI.

La ingeniería de requisitos plantea la división de la fase de requisitos en cuatro etapas: identificación, análisis, especificación y validación de requisitos [15]. Para la identificación de requisitos del proyecto, se establecieron cuatro fuentes principales de información: primera, la recolección de datos sustentada sobre la caracterización del sector y la industria, específicamente en las Mipymes [16], [5]; segundo, identificación de fuentes documentales de trabajos previos y el estado de la técnica [10]; tercero, exploración de antecedentes relacionados con investigaciones y propuestas de mejora de procesos para la industria del software, y cuarto, el análisis y experiencia del grupo de trabajo, junto con la vinculación del sector educativo (estudiantes y graduados de las carreras relacionadas con las TIC y el software).

La Figura 3 contempla el trabajo colectivo y colaborativo dispuesto sobre las fuentes para determinar los requerimientos del sistema de información. Seguidamente, se presenta una breve explicación de cada una de las fuentes de información y los instrumentos empleados en algunas de ellas.

Fig. 3. Fuentes de los Requerimientos

1) Caracterización de la industria: el instrumento de medición diseñado y aplicado a las empresas Mipymes dedicadas a la producción de software en la ciudad de Pereira (Colombia), tuvo el objetivo de determinar y caracterizar su proceso de gestión en la etapa de requerimientos, y que además permitiera recolectar la información suficiente para generar preguntas a la industria del software sobre la manera como están siendo tratados los requisitos, y en consecuencia conocer el estado de la apropiación del aseguramiento de la calidad del producto software [12]. De esta manera, se tomaron como entradas las necesidades representadas por el sector, que se convirtieron en una de las fuentes de requerimientos más determinantes para el desarrollo del sistema de información CHAMI.

2) Estado de la técnica y caracterización de herramientas: mediante una metodología de exploración teórica se hizo hincapié en un aspecto de gran importancia, como es la fiabilidad, con el fin de no desconocer el ejercicio realizado por autores y organizaciones que son autoridad en temas de ingeniería de software, ingeniería de requisitos y aseguramiento de la calidad del software. En tal sentido, en este estudio se tuvieron en cuenta expertos en la disciplina como Pressman

[17], Sommerville [18], Wiegers [19], entre otros, y organizaciones como International Requirements Engineering Board (IREB) [20], International Standards Organization (ISO) [21], IEEE Computer Society [4] y Software Engineering Institute (SEI).

Respecto a la caracterización de herramientas para el desarrollo y la gestión de requisitos, se hace una exploración teórica y práctica de las herramientas más usadas en el entorno mundial por las compañías desarrolladoras de software a la hora de desarrollar y gestionar los requerimientos. Esta caracterización se realiza considerando factores como tipo de licenciamiento (de pago y libres), plataforma, facilidad de soporte o contacto con el fabricante, fecha de la última actualización, vigencia en el mercado, posibilidad de prueba, etc.

Finalmente, se hace un análisis comparativo para determinar sus principales características, funcionalidades y limitaciones, además de identificar los elementos de entrada requeridos, las restricciones, los detalles del proceso que se ha de seguir, los atributos que permite definir, sus ventajas y desventajas, entre otros aspectos [12].

3) Trabajos previos: por medio de una exploración teórica de antecedentes, a partir de la cual las deficiencias encontradas por [10] y la caracterización de la industria reportada en el punto número 1, respecto a un deficiente o nulo aseguramiento de la calidad de los requisitos en la industria local del software, se realiza una exploración de antecedentes regionales, nacionales e internacionales, con la intención de evidenciar proyectos de investigación, desarrollos e iniciativas propuestos para contribuir a una solución a este problema. A partir de allí, se logran adquirir algunos requerimientos y lecciones aprendidas importantes para el desarrollo del sistema de información.

4) Consideraciones del grupo de investigación: basados en la experiencia del equipo del proyecto de investigación y la articulación con otros actores vinculados con el proyecto por parte del grupo de Investigación Entre Ciencia e Ingeniería de la Universidad Católica de Pereira, y a partir de las investigaciones previas y desarrollos realizados, se establece un instrumento para la inclusión de requerimientos y consideraciones prácticas que deben ser contempladas por el sistema que se va a desarrollar. De esta manera, se logran requerimientos adicionales que, desde la práctica y la experiencia, aportan al mejoramiento de la propuesta de desarrollo.

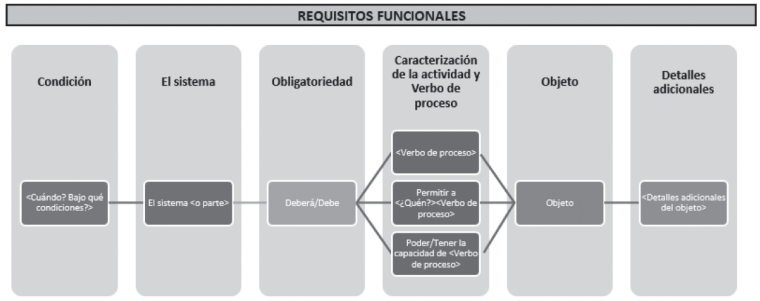
C. Fase 3: Adaptación del modelo PEVReS para la especificación y validación de requisitos de software

Una vez identificados los requerimientos, se efectúan su

especificación y validación de acuerdo con las técnicas propuestas por el Procedimiento para especificar y validar requisitos de software (PEVReS) [8], tal como se muestra en la Figura 4, el cual facilita la creación de un documento SRS que cumpla con las características de una buena especificación [22], [10], y además se validen mediante un trabajo conjunto entre el analista y el cliente y/o usuario del sistema para asegurar que los requisitos que se identificaron y se especificaron reflejen realmente lo que los usuarios necesitan [23].

Fig. 4. Procedimiento PEVReS [24]

El procedimiento aporta además los formatos y plantillas necesarios para lograr la aplicación de cada una de las técnicas, así como el documento SRS basado en una ampliación del estándar IEEE Std. 830. En la Figura 5, por ejemplo, se representa la técnica de especificación de requisitos de usuario (DRU) en lenguaje natural sugerida por el procedimiento, mediante la cual se establece la sintaxis adecuada de escribir los requerimientos funcionales y no funcionales que luego serán llevados al documento SRS, de tal manera que se cumplan los criterios de evaluación de un buen requisito definidos por [19]: completo, correcto, realizable, necesario, priorizable, no ambiguo y verificable.

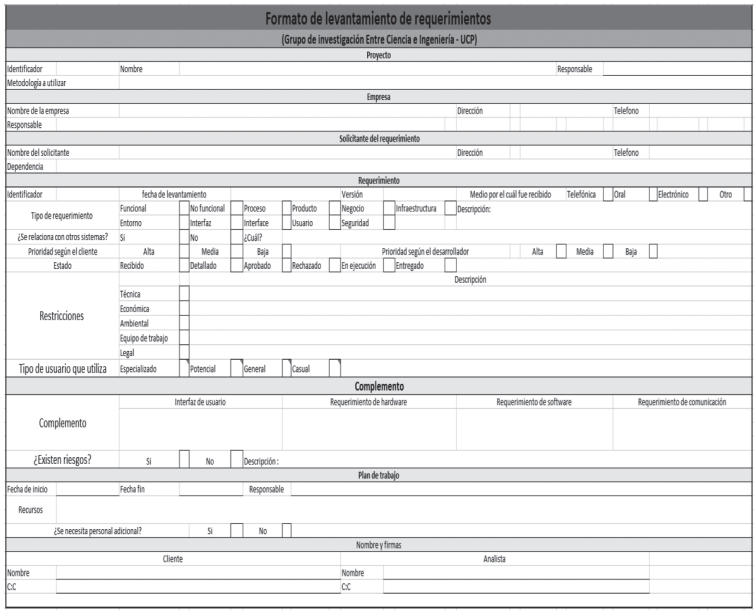


Requisitos funcionales - Ejemplo

Escenario condicional	Obligatoriedad	Interacción	Objeto	Detalles
	El sistema debe	Permitir a los estudiantes <matricular>	Un curso	Con el cual cumplan los prerrequisitos
	El sistema debe	Permitir a los profesores <consultar>	La lista de estudiantes	Registrado en cada uno de sus cursos
Cuando se ha cumplido fecha de inscripción a un curso o cuando se ha completado el cupo máximo de estudiantes	El sistema debe	<Cerrar>	El curso	
	El sistema debe	Tener la capacidad de <notificar>	Al sistema de pagos	Cuando un estudiante haya realizado el proceso de pago de matrícula
	El sistema debe	Permitir a los administradores <crear>	Cursos	
Tan pronto como un estudiante haya realizado el pago de matrícula		<Imprimir>	La factura del curso	

Fig. 5. Esquema de especificación de requisitos funcionales en lenguaje natural [25]

De igual forma, se utilizan instrumentos para la gestión de requisitos, con la intención de determinar aspectos que ayuden con la tipificación, identificación, priorización, descripción y control de los requerimientos, una vez se haya capturado por parte del cliente, lo que garantiza la documentación y trazabilidad. Uno de estos instrumentos, elaborado por el grupo de investigación, se presenta en la Figura 6.



Formato de levantamiento de requerimientos

(Grupo de investigación Entre Ciencia e Ingeniería - UCP)

Proyecto

Identificador	Nombre	Responsable	
Metodología a utilizar			

Empresa

Nombre de la empresa		Dirección	Telefono
Responsable			

Solicitante del requerimiento

Nombre del solicitante		Dirección	Telefono
Dependencia			

Requerimiento

Identificador	fecha de levantamiento				Versión				Medio por el cuál fue recibido	Telefónica		Oral		Electrónico		Otro	
Tipo de requerimiento	Funcional		No funcional		Proceso		Producto		Negocio		Infraestructura		Descripción:				
	Entorno		Interfaz		Interface		Usuario		Seguridad								
¿Se relaciona con otros sistemas?	Si		No		¿Cuál?												
Prioridad según el cliente	Alta		Media		Baja		Prioridad según el desarrollador						Alta		Media		Baja
Estado	Recibido		Detallado		Aprobado		Rechazado		En ejecución		Entregado						
													Descripción				
Restricciones	Técnica																
	Económica																
	Ambiental																
	Equipo de trabajo																
	Legal																
Tipo de usuario que utiliza	Especializado		Potencial		General		Casual										

Complemento

	Interfaz de usuario	Requerimiento de hardware	Requerimiento de software	Requerimiento de comunicación
Complemento				

¿Existen riesgos? Si No Descripción:

Plan de trabajo

Fecha de inicio	Fecha fin	Responsable	
Recursos			

¿Se necesita personal adicional? Si No

Nombre y firmas

Cliente		Analista	
Nombre		Nombre	
CC		CC	

Fig. 6. Formato para identificación de requisitos.
Fuente: Grupo de investigación Entre Ciencia e Ingeniería

D. Fase 4: Análisis y Diseño del sistema de información

Se presenta en esta sección un resumen, a manera de ejemplo demostrativo, de algunos de los diagramas UML y modelos desarrollados como parte del análisis estructural y dinámico del sistema de información CHAMI. Con ellos, se pretende adquirir un mayor dominio del problema y de la solución, de tal manera que los resultados correspondan con los requerimientos y cumplan las expectativas del proyecto.

1) Modelo de datos: la mayoría de los sistemas de software robustos utilizan bases de datos de información de altos volúmenes de datos. En algunos casos, la base de datos es independiente del sistema de software, en otros se crea para el sistema que se está desarrollando. Es una parte importante del modelado de sistemas para la definición de la forma lógica de los datos procesados por el sistema. Estos se denominan a menudo modelos semánticos de datos [17]. La Figura 7 muestra el esquema general del modelo de datos elaborado para una arquitectura multi-tenant, capaz de soportar múltiples equipos de trabajo, laborando en múltiples proyectos para múltiples empresas del sector, donde los diferentes usuarios no dependen de la empresa y pueden ir y venir entre diferentes empresas, ligados a ellas solamente por sus proyectos de interés.

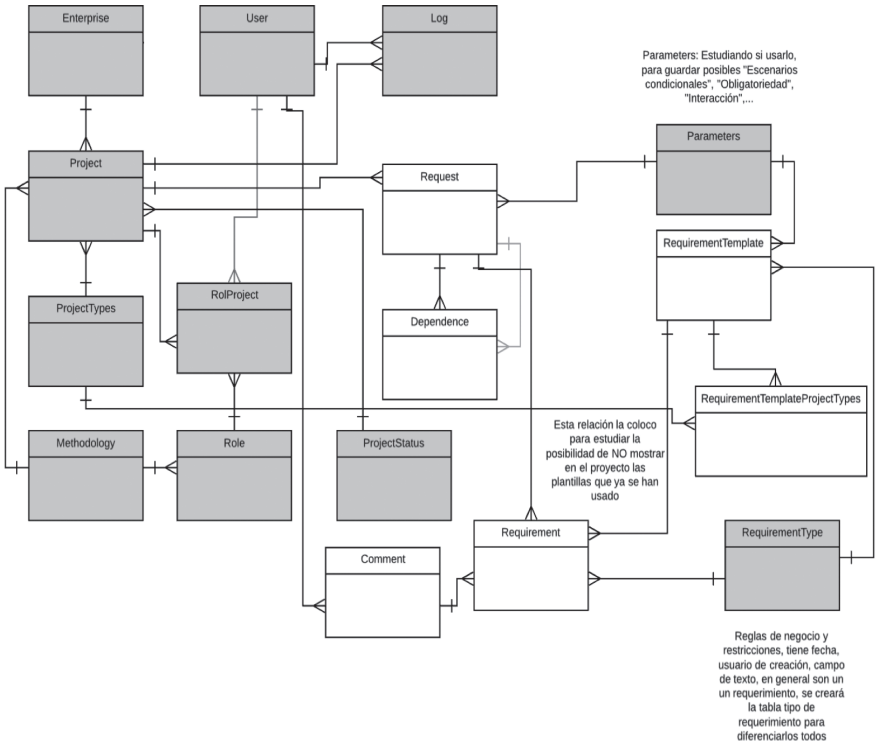


Fig. 7. Diagrama del Modelo de Datos

2) Modelo de clases: representa la estructura estática del sistema de información, indicando las clases que lo componen junto con sus atributos, operaciones, y las relaciones entre los objetos, tales como asociaciones, herencias, composición o agregación, entre otras. De acuerdo con las técnicas de diseño, para este tipo de proyectos se hace indispensable considerar el diagrama de clases, por lo que en la Figura 8 se presenta el diagrama respectivo.

Fig. 8. Diagrama de clases

A continuación se presenta una breve descripción de algunas de las clases que son indispensables dentro del propósito del modelo de aseguramiento de la calidad de los requerimientos, con la intención de ejemplificarlas:

- *RequerimentTemplate:* permite identificar información de los requerimientos de acuerdo con el formato del modelo llamado "*identificación de requerimientos*", por medio del cual se almacena información relevante mediante atributos como ID del requerimiento, responsable, versión, descripción, tipificación, además de otros atributos que permiten mantener la trazabilidad de los requerimientos y garantizar la calidad en el proceso de gestión.

- *Requirement:* cumple un papel indispensable dentro del modelo, puesto que caracteriza y consolida el listado de requerimientos una vez realizada su identificación mediante el proceso de elicitación (identificación) con el cliente, como primera fase del proceso.

- *Request:* se encarga de la especificación del requisito de acuerdo con los parámetros entregados por la técnica de especificación en lenguaje natural antes presentada, y garantiza que la descripción del requerimiento siga la sintaxis adecuada y que todos los atributos relacionados con su estructura hayan sido correctamente diligenciados (escenario condicional, obligatoriedad, caracterización de la actividad y verbo de proceso, objeto, detalles adicionales).
- *Dependence:* se considera como fuente de control que justifique y sustente la dependencia entre grupos de requerimientos, que dentro de un esquema tradicional podrían agruparse por Caso de Uso o por disponibilidad del cliente y/o Usuario, ayudando a garantizar la trazabilidad de los requerimientos.

Sobre estas clases se establecen los métodos que se pueden ejecutar sobre los objetos pertenecientes a cada clase, los cuales representarán funciones de la aplicación que estarán disponibles para el usuario mediante la aplicación.

3) Modelo relacional: dentro del proceso de diseño se considera indispensable la realización del modelo relacional, el cual permite diseñar la base de datos con las tablas en las cuales será almacenada la información del sistema. Dicha estructura, que puede entenderse como un conjunto de relaciones bidimensionales entre tablas constituidas por filas (tuplas) y columnas (atributos) [26], contempla principalmente las entidades **Request** y **Requirement**, en las cuales reposan los registros correspondientes a cada requerimiento y su especificación.

Asimismo, se incluyen otras tablas en función de la operatividad y parametrización del sistema, que permiten obtener un conjunto de elementos que aportan a la automatización del proceso de desarrollo y gestión de cada requisito, por supuesto al interior de un determinado proyecto que a su vez pertenece a una empresa y posee usuarios particulares. La estructura mencionada se presenta a continuación en la Figura 9.

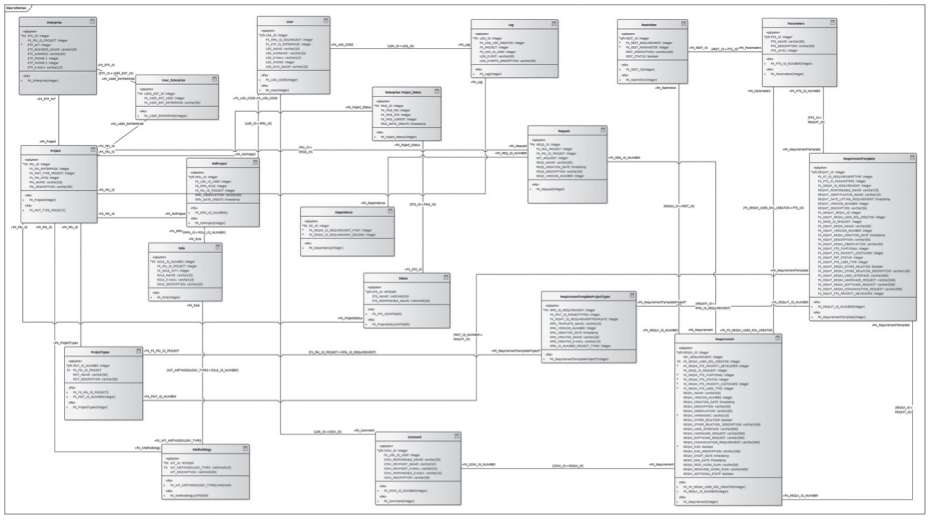


Fig. 9. Modelo Relacional

4) Diagramas de Casos de Uso: existen diferentes maneras de clasificar los requisitos, sin embargo, de acuerdo con [24], basta una clasificación en requisitos de negocio, requisitos de usuario y requisitos del sistema para describir de forma suficiente las funcionalidades del sistema que representan las necesidades del cliente. Asimismo, para cada tipo de requisito existe un conjunto de técnicas que ayudan a representarlo. El diagrama de casos de uso es adoptado en este proyecto como la técnica para representar los requisitos funcionales que corresponden al nivel usuario y su comportamiento dinámico [27].

Este diagrama permite representar la forma en que un cliente (Actor) opera con el sistema, y además la forma, el tipo y orden en que los elementos interactúan entre sí. Un ejemplo de los diagramas realizados durante el diseño del sistema se muestra en la Figura 10.

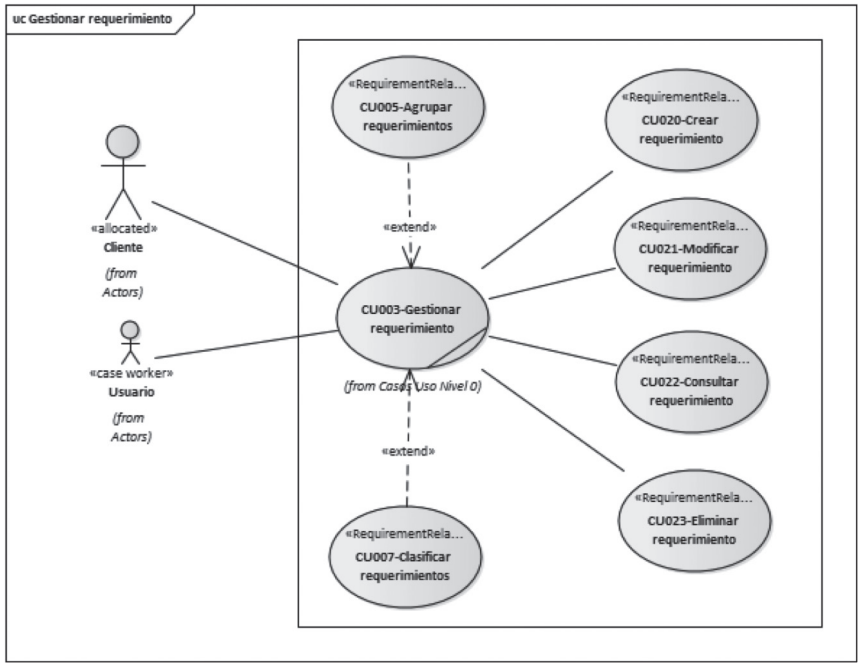


Fig. 10. Caso de uso Gestionar Requerimiento

5) Diagramas de Secuencia: el diagrama de secuencia es un tipo de diagrama de interacción cuyo objetivo es describir el comportamiento dinámico del sistema de información haciendo énfasis en la secuencia de los mensajes intercambiados por los objetos [27]. Estos diagramas ayudan a definir los métodos u operaciones que se deben codificar al interior de las clases, así como los mensajes de interacción con el usuario y la secuencia lógica de las funcionalidades.

El diagrama presentado en la Figura 11 muestra una interacción del actor Mipyme con la interfaz gestión de proyectos, la cual a su vez se comunica con el controlador y la entidad responsable de los proyectos, de tal manera que pueda crear y consultar información de un determinado proyecto.

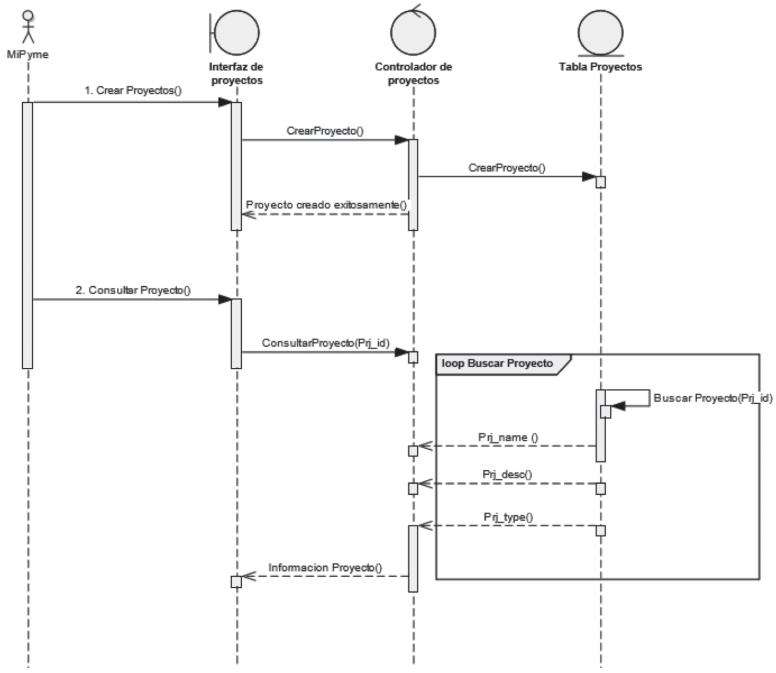


Fig. 11. Diagrama de secuencia Gestión de Proyecto

III. RESULTADOS

En esta sección se presentan los artefactos software desarrollados de acuerdo con la Fase 5 de la metodología propuesta (desarrollo del sistema de información), presentada en la Figura 1, así como las etapas 2 y 3 del modelo de ciclo de vida guiado por prototipos, presentado en la Figura 2. Esta labor se realizó a partir de los requerimientos especificados y corresponden a los *mockups* entregados al ingeniero desarrollador como parte del proceso de maquetación de vistas.

En esta etapa del proyecto, el documento SRS cumple un papel supremamente importante, ya que es el insumo principal para el proceso de desarrollo, toda vez que es el documento en el cual se describen con claridad y sin ambigüedades las necesidades y expectativas del cliente, de cuya calidad depende en gran medida la calidad del producto software resultante.

En la Figura 12 se puede observar un ejemplo de requisitos de usuario escritos mediante la técnica de lenguaje natural (ver Figura 5). Posteriormente, estos requisitos se clasifican, ponderan, priorizan y especifican mejor mediante la plantilla de identificación y análisis (ver Figura 6) con el propósito de facilitar su gestión y trazabilidad a lo largo del proyecto.

REQUERIMIENTOS FUNCIONALES					
Identificador	**Escenario condicional**	**Obligatoriedad**	**Interacción**	**Objeto**	**Detalles**
REQF001		El sistema debe	permitir al administrador <gestionar*>	un usuario	
REQF002	Siempre y cuando no exista un proyecto con el mismo nombre	El sistema debe	permitir al administrador <gestionar>	un proyecto	
REQF003	Siempre y cuando no exista un requerimiento con el mismo nombre en el proyecto	El sistema debe	permitir solo a un usuario <gestionar>	un requerimiento	
REQF004		El sistema debe	permitir al usuario cliente <describir>	un requerimiento	en un máximo de 150 caracteres
REQF005	Al crear un requerimiento	El sistema debe	<asignar>	fecha del requerimiento	usando la fecha actual del sistema
REQF006	Al modificar un requerimiento	El sistema debe	<asignar> una fecha de última actualización	al requerimiento	de acuerdo con la fecha del sistema
REQF007		El sistema debe	permitir al usuario <asignar> prioridad	a un requerimiento	de acuerdo con las siguientes categorías: alta, media y baja
REQF008	Al crear o modificar un requerimiento	El sistema debe	<versionar>	el requerimiento	

Fig. 12. Ejemplo de requerimientos funcionales

A. Elaboración de prototipos

Una vez finalizadas las etapas de educción, especificación y documentación de requerimientos, se procedió a la maquetación del sistema mediante la elaboración de los primeros *Mockups* y la promoción de una comunicación continua con el cliente que permitiera una oportuna retroalimentación, lo cual aportó algunas mejoras al diseño del sistema y facilitó al cliente una mejor comprensión de requisitos. Este involucramiento del cliente en el proceso de diseño (Diseño Centrado en el Usuario —DCU—) y desarrollo constituye una característica fundamental de la metodología de prototipos, no hacerlo se podría entender, según [10], como una de las principales causas de fracaso de los proyectos de software atribuibles a los requisitos. La Figura 13 muestra uno de los primeros prototipos de alta fidelidad realizados como parte del proyecto.

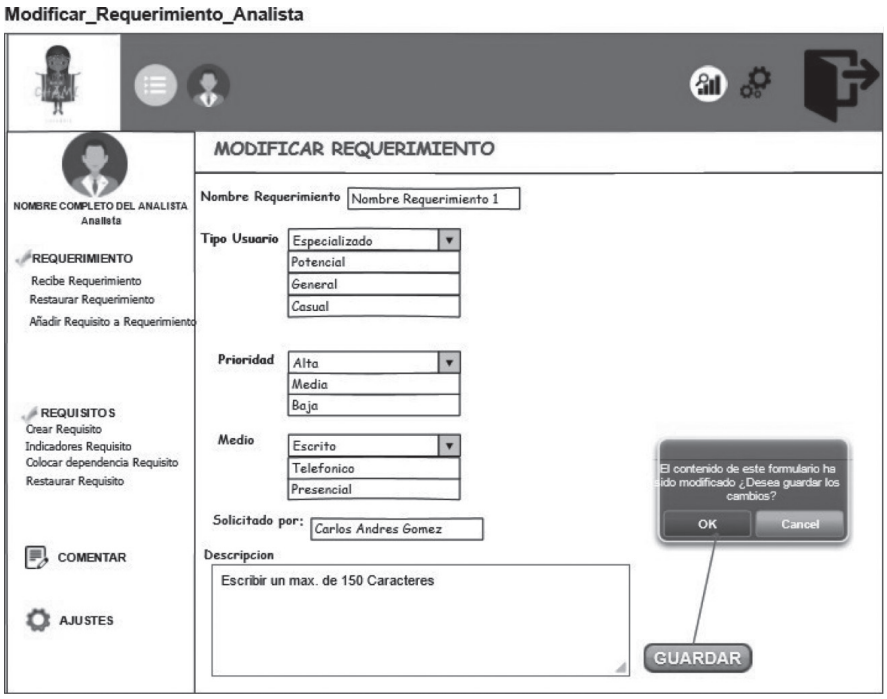


Fig. 13. *Mockup* Modificar Requerimiento Analista

B. Desarrollo del sistema de información CHAMI

El sistema de información desarrollado se comporta en forma de plataforma web, permite ingresar y gestionar los requisitos de un proyecto de software y llevar a cabo sobre ellos un conjunto de actividades que permiten asegurar su calidad (ver Figura 14), esto es, cumplir los criterios de evaluación de un requisito para que sea considerado como de calidad en el proceso de gestión del proyecto. El sistema fue desarrollado con la intención de facilitar a las Mipymes desarrolladoras de software de la región el proceso de desarrollo y gestión de los requisitos para sus proyectos.

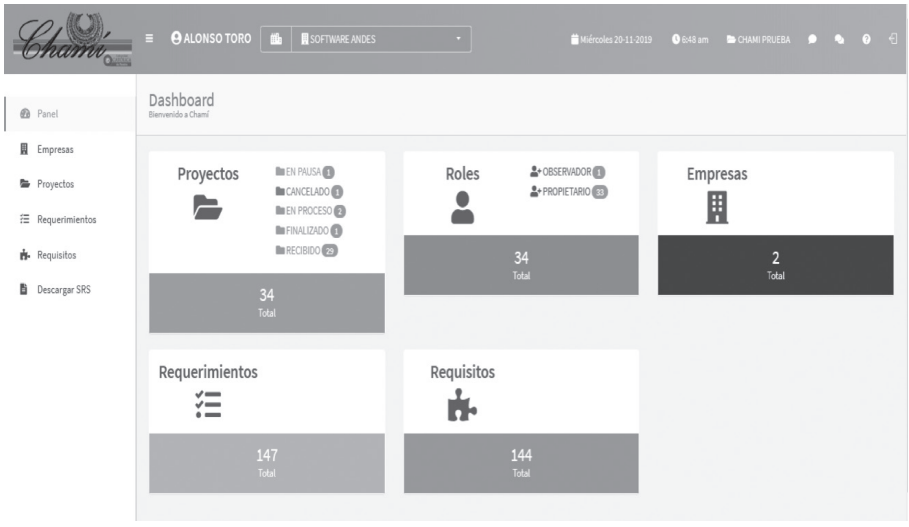


Fig. 14. Vista general de la Plataforma CHAMI

El sistema está conformado por seis diferentes módulos, mediante los cuales se puede realizar toda la gestión de las empresas, los usuarios y proyectos y los requerimientos/ requisitos asociados a ellos. De igual forma, el sistema permite la generación del documento SRS de un proyecto una vez finalizado el proceso de desarrollo y gestión de requisitos. A continuación, se presenta una descripción general de algunos de los módulos principales:

1) *Panel:* corresponde al *Dashboard* de la aplicación, donde se presentan estadísticas generales de los módulos subsecuentes.

2) *Empresa:* permite el ingreso al administrador de empresas, donde el usuario podrá gestionar las empresas que estarán asociadas a un determinado proyecto.

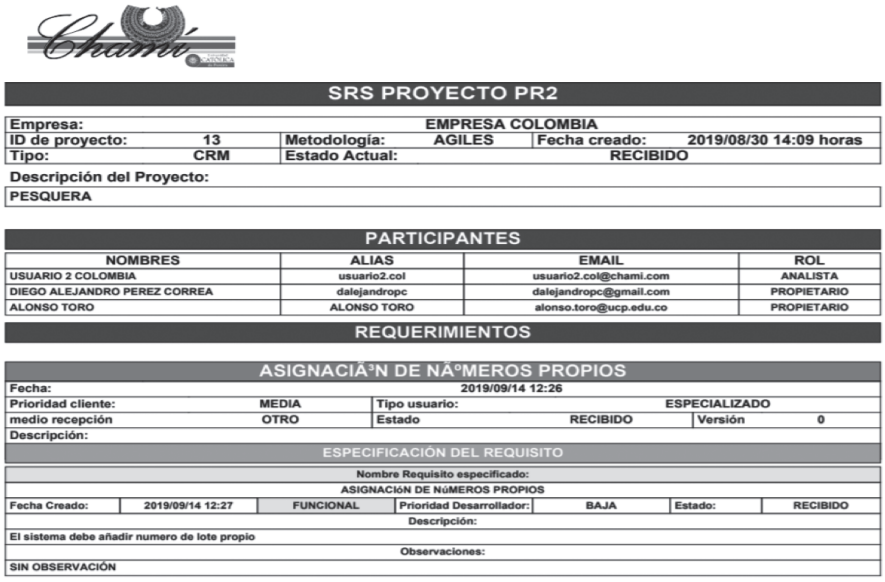
3) *Proyectos:* este módulo permite administrar los proyectos de una empresa.

4) *Requerimientos:* permite ingresar a la administración de requerimientos de un proyecto; se mostrarán únicamente los requerimientos de los proyectos de los cuales el usuario sea propietario o analista.

5) *Requisitos:* una vez ingresados los requerimientos a un proyecto, mediante este módulo se puede gestionar la especificación de requisitos.

6) Descargar SRS: esta funcionalidad de la aplicación permite generar y descargar en formato PDF el documento SRS del proyecto en el cual se esté trabajando.

La Figura 15 corresponde a una parte del documento SRS que permite generar la plataforma.



SRS PROYECTO PR2

Empresa:		EMPRESA COLOMBIA			
ID de proyecto:	13	Metodología:	AGILES	Fecha creado:	2019/08/30 14:09 horas
Tipo:	CRM	Estado Actual:		RECIBIDO	

Descripción del Proyecto:

PESQUERA

PARTICIPANTES

NOMBRES	ALIAS	EMAIL	ROL
USUARIO 2 COLOMBIA	usuario2.col	usuario2.col@chami.com	ANALISTA
DIEGO ALEJANDRO PEREZ CORREA	dalejandropc	dalejandropc@gmail.com	PROPIETARIO
ALONSO TORO	ALONSO TORO	alonso.toro@ucp.edu.co	PROPIETARIO

REQUERIMIENTOS

ASIGNACIÃ³N DE NÃºMEROS PROPIOS

Fecha:	2019/09/14 12:26				
Prioridad cliente:	MEDIA	Tipo usuario:		ESPECIALIZADO	
medio recepción	OTRO	Estado	RECIBIDO	Versión	0

Descripción:

ESPECIFICACIÓN DEL REQUISITO

Nombre Requisito especificado:					
ASIGNACIÓN DE NúMEROS PROPIOS					
Fecha Creado:	2019/09/14 12:27	FUNCIONAL	Prioridad Desarrollador:	BAJA	Estado: RECIBIDO
Descripción:					
El sistema debe añadir numero de lote propio					
Observaciones:					
SIN OBSERVACIÓN					

Fig. 15. Módulo de generación del documento SRS

Los módulos "Requerimientos" y "Requisitos" permiten, además de la gestión, la especificación de algunos procedimientos adicionales que se mencionan a continuación:

– *Dependencias:* el sistema facilita el reporte de dependencias entre requerimientos, con la finalidad de facilitar su priorización y posterior desarrollo.

– *Gestión de la configuración:* permite el control de versiones de un requerimiento o requisito, de tal manera que se puedan bajar o subir sus versiones de acuerdo con las actualizaciones realizadas sobre el mismo.

– *Exportar e imprimir:* permite exportar en formatos CSV y PDF, o imprimir el listado de requerimientos o requisitos correspondientes a un determinado proyecto.

– *Buscar:* en todos los módulos se tiene la posi bilidad de realizar una búsqueda por palabra clave, de tal manera que se pueda ubicar fácilmente un elemento cuando el listado es muy grande.

– *Gestionar estados:* permite controlar los estados de avance de un determinado requerimiento o requisito. Al cambiar de un estado a otro se genera una nueva versión del elemento.

– *Añadir sistemas foráneos:* este procedimiento, que aplica solo a los requisitos, permite indicar la existencia de un sistema externo al sistema en desarrollo con el cual se requiera interactuar.

– *Añadir restricción:* permite indicar las restricciones (*constraints*) relacionadas con el desarrollo de un requisito en particular.

– *Comentarios:* permite adicionar notas al administrador del proyecto o analista asignado, sobre un determinado requerimiento, requisito o proyecto en particular. Son visibles en todo el *workspace* y facilitan la realización de ajustes o cambios sobre los elementos en desarrollo.

IV. DISCUSIÓN

Los resultados obtenidos en este proyecto permiten describir los logros alcanzados en las etapas de la metodología propuesta, en lo que tiene que ver con el diseño del sistema de información [3] realizada como parte del proyecto de investigación denominado "Procedimiento para especificar y validar requisitos de software en MIPYMES desarrolladoras de software, basado en estudios previos en la región". Lo anterior responde a la necesidad de indagar sobre los problemas de calidad del sector del software relacionado con el tratamiento que se les hace a los requisitos como un punto clave para lograr productos y procesos de calidad, con el ánimo de desarrollar propuestas que contribuyan a mejorarla, inicialmente en la industria que se representa desde la ciudad de Pereira. Para ello, se hace un recorrido por algunos trabajos que permiten dar cuenta de los logros que se han alcanzado en investigaciones similares y que tienen como objeto de estudio los requisitos, reconociendo, a partir de unos antecedentes regionales, nacionales e internacionales los orígenes de ese problema y las propuestas que se han desarrollado para ayudar a corregirlo. Finalmente, y como conclusión, se encuentra que han sido muchas las propuestas que de una u otra manera pretenden contribuir al mejoramiento de la calidad del software a partir de los requisitos, pero en ninguna de las revisadas se ha propuesto un método, proceso o lineamiento claramente definido para elaborar una completa, consistente, modificable y trazable, específicamente de requisitos de software en las micro, pequeñas y medianas empresas (MIPYMES. La primera de ellas tuvo como propósito implementar un modelo de ciclo de vida que guiara el desarrollo mediante el uso de prototipos, lo cual fue indispensable para tener una línea clara en la ejecución del proyecto. Si bien no se presenta en este documento el seguimiento de todas las fases, es

importante al menos indicar cuál fue el transcurso metodológico del proyecto de diseño que implementaría el modelo de aseguramiento de la calidad de los requisitos ya elaborado por el grupo de investigación referenciado [28].

Algunos pequeños productores de software llevan a cabo sus proyectos según se lo indican las herramientas que utilizan y acompañados de sus ideas de lo que es mejor para el proyecto; sin embargo, es necesario comprender los principios y el conocimiento que hay alrededor de cada fase de un proyecto de desarrollo de software para, de esta manera, reconocer y apropiar los métodos y las metodologías existentes en cada proyecto, más allá de la herramienta [29], por lo que llegar a este nivel de detalle con los requerimientos puede entregar ideas diferentes a las Mipymes de cómo hacerlo para asegurar mayor éxito en sus procesos y productos.

De igual forma, se presentan las fases propuestas por el modelo de ciclo de vida seleccionado, pero en este documento solo se desarrollan aquellas relacionadas con el diseño del sistema.

Un segundo aspecto relevante tiene que ver con las fuentes de información empleadas para la identificación de requisitos [30], pues el proyecto no se limitó únicamente al punto de vista de los integrantes del equipo, sino que consideró el estado de la técnica, los antecedentes del orden regional, nacional e internacional y, aún más importante, se involucraron los sectores productivo (representado por la industria del software) y académico [5], logrando una sinergia que difícilmente se logra encontrar en el desarrollo de un sistema de software.

Por otro lado, la adaptación de técnicas y buenas prácticas sugeridas por el Procedimiento para especificar y validar requisitos de software (PEVReS) constituyó un insumo importante para facilitar las actividades correspondientes en las fases de requisitos y modelado del proyecto (Fase 4 de la metodología), a partir del cual se elaboraron los modelos conceptuales (estáticos y dinámicos) necesarios para lograr un mejor entendimiento del problema y de la solución para desarrollar.

Teniendo presente que los proyectos, en este caso los de software, deben concluir en un valor de negocio para quienes los desarrollan y para quienes los utilizan [31], abordar este proyecto rigurosamente desde la modelación, también se propone dejar un valor agregado para la comunidad académica y para la industria del software.

Igualmente, se entrega información que describe aspectos del desarrollo y resultados finales del proyecto, pasando por el proceso de educción y especificación de requisitos y la elaboración de prototipos relacionados

con la maquetación de las vistas, y el sistema de información desarrollado junto con algunas de sus principales funcionalidades.

Finalmente, la aplicación web resultante exhibe características particulares que la hacen diferente de otras —cuyo enfoque también es la gestión de requisitos—, puesto que implementa un modelo de RQA producto de varios procesos de investigación, además de varias de las técnicas propuestas por el procedimiento PEVReS para guiar a las Mipymes en el proceso de definición y especificación de requisitos de usuario que cumplan con criterios de calidad, para, finalmente, generar de manera automática un documento SRS ajustado a las necesidades de cada proyecto y empresa. De igual forma, la aplicación simplifica notablemente la redacción de los requisitos al entregar al usuario elementos gramaticales preestablecidos reutilizables y reducir las posibilidades de ambigüedad debidas al lenguaje natural, a la vez que lo guía por una secuencia de fases en las que, además de lograr que el requisito sea completo, correcto y priorizable, pase por un proceso de gestión de la configuración y se establezcan las dependencias, sistemas foráneos y restricciones a las que haya lugar. Lo anterior permite, entonces, identificar estos elementos diferenciadores como rasgos que distinguen esta aplicación de las citadas y que existen para labores similares.

V. CONCLUSIONES

Se presentó en este capítulo el resumen de un conjunto de actividades de ingeniería de software relacionadas con el diseño de un sistema de información, desarrollado por el Grupo de Investigación Entre Ciencia e Ingeniería de la Universidad Católica de Pereira, que implementa un modelo *cuasi*-automatizado para el aseguramiento de la calidad de los requisitos en empresas de software, el cual pretende ser un instrumento útil para las Mipymes desarrolladoras de software a la hora de llevar a cabo los procesos de desarrollo y gestión de requisitos. El concepto de *cuasi*-automatizado hace alusión a su diferenciación de los procesos automatizados en sentido estricto, dada la necesidad de intervención humana en la incorporación de los datos de entrada y la búsqueda de las salidas.

Los resultados del proyecto permiten obtener valiosa información de una experiencia de hacerle ingeniería de software a la ingeniería de software, con el pretexto de buscar asegurar la calidad del proceso redundando en la calidad del producto.

Asimismo, la aplicación de buenas prácticas de ingeniería de software, RQA y SQA, contribuye a adoptar un enfoque sistemático y disciplinado, lo cual materializa el propósito de diferentes metodologías para la gestión

de proyectos que están en mora de aportar herramientas para proyectos muy específicos, como los de desarrollo de software. Esto contribuye a elevar el nivel de calidad del proceso desde las primeras etapas del proyecto y, en consecuencia, la calidad del producto software resultante.

En este caso, la aplicación de una propuesta de aseguramiento de calidad de requerimientos sobre una propuesta de un MCVS, donde ambos se convierten en alternativas para mejorar la calidad del proceso de desarrollo, en una simbiosis interesante de actualizar y sistematizar mejor las metodologías y las propuestas clásicas para hacer software.

Por otro lado, se logra el desarrollo de un sistema de información adecuado a las necesidades de los productores de software de la región, que tiene la capacidad de acompañar los proyectos de software en la gestión de requisitos de calidad, y que desde etapas tempranas del proyecto asegura la elaboración de especificaciones de software acordes con las características de calidad establecidas por los autores y organizaciones internacionales expertas en la disciplina.

Entre los industriales consultados, la gestión de requerimientos constituye, en algunos casos, una tarea dentro del ejercicio de análisis y diseño. Sin embargo, el trabajo elaborado permite concluir que, al darle un papel más protagónico a la gestión de requerimientos, se logra una modelación más acertada y en correspondencia con el problema abordado, lo que, necesariamente, redundará en que las fases siguientes tengan un mayor nivel de atino.

También, resulta muy importante involucrar al cliente/usuario en las etapas de requisitos y modelado del proyecto, ya que de esta manera se logra comprender mejor las necesidades del cliente, su retroalimentación inmediata y la garantía en el cumplimiento de sus expectativas a través del desarrollo de un software que satisface sus necesidades, y que además exhibe atributos de calidad competitivos. Esto no solo se logra por el MCVS elegido, sino también por la plataforma que permite la automatización de los requerimientos y exige la interacción entre los especialistas en la gestión de estos y el usuario.

En la misma línea, se describen los logros alcanzados en cada una de las etapas de la metodología propuesta, haciendo énfasis en el diseño del sistema de información, y los resultados permiten evidenciar el cumplimiento de los objetivos del proyecto. Asimismo, se da cuenta del proceso de gestión de requisitos funcionales y no funcionales con la ayuda del procedimiento PEVReS, de la realización de los modelos conceptuales bajo el estándar UML

haciendo uso de herramientas CASE, y del desarrollo del sistema de información web a partir de prototipos.

Finalmente, es importante mencionar que ya se hizo la validación del modelo, a manera de experimento, y se encuentra en proceso de publicación, razón por la cual no se presenta en este capítulo.

VI. REFERENCIAS

[1] R. S. Pressman, Ingenier*á* del Software. Un Enfoque Pr*á*ctico. México: McGrawHill, 2010.

[2] S. R. Gómez Palomino y E. Moraleda Gil, Aproximación a la ingeniería del software. Madrid: Editorial Centro de Estudios Ramón Areces S. A., 2014.

[3] A. Toro y J. G. Gálvez, "Especificación de requisitos de software: una mirada desde la revisión teórica de antecedentes", Entre Cienc. e Ing., vol. 10, n.º 19, pp. 108-113, 2016.

[4] IEEE Computer Society, Swebok V3.0, n.º V3.0, 2014.

[5] L. Peláez, A. Toro, J. López, y A. Medina, "Caracterización del proceso de desarrollo de software en Colombia: una mirada desde las PYMES productoras", Rev. PÁGINAS, vol. 92, pp. 89-98, 2012.

[6] P. Bourque, R. E. Richard, E. Fairley, and IEEE Computer Society, Guide to the software engineering ody of knowledge (SWEBOK), 2014.

[7] F. Ramírez García y P. Puello Marrugo, "Gestión de Requisitos en la Ingeniería de Software", Ingeniator Rev. Virtual los Programas Ing., vol. 1, n.º 1, pp. 57-65, 2009.

[8] A. Toro Lazo, J. G. Gálvez Botero, y L. E. Peláez-Valencia, "Procedimiento para especificar y validar requisitos de software", en Investigación e Innovación en Ingeniería de Software, 2.ª ed., Medellín: Sello Editorial TdeA, 2018.

[9] P. Bourque, R. Dupuis, A. Abran, J. W. Moore, and L. Tripp, "Guide to the software engineering body of knowledge, Version 3.0", Piscataway, NJ, USA, 2014.

[10] A. Toro Lazo, J. G. Gálvez Botero, y S. V. Hurtado Gil, "Procedimiento para especificar y validar requisitos de software en mipymes desarrolladoras de software de la ciudad de Pereira, basado en estudios previos en la región", Universidad Autónoma de Manizales, 2017.

[11] U. J. Bonaparte, "Proyectos UML Diagramas de clases y

aplicaciones JAVA en NetBeans 6.9.1.", Cátedra Paradigmas de Programación, Univ. Tecnológica Nacional, Argentina, 2012.

[12] L. E. Peláez-Valencia, A. Toro-Lazo, L. D. Ledesma-Mariño, D. E. Rodríguez-Franco, y E. H. Largo-Bueno, "Revisión de herramientas para la gestión de requisitos: un aporte al aseguramiento de la calidad del software", Vent. Inform., n.º 38, Ene. 2018. doi: 10.30554/ventanainform.38.2865.2018

[13] E. M. Méndez Nava, "Modelo de evaluación de metodologías para el desarrollo de software", Universidad Católica Andrés Bello, 2006.

[14] M. Arias Chávez, "La ingeniería de requerimientos y su importancia en el desarrollo de proyectos de software", InterSedes Rev. las Sedes Reg., vol. 6, n.º 10, pp. 1-13, 2011.

[15] IEEE Computer Society, SWEBOK V3.0. Guide to the Software Engineering Body Of Knowledge. NJ, 2014.

[16] L. E. Peláez Valencia, A. Toro Lazo, D. E. Rodríguez Franco, y L. D. Ledesma Mariño, "El aseguramiento de la calidad de los requerimientos en la industria local del software : caso Pereira", en Nuevas tendencias investigativas en computación, informática y educación en ingeniería, 2019, pp. 48-69.

[17] R. Pressman y B. R. Maxim, Ingeniería de Software, un enfoque práctico, 8.ª ed. Nueva York: McGraw-Hill, 2015.

[18] I. Sommerville, Software Engineering, 10th Ed. London: Pea, 2016.

[19] K. Wiegers and J. Beatty, Software Requirements. Washington, 2013.

[20] International Requirements Engineering Board (IREB), International Requirements Engineering Board, 2020.

[21] ISO, "International Organization for Standardization", The International Organization for Standardization, 2020.

[22] A. N. Camacho Zambrano, "Herramienta para el análisis de requerimientos dentro de la pequeña empresa desarrolladora de software en Bogota", Pontificia Universidad Javeriana, 2005.

[23] O. C. Medina, M. M. Marciszack, y M. A. Groppo, "Trazabilidad y validación de requerimientos funcionales de sistemas informáticos mediante

la transformación de modelos conceptuales", ReCIBE. Revista Electrónica de Computación, Informática, Biomédica y Electrónica, vol. 5, n.º 1, 2016 [En línea]. Disponible en: http://recibe.cucei.udg.mx/ojs/index.php/ReCIBE/article/view/53

[24] A. Toro Lazo y J. G. Gálvez Botero, "Procedimiento PEVReS", 2017.

[25] J. G. Gálvez, A. Toro, y S. V. Hurtado, "Procedimiento para especificar y validar requisitos de software en MiPymes desarrolladoras de software, basado en estudios previos de la región", Universidad Autónoma de Manizales, 2017.

[26] J. Quiroz, "El modelo relacional de bases de datos", Boletín de Política Informática, n.º 6, pp. 53-61, 2010.

[27] B. Unhelkar, Software Engineering with UML. Boca Raton, Florida: Taylor & Francis Group, 2018.

[28] L. E. Peláez, "Validación de un modelo propio para el desarrollo y la gestión de proyectos de software", Pereira, 2014.

[29] IEEE, Guide to Enterprise IT Body of Knowledge (EITBOK). 2017.

[30] S. L. Buitrón, B. L. Flores-Ríos, y F. J. Pino, "Elicitación de requisitos no funcionales basada en la gestión de conocimiento de los stakeholders", Ingeniare. Rev. Chil. Ing., vol. 26, n.º 1, pp. 142-156, 2018. doi: 10.4067/S0718-33052018000100142.

[31] Project Management Institute (PMI), Guía de los FUNDAMENTOS PARA LA DIRECCIÓN DE PROYECTOS (Guía del PMBOK) 6.ª ed., Newtown Square, Pennsylvania: PMI, 2017.

Capítulo 3
Formalization of domain problems described in the cause-and-effect diagram in the context of the software development process

Fabio Alberto Vargas Agudelo - Fvargas@tdea.edu.co
Tecnológico de Antioquia
Carlos Mario Zapata Jaramillo - Cmzapata@unal.edu.co
Universidad Nacional de Colombia, Medellín

I. INTRODUCTION

The system analyst is responsible—in the early requirements elicitation process—for identifying the problems to be solved with a software application. Such identification is based on: (i) the analyst experience and knowledge about the domain area; and (ii) the stakeholder support. Problems are collected by the system analyst during the domain analysis for using them as the main input of the requirements specification. Such specification is consistent with the problems identified and the stakeholder needs.

The system analyst detects and describes problems from the domain discourse recognition and eventually presents them in a formal way by using a representation diagram. Some methods of the software development process, goal-oriented software engineering (GORE) [1], and organizational analysis have diagrams for representing problems. For instance, UNC-Method [2] includes the cause-and-effect diagram for specifying problems and relating them to the system and the stakeholder goals; logical framework [3] has objective trees and problem trees for linking goals and problems during the project formulation and the decision-making process; Business Modeling with UML [4] has a goal schema for graphically linking goals and problems; NFR Framework [5] has problem frameworks for specifying problems in the non-functional requirements elicitation process.System analysts specify the problems in diagrams as textual descriptions, based on their experience and knowledge. However,

no formal process is driven for guaranteeing the problems described are actual problems, and such problems are clearly understood by the stakeholders. The aforementioned facts lead to some gaps still remaining in the early software requirements elicitation process: relation among early and late requirements is poor [6]; requirements and business goal are commonly unrelated to each other [7]; information systems use to misrepresent the requirements captured from the business model [8]; functionality expected from business processes is unrelated to the software system functionality [7]; traceability among expectations, needs, and their representation in a goal diagram is poor [9]; business goals are misused for assuring the requirements specification completeness and sufficiency [10] [11]; an initial company model is insufficient for getting relevant information based on the context of the company [6]; problems are poorly described, making difficult to link and trace them [12]; problems identified during requirements elicitation are described in a positive way [12]; no formal methods are established for goal and problem definition [13]; problem definition is a hand-made process, since analysts commonly draw up goal and problem diagrams in a subjective way [14]; analyst experience and stakeholder knowledge are useful for determining goals, but no validation against problems is included in such a process [15].

In this chapter, we propose a method to formalize the problems of the cause-and-effect diagram drawn up by the analyst in the domain context where the software application will be used. We define a set of syntactic rules for specifying problems, and we allow analysts and stakeholdersto understand the problems better. Improved traceability and consistency related to the business goals, the system goals, and the requirements can be achieved by using the formalized problem schemas.

This chapteris organized as follows: in Section II, we present the theoretical framework with some definitions; in Section III, we present some methods for specifying problems in the requirements elicitation process and the organizational analysis; in Section VI, we propose a method to formalize the problems of the cause-and-effect diagram during the software development process; in Section V, we provide an example for applying our proposal into a lab study; finally, in Section VI, we discuss some conclusions and future work.

II. THE ORETICAL FRAME-WORK

Goal-Oriented Requirement Engineering (GORE) is an approach for promoting the use of goals as the

basis of the software requirements elicitation. GORE includes a point of view related to the purpose of the system—intentional in nature. The introduction of an intentional point of view allows stakeholders to expresstheir needsmore naturally, focusing on what they want—their goals—*versus* the way to achieve them—conventional requirements. Requirements can be derived from goals [16]. Some approaches to GORE differ in two main factors: the focused requirements engineering activity—*e.g.*, requirements elicitation, modeling, and analysis—and the supported degree of formalism. For example, KAOS [17] is focused on formal requirements modeling. NFR Framework [5] is also focused on modeling but targeted on non-functional requirements with a less formal approach. I* is a methodology based on the NFR Framework [18] focused on the initial phases of the requirements elicitation—particularly, the business modeling. GBRAM [19] is intended to integrate scenarios into the context of goal modeling.

Cause-and-effect diagram is used by organizations during the requirements elicitation process in order to think about the actual causes of the potential problems and then to establish corrective actions. In this context, the cause-and-effect diagram can be used to guide the analysis and reflection about the problem understanding, the identification of root causes and possible solutions, and the decision- making process.

Problems are specific issues requiring a solution in a specific domain. Perales [20] defines a problem as any planned/spontaneous situation producing, on the one hand, a certain degree of uncertainty and, on the other, a behavior aimed at finding a solution.

Formalization is a set of rules, expressions, and meanings intended to characterize a language, allowing for a step-by-step interpretation of a particular process [21].

Pre-conceptual schema is a conceptual-graph-like knowledge representation for requirements elicitation (see the main elements in Fig. 1). Such a representation can be obtained from controlled natural language discourses, and it can be then converted into standard UML diagrams. Pre-conceptual schemas are intermediate models for obtaining UML diagrams from natural language discourses [22].

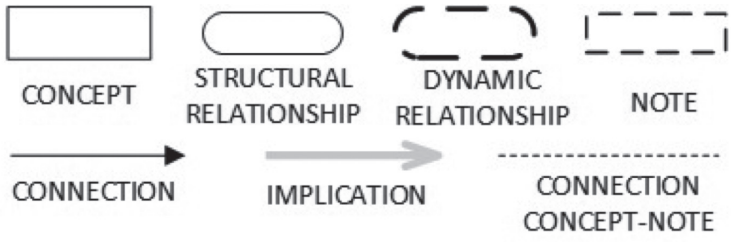


Fig. 1. Syntax of the pre-conceptual schemas. Source [22]

III. BACKGROUND

The UNC-Method [2] includes—in the software requirements elicitation process—the cause-and-effect diagram as a tool for identifying the domain problems and their relation to both organizational and system goals. Such identification is hand-made by the analyst and the stakeholder for specifying problems and creating a textual representation of the cause-and-effect diagram. Traceability and consistency among the organizational goals and the functional requirements of the software application are difficult to achieve at this level.

Business Modeling with UML [4] has a goal/problem diagram to relate the organizational goals to the problems identified in the domain. Problems are intended to be obstacles to the achievement of goals. Problem specification is manually written by using textual descriptions, and relationships of problems and the different actors involved in the process are difficult to achieve. Also, syntactic and semantic structures to formalize the problems are underspecified.

In the NFR Framework [5], problems are specified by using problem frameworks in the non-functional requirements elicitation process. Informal representation of the problem domain can lead to traceability and consistency problems when reviewed against the original goals.

Zapata, Acevedo, and Moreno [13] make a representation of the semantic relations among problems and goals by using predicate logic. The authors propose formal methods as a way to plan goals and domain problems.

Lin and Zhi [15] establish I* techniques for combining goal analysis and problem analysis. They argue goal decomposition requires domain knowledge and, consequently, problems can be used for generating requirements. However, no structures are used for specifying either goals or problems.

Vargas [12] defines grammar rules for specifying goals and problems,

and he applies them to the cause-and-effect diagram. He also establishes an approximation to a consistency relation based on such rules, but he uses syntactic structures with explicit common words among goals and problems. This fact leads to some drawbacks since words used to express problems are different from those used to express related goals.

In terms of organizational analysis, Sanchez [3] argues the logical framework method includes an objective tree—future outcomes you want to accomplish—and a problem tree—what you want to solve. The problem tree is generated from the objective tree by using some rules. Basically, problems are expressed as negative states related to the objectives. When the central objective of the process is expressed in a negative form, you can have the central problem and so on. Even though the rules are defined, the whole process should be manually carried out by the organizational analyst.

IV. RESULTS

During the early software requirements elicitation, the specification of the problems to be solved by the software application is crucial. The analyst and the stakeholder are directly involved in this task, and they complete it based on their experience and organizational knowledge. In some methods for eliciting requirements, a diagram is often used for drawing the problems. The formal method we propose in this section for expressing the problems in the cause-and-effect diagram is based on a set of syntactic and semantic rules and pre-conceptual schemas for graphically representing context information. We use such rules and schemas due to their proximity to the natural language of the stakeholder and the technical language of the analyst. The steps of the formal method for expressing problems in the cause-and-effect diagram are the following:

Step 1. Recognizing the syntactic-semantic structure for specifying problems. In Fig. 2 we propose a preconceptual-schema-based structure for problems. An organizational analyst should be familiarized with this structure after the first step.

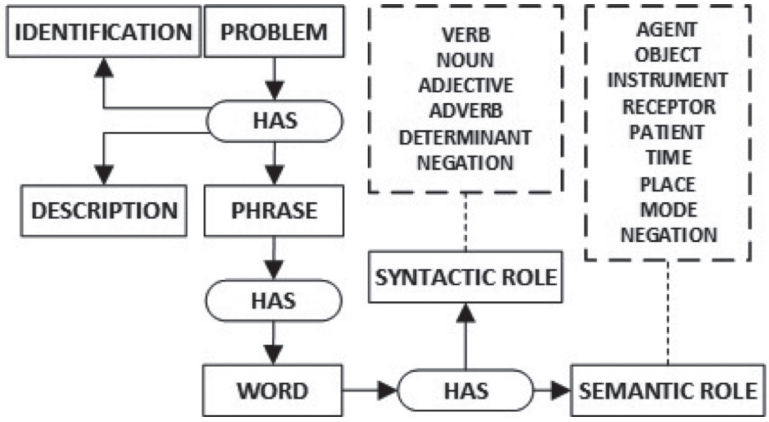


Fig. 2. Problem structure based on pre-conceptual schemas. Source: the authors

Step 2. Using a set of syntactic-semantic rules for formalizing problems. We define three rules to be formally used for specifying the problems described in the cause-and-effect diagram; we allow the analyst to specify problems in a clear and easy way to be understood by the stakeholders. By using the rules, relationships among problems, and organizational and system goals are guaranteed by traceability and consistency. In the rules, the syntactic structures of the problems are adapted from Vargas [12], while the graphical schemas are proposed in this paper. We introduce the red slash as a new symbol for denoting the negation of an element of the pre-conceptual schemas. Red slash is supposed to be put over the element we want to negate. The abbreviations used in the rules are: S=Sentence, V=Verb, Ad=Adjective, NP= Noun Phrase, Adv=Adverb, N=Noun.

Rule No. 1. Enunciating the problem by using a negative connotation adverb. The structures are shown in Tables 1 and 2.

Rule No. 2. Enunciating the problem by using a negative connotation adjective. The structures are shown in Table 3.

Rule No. 3. Enunciating the problem by using a negative connotation noun. The structures are shown in Table 4.

V. LABORATORY EXAMPLE

Zapata and Arango [2] present a cause-and-effect diagram related to the selling process of a company (See Fig. 3).

Table 1. Negative-connotation adverb type 1. Source: the authors

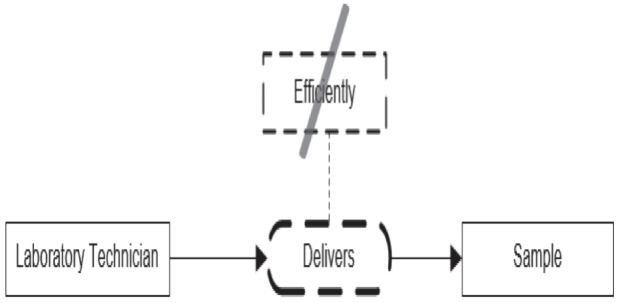
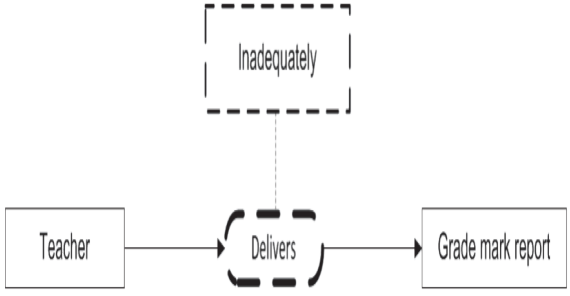
Description	Restrictions	Example
S=NP1+Adv1+Adv2+V+NP2	Adv1=Not; V = Action verb; Adv2 = positive connotation adverb	The laboratory technician does not efficiently deliver samples
Graphical schema for formalizing problems		
		
S=NP1+Adv+V+NP2	Adv = negative connotation	The Teacher inadequately delivers the grade mark report
Graphical schema for formalizing problems		
		

Table 2. N egative-connotation adverb type 2. Source: the authors

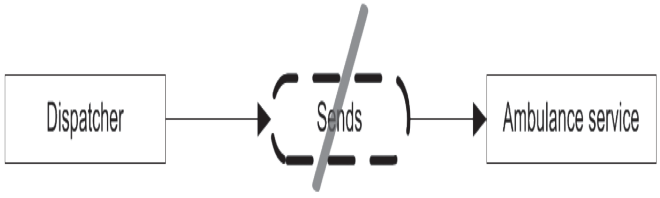
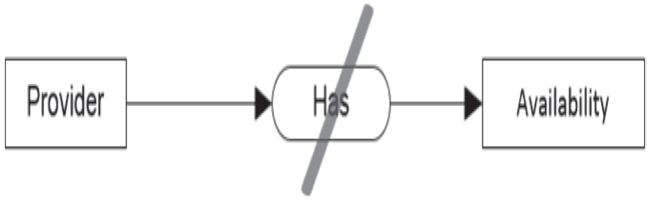
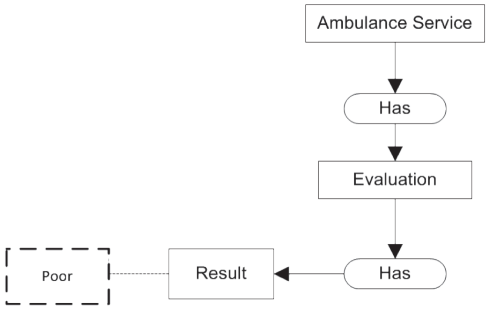
Description	Restrictions	Example
S=NP1+Ad-v+V+NP2	Adv1=Not; V = action verb	The dispatcher does not send the ambulance service
Graphical schema for formalizing problems		
		
S=NP1+Ad-v+V+NP2	Adv = not; V= verb to have	The provider does not have availability
Graphical schema for formalizing problems		
		

Table 3. Negative-connotation adjective type 1. Source: the authors

Description	Restrictions	Example
S=NP+V+Ad	Ad= negative connotation	The Ambulance service is poor
Graphical schema for formalizing problems		
		

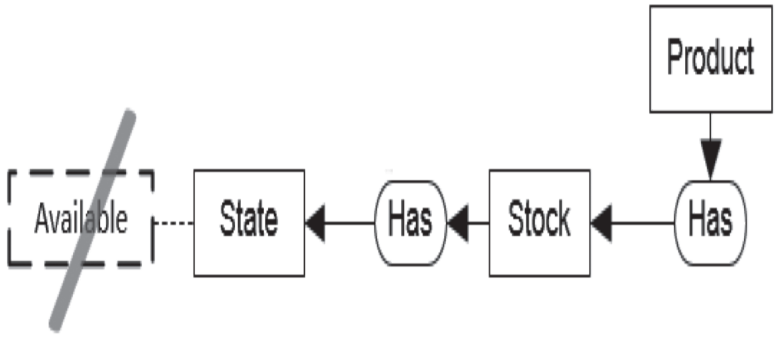
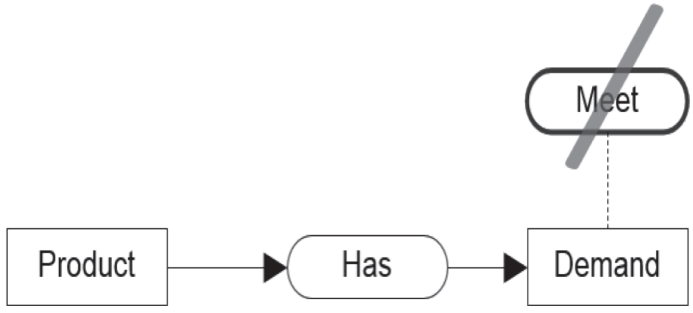
Description	Restrictions	Example
S=NP+V+Adv+Ad	Ad = positive connotation Adv = Not V= verb to be	The Stock product is not available
Graphical schema for formalizing problems		
		
S=NP+V1+Adv+V2	V1= verb to be Adv = not V2= achievement verb	The product demand is not met
Graphical schema for formalizing problems		
		

Table 4. Negative-connotation noun type 1. Source: the authors

Description	Restrictions	Example
S=NP1+V+NP2	NP2= negative connotation	Document management has delays
Graphical schema for formalizing problems		
S=N-P1+V+NP2	NP2 = positive connotation; V= verb to have in negative form	The Project has no papers
Graphical schema for formalizing problems		

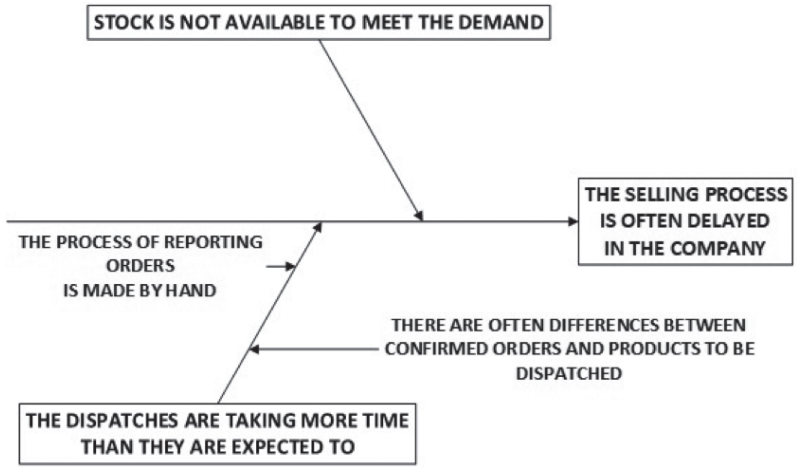


Fig. 3. Cause-and-effect diagram related to the selling process of a company. Source [2]

Problems described in the diagram (see Fig. 3.) are ambiguously generated since both their structure and specification are unclear. In Table 5, we summarize some of the detected ambiguities.

According to our proposed method, we need to re-write and formalize each problem by using the rules defined in this paper. In Table 6, we present the problems formalized. The modified cause-and-effect diagram is presented in Fig. 4.

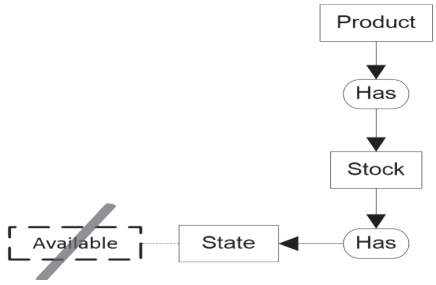
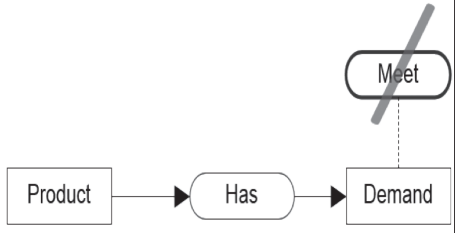
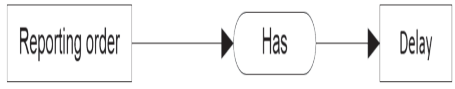
Table 5. Ambiguities detected in the cause-and-effect diagram. Source: the authors

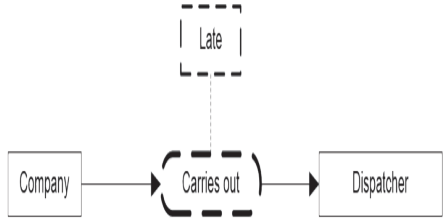
Problems	Ambiguity detected
Stock is not available to meet the demand	The problem specification does not refer to a single problem, but rather many problems can be deduced.
The selling process is often delayed in the company	The problem specifi cation is clear, but the syntactic structure of the problem can be improved
The dispatches are taking more time than they are expected to	The speci fication is not conducive to the clear statement of the problem since the sentence does not contain a negative connotation
The process of reporting orders is made by hand	The specifi cation is not conducive to the clear statement of the problem, since the sentence does not contain a negative connotation.
There are often differences between confi rmed orders and products to be dispatched.	The specifi cation does not defi ne the problem clearly. Again, no negative connotation is detected.

VI. CONCLUSIONS AND FUTURE WORK

Many of the problems detected and further specified in software development processes are exposed in an unclear way, and in some cases, they avoid the expression of a problem by themselves. This fact implies misinterpretations made by analysts and stakeholders, causing reprocesses and irrelevant software requirements discovery to an organization.

Table 6. Problems formalized. Source: the authors

Problems	Problems formalized
Stock is not available to meet the demand	The Stock product is not available (Rule No 2)  The Product demand is not met (Rule nro2) 
The process of reporting orders is made by hand	The reporting order has delays (Rule No 3) 
There are often differences between confi rmed orders and products to be dispatched	The Stock has no products (Rule No 3)
The selling process is often delayed in the company.	The company selling process has delays (Rule No 3)

The dispatches are taking more time than they are expected to do.	The company carries out the dispatches late (Rule No 1) 

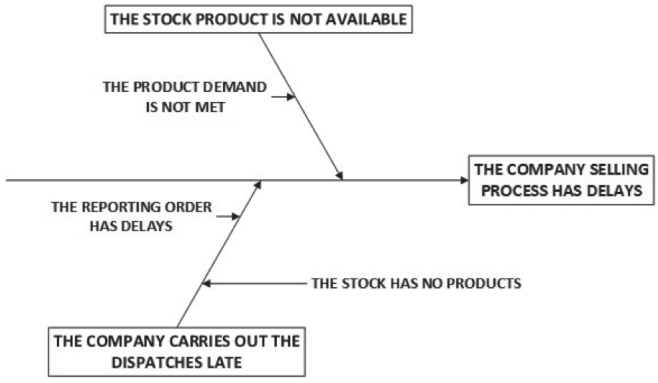


Fig. 4. Final cause-and-effect diagram. Source: the authors

The formalization of the problems allows for generating traceability and consistency during the initial software requirements elicitation, leading to a less ambiguous representation of problems, so thestakeholders and analysts can understand the problems more easily .

In this chapter we proposed structures for formalizing problems in order to support system analysts in the early software requirements elicitation process. We expect to generate a higher degree of reliability in the software solutions raised. In fact, problems should be aligned to the organizational context, and they should be relevant for understanding the stakeholder needs.

As future work, we propose the extension to the ruleset used for formalizing problems as a way to generate traceability with different actors in processes of organizational and requirements analysis. We also propose the automation of the process with the generation of ontologies and syntactic and semantic relationships with the organization goals for validating the problems detected.

VII. REFERENCES

[1] P. P. Negri, V. E. S. Souza, A. L. de Castro Leal, R. de Almeida Falbo, G. Guizzardi. "Towards an Ontology of Goal-Oriented Requirements". *In CIbSE*. 2017. pp. 469-482.

[2] C. M. Zapata, F. Arango. "The UNC-Method: a problem based software development method". *Ingeniería e Investigación*. Vol. 29(1). 2009. pp. 69-75.

[3] N. Sánchez. "El marco lógico. Metodología para la planificación, seguimiento, y evaluación de proyectos". *Revista Visión Gerencial*. Vol. 6(2). 2006. pp. 328-343.

[4] H. Eriksson, M. Penker. "Business modeling with UML: Business patterns at work". *OMG Press*, 2000.

[5] M. Marinho, D. Arruda, F. Wanderley, & A. Lins. "A systematic approach of dataset definition for a supervised machine learning using NFR framework". In 2018 11th International Conference on the Quality of Information and Communications Technology (QUATIC). 2018. pp. 110-118.

[6] A. Martínez, O. Pastor, J. Mylopoulos, P. Giorgini. "From Early Requirements to Late Requirements: A goal-based approach". *Eight International Bi-Conference Workshop on Agent-Oriented Information System (AOIS)*. 2006. pp. 1-12.

[7] D. F. Mendonça, G. N. Rodrigues, R. Ali, V. Alves & L. Baresi. "GODA: A goal-oriented requirements engineering framework for runtime dependability analysis". *Information and Software Technology*. Vol. 80. 2016. Pp. 245-264.

[8] H. Estrada, A. Martínez, O. Pastor, J. Sánchez. "Generación de especificaciones de requisitos de software a partir de modelos de negocios: un enfoque basado en metas". *V Workshop de engenharia de requisites*. 2002. pp. 179-193.

[9] J. Jureta, S. Faulkner, P-Y. Schobbens. "Clear Justification of Modeling Decisions for Goal-Oriented Requirements Engineering". *Requirements Engineering*. Vol. 13(2). 2008. pp. 87-115.

[10] C. Banerjee, A. Banerjee & S. K. Sharma. "Estimating influence of threat using Misuse Case Oriented Quality Requirements (MCOQR) metrics: Security requirements engineering perspective". International Journal of Hybrid

Intelligent Systems. Vol. 14(1-2). 2017. pp. 1-11.

[11] U. Zafar, M. Bhuiyan, P. W. C. Prasad & F. Haque. "Integration of Use Case Models and BPMN Using Goal-Oriented Requirements Engineering". *JCP*. Vol. 13(2). 2018. Pp. 212-221.

[12] F. Vargas. "Método para establecer la consistencia de los problemas en el diagrama causa-efecto con el diagrama de objetivos de KAOS". *Tesis de maestría*. Universidad Nacional de Colombia. 2010.

[13] C. Zapata, J. Acevedo, D. Moreno. "Representación de relaciones semánticas entre problemas y objetivos mediante lógica de predicados". *Revista EIA*. 2011. Vol. 8(15). pp. 61-72.

[14] C. Zapata, L. Lezcano. "Caracterización de los verbos usados en el diagrama de objetivos". *Dyna.* Vol. 76(158). 2009. pp. 219-228.

[15] C. Menghi, P. Spoletini & C. Ghezzi. "Integrating goal model analysis with iterative design". *In International Working Conference on Requirements Engineering: Foundation for Software Quality*. pp. 112-128. 2018.

[16] L. Olsina & P. Becker. "Linking Business and Information Need Goals with Functional and Non-functional Requirements". In CIbSE. 2018. pp. 381-394.

[17] A. Dardenne, V. Van Lamsweerde, S. Fickas. "Goal-directed requirements acquisition". *Science of computer programming*. Vol. 20(1). 1993. pp. 3-50.

[18] E. Yu. "Modelling Strategic Relationships for Process Reengineering". *Ph.D. Thesis*. University of Toronto. 1995.

[19] A. I. Antón, C. Potts. "The use of goals to surface requirements for evolving systems". *20th IEEE International Conference on Software Engineering*. 1998. pp. 157-166.

[20] F. Vargas Agudelo. "Modelo para la especificación de requisitos iniciales de software a partir de la relación sintáctica y semántica entre objetivos y problemas". *Tesis de doctorado*. Universidad Nacional de Colombia. 2016.

[21] Y. Sermet & I. Demir. "Towards an information centric flood ontology for information management and communication". *Earth Science Informatics"*. Vol. *12*(4). 2019. Pp. 541-551.

[22] C. M. Zapata, A. Gelbukh, F. A. Arango. "Pre-conceptual schema: A conceptual-graph-like knowledge representation for requirements elicitation". *5th Mexican International Conference on Artificial Intelligence*. 2006. pp. 27-37.

[23] B. Manrique-Losada, C. M. Zapata-Jaramillo & D. A. Burgos. "Re-expressing business processes information from corporate documents into controlled language". *In International Conference on Applications of Natural Language to Information Systems*. 2016. pp. 376-383.

Capítulo 4
Caracterización Geomorfológica de la Cuenca del Río Negro Utilizando Sistemas de Información Geográfica (SIG)

Mauricio EscobarTobón - mauricio_escobar54181@elpoli.edu.co
Juan Carlos Valdés Quintero - jcvaldes@elpoli.edu.co
Laura María Marín Arroyave - aura_marin95162@elpoli.edu.co
Politecnico Colombiano Jaime Isaza Cadavid

I. INTRODUCCIÓN

La cuenca del río Negro la conforman 219 veredas que se encuentran repartidas en diez municipios de la región del Oriente antioqueño: Rionegro, Marinilla, Santuario, San Vicente, El Retiro, El Peñol, Envigado, El Carmen de Viboral, Guarne y La Ceja del Tambo, como se observa en la Figura 1. [1].

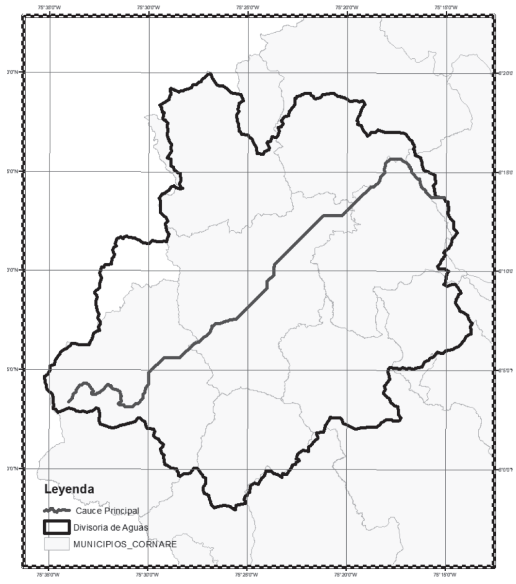


Fig. 1. Cuenca del río Negro con su cauce principal.

Esta cuenca es por definición la porción de terreno que drena al punto de cierre en la entrada del río Negro al embalse Peñol-Guatapé, a una altura sobre el nivel del mar de 1.900 m; con su inicio en la parte alta del cerro Pantanillo ubicado en el municipio de El Retiro, Antioquia, a una altitud aproximada de 3.100 metros sobre el nivel del mar (m s.n.m) [2].

Analizar la geomorfología de una cuenca hidrográfica permite conocer con anterioridad su comportamiento hidrológico, hidráulico e hidrodinámico, los cuales tienen relación directa con los fenómenos de inundación, avenidas torrenciales y deslizamientos. Lo que indica que el entendimiento de estas variables aporta al conocimiento de la gestión del riesgo de la cuenca.

Es aquí donde se hace interesante el estudio, ya que la cuenca del río Negro es una de las zonas con mayor presión demográfica en el Oriente antioqueño, con una población aproximada, de acuerdo con la proyección 2020 del DANE, de 398.734 [3].

El Capítulo de Libro se estructura de la siguiente forma: En la sesión II Materiales y Métodos; en la sesión III Resultados; en la sesión IV Discusión; y finalmente en la sesión V Conclusiones.

II. MATERIALES Y MÉTODOS

A. Materiales

Comportamiento histórico de la cuenca, visitas de campo, histórico de fenómenos ocurridos, Software ArcGis y Microsoft Excel.

B. Métodos

En la delimitación de la cuenca, se tomaron las curvas de nivel del Modelo de Elevación Digital (DEM) encontradas en estudios realizados por Cornare para la cuenca del río Negro, luego se ejecutaron algunos ajustes, como la corrección en los vacíos de información resultantes en el DEM [4].

Con estos datos y el uso del software ArcGis 10.6.1, se hallaron los parámetros principales para la caracterización de la cuenca, como se explica a continuación:

Con la Herramienta *Fill*, se realizan correcciones de los vacíos y picos (celdas con elevación mayor a la que se espera dada la tendencia de la superficie circundante). Lo cual puede ser visualizado en las Figuras 2 y 3.

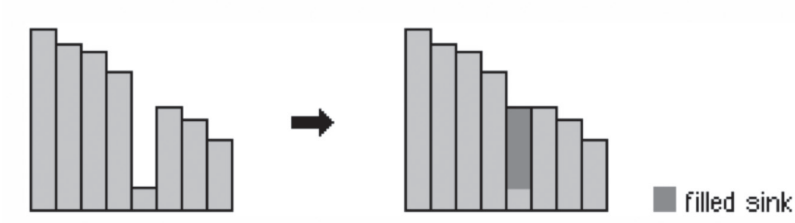


Fig. 2. Vista de perfil de un sumidero antes y después de ejecutar la herramienta Fill.

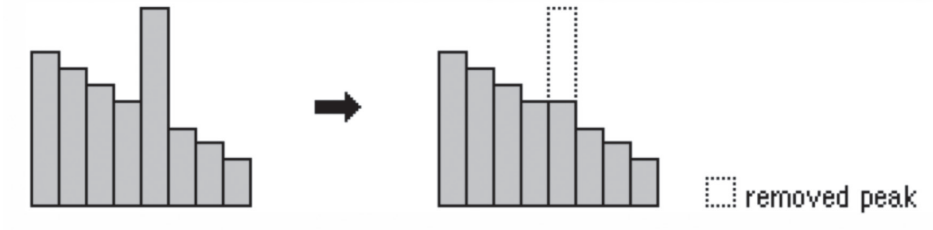


Fig. 3. Vista de perfil de un pico antes y después de ejecutar la herramienta Fill.

Con la herramienta *Flow Direction*, se encuentran los flujos de dirección del sistema de drenaje de la cuenca. Esta función utiliza una superficie como entrada y proporciona como salida un ráster que permite visualizar la dirección del flujo que sale de cada celda. Luego, se utiliza la herramienta *Flow Accumulation,* la cual calcula el flujo acumulado como el peso acumulado de todas las celdas que fluyen en cada celda de pendiente descendente en el ráster de salida [5]. Teniendo esto, se hace necesario definir el punto de entrega de las aguas recogidas por la cuenca; para el caso de estudio, dicho punto está representado por el lugar donde el río Negro deposita sus aguas al embalse El Peñol, para lo cual se utiliza la herramienta que proporciona el software para la creación de un nuevo *Shapefile* y se realiza la respectiva interpolación del Sumidero [1].

A partir de las direcciones de flujo obtenidas y del punto sumidero creado, se proyecta la divisoria de aguas de la cuenca. Se debe tener en cuenta que dicha divisoria no interfiera con intersecciones de otro sistema ya que podrían presentarse errores en los cálculos [5]. Para efectos de cálculos de área, perímetro y otras variables importantes, se hace la conversión de la cuenca de dato ráster a dato vectorial, como se observa en la Figura 4.

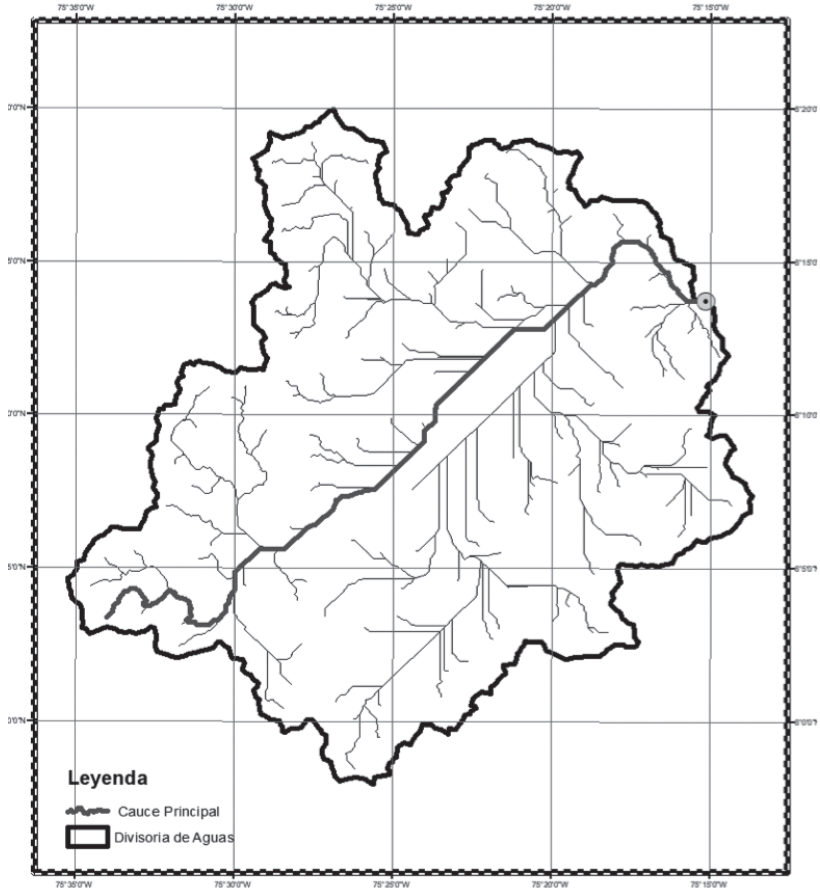


Fig. 4. Sistema de drenajes de la cuenca del río Negro.

Para el cálculo de las pendientes, se usa la herramienta *Slope,* la cual calcula la tasa máxima de cambio del valor de cada celda con sus respectivas celdas vecinas [6] (ver Figura 5), las cuales, con ayuda de la herramienta *Reclassify*, se redistribuyen en nueve categorías y con los datos arrojados y el uso de Excel, se construye la curva hipsométrica de la cuenca y el histograma de pendientes.

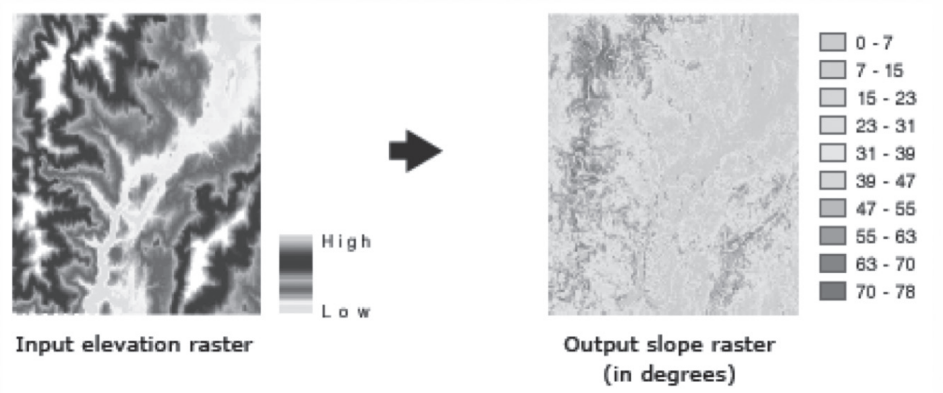


Fig. 5. Cálculo de pendientes utilizando la herramienta Slope.

III. RESULTADOS

Con los gráficos y tablas obtenidos del procesamiento de datos con ArcGis, se procede a calcular los parámetros necesarios para la caracterización geomorfológica de la cuenca. Estos parámetros son:

Área de la cuenca (A): Es la porción de terreno al interior de la divisoria de aguas, donde todas las aguas de escorrentía van al mismo punto de desagüe o punto sumidero. Se calcula a través de la sumatoria de las áreas comprendidas entre las curvas de nivel y los límites de la cuenca.

$$A = 928{,}150534 \text{ Km}^2$$

Perímetro de la cuenca (P): Es la longitud total de los límites de la cuenca o divisoria de aguas.

$$P = 173{,}574782 \text{ Km.}$$

Longitud del cauce principal (Lc): Es la longitud que recorre el cauce principal desde su nacimiento hasta el punto sumidero.

$$Lc = 53{,}340487 \text{ Km}$$

Pendiente del cauce principal (PLc):

$$PLc\ (\%) = ((\text{Cota Máx} - \text{Cota Min}) / Lc*1000))*100$$
$$PLc = ((3100-1900) / (53{,}340487*1000))*100$$
$$PLc = 2{,}249698245\ \%$$

Longitud de la cuenca (L): Para calcular esta distancia, debe proyectarse el cauce principal de forma perpendicular a las curvas de nivel hasta que se intercepte con la divisoria de aguas. La longitud de la cuenca es entonces la distancia desde el intercepto del cauce principal con la divisoria, siguiendo el recorrido del mismo, hasta el punto sumidero [7].

$$L = 56{,}5238 \text{ Km}$$

Ancho medio de la cuenca (W): Es la relación que existe entre el área de la cuenca y la longitud del cauce principal.

$$W = A / Lc$$
$$W = 928{,}150534\ Km^2 / 53{,}340487\ Km$$
$$W = 17{,}40048856\ Km$$

Índice de compacidad o Coeficiente de **Gravelius (Kc):** Es la relación entre el perímetro de la cuenca y el perímetro de una circunferencia con área igual a la de la cuenca (ver Tabla 1) [6]. Luego de algunos procedimientos matemáticos simples, se llega a la siguiente fórmula:

$$Kc = P / 2\sqrt{\pi A}$$
$$Kc = 173{,}574782\ Km / 2*\sqrt{\pi}*928{,}150534\ Km2$$
$$Kc = 1{,}067209486$$

Tabla 1. Rangos aproximados del Índice de compacidad

Rango	Descripción
1 < Kc < 1,25	Forma casi redonda a oval-redonda
1,25 < Kc < 1,5	Forma oval-oblonga
Kc > 1,5	Forma oval-oblonga a rectangular-oblonga

Factor de forma: Es la relación entre el ancho medio de la cuenca y la longitud del cauce principal elevado al cuadrado (ver Tabla 2) [8], [9].

$$Kf = A / Lc^2$$
$$Kf = 928{,}150534\ Km^2 / (53{,}340487\ Km)^2$$
$$Kf = 0{,}326215404$$

Tabla 2. Rangos aproximados del Factor de forma (Kf)

Kf (valores aproximados)	Forma de la cuenca
< 0,22	Muy alargada
0,22 a 0,3	Alargada
0,3 a 0,37	Ligeramente alargada
0,37 a 0,45	Ni alargada ni ensanchada
0,45 a 0,60	Ligeramente ensanchada
0,60 a 0,80	Ensanchada
0,80 a 1,20	Muy ensanchada
> 1,20	Rodeando el desagüe

Densidad de drenaje (Dd): Es la relación entre la sumatoria de las longitudes de todos los cauces de la cuenca y el área de esta (ver Tabla 3) [10].

$$Dd = \Sigma L / A$$
$$Dd = 528{,}676963\ km / 928{,}150534\ km^2$$
$$Dd = 0{,}569602606\ km/km^2$$

Tabla 3. Valores descriptivos de la Densidad de drenaje.

Densidad de drenaje (km/km^2)	Categoría
< 1	Baja
1 a 2	Moderada
2 a 3	Alta
> 3	Muy alta

Relación de relieve (Rr): Es la relación entre la longitud de la cuenca

y la diferencia de altura entre el punto sumidero y el punto más alto de la divisoria de aguas. El relieve de la zona influye directamente en la respuesta hidrográfica de la cuenca, ya que entre más escarpado sea el terreno, mayor será la velocidad del flujo y menores los tiempos de su concentración.

Cota Máxima = 3100 m = 3,1 km
Cota Mínima = 1900 m = 1,9 Km
Rr = h / L
Rr = (3,1 Km - 1,9 Km) / 56,5238 Km
Rr = 0.0212299952

Pendiente media de la cuenca (Pm): Es el valor que controla la velocidad con la que se dará la escorrentía superficial en la cuenca. Se obtiene a partir de la herramienta *Slope* de Arcgis. Este dato tiene una relación directa con la relación de relieve (ver Tabla 4) [8].

Pm = 12,88677%

Tabla 4. Valores descriptivos de la pendiente media de la cuenca

Pendiente media (%)	Tipo de relieve
0 a 2	Llano
2 a 5	Suave
5 a 10	Accidentado medio
10 a 15	Accidentado
15 a 25	Fuertemente accidentado
25 a 50	Escarpado
> 50	Muy escarpado

Curva hipsométrica: Es una representación gráfica del promedio de relieve de la cuenca. En el eje de las abscisas se ubica el porcentaje de área acumulada entre curvas, y en el eje de las ordenadas se ubican las cotas de las curvas de nivel consideradas (ver Tabla 5 y Figura 6).

Tabla 5. Datos para la construcción de la curva hipsométrica

Tabla Curva Hipsométrica							
Intervalos	COTAS			Área Km^2			
	Min	Máx	Promedio	Área entre curvas	Acumulado	% Acumulado	% Área entre curvas
1	1900	2064.0293	1982.01	18926798.62	928119968.45	100	2.03926208
2	2064.102295	2167.1853	2115.64	417320447.52	909193169.83	97.96073792	44.964063
3	2167.188232	2256.24585	2211.72	173345240.97	491872722.31	52.99667489	18.6770296
4	2256.259521	2359.36523	2307.81	129145896.48	318527481.33	34.31964532	13.9147848
5	2359.383545	2457.80957	2408.60	79506438.48	189381584.85	20.4048605	8.56639671
6	2457.818115	2551.51929	2504.67	57768073.08	109875146.37	11.83846379	6.22420323
7	2551.57251	2663.91943	2607.75	31637226.24	52107073.29	5.614260555	3.40874319
8	2664.117188	2804.66797	2734.39	14117788.77	20469847.06	2.205517363	1.5211168
9	2804.688965	3100	2952.34	6352058.29	6352058.29	0.684400563	0.68440056
Total				928119968.45			

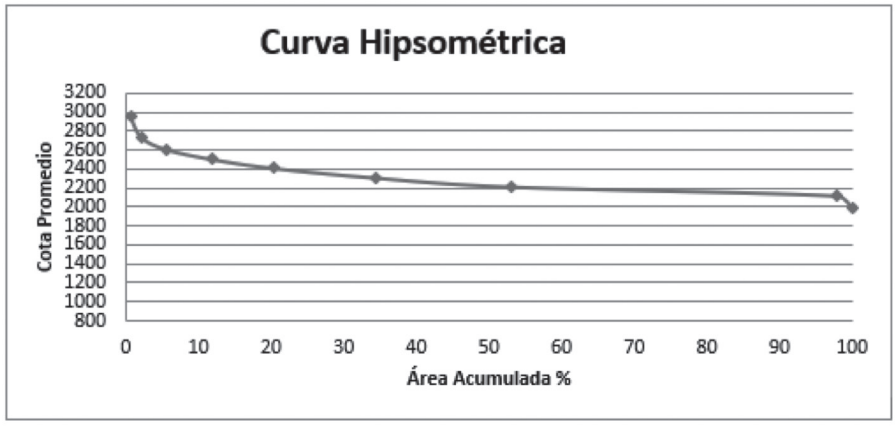


Fig. 6. Curva hipsométrica de la cuenca del río Negro.

Histograma de pendientes: A partir de los datos con los que se realizó la curva hipsométrica, y propiamente de su gráfica, se adiciona una serie cuyo valor en las abscisas es el porcentaje de las áreas entre las curvas de nivel. Dicha serie se cambia de gráfico por uno de barras horizontales, y así se obtiene el histograma de pendientes de la cuenca (ver Figura 7).

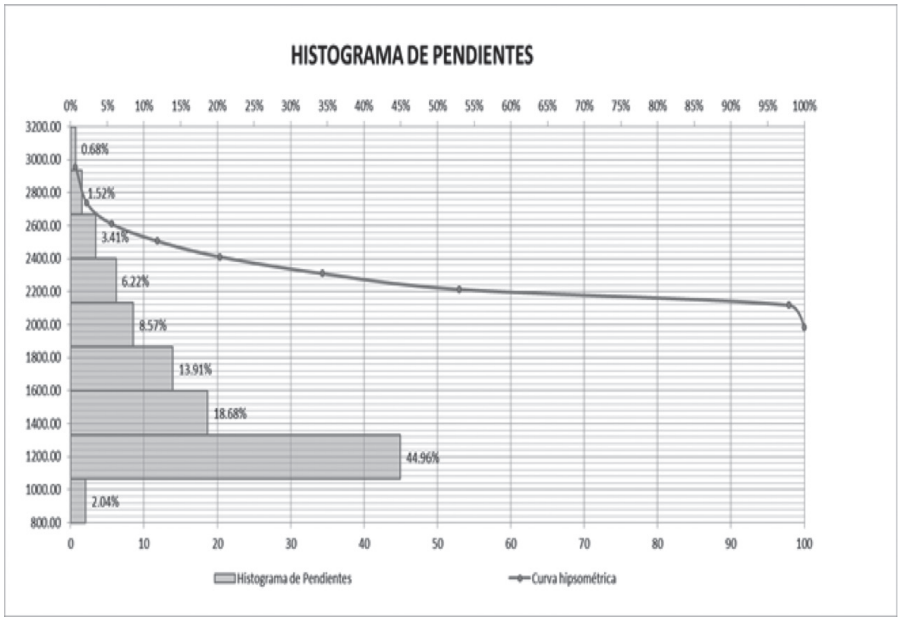


Fig. 7. Histograma de pendientes de la cuenca del río Negro.

Índice asimétrico: Es la relación del área de las vertientes de la cuenca, las cuales estarán separadas por el cauce principal. Este índice es de gran importancia cuando se desea evaluar la homogeneidad de la red de drenaje y la probabilidad de la ocurrencia de crecidas hacia un lado u otro del cauce principal.

Si el índice asimétrico es mayor que 1, indica que el cauce principal está bastante recargado hacia una de las vertientes; si es igual a 1, habrá una distribución uniforme del cauce principal (ver Figura 8).

Área Mayor (Ama): A la derecha del cauce principal

Área Menor (Ame): A la izquierda del cauce principal

$$Ias = Ama / Ame$$

$$Ias = 504{,}233247\ km^2 / 423{,}17288\ km^2$$

$$Ias = 1{,}189461391$$

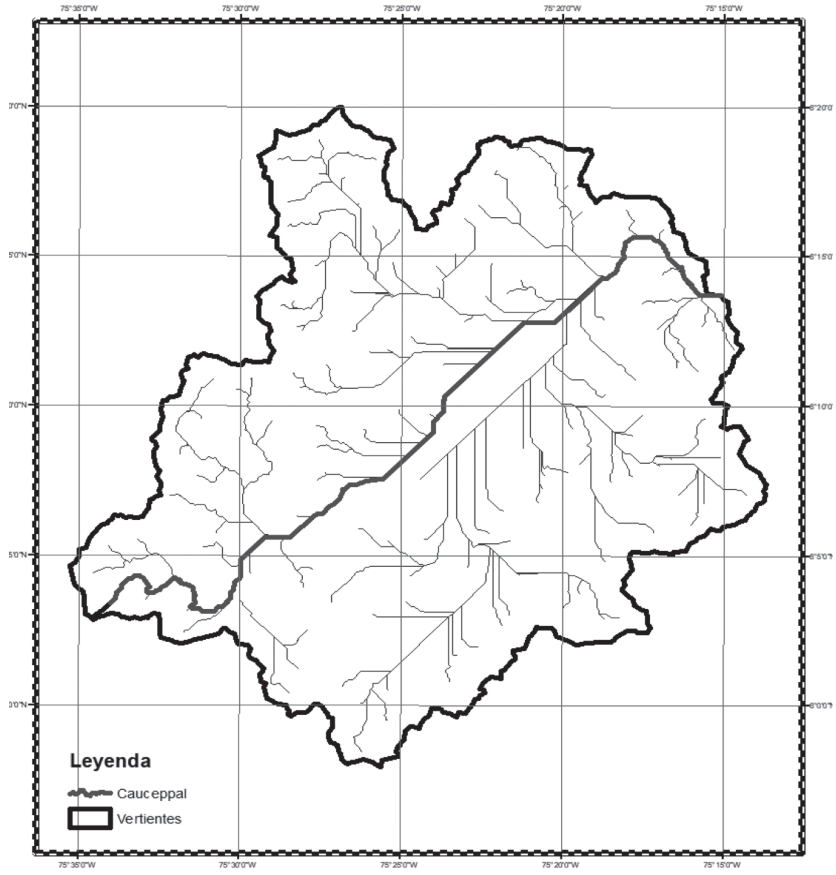


Fig. 8. Vertientes de la cuenca del río Negro.

IV. DISCUSIÓN

Área de la cuenca: El área de la cuenca es de 928,1505 km^2, la cual, a juicio de los autores, comprende una cuenca grande no solo por su área, sino por la zona que la comprende y la administración política que la compone, lo cual hace su estudio complejo, y para estudios posteriores se hace necesario la división en varias subcuencas y/o microcuencas.

Longitud del cauce principal (Lc): El cauce está compuesto por diferentes tramos donde se tienen varios comportamientos. La longitud se hace suficiente para que el río realice los procesos de erosión, transporte y depósito de material.

Índice de compacidad o Coeficiente de **Gravelius (Kc):** El índice de compacidad para la cuenca es de 1,067209486, lo que indica que su forma es casi redonda a oval redonda. Este valor de Kc determina que la cuenca tiene una tendencia a concentrar volúmenes de agua de escorrentía bastante altos; comportamiento que llevaría a la ocurrencia de inundaciones lentas pero prolongadas a lo largo del tiempo, cuando se presenten fuertes precipitaciones en toda el área de la cuenca [8].

Factor de forma (Kf): El factor de forma de la cuenca del río Negro es de 0.326215, valor que lleva a determinar que dicha cuenca es ligeramente alargada.

Densidad de drenaje (Dd): En la cuenca del río Negro, por su forma oval redonda y el resultado obtenido, se contempla que su densidad de drenajes es baja, lo cual obedece a que la mayor parte de la cuenca está compuesta por un valle, formando así pocos cauces, pero con caudales altos [11].

Relación de relieve (Rr): Al realizar este cálculo se comprueba lo dicho en los literales anteriores, lo que demuestra aún más el comportamiento de inundaciones lentas y prolongadas en el tiempo.

Pendiente media de la cuenca (Pm): Responde a la discusión presentada en los literales anteriores al definir la cuenca como un llano.

Curva hipsométrica: Con respecto a la Figura 6 y la Tabla 5, se puede determinar que la cuenca del río Negro es una cuenca en fase de madurez. En la Figura 6 se observa una curva cuyos mayores cambios se dan al inicio (punto más alto de la divisoria de aguas) y al final de la cuenca (punto sumidero); en la parte media, la curva no presenta cambios significativos en el transcurso de esta. Este comportamiento también se evidencia en el histograma de pendientes.

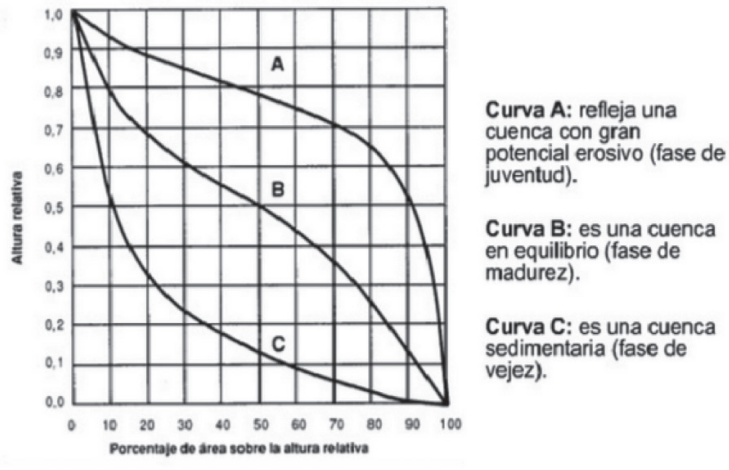


Fig. 9. Clasificación de la curva hipsométrica

Índice Asimétrico (**Ias):** Para la cuenca del río Negro se tiene una carga hacia la vertiente derecha, indicando que para esta vertiente la probabilidad de inundaciones es más alta que para la vertiente izquierda. Esto puede comprobarse de manera práctica al momento de realizar visitas de campo o al comparar datos históricos [12].

V. CONCLUSIONES

La cuenca del río Negro presentará en su mayor parte inundaciones lentas y prolongadas en el tiempo, debido a la tendencia a concentrar volúmenes de agua bastante altos y aumentar su tiempo de concentración.

Debido a su comportamiento de inundaciones lentas y prolongadas, se hace factible encontrar modelos predictivos e implementar sistemas de alerta temprana que ayuden a la disminución de la vulnerabilidad con respecto a la amenaza por inundación en los centros poblados directamente relacionados a los drenajes de la cuenca.

La vertiente derecha compuesta por los cauces de la quebrada La Pereira, el río Pantanillo, la quebrada La Cimarrona, entre otros, presenta históricamente mayor número de inundaciones, lo cual se puede evidenciar de forma teórica al momento de calcular el índice asimétrico.

Los mayores procesos erosivos de la cuenca se presentan hacia la divisoria de aguas, en las partes alta y media utilizando los afluentes principales como transporte y depositando el material erosionado sobre el valle de San Nicolás.

VI. REFERENCIAS

[1] Consorcio Pomcas Oriente Antioqueño, "Formulación del Plan de Ordenación y Manejo de la Cuenca Hidrográfica de Río Negro", vol. 1, p. 448, 2016.

[2] Universidad Nacional de Colombia y Corporación Autónoma Regional de las cuencas de los Ríos Negro y Nare "Cornare", "Plan de Ordenación y Manejo de la Cuenca Pantanillo, Municipio de El Retiro", 2006.

[3] DANE, "DANE, Información para todos", 2018. Disponible en: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion.

[4] M. Escobar Tobón, S. Mateus Santiago, y J. C. Valdés Quintero, Investigación e Innovación en Ingeniería de Sotfware, Capitulo X: Estimación de precipitación y evapotranspiración utilizando Sistemas de Información Geográfica a partir de datos obtenidos en estaciones metereológicas ubicadas en la cuenca del río Negro. Medellín, 2019.

[5] J. C. Valdés Quintero, C. C. Castro C., y J. F. Escobar M., "Geoinformática aplicada con Aprendizaje Basado en Problemas. Capítulo 12: Análisis de datos espaciales para la caracterización de parámetros geomorfométricos en cuencas hidrográficas". Instituto Antioqueño de Investigación, 2017.

[6] ESRI, "ArcMap," 2020. Disponible en: https:// desktop. arcgis. com/ es/ arcmap/ 10.3/ tools/ spatial-analyst-toolbox/ how-slope-works.htm.

[7] R. Guerrero, D. Serpa, y D. Vázquez, "Análisis geomorfológico de la cuenca del río Mazán". Working paper, diciembre 2015. doi: 10.13140/RG.2.1.3101.0165.

[8] J. Pérez, *Fundamentos del ciclo hidrológico*. Universidad Central de Venezuela. Facultad de Ingeniería Departamento de Meteorología e Hidrología, 1979.

[9] J. S. I and M. Babar, "Morphometric Analysis with Reference to Hydrogeological Repercussion on Domri River Sub-basin of Sindphana River Basin, Maharashtra, India", *J. Geosci. Geomatics*, vol. 1, n.° 1, pp. 29-35, 2013. doi: 10.12691/jgg-1-1-5.

[10] M. Dugarte, C. Ferrer, and A. Delgadillo, "Respuesta hidrogeomórfica de una

cuenca altamente asimétrica. La tormenta del 11 de febrero del 2005 y su impacto en el río Mocotíes. Mérida-Venezuela", *Rev. Geogr. Venez.*, vol. 56, n.° 1, pp. 11-40, 2015.

[11] Universidad Nacional de Colombia y Corporación Autónoma Regional de las cuencas de los Ríos Negro y Nare "Cornare", "Plan de Ordenación y Manejo de la Cuenca La Pereira, Municipio de La Ceja del Tambo", 2006.

[12] V. Restrepo, "Invierno causó estragos en el Oriente antioqueño", *El Colombiano*, Medellín, 19-Marzo 2017.

Capítulo 5
COVID-19: Un Análisis Basado en Minería de Texto

Victor Daniel Gil vera - Victor.gilve@amigo.edu.co
Catalina Quintero Lopez - Catalina.quinterolo@amigo.edu.co
Universidad Católica Luis Amigó

I. INTRODUCCIÓN

El COVID-19 declarada pandemia mundial por la OMS en mayo de 2020, surge en China a finales de 2019 [1] generando una emergencia sanitaria a nivel mundial por ser altamente infeccioso. A cifras de noviembre de 2020, en el mundo se han infectado 55,6 millones de personas, se han recuperado 35,8 millones y han muerto 1,34 millones. Esta problemática ha obligado a tomar medidas a los diferentes gobiernos a nivel mundial; cuarentenas obligatorias, medidas de aislamiento social, toques de queda y restricciones sociales para reducir la propagación y evitar el colapso hospitalario, lo que ha impactado directamente en las dinámicas socioculturales y económicas, laborales y demás.

El objetivo de este trabajo fue emplear la técnica de la minería de texto para realizar un análisis de sentimientos para analizar la percepción que tienen las personas sobre las medidas que se han implementado para evitar la propagación del virus a partir de comentarios en la red social "Twitter". Se emplearon 3000 comentarios, se creó un conjunto de datos con texto y usuario, se midieron las puntuaciones de influencia en función de diversos parámetros. El trabajo se divide en las siguientes secciones: la sección dos presenta una contextualización general sobre el COVID-19, en la tres se presenta una contextualización sobre la minería de texto y el análisis de sentimientos; en la cuatro se presenta la metodología; en la cinco los resultados y la discusión. Finalmente se concluye.

II. COVID-19

Este virus comenzó en la ciudad de Wuhan, China, y se ha propagado rápidamente en la mayoría de países del mundo [2]. A la fecha ha dejado

más de 1,34 millones víctimas mortales. La OMS declaró el 30 de enero de 2020 el brote del COVID-19 como una emergencia de salud pública y en el mes de mayo del mismo año lo declara pandemia mundial [3].

Los primeros casos estaban relacionados con un mercado de animales vivos y mariscos, o el llamado "húmedo", en Wuhan, provincia de Hubei, que se cerró rápidamente [2]. En promedio, cada individuo con el virus está infectando a otros dos [3]. La transmisibilidad del COVID-19 ha sido similar en todos los países del mundo, los contagios de personas en las principales ciudades del mundo se han vuelto inevitables debido a la exportación sustancial de casos asintomáticos y a la ausencia de intervenciones de salud pública a gran escala desde el inicio de la pandemia [2].

La medida en que los pacientes asintomáticos y subclínicos pueden transmitir el virus es alta, especialmente a personas de edad avanzada con comorbilidades [4]. En las enfermedades respiratorias, por lo general, se alojan una gran cantidad de virus o bacterias en la faringe y al tracto respiratorio [4]. Este virus se transmite principalmente de persona a persona por vía aérea, los síntomas más habituales que se presentan en personas contagiadas son; fiebre, tos seca y cansancio, dolores y molestias, congestión nasal, dolor de cabeza, conjuntivitis, dolor de garganta, diarrea, pérdida del gusto o el olfato, erupciones cutáneas o cambios de color en los dedos de las manos o los pies [5], [6] including elderly COVID-19 patients admitted to Wuhan Huoshenshan Hospital between February 10 and 13, 2020. Demographic data, medical history, signs and symptoms, and laboratory parameters were collected and analysed. Results We included 131 elderly COVID-19 patients (50 patients with diabetes, [7]patients with acute myocardial infarction (AMI.

Estos síntomas suelen ser leves y comienzan gradualmente, algunas de las personas infectadas solo presentan síntomas leves. La mayoría de las personas (cerca al 80%) se recuperan de la enfermedad sin necesidad de tratamiento hospitalario. Alrededor de 1 de cada 5 personas que se infectan acaba presentando un cuadro grave y experimenta dificultades para respirar [8]positive with PCR confirmation, patients who had surgery between February 17, 2020 and April 26, 2020 at a major New York City hospital. Clinical characteristics and outcomes including ICU admission, ventilator requirement, and mortality were analyzed. Results Thirty-nine COVID-19 surgical patients were identified. Mean age was 53.9 years and there were more males than females in the cohort (56.4% vs 43.6%. Las personas mayores y las que

padecen afecciones médicas previas como hipertensión arterial, problemas cardiacos o pulmonares, diabetes o cáncer tienen más probabilidades de presentar cuadros graves. Sin embargo, cualquier persona puede infectarse y caer gravemente enferma [9]risk perception, and behavioral adherence during the COVID-19 epidemic. Methods A national cross-sectional survey was conducted in 31 provinces in China with a total number of 5039 effective questionnaires collected. The questionnaire included sociodemographic characteristics, COVID-19 risk communication factors, mask and soap supply, and engagement in preventive behaviors during the epidemic. Multivariable Logistic regression was used. Results An overwhelmingly high prevalence of Chinese people was exposed to COVID-19 related risk communication messages (86.5%. Son grandes los esfuerzos realizados por entes gubernamentales para detener su propagación. Día tras día, científicos en todo el mundo aúnan esfuerzos para desarrollar tratamientos y una vacuna efectiva para la inmunización de adultos, jóvenes y niños. La Tabla 1 presenta las vacunas se están desarrollando diferentes compañías farmacéuticas a nivel mundial:

Tabla 1. Vacunas en Desarrolla (Elaboración del autor)

	Biontech (Pfizer)	**AstraZeneca**	**Moderna**	**Janssen**	**Sanofi**	**Curevac**
País	**EEUU y Alemania**	**Reino Unido**	**EEUU**	**EEUU**	**Francia y Reino Unido**	**Alemania**
Fase actual de ensayos clínicos	III	III	III	III	II	II
Dosis necesarias por persona	2	2	2	1	2	2

En muchos países debido a la ausencia de una vacuna oficial han recurrido a las cuarentenas y al distanciamiento físico para reducir la tasa de infección y evitar que los sistemas nacionales de salud (clínicas y hospitales) se vean colapsados. Sin embargo, el reto de hacer frente a la pandemia en el ámbito de la salud pública va mucho más allá de estas medidas. Hay problemas crónicos y profundamente arraigados de enfermedad, pobreza y educación, especialmente en América Latina, que complican tanto la respuesta

inmediata a la crisis como el eventual levantamiento de las restricciones. Estos problemas en cierta medida complican tanto la respuesta inmediata a la crisis como el eventual levantamiento de las restricciones [10] many countries have implemented policies to minimize the spread of COVID-19 infection. However, some of these policies prevent people from physical contact. Consequently, many individuals may rely on social media to obtain information concerning COVID-19. Unfortunately, social media use (especially problematic social media use, [11], [12]public confidence in government sources of health information and general concern for the welfare of others support health advisories during a pandemic and save lives. We tested this assumption through a time-series analysis of cross-national differences in COVID-19 mortality during an early phase of the pandemic. Country data on income inequality and four dimensions of social capital (trust, group affiliations, civic responsibility and confidence in public institutions.

Existen interrogantes y aspectos inciertos a los cuales no se les conoce la respuesta exacta; como por ejemplo la manera en que se replica la enfermedad en diferentes sitios, el tiempo durante el cual los pacientes permanecen infecciosos, el tiempo que deben estar aislados. El virus está mutando en formas más o menos transmisibles [13].

Los adultos mayores con comorbilidades tienen mayor riesgo de desarrollar una enfermedad grave como el síndrome respiratorio agudo severo (SARS-CoV-2). Según las cifras de China, la letalidad ronda cerca al 5%, pero es concebible que sea mucho mayor [14]. Es difícil para los lugares sin instalaciones de diagnóstico diferenciar entre el coronavirus y la influenza estacional [4]. El alcance de la propagación más allá de China es inminente, situación que se ve reflejada en la cantidad de contagios y muertes que el virus ha generado en países europeos y en EEUU [15]. El reservorio animal del virus no ha sido confirmado, pero el análisis filogenético ha apuntado hacia los murciélagos después de algunas especulaciones no confirmadas de una participación de serpientes y pangolines [16].

En cualquier caso, parece probable que el SARS-CoV-2 se originó en un mercado húmedo [17]. Los mercados húmedos son un riesgo de transmisión de enfermedades zoonóticas a los humanos [18]. China y otros países asiáticos pueden querer revisar su política en estos mercados para intentar minimizar el riesgo, esto podría significar imponer medidas más estrictas de control de infecciones o incluso prohibirlas por completo [11]. Los consumidores en los países asiáticos donde abundan los mercados húmedos esperan comprar

animales vivos y usarlos en su cocina. En ese caso, las consideraciones socioculturales harán que abogar por su eliminación sea una propuesta difícil, que deberá adaptarse cuidadosamente a la población asiática. Los mercados húmedos son un lugar de intercambio de materiales genéticos de una variedad de animales diferentes, por lo cual, se deben vigilar los estándares de higiene y hacer un seguimiento de los microbios que circulan dentro de estos [19]. La Tabla 2 presenta el resumen de casos, personas recuperadas y decesos a causa del COVID-19 a fecha de noviembre de 2020:

Tabla 2. Cifras oficiales (OMS)

Ubicación	Total de casos	Personas recuperadas	Muertes
Antioquia	200 k	187 k	3785
Colombia	1,23 M	1,14 M	34.929
Mundo	55,6 M	35,8 M	1,34 M

En la revisión del estado del arte diversas investigaciones afirman que las medidas preventivas adoptadas por los entes gubernamentales pueden ser percibidas como experiencias traumáticas y generan un impacto psicológico amplio y duradero para quienes la sobrellevan, afectando a diferentes grupos etarios como niños, estudiantes y adultos mayores [20]. Las restricciones establecidas pueden crear una sensación de pérdida de la libertad por la limitación en los desplazamientos, el alejamiento de los seres queridos, la repentina desaceleración de la economía y la incertidumbre constante del posible contagio de la enfermedad; ocasionado la aparición de alteraciones psicológicas [21].

III. MINERÍA DE TEXTO

Proceso de derivar información de alta calidad de información textual para extraer información útil para la toma eficiente de decisiones [22]. Básicamente, es un proceso de análisis y exploración de grandes volúmenes de información textual no estructurada con el fin de identificar conceptos, patrones, temas, palabras claves, tendencias, gustos, entre otros.

Como rama de la analítica de datos, permite realizar análisis de manera masiva, identificar tendencias, clasificar comentarios y opiniones, clasificar tipos de sentimientos y construir nubes de palabras que permiten obtener una visión general de la información.

En el último quinquenio ha sido empleada y utilizada por científicos y académicos a nivel mundial, debido al desarrollo de plataformas de "Big-Data" y algoritmos de aprendizaje profundo que tienen la capacidad de analizar conjuntos masivos de datos no estructurados.

La minería de texto puede definirse en términos generales como un proceso intensivo en conocimiento en el que un usuario interactúa con una colección de documentos a lo largo del tiempo mediante herramientas de análisis [23]. La Figura 1 presenta el proceso de minería de texto:

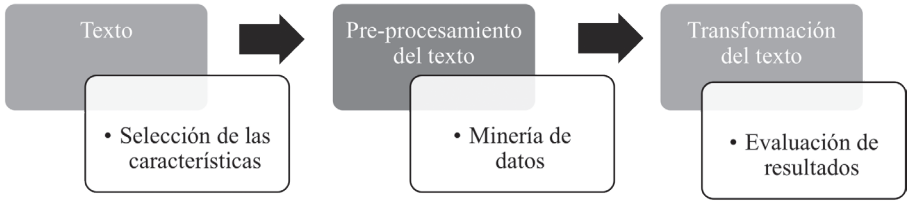


Figura 1. Proceso de Minería de Texto

3.1 Análisis de Sentimientos

Es una rama de la minería de texto, la cual se ha aplicado ampliamente en diferentes contextos; políticos, socioculturales, económicos, entre otros. Diferentes investigaciones la han empleado para comprender los pensamientos de las personas. Las palabras clave extraídas pueden convertirse en conceptos a partir de datos desordenados y transformarse en información valiosa [24].

La información disponible en las redes sociales son una fuente de información destacada durante la crisis generada por la pandemia del COVID-19. Los mensajes generados por los usuarios permiten ver un panorama general de lo que piensan las personas, conocer estados de ánimo y opiniones. Debido a la gran cantidad de mensajes emitidos en la redes sociales, es posible realizar un análisis a gran escala de la evolución que ha tenido la pandemia y la percepción de todos los factores asociados a ella [25].

IV. MÉTODOLOGÍA

En este trabajo se empleó el software estadístico R 3.6.2 para analizar 3000 comentarios de la red social Twitter sobre las medidas adoptadas para evitar la propagación del virus COVID-19. Los paquetes y librerías empleadas en el análisis se presentan a continuación:

```
# Install
install.packages("tm")  # for text mining
install.packages("SnowballC") # for text stemming
install.packages("wordcloud") # word-cloud generator
install.packages("RColorBrewer") # color palettes
install.packages("dplyr") # color palettes
install.packages("tidyr") # color palettes
install.packages("tidytext") # color palettes
# Load
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
library("corpus")
library(rtweet")
library("dplyr")
library("tidyr")
library("tidytext")
```

Los *Tweets* fueron seleccionados a través de una API. Se empleó la librería *rtweet*, un *wrapper R* que se comunica con la API de Twitter. Para la extracción de comentarios se creó una *Twitter App*, mecanismo que proporciona *Twitter* para desarrolladores que quieran acceder a sus contenidos a través de programas, para lo cual proporciona una serie de claves y *tokens* personalizadas. El límite de extracción de comentarios fueron 100 cada 15 minutos (rate limiting). La Tabla 3, presenta algunos de los comentarios empleados en el análisis:

Tabla 3. Comentarios de Twitter (Elaboración del autor)

Usuario	Comentario
@NCEAatUSC	COVID-19 has exacerbated feelings of isolation and loneliness in older adults. Practice physically distanced social interactions with older adults to help decrease social isolation and promote connections.
@DanCorcoranTV	Many CT families are worried that social isolation is having a devastating effect on the physical and mental health of their loved ones. Experts say it is a hidden crisis amid the COVID-19 pandemic
@HSScomms	Human–dog relationships during the COVID-19 pandemic: booming dog adoption during social isolation
@cnni	The French health minister said on Tuesday there is a “credible risk” the novel coronavirus outbreak could turn into a pandemic.
@JacariOxford	Local lockdowns and restrictions are leading to increased social isolation for children with English as an additional language and limited opportunities to practise their English..

Una vez seleccionados los “Tweets” para realizar el análisis, se almacenaron en un archivo de texto plano .txt, el cual se importó al programa para realizar el proceso de *tokenización*, es decir, el proceso de limpieza en el cual se sustituyen de manera masiva caracteres especiales (“/”, “@” y “|”), espacios en blanco, palabras comunes, etc. Algunas palabras fueron reducidas a su forma raíz y se eliminaron los sufijos. El código empleado para la depuración del texto se presenta a continuación:

```
# Load the data as a corpus docs <-
Corpus(VectorSource(text))
# Tokenization toSpace <- content_
transformer(function (x , pattern ) gsub(pattern,
“ “, x)) docs <- tm_map(docs, toSpace, “/”) docs
<- tm_map(docs, toSpace, “@”) docs <- tm_
map(docs, toSpace, “\\|”)
# Convert the text to lower case
docs <- tm_map(docs, content_
transformer(tolower))
# Remove numbers
docs <- tm_map(docs, removeNumbers)
# Remove english common stopwords
docs <- tm_map(docs, removeWords,
```

```
stopwords("english"))
# Remove your own stop word
# specify your stopwords as a character vector
docs <- tm_map(docs, removeWords, c("blabla1",
"blabla2"))
# Remove punctuations
docs <- tm_map(docs, removePunctuation)
# Eliminate extra white spaces
docs <- tm_map(docs, stripWhitespace)
# Text stemming
# docs <- tm_map(docs, stemDocument)
```

V. RESULTADOS Y DISCUSIÓN

Una vez depurados los comentarios se procedió a construir la matriz de documentos, en la cual aparece la frecuencia de aparición de las palabras de los comentarios. El código empleado y la matriz de clasificación para las 10 palabras más frecuentes se presentan a continuación:

```
dtm <- TermDocumentMatrix(docs)
m <- as.matrix(dtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)
```

	word	freq
distancing	distancing	180
quarantine	quarantine	169
social	social	157
public	public	149
people	people	120
pandemic	pandemic	110
infectious	infectious	97
diseases	diseases	83
crowed	crowed	75
understand	understand	69

Posteriormente, se procedió a generar la nube de palabras o *wordcloud*, en la cual aparecen las palabras más frecuentes y cuyo tamaño de fuente varía dependiendo de la frecuencia de aparición de las palabras. El código empleado y el *wordcloud* se presentan a continuación:

```
# wordcloud
set.seed(1234)
wordcloud(words = d$word, freq = d$freq,
min.freq = 1,
max.words=200, random.order=FALSE, rot.
per=0.35, colors=brewer.pal(8, "Dark2"))
```

Figura 2. Nube de Palabras

El resultado de la nube de palabras muestra claramente que las palabras *"distancing", "quarantine", "social*", "*public"* y "*people"* son las más importantes de todos los *tweets* seleccionados sobre "COVID-19". Se debe aclarar que, los argumentos que componen la función fueron definidos de la siguiente manera:

- *min.freq*: palabras con una frecuencia inferior a "min.freq" no se graficarán (frecuencia 1)
- *max.words*: número máximo de palabras que se graficarán (200 palabras)
- *random.order*: trazar palabras en orden aleatorio. Si es falso, se grafican en frecuencia decreciente (en este trabajo se adoptó el valor FALSO)
- *rot.per*: proporción de palabras con rotación de 90 grados (texto vertical) (0.35)
- *colors:* palabras de color de menor a mayor frecuencia. En nuestro caso adoptamos los valores de la paleta (brewer.pal(8, "Dark2"))

Adicionalmente, se clasificaron los términos más frecuentes, en este caso las que aparecían al menos 5 veces:

```
# Frequency.Terms
findFreqTerms(dtm, lowfreq = 5)
[2] "isolation" "economy" "crisis"
"work"
[9] "covid" "fatal" "conflict" "diagnosed"
[25] "poverty" "health" "infections"
"masks" "human"
```

Finalmente, se asociaron términos frecuentes utilizando la función *findAssocs()*. Se descartaron

correlaciones inferiores a 0.3. Se identificaron las palabras que estaban asociadas con la palabra "COVID-19" en los *tweets* seleccionados, dando los siguientes resultados:

findAssocs(dtm, terms = "covid", corlimit = 0.3), $covid		
Isolation 0.45	Confinement 0.48	Distancing 0.48
Public 0.40	Health 0.45	Virus 0.43
Keep 0.39	Masks 0.40	Infect 0.39
Economy 0.39	Love 0.39	Social 0.39
Measures 0.39	Rate 0.39	Need 0.39

Finalmente, se clasificaron los Tweets por tipos de sentimientos, para ello empleamos el código:

```
sentiments %>%
  filter(word_count %in% c("very postive",
"positive", "neutral","negative", "very neg-
ative")) %>%
  arrange(word) %>% #sort
  select(-score) %>% #remove this field
```

Lo que nos permitió obtener la siguiente clasificación final:

sentiment	total
very positive	198
positive	443
neutral	224
negative	1683
very negative	452

VI. CONCLUSIONES

Gracias a la minería de texto y al análisis de sentimientos es posible conocer de manera masiva la percepción que tienen las personas sobre un fenómeno determinado. Esta investigación permitió analizar los comentarios en de 3000 personas acerca de las medidas adoptadas para prevenir la propagación del COVID-19, identificar los términos más frecuentes, construir una nube de palabras ("*Wordcloud*") y clasificarlos en cuatro tipos de sentimientos.

La minería de texto, a diferencia de otras técnicas de análisis de información cualitativa, tiene la ventaja de que permite descubrir información nueva, permite analizar grandes bases de datos de información textual mediante la tecnología de *Data mining*, generando resultados que son fáciles de entender, lo que contribuye a la toma eficiente de decisiones, ya que permite detectar la información relevante, tendencias, patrones, entre otros. En efecto, la minería de datos se constituye como una poderosa herramienta para el análisis de información cualitativa.

A pesar de que la mayoría de medidas implementadas para evitar la propagación del virus han afectado considerablemente las dinámicas sociales, culturales, económicas, educativas y demás, estás medidas han ayudado a disminuir los contagios y en

efecto las muertes. Hasta que no haya una vacuna oficialmente reconocida por las autoridades mundiales de salud, las personas deben acogerse a estas medidas y buscar apoyo psicológico, económico y demás en caso tal de verse sumamente afectados.

VII. REFERENCIAS

[1] R. Habibi *et al.*, "Do not violate the International Health Regulations during the COVID-19 outbreak," *Lancet*, 2020.

[2] T. Burki, "Outbreak of coronavirus disease 2019," *Lancet Infect. Dis.*, 2020.

[3] D. Chang, H. Xu, A. Rebaza, L. Sharma, and C. S. Dela Cruz, "Protecting health-care workers from subclinical coronavirus infection," *Lancet Respir. Med.*, 2020.

[4] J. Zhang, L. Zhou, Y. Yang, W. Peng, W. Wang, and X. Chen, "Therapeutic and triage strategies for 2019 novel coronavirus disease in fever clinics," *Lancet Respir. Med.*, 2020.

[5] A. Tsatsakis *et al.*, "SARS-CoV-2 pathophysiology and its clinical implications: An integrative overview of the pharmacotherapeutic management of COVID-19," *Food Chem. Toxicol.*, vol. 146, p. 111769, 2020.

[6] P. Zhang *et al.*, "Risk factors associated with the progression of COVID-19 in elderly diabetes patients," *Diabetes Res. Clin. Pract.*, p. 108550, 2020.

[7] K. D. Lee *et al.*, "Providing essential clinical care for non-COVID-19 patients in a Seoul metropolitan acute care hospital amidst ongoing treatment of COVID-19 patients," *J. Hosp. Infect.*, vol. 106, no. 4, pp. 673–677, 2020.

[8] N. M. Vranis, J. M. Bekisz, D. A. Daar, E. S. Chiu, and S. C. Wilson, "Clinical outcomes of COVID-19 positive patients who underwent surgery: A New York City experience," *J. Surg. Res.*, 2020.

[9] X. Wang, L. Lin, Z. Xuan, J. Xu, Y. Wan, and X. Zhou, "Risk communication on behavioral responses during COVID-19 among general population in China: A rapid national study," *J. Infect.*, 2020.

[10] C.-Y. Lin, A. Broström, M. D. Griffiths, and A. H. Pakpour, "Investigating mediated effects of fear of COVID-19 and COVID-19 misunderstanding in the association between problematic social media use, psychological distress, and insomnia," *Internet Interv.*, vol. 21, p. 100345, 2020.

[11] E. A. Severo, J. C. F. De Guimarães, and M. L. Dellarmelin, "Impact of the COVID-19 pandemic on environmental awareness, sustainable consumption and social responsibility: Evidence from generations in Brazil and Portugal," *J. Clean. Prod.*, p. 124947, 2020.

[12] F. J. Elgar, A. Stefaniak, and M. J. A. Wohl, "The trouble with trust: Time-series analysis of social capital, income inequality, and COVID-19 deaths in 84 countries," *Soc. Sci. Med.*, vol. 263, p. 113365, 2020.

[13] W. Liang *et al.*, "Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China," *Lancet Oncol.*, 2020.

[14] Z. Xu, S. Li, S. Tian, H. Li, and L. Kong, "Full spectrum of COVID-19 severity still being depicted," *Lancet*, 2020.

[15] X. Gu, B. Cao, and J. Wang, "Full spectrum of COVID-19 severity still being depicted – Authors' reply," *Lancet*, 2020.

[16] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges," *Int. J. Antimicrob. Agents*, p. 105924, 2020.

[17] D. MacKenzie, "How bad will it get?," *New Sci.*, vol. 245, no. 3269, p. 7, 2020.

[18] T. Solomon, P. Lewthwaite, D. Perera, M. J. Cardosa, P. McMinn, and M. H. Ooi, "Virology, epidemiology, pathogenesis, and control of enterovirus 71," *Lancet Infect. Dis.*, vol. 10, no. 11, pp. 778–790, 2010.

[19] D. MacKenzie, "Wuhan-like virus discovered seven years ago," *New Sci.*, vol. 245, no. 3269, p. 9, 2020.

[20] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[21] S. K. Brooks *et al.*, "The psychological impact of quarantine and how to reduce it: rapid review of the evidence," *Lancet*, 2020.

[22] X. Xie, Y. Fu, H. Jin, Y. Zhao, and W. Cao, "A novel text mining approach for scholar information extraction from web content in Chinese," *Futur. Gener. Comput. Syst.*, 2019.

[23] R. Feldman and J. Sanger, *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press, 2007.

[24] W.-L. Chang and J.-Y. Wang, "A 2020 perspective on 'Mine is yours? Using sentiment analysis to explore the degree of risk in sharing economy,'" *Electron. Commer. Res. Appl.*, p. 100934, 2020.

[25] A. Kruspe, M. Häberle, I. Kuhn, and X. X. Zhu, "Cross-language sentiment analysis of European twitter messages duringthe COVID-19 pandemic," *arXiv*, 2020.

Capítulo 6
Propuesta de un Método de Evaluación Multicriterio como Herramienta para Determinar Zonas Óptimas de Ubicación de Rellenos Sanitarios Utilizando Sistemas de Información Geográfica

María Isabel Duque Cano - Isabelduquecano@gmail.com
Claudia Elena Durango Vanegas - Claudia.durango@usbmed.edu.co
Valeria Carvajal Hurtado - Valeria.carvajal06@gmail.com
Universidad de San Buenaventura, Medellín

I. INTRODUCCIÓN

Mundialmente el manejo de los residuos sólidos se ha convertido en un problema de importancia debido al exponencial aumento de la concentración de población en los centros urbanos y el actual modelo de consumo insostenible. Además, la falta de planificación en la búsqueda de sitios que permitan su adecuado manejo. De acuerdo con el informe presentado por el Banco Mundial titulado What a Waste 2.0 (2018) "*en el mundo se generan anualmente 2010 millones de toneladas de desechos sólidos municipales, y al menos el 33 % de ellos no se gestionan sin riesgo para el medio ambiente. ... el crecimiento de la población y el desarrollo económico harán que la cantidad de desechos a nivel mundial aumente 70 % en los próximos 30 años y llegue a un volumen asombroso de 3400 millones de toneladas de desechos generados anualmente*". En el caso concreto de Colombia, según el Departamento Nacional de Planeación (DNP) [1]Alcantarillado y Aseo, con el objetivo de mostrar el estado de la disposición final de residuos sólidos a nivel nacional para la vigencia 2018. Este documento presenta información prioritaria para el análisis de las entidades del orden nacional con competencias en el sector como el Ministerio de Vivienda, Ciudad

y Territorio (MVCT, la mayoría de los departamentos que reportan más toneladas se localizan en las zonas del país donde existe mayor densidad poblacional como se evidencia (ver Fig. 1) y así como se presentó en el Informe de Disposición Final 2017 (ver Fig. 2):

En el proceso de selección de sitios aptos para la localización de rellenos sanitarios se debe considerar una cantidad significativa de variables (ver Tabla I) y restricciones que según la ubicación del área de estudio se presentan en la dimensiones social, económica y ambiental. Por ello, los Sistemas de Información Geográfica (SIG) se presentan como una herramienta que permiten contemplar y contrastar diferentes escenarios en función de criterios propios del área de estudio hasta generar un resultado que satisfaga condiciones ambientales y socioeconómicas. Por otro lado, se considera la idea de utilizar el Método Decisión Multicriterio para ordenar y establecer prioridades en la evaluación de las variables criterio en la selección de sitios óptimos para la ubicación de rellenos sanitarios, permitiendo determinar prioridades según los escenarios sociales, económicos y ambientales.

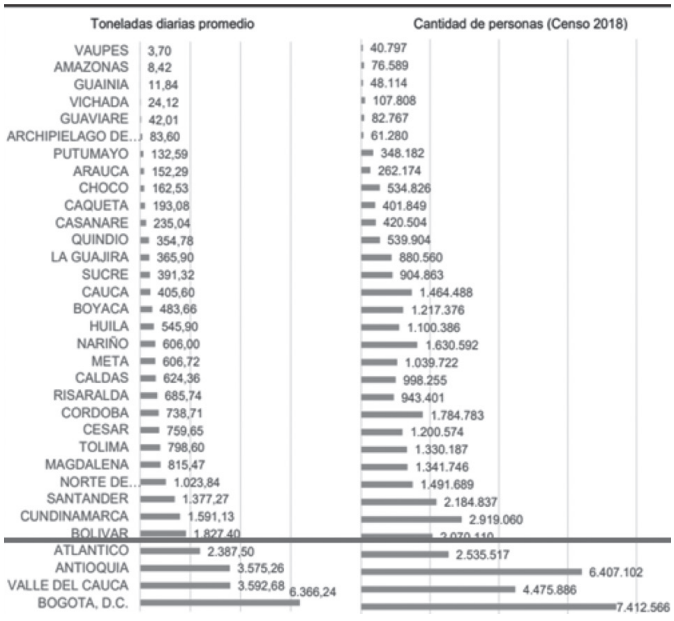


Fig. 1 *Toneladas diarias promedio presentadas al servicio público de aseo por departamento y para el Distrito Capital VS Cantidad de personas en cada departamento*

Fuente: *Tomado de Disposición Final de Residuos Sólidos (DNP, 2019)*

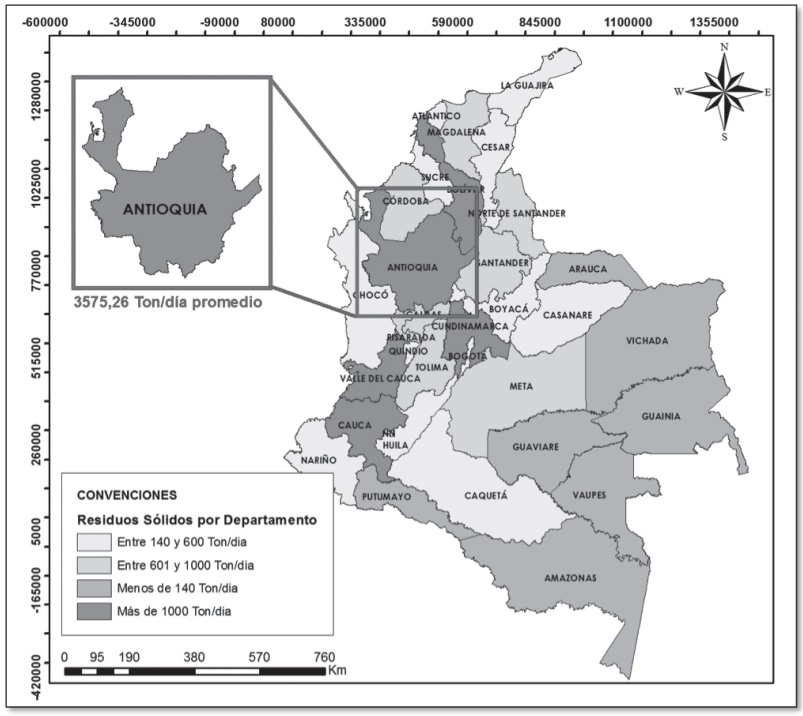


Fig. 2 *Toneladas diarias promedio de residuos por departamento*
Fuente: *Adaptado del Informe de Disposición Final de Residuos Sólidos (DNP, 2019)*

Por lo anterior, en esta propuesta se busca probar que el Método de Decisiones Multi-Criterio es una herramienta de apoyo en la determinación de zonas óptimas para la ubicación de rellenos sanitarios, utilizando SIG en base a los criterios establecidos en la legislación ambiental colombiana.

El capítulo se organiza de la siguiente manera: en la Sección 2, se presentan los antecedentes encontrados en trabajos similares. En la Sección 3, se presenta el método de evaluación multicriterio que se propone. En la Sección 4, se presentan los resultados de la investigación. En la Sección 5, se presentan la discusión de los resultados obtenidos en la investigación. Por último, se presentan las conclusiones.

II. ANTECEDENTES

El legislador colombiano con el propósito de proteger la naturaleza, la conservación de los recursos y la garantía de los derechos individuales a promulgado diversas normas relacionadas con los residuos sólidos, estableciendo pautas para su adecuado manejo y dar soluciones a las

situaciones que se puedan propiciar por la inadecuada práctica, consagrando como finalidad la prestación de servicios público de aseo [2].

La expedición del Código de Recursos Naturales en 1974 (Decreto 2811 de 1974, "Por el cual se dicta el Código Nacional de Recursos Naturales Renovables y de Protección al Medio Ambiente") da la iniciativa a políticas y un marco normativo ambiental que se ocupen de los impactos por el inadecuado manejo de los residuos sólidos generados a partir del Título III de la Parte IV del Libro I donde se encuentra todo lo relacionado con los residuos sólidos, las basuras, los desechos y los desperdicios (arts. 34 a 38). En este documento se consagran las obligaciones a las entidades territoriales y a los productores de residuos sólidos para la gestión de estos, se da a entender que el adecuado manejo comprende la recolección, el tratamiento, el procesamiento y la disposición final haciendo uso de mejores métodos, de acuerdo con los avances de la ciencia y tecnología. Así mismo, se determinó que la investigación científica y técnica se fomentarían, entre otros, para reintegrar al proceso natural y económico los residuos provenientes de las actividades domésticos o núcleos humanos generales [3].

El código prohíbe la descarga, sin autorización, de los residuos, basuras y desperdicios, en general, desechos que deterioren los suelos o causen daño o molestia a individuos o núcleos humanos. El Decreto 2811 del 1974 determina el uso de medios que permitan evitar el deterioro del ambiente y de la salud humana en su disposición final, obligando a los municipios a organizar servicios adecuados de recolección, transporte y disposición final de basuras, lineamientos que quedaron incluidos en la Ley 142 de 1994. [3]

La Ley 9 de 1979, conocida como el Código Sanitario Nacional, dedica sus artículos 22 a 35 a regular lo relacionado con las basuras haciendo énfasis en su recolección, transporte y disposición final, prohibiendo el sistema de quemas al aire libre como método de eliminación de basuras, de separación y de clasificación en las vías públicas, el almacenamiento al aire libre y su abandono en predios no autorizados [4].

La Ley 99 de 1993 se limita a fijar las funciones de las autoridades ambientales, buscando garantizar la adecuada gestión de los residuos. En dicha Ley se establece que la correspondería a los municipios y distritos para ejecutar programas de disposición, eliminación y reciclaje de residuos sólidos, de igual forma decreta la responsabilidad del Ministerio de Ambiente de promover la formulación de planes de reconversión

industrial ligados a la implementación de tecnologías ambientalmente sanas y a la realización de actividades de descontaminación, de reciclaje y de reutilización de residuos, delegando en las Corporaciones Autónomas Regionales la evaluación, el control y el seguimiento ambiental del recurso suelo, frente al vertimiento o incorporación de residuos sólidos al agua o suelo [5].

Con la expedición de las leyes, anteriormente referenciadas, se ha desarrollado en Colombia un marco normativo y reglamentario para la gestión integral de residuos. De este modo, nace la necesidad de ubicar áreas óptimas para los rellenos sanitarios que cumplan dicho marco jurídico y contemplen las pautas y parámetros para la disposición final de los residuos. A continuación, se relacionan casos de estudio en la determinación de áreas optimas de rellenos sanitarios con herramientas SIG.

Vatalis y Manoliadis (2002) realizan un enfoque multicriterio de dos etapas utilizando SIG para la selección de sitios de rellenos sanitarios. La propuesta consiste en generar una zonificación para la posterior selección de los sitios por medio de la programación de compromiso. El método clasifica los sitios utilizando varios escenarios con diferentes ponderaciones y puntos de equilibrio, lo que permite al usuario evaluar la solidez de los sitios individuales y proponer como solución aquella área que ocupa el primer lugar en todas las opciones de evaluación del método. De acuerdo con los autores, este método es apropiado cuando se requiere una evaluación rápida de sitios para rellenos sanitarios [6]. Sin embargo, este método de evaluación multicriterio presenta problemas para introducir un enfoque de programación en la selección adecuada para rellenos sanitarios, esto considerando que dos de los objetivos principales a menudo entran en conflicto en el desarrollo sostenible, tales como: la economía de aplicación y el desempeño ambiental. La decisión final a menudo implica una compensación entre estos dos objetivos.

Mena *et al*. (2006) explican un modelo de capacidad de acogida que facilita la localización de los sitios adecuados para albergar un relleno sanitario. Para ello, proponen una metodología basada en la integración de la geomática y las técnicas de evaluación multicriterio, proceso que genera valiosa información que ayuda a la toma de decisiones, particularmente en problemas de asignación de actividades y gestión territorial. La aplicación de las técnicas de evaluación multicriterio en el ámbito de la Geomática requiere que cada variable o factor temático deben ser estructurados como una matriz con el propósito de comparar

la importancia de cada uno de los factores y la consistencia de los juicios de valor [7]. Es importante considerar que, entre más restricciones espaciales sean incluidas en la modelación, la región factible y las superficies tienden a disminuir.

Chang *et al*. (2008) proponen una combinación de SIG con Análisis de Decisión Multicriterio Difuso para Localización de Rellenos Sanitarios en Regiones de Alto Crecimiento Urbano. Los autores realizan un análisis de dos etapas sinérgicamente para formar un sistema de apoyo de decisión espacial (DSS) para la gestión de residuos en una región urbana de rápido crecimiento. En este método se realiza una selección inicial para eliminar tierras inadecuadas, seguido del método de decisiones difusas de múltiples criterios (FMCDM) para identificar el sitio más adecuado utilizando la información proporcionada por los expertos regionales con referencia a cinco criterios elegidos. La toma de decisiones difusa multicriterio (FMCDM) difiere de los métodos convencionales de integración de SIG con MCDM, porque el enfoque sigue dos pasos secuenciales en lugar de un esquema totalmente integrado. La evaluación multicriterio se ocupa principalmente de cómo combinar varios criterios para formar un único índice de evaluación [8] complex, tedious, and protracted process requiring evaluation of many different criteria. This paper presents a fuzzy multicriteria decision analysis alongside with a geospatial analysis for the selection of landfill sites. It employs a two-stage analysis synergistically to form a spatial decision support system (SDSS. La ventaja de este método radica en la capacidad de incorporar el conocimiento de los expertos en el dominio en el incierto proceso de toma de decisiones cuando existe una falta de información nítida relacionada con ciertos criterios. Sin embargo, la disponibilidad de información puede ser un factor limitante en la aplicación del análisis espacial y la selección del mejor sitio depende de los juicios de los expertos de dominio y puede ser sensible a los cambios en los pesos de decisión asociados con los criterios. Lo cual representa una desventaja en el método, cuando se considera que, en ciertas situaciones, dos expertos pueden tener juicios contradictorios sobre la idoneidad de un sitio candidato. Por lo tanto, se requiere evaluar el grado de diferencia o similitud entre los dos expertos en asociación con los pesos de decisión.

Zafra *et al*. (2012) presentan una metodología para la localización de rellenos sanitarios mediante Sistemas de Información Geográfica, aplicado a un caso regional colombiano. En el estudio se plantea el desarrollo metodológico para la localización de rellenos sanitarios mediante la combinación de SIG, el proceso

analítico jerárquico (PAJ) y el método de ponderación aditiva simple (MPAS). Los autores proponen una estructura jerárquica que permite identificar y predecir apropiadamente el peso de importancia de cada variable según un rango que va desde 0 hasta 1000, donde los valores más altos corresponden a mayor compatibilidad con áreas apropiadas para la localización de rellenos sanitarios y son asignados utilizando diversos criterios de otros autores. Los autores consideran para el estudio las siguientes variables: hidrología, accesibilidad de vías, distancia a perímetros urbanos y reservas indígenas, disponibilidad de material de cubrimiento, condición del suelo y topografía, ocupación actual del área, densidad de población, dirección del viento, restricciones, capacidad y prohibiciones. Mientras que, para procesar información básica para generar los buffer se considera la ubicación de aeropuertos, bosques, zonas centro, capacidad del sitio, clases agrológicas, vegetación, cuencas, Modelo de Elevación Digital (DEM), drenajes, fallas geológicas, formación geológica, relieve, humedales, inestabilidad del suelo, lagos, capa freática, parques naturales, precipitación, reservas indígenas, riesgo sísmico, ríos, dirección del viento prevalente, pueblos, vías principales y secundarias. Los autores consideran que es necesario el modelado espacio-temporal de la población, la cantidad de residuos producidos y el volumen anual a manejar para determinar el área precisa para la instalación del relleno sanitario. Seguidamente, se utiliza el método de ponderación aditiva y con la ayuda de herramientas SIG sobreponen las áreas encontradas hasta generar un mapa final en el cual se visualiza el rango de pesos y las áreas prohibidas, donde las zonas con pesos de mayor importancia representan los lugares más compatibles para la localización de rellenos sanitarios. Adicionalmente, se genera un mapa tomando en cuenta los criterios definidos por la legislación colombiana, los cuales no incluyen la variable hidrología [9]. Con esta propuesta se encuentra que la metodología implementada por los autores es 81% más restrictiva que la legislación debido a la gran cantidad de variables que se consideran para el modelado.

Govind *et al*. (2016) proponen la combinación de diferentes técnicas para toma de decisiones Multicriterio en la selección de sitios aptos para la localización de rellenos sanitarios en el artículo titulado "Modelo para la selección de sitios aptos para relleno sanitario utilizando un enfoque MCDM difuso integrado". Los autores utilizan el Proceso de Análisis Jerárquico-Difuso (PAJ-D) para realizar comparaciones por pares de criterios con el fin de asignar pesos y clasificarlos. La Técnica para

Ordenar Preferencias por Similitud hasta encontrar la Solución Ideal-Difusa (TOPSSI-D) es empleada para clasificar las alternativas; razón por la cual se usa el método difuso en cada técnica para desenmascarar valores imprecisos al final del proceso. Los autores calculan la importancia de los pesos y la clasificación de cada criterio utilizando una matriz de comparación de criterios, luego, ejecutan el proceso de clasificación de alternativas, dónde basados en una revisión de la literatura, identifican los criterios de evaluación relacionados con la selección del sitio del vertedero y se utilizan cuestionarios de expertos para examinar los índices aptos para el proceso de selección. Los autores descubren que la metodología empleada puede ser una poderosa herramienta guía para la gestión de rellenos sanitarios puesto que la precisión de procedimiento de selección de vertederos puede ser significativamente mejorada en comparación con los métodos tradicionales, reduciendo el costo y el tiempo requerido [10]. La ventaja principal de la metodología es que ayuda a representar y analizar la incertidumbre y error en el proceso de toma de decisiones en cuanto a la selección de sitios para localización de rellenos sanitarios.

Barzehkar *et al*. (2019) presentan un estudio de selección de sitios para relleno sanitario usando SIG Basado en Evaluación Multicriterio, aplicado a una región SaharKhiz, Iran. Los autores proponen el reconocimiento de sitios potenciales para vertederos realizando una comparación entre la lógica Difusa y la lógica Booleana. Además, utilizan la lógica Booleana para reconocer zonas potenciales y excluidas, basadas en un sistema de valor cero y uno con ayuda de herramientas SIG. Mientras, la lógica difusa se utiliza para estandarizar las capas de información en función de su tipo (aumentando o disminuyendo) utilizando valores entre cero y uno [11]. Los autores estiman el peso final de cada capa utilizando el Proceso de Jerarquía Analítica, posteriormente, el método de ponderación de pesos integra las capas en el entorno SIG para proporcionar el mapa final de idoneidad del sitio apto para relleno sanitario. Los resultados muestran que la lógica Difusa tiene mayor flexibilidad para resolver conflictos de juicio humano y tiene una mayor precisión que la lógica Booleana en la selección de localización óptima de vertederos.

Como se evidencia en [6], [7], [8] complex, tedious, and protracted process requiring evaluation of many different criteria. This paper presents a fuzzy multicriteria decision analysis alongside with a geospatial analysis for the selection of landfill sites. It employs a two-stage analysis synergistically to form a spatial decision support system (SDSS, [9], [10], [11] en la literatura

se encuentran varios Métodos de Análisis de decisiones Multi-Criterio aplicadas a la ubicación optima de rellenos sanitarios, por lo que, en el desarrollo de esta investigación se emplea una combinación de diferentes métodos que permiten encontrar la solución óptima al objeto del presente documento, considerando las características particular del Departamento de Antioquia en Colombia.

III. MÉTODO DE EVALUACIÓN MULTICRITERIO

A. Área de Estudio

El Departamento de Antioquia se encuentra ubicado al noreste de Colombia sobre la región Andina y cuenta con un área de 63.612 km^2 aproximadamente. Limita al norte con el mar Caribe y los departamentos de Córdoba y Bolívar; por el este con Bolívar, Santander y Boyacá. Por el sur con Caldas y Risaralda y por el oeste con el departamento del Chocó. El Departamento de Antioquia se caracteriza por un relieve variado con alturas que oscilan entre los 1.000 y los 4.080 metros sobre el nivel del mar.

De igual manera, las temperaturas varían alrededor del territorio debido a su topografía; sobre las riveras del Magdalena, así como en el sector de Caucasia, las temperaturas superan los 28°C, constituyéndose en los sectores más calientes del Departamento. En el Urabá antioqueño y en el valle de río Cauca, las temperaturas oscilan entre 26 y 28°C. En el resto de la montaña antioqueña, las temperaturas dependen estrechamente de la elevación generando pisos térmicos templados y fríos [12]. La precipitación en Antioquia oscila entre los 1500 y los 4000 mm anuales, presentándose una distribución de las lluvias durante el año de tipo bimodal entre temporadas secas y lluviosas en la mayoría del departamento (véase la Fig. 1)

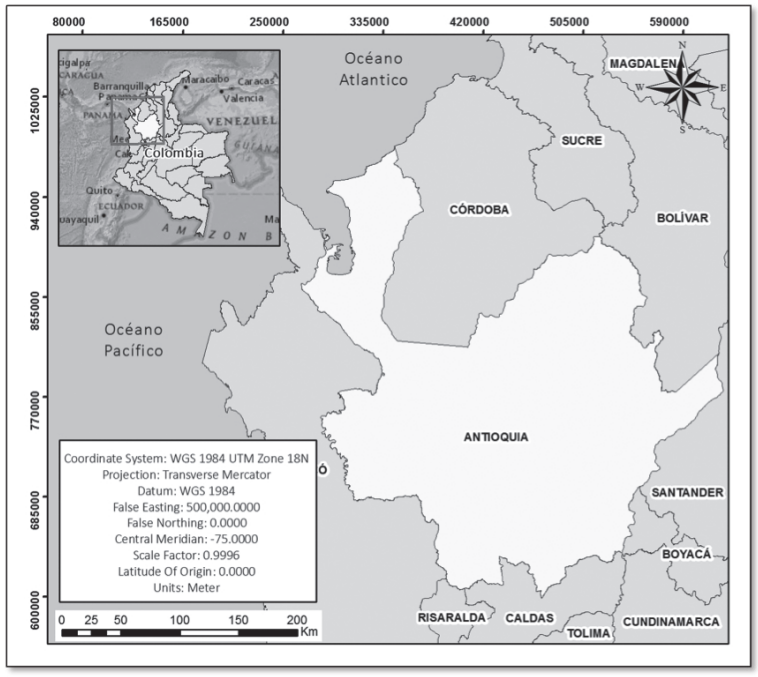


Fig. 3 *Área de Estudio – Antioquia*
Fuente: *Construcción propia*

B. Metodología propuesta

Con el objetivo de proponer un Método de Decisión Multicriterio como herramienta de apoyo para determinar zonas óptimas para la ubicación de rellenos sanitarios, se desarrollaron cinco (5) fases características de los proyectos SIG. La primera fase es la Educción de Requisitos, en la cual se realiza una caracterización del Método de Análisis de Decisiones Multi-Criterio en función de determinar zonas óptimas para la ubicación de rellenos sanitarios. En esta fase se construyen los antecedentes e identifican los requisitos y productos del proyecto.

La segunda fase es el Diseño del Proyecto, en la cual se definen el plan de adquisiciones de geodatos, la estructura de estos, la identificación de las herramientas SIG y los geoprocesos a desarrollar. Para ello, se cuenta con dos etapas, la primera determina las condiciones de las áreas aptas para la localización de los vertederos de residuos sólidos, de acuerdo con las restricciones y prohibiciones establecidas en el marco legal vigente colombiano. Y la segunda consta del diseño de evaluación multicriterio utilizando el proceso de análisis jerárquico y la programación de compromiso, identificando y

definiendo las variables a considerar en el Método de Análisis de Decisiones Multi-Criterio.

La tercera fase consiste en el Desarrollo del Proyecto, en donde se implementan los geoprocesos y la estructura de geodatos diseñados en la etapa anterior. Para ello, se requiere definir el área de estudio y la adquisición de geodatos. Para llevar a cabo esta etapa se realiza la consulta a expertos del tema y se consideran las características ambientales, físicas y socioeconómicas para ponderar la importancia relativa para cada una de las variables por medio del Proceso Analítico Jerárquico, propuesto por Thomas L. Saaty, en donde se utilizan una escala de prioridades basadas en la preferencia de un elemento sobre otro, que sintetiza los juicios emitidos por los evaluadores y entrega un ranking u ordenamiento de las alternativas de acuerdo con los pesos obtenidos (prioridades).

En tal sentido, se comparan y evalúan las variables de cada uno de los componentes a través de la construcción de una matriz cuadrada de comparación por componente, donde la diagonal principal marca valores de 1 por tener la misma importancia. Con los resultados obtenidos se construye una matriz cuadrada que compara la importancia de cada uno de los factores con los demás bajo la siguiente escala (Tabla I).

Tabla 1. Escala de ponderación de Satty para la *ponderación de la importancia relativa de las variables empleadas en el método*

Importancia	Definición	Explicación
1	Igual importancia	Dos variables contribuyen igualmente al objetivo
3	Importancia moderada	La experiencia y el juicio están moderadamente a favor de una variable sobre la otra.
5	Importancia fuerte	La experiencia y el juicio están fuertemente a favor de una variable sobre la otra
7	Importancia muy fuerte	Una variable está muy fuertemente favorecida y su dominio ha sido demostrado en la práctica.
9	Importancia extrema	Es máxima la importancia de una variable sobre la otra.
2,4,6,8	Valores intermedios entre los dos juicios contiguos	Cuando un término medio es necesario

Importancia	Definición	Explicación
Recíproco de los números de arriba	Si al elemento i le fue asignado alguno de los números de arriba al compararse con el elemento j, entonces j tiene el valor recíproco cuando se compara con el elemento i.	

Fuente: Tomado de Saaty (1970)

Luego de asignarse importancia para cada variable se procede a estimar los correspondientes pesos relativos de los criterios, que representan la importancia relativa de cada criterio o las prioridades de las diferentes alternativas respecto a un determinado criterio. La comparación de las diferentes alternativas respecto al criterio del nivel inferior de la estructura jerárquica, como la comparación de los diferentes criterios de un mismo nivel jerárquico dan lugar a una matriz cuadrada denominada matriz de decisión.

El proceso matemático preciso que se requiere para realizar la priorización de cada uno de los elementos que se comparar sigue el procedimiento de tres pasos que a continuación se detalla (i) sumar los valores de cada columna de la matriz de decisión, (ii) dividir cada elemento de tal matriz entre el total de su columna, (iii) calcular el promedio de los elementos de cada región de las prioridades relativas de los elementos que se comparan.

Una vez implementado el Método se procede a la cuarta y quinta etapa, la verificación de la calidad y el mantenimiento, con el propósito de definir e implementar las políticas de calidad de los geodatos, presentar el registro de ejecución pruebas de funcionalidad y no funcionales del proyecto SIG y la política de mantenimiento y operación del Método. En la Fig. 2 se presenta el proceso metodológico que se lleva a cabo en el presente documento.

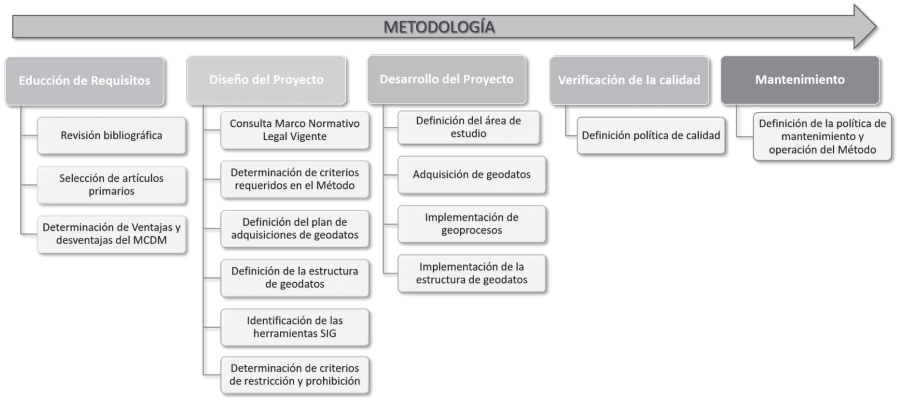


Fig. 4 Metodología
Fuente: Construcción propia

1) Entorno de geoprocesamiento

Como requisito previo a la ejecución del Método se configura el entorno de geoprocesamiento en ArcGIS 8.0, software SIG utilizado para el desarrollo de este, donde se parametrizan (i) la extensión de salida, para restringir específicamente el proceso al área de estudio. Seguidamente, se establece (ii) el tamaño de celda o resolución igual a 30m y (iii) una máscara con el propósito de que el análisis solo ocurra en el límite político-administrativo de Antioquia. En la Fig. 2 se presenta la ventana configuración del entorno de geoprocesamiento del Método.

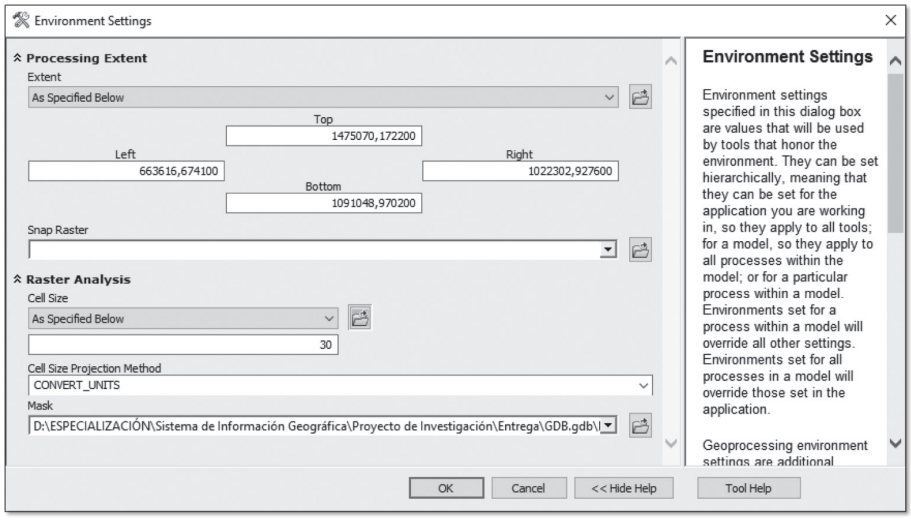


Fig. 5 Configuración de entorno de Geoprocesamiento

C. Adquisición de datos

En esta etapa se realiza la adquisición de los geodatos requeridos para proyecto, los cuales consisten en información secundaria de diferentes entidades oficiales a nivel regional y nacional, tales como: las Corporaciones Autónomas Regionales de la jurisdicción, la Gobernación de Antioquia, el Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM, el Instituto de Investigación de Recursos Biológicos Alexander Von Humboldt – IavH, el Instituto Geográfico Agustín Codazzi – IGAC y Sistema de Información Ambiental de Colombia – SIAC.

Es importante resaltar que los geodatos difieren en lo que respecta a temporalidad y nivel de detalle debido al Estado de Emergencia Económica, Social y Ecológica (COVID–19) que se presentaba al momento del desarrollo de la presente investigación, lo cual dificultó la gestión y acceso a los mismos. Sin embargo, los datos adquiridos no fueron editados para homogenizar tales diferencias teniendo en cuenta que la información fue tomada de fuentes secundarias.

Por otra parte, esta investigación tiene como objetivo principal **proponer un método**, que se espera sea un marco de referencia para futuros estudios. Por lo anterior, para este estudio no se generan resultados que puedan ser utilizados como insumo para la generación de planes de desarrollo de rellenos sanitarios.

En la Tabla I se presenta la fuente y descripción de los geodatos adquiridos (variables) para el desarrollo de la propuesta de un Método de Evaluación Multicriterio como herramienta de apoyo para determinar zonas óptimas para la ubicación de rellenos sanitarios, utilizando Sistemas de Información Geográfica.

Tabla 2. *Geodatos adquiridos para el Método de Análisis Multi - Criterio*

Shape/Raster	**Descripción**
Amenazas (*EnvThreat*)	Mapa nacional integrado de amenaza por movimiento en Masa a escala 1:100.000 generado por Servicio Geológico Colombiano.
Acuíferos (*Aquifer*)	Esta capa de información contiene la localización de Sistemas Acuíferos de Colombia del Estudio Nacional de Aguas - ENA2014, realizado por el Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM a escala 1:100.000
Aeropuertos (*Airport*)	La información es tomada de la cartografía IGAC (2017) a escala 1:25.000 y corroborada a partir del Índice de Aeródromos y Helipuertos de la Aerocivil de Colombia
Antioquia (Antioquia)	Delimitación político – administrativa del departamento de Antioquia definidos por la Comisión Conjunta "Área Metropolitana del Valle de Aburrá, Corantioquia, Cornare, Corpouraba y Gobernación de Antioquia a escala 1:100.000.
Áreas de Manejo Especial (*SpManagArea*)	Es la intersección de las siguientes capas de información: Límites de páramo propuestos en el Atlas de Páramos Morales realizados por el Instituto de Investigación de Recursos Biológicos Alexander von Humboldt -IavH Humedales, Parques Naturales Nacionales y Parques Naturales Regionales, Manglares, Reserva Forestal 2nda, Resguardo Indígena, tomados del Sistema de Información Ambiental de Colombia -SIAC a escala 1:100.000.
Cabeceras municipales (*HumanSettl*)	Delimitación de los más relevantes centros poblados del departamento de Antioquia sean o no cabeceras municipales, en escala 1:25.000. Definidos por la Comisión Conjunta "Área Metropolitana del Valle de Aburrá, Corantioquia, Cornare, Corpouraba y Gobernación de Antioquia
Coberturas del suelo (*LandCover*)	Elaboración propia basado en los usos de suelo del Instituto Geográfico Agustín Codazzi- Subdirección de Agrología – Grupo Interno de Trabajo Geomática a escala 1:1000.000
Composición de la tierra (*LandComposit*)	Shape temático que representa la distribución de las características del suelo, determinadas mediante el levantamiento general de suelos del departamento de Antioquia a escala 1:100.000, publicado el año 2007. Suministra información importante acerca del recurso suelo, a través de la descripción e interpretación de sus ambientes edafogenéticos, sus características físicas, químicas, mineralógicas y morfológicas, su taxonomía y distribución espacial, como base para la determinación de sus potenciales productivos, describiendo las limitantes de uso. Información realizada por el Instituto Geográfico Agustín Codazzi - IGAC

Shape/Raster	**Descripción**
Paisaje (*LandScape*)	Shape temático que representa las características del paisaje, determinadas mediante el levantamiento general de suelos del departamento de Antioquia a escala 1:100.000, publicado el año 2007. Información realizada por el Instituto Geográfico Agustín Codazzi - IGAC
Curvas de nivel (*LevelCurve*)	Curvas de Nivel escala 1:100.000 generadas a partir de la restitución vectorial dentro de convenio interadministrativo Cartografía Antioquia - Instituto Geográfico Agustín Codazzi
Drenaje Doble (*Doble_Waterbody*)	Drenaje Doble escala 1:500.000 de Antioquia generadas a partir de restitución vectorial. Información realizada por el Instituto Geográfico Agustín Codazzi y entregada a la Gobernación de Antioquia
Drenaje Sencillo (*Simple_WaterBody*)	Drenaje Sencillo escala 1:500.000 generadas a partir de la restitución vectorial. Información que realizada por el Instituto Geográfico Agustín Codazzi y entregada a la Gobernación de Antioquia
Embalse (*Reservoir*)	Embalses escala 1:100.000 generadas a partir de restitución vectorial. Información que realizada por el Instituto Geográfico Agustín Codazzi y entregada a la Gobernación de Antioquia
Escorrentía (*AverageRunoff*)	Información de las curvas de Escorrentía Promedio 1:100.000 del departamento de Antioquia y zonas aledañas. Cartografía realizada en el Proyecto "Un Modelo para la Administración de la Demanda del Recurso Hídrico Superficial en la Jurisdicción de Corantioquia (SGDH - DUBERDICUS)" Corporación Autónoma Regional del Centro de Antioquia "Corantioquia" - Subdirección de Recursos Naturales.
Fallas geológicas (*GeoFlaw*)	Información de las Estructuras Geológicas 1:400.000 de la jurisdicción de Corantioquia. Cartografía generada por la Universidad Nacional de Colombia, tomando como fuente las cartas análogas del INGEOMINAS
Hidrocarburos (*HidrocarExplo*)	Detalla el estado de la actividad hidrocarburífera del país, en él se muestran las áreas que se encuentran en Exploración y Explotación y las áreas disponibles para la celebración de Contratos en las diferentes cuencas sedimentarias del país. Elaborado por la Agencia Nacional de Hidrocarburos a escala 1:25.000
Usos de suelo (*LandUse*)	Establece el mejor uso de las tierras, donde se analizan y evalúan una serie de características biofísicas estables en el tiempo y en el espacio; que influyen en la selección y desempeño de los usos agropecuarios y forestales, principalmente, con requerimientos implícitos de protección y conservación de los recursos naturales. Información tomada del Instituto Geográfico Agustín Codazzi-Subdirección de Agrología - Grupo Interno de Trabajo Geomática a escala 1:100.000

Shape/Raster	**Descripción**
Títulos mineros (*Mining*)	Información correspondiente a los Títulos Mineros y las Solicitudes de Títulos Mineros tramitadas ante el Instituto Colombiano de Geología y Minería-INGEOMINAS para todo el territorio nacional a escala 1:100.000
Vías (*Road*)	Vías es el producto cartográfico básico actualizado a escala concertada (1:25.000), que contiene elementos altimétricos y planimétricos de cartografía existente que se editan de acuerdo con la interpretación de objetos en una imagen ortorectificada provenientes de plataformas satelitales o aerotransportadas. Generado por el Instituto Geográfico Agustín Codazzi - IGAC

*La denominación entre paréntesis bajo el nombre de cada Shape, corresponde al nombre nemónico utilizado en modelos de geoprocesamiento.

Fuente: Construcción propia

IV. RESULTADOS

A. Diseño del método

Para determinar la ubicación optima de los rellenos sanitarios se deben cumplir dos etapas. La primera consta de la superposición de mapas que contemplan las prohibiciones y restricciones estipulados en el Decreto 838 de 2005 del Ministerio de Ambiente, Vivienda y Desarrollo Territorial, para posteriormente cumplir con la segunda etapa que desarrolla el Método de Análisis de Decisiones Multi-Criterio.

1) Primera etapa

A continuación, se exponen las clases temáticas que se tienen en cuenta para la definición de prohibiciones asociadas a la ubicación de áreas para rellenos sanitarios, las cuales corresponde a los criterios determinados en el artículo 6 del Decreto 838 de 2005, "por el cual se modifica el Decreto 1713 de 2002 sobre disposición final de residuos sólidos y se dictan otras disposiciones" [19]:

Cuerpos de agua. Distancia de treinta (30) metros de ancho de la faja paralela del cauce permanente de drenajes sencillos y embalses; cien (100) metros para drenajes dobles.

Fuentes subterráneas: Zonas de recarga de acuíferos.

Áreas con fallas geológicas. Distancia de (60) metros de zonas de la falla geológica.

Áreas pertenecientes al Sistema de Parques Nacionales Naturales y demás áreas de manejo especial como resguardos indígenas y de ecosistemas especiales como humedales, páramos y manglares.

Adicionalmente, se contemplan ciertas especificaciones y requisitos

particulares, que se deben cumplir para la ubicación de rellenos sanitarios. Lo anterior, se constituye como las restricciones enmarcadas en el Decreto 838 de 2005.

Proximidad a aeropuertos y aeródromos: se debe respetar una distancia de 13 km a la redonda con respecto a aeropuertos y aeródromos [20].

Zonas de amenaza por inestabilidad del terreno o inundación: se excluyen zonas de alta amenaza por inestabilidad del terreno o por inundación.

Zonas de riesgo sísmico, se excluyen zonas de alto riesgo sísmico.

2) Segunda etapa

En la segunda etapa se utilizan criterios de evaluación local en los componentes ambientales, socioeconómico y técnico-operativo. Estos criterios son tomados de la revisión bibliográfica y de los parámetros y metodología establecidos en el Decreto 838 de 2005. En la Tabla II se relacionan las variables a considerar:

Tabla 3. *Variables de análisis de decisiones multi – criterio*

	ID	VARIABLE	DESCRIPCIÓN
AMBIENTAL	A1	Deterioro en la calidad del agua superficial	Relación que tendrá el área en la que se efectúa la disposición final de residuos, respecto a las fuentes hídricas permanentes y superficiales
	A2	Deterioro a la calidad del agua subterránea, infiltración	Relación que tiene el área en la que se efectúa la disposición final de residuos, respecto a los acuíferos y zonas de recarga
	A3	Cobertura del suelo	Características ecológicas
	A4	Molestias por olor	Incidencia que puede tener la dirección de los vientos con respecto al casco urbano, en la operación de la infraestructura instalada en el área en que se efectúa la disposición final de residuos.
	A5	Proximidad a lugares protegidos	Hace referencia al impacto que se genera en las áreas pertenecientes al Sistema de Parques Nacionales Naturales y demás áreas de manejo especial y de ecosistemas especiales, tales como: humedales, páramos y manglares.
	A6	Ausencia de intrusión óptica	Geoformas del área respecto al entorno, es decir, la incidencia que puede tener el relleno sanitario sobre el paisaje y el entorno

	ID	VARIABLE	DESCRIPCIÓN
AMBIENTAL	A7	Composición del suelo	Presencia de arcillas, limos, arenas
	A8	Escorrentía promedio	Precipitación promedio
	A9	Usos del suelo	Uso principal del suelo
	A10	Pendientes	Pendiente del terreno
SOCIOECONÓMICO	B1	Uso económico de la tierra y valor social que consiste en el valor presente neto	Determina las actividades que actualmente se vienen realizando con el objeto de prever posibles impactos sobre la comunidad
	B2	Depreciación	Es la incidencia del valor en los predios aledaños al relleno sanitario y el desarrollo económico.
	B3	Distancia de los asentamientos humanos	Distancia entre el perímetro urbano, respecto del área para la disposición final de residuos sólidos
	B4	Explotación o reserva industrial y mineral	Títulos mineros otorgados cercanos al área de rellenos sanitarios.
	C1	Carreteras de acceso	Corresponde a la facilidad y economía que la persona prestadora de servicio público de aseo en el componente de recolección y transporte
	C2	Disponibilidad de material de cobertura	Se asocia a los costos de transporte en que incurre la persona prestadora del servicio para obtener y llevar el material de cobertura necesario para dar cumplimiento a las especificaciones técnicas y ambientales en los procesos de operación diaria y cierre y clausura del relleno sanitario.
	C3	Evaluación hidrológica	Infraestructura del sistema de drenaje a interceptar y desviar las aguas lluvias que caen o penetran al predio del relleno
	C4	Distancia desde la fuente principal de consumos de residuos	Distancia hacia los centros urbanos y hacia las agrupaciones pobladas
	C5	Pendientes promedio del terreno	Facilidades de construcción y operación por las condiciones de suelo y topografía

Fuente*: Tomado del Decreto 838 de 2005* [13] *y Vatalis y Manoliadis (2002)* [6].

La calificación de los indicadores anteriores se realiza de acuerdo con el desempeño de los criterios de la siguiente manera: Muy bueno (5), Bueno (4), Medio (3), Malo (2), Muy malo (1). La metodología propuesta utiliza técnicas de clasificación multicriterio para combinar los datos del sitio disponibles, a menudo conflictivos, en un sitio general final.

B. Desarrollo del método

1) Primera Etapa

La primera etapa consiste en hallar las restricciones y prohibiciones dentro del área de estudio para el establecimiento de las áreas aptas para la localización de rellenos sanitarios. Para ello, se contrastan los criterios de prohibición y restricción definidos en la etapa de diseño por medio de lógica booleana.

Bajo este método se clasifican cómo áreas prohibidas aquellas superficies que pertenecen a la faja paralela de los cauces permanentes y embalses, aquellas que están localizadas sobre acuíferos y sobre fallas geológicas, además de las áreas de manejo especial. En la Fig. 4 se presenta el mapa resultante de la superposición de los criterios de prohibición y el Fig. 5 se detalla el geoproceso realizado para llegar a este.

Para la determinación de las áreas restringidas se cuantifican aquellas superficies cerca a los centros urbanos, aeropuertos y las zonas que cuentan con una calificación alta o muy alta en amenazas por estabilidad de terreno o inundación y riesgo sísmico. En la Fig. 6 se muestra el mapa resultante de la superposición de los criterios de restricción y el Fig. 7 se detalla el geoproceso realizado para llegar a este.

Al sobreponer las superficies resultantes de las áreas prohibidas y áreas restringidas se obtiene el mapa de las zonas aptas para la construcción de relleno sanitario, las cuales corresponden a aquellos espacios con valor de cero (0), como consecuencia de no cumplir ninguno de los criterios de prohibición o restricción establecidos. En la Fig. 9 se presenta el mapa de áreas aptas para la construcción de rellenos sanitarios y el Fig. 8 se detalla el geoproceso realizado para llegar a este.

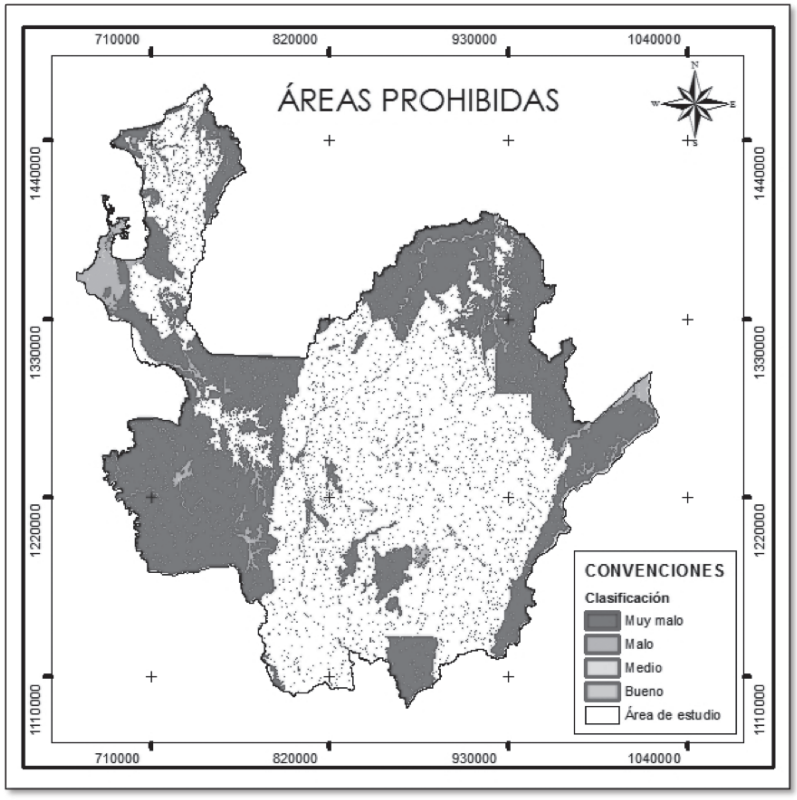


Fig. 6 *Áreas prohibidas.*
Fuente: *Construcción propia*

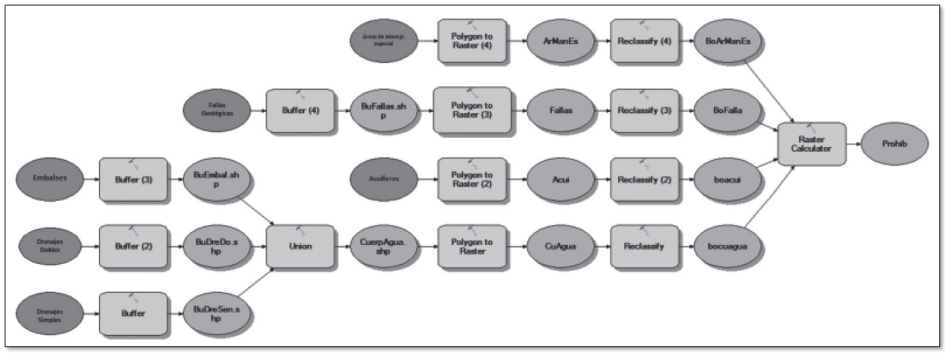


Fig.7 *Geoproceso para determinar áreas prohibidas*
Fuente: *Construcción propia*

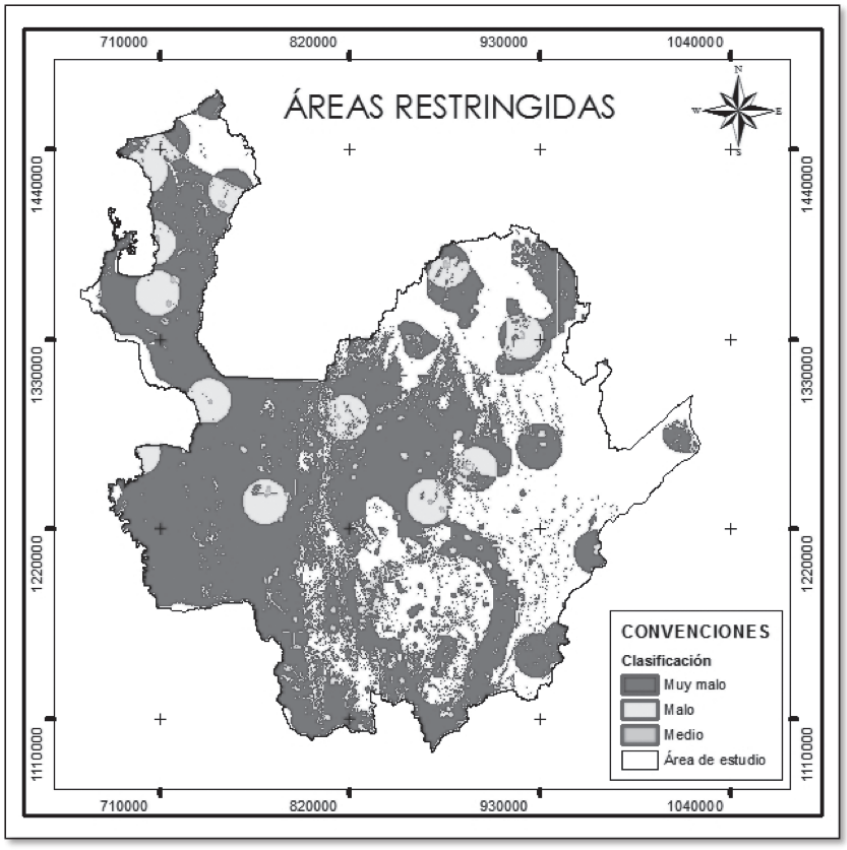


Fig. 8 *Áreas restringidas*
Fuente: *Construcción propia*

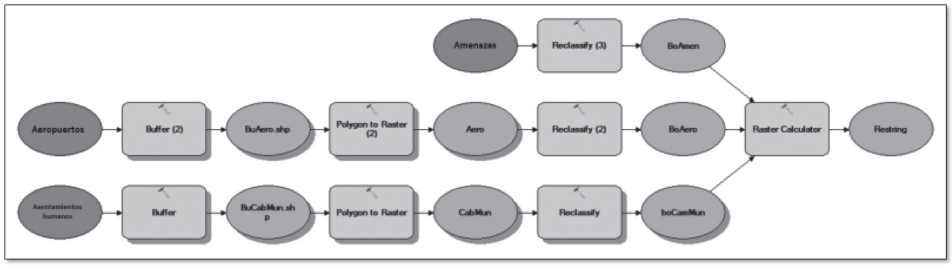


Fig. 9 Geoproceso para determinar áreas restringidas
Fuente: *Construcción propia*

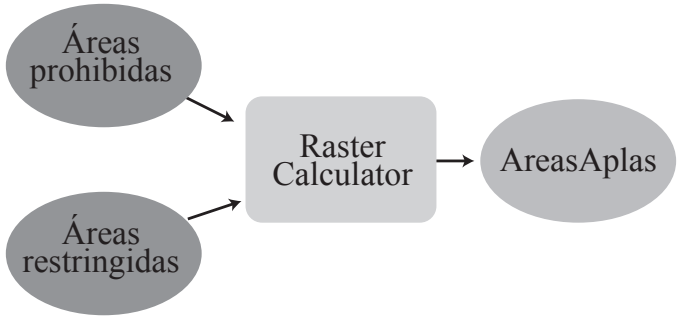


Fig. 10 Geoproceso para determinar áreas aptas para la construcción de rellenos sanitarios.
Fuente: Construcción propia

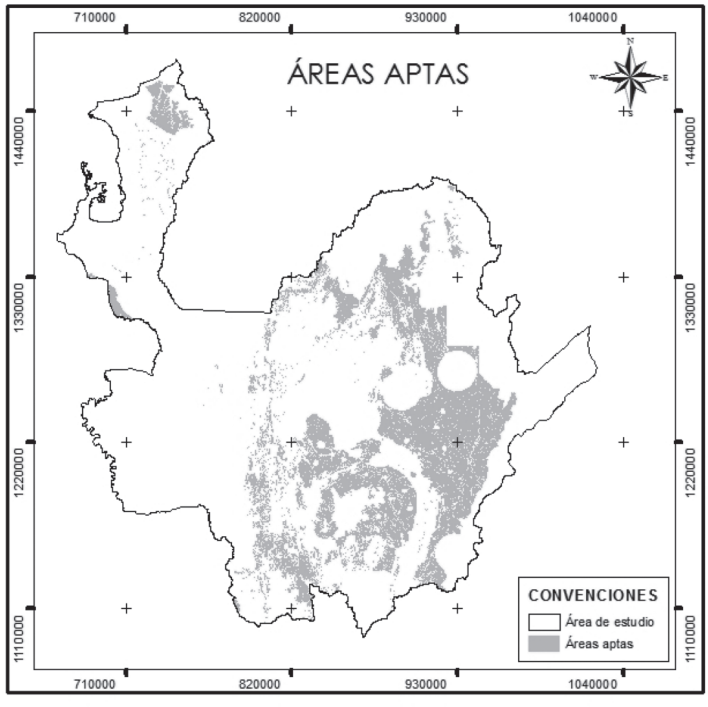


Fig. 11 Áreas aptas para la construcción de rellenos sanitarios
Fuente: Construcción propia

2) Segunda etapa

En la segunda etapa se procede a aplicar el Método de Decisión Multi-criterio evaluando el peso de las variables (Tabla II). Estas se ponderan según el rango de evaluación presentada anteriormente (1-5, Muy malo – Muy bueno respectivamente) mediante el uso de herramientas de apoyo en ArcGIS; como se muestra a continuación:

A1, Deterioro en la calidad del agua superficial: Se clasificaron las distancias a los drenajes tomando como información de entrada la capa de Buffer de drenajes de la primera fase, bajo la lógica de que entre mayor distancia hay mayor potencial ambiental para establecer el vertedero.

A2, Deterioro a la calidad del agua subterránea: Se clasificaron las distancias a los acuíferos, donde entre mayor distancia hay mayor potencial ambiental para establecer el vertedero.

A3, Cobertura del suelo (Tabla III):

Tabla4. *A3 Cobertura del suelo*

Criterio	Ponderación
Conservación, Cuerpos de agua, Zonas urbanas	1
Forestal	2
Agroforestal, agrícola	3
Silvopastoril	4
Ganadería	5

Fuente: Construcción propia

A4, Molestias por olor: se clasificaron las distancias a los asentamientos humanos (cabeceras municipales) donde entre mayor distancia hay mayor potencial ambiental para establecer el relleno sanitario.

A5, Proximidad a lugares protegidos: se clasificaron las distancias a las áreas definidas como protegidas donde entre mayor distancia hay mayor potencial ambiental para establecer el relleno sanitario.

A6, Ausencia de intrusión óptica: se clasificó el peso considerando el lugar donde según el paisaje, la ubicación del relleno sanitario causaría menor o mayor intrusión óptica (Tabla IV).

Tabla 5. *A6 Ausencia de intrusión óptica*

Criterio	Ponderación
Zona urbana	1
Planicie aluvial, Planicie fluvio lacustre	2
Lomerío, Misceláneo rocoso, Montaña	3
Piedemonte, Valle aluvial	4
Altiplanicie	5

Fuente: *Construcción propia*

A7, Composición del suelo: considerando la permeabilidad del suelo, fueron clasificados los sitios apropiados para la ubicación del relleno en pro del cuidado del suelo en cuanto a la infiltración de lixiviados

(Tabla V):

Tabla 6. *A7 Composición del suelo*

Criterio	Ponderación
Zona urbana, cuerpos de agua	1
Gravas	2
Arenas	3
Limos	4
Arcillas	5

Fuente: *Construcción propia*

A8, Escorrentía promedio: se rasterizaron las isolíneas correspondientes a escorrentía promedio, clasificando la interpolación resultante bajo la lógica de que entre menor escorrentía promedio hay mayor potencial ambiental para establecer el relleno sanitario.

A9, Usos del suelo (Tabla VI): los suelos con condiciones más apropiadas para la ubicación del relleno sanitario obedecen a los más deteriorados. Dado que el impacto sobre el suelo existe previamente, la afectación del relleno no propiciará un impacto no sea significativo desde el punto de vista ambiental.

Tabla 7. *A9 Usos del suelo*

Criterio	Ponderación
Zonas urbanas, Forestal de protección, Protección-producción, Cuerpo de agua, Conservación de suelo y recursos hídricos	1
Forestal de producción	2
Cultivos	3
Agrosilvícola y agropastoril con cultivos	4
Silvopastoril y pastoreo	5

Fuente: *Construcción propia*

A10, Pendientes (Tabla VII): se determina la facilidad de construcción, operación y desarrollo en el área en que se efectuará la ubicación del relleno, por lo tanto a menor pendiente mejores condiciones, evitando impactos ambientales significativos debidos al transporte y movilización de vehículos en las etapas de construcción y operación del relleno.

Tabla 8. *A10 PENDIENTES*

Criterio	Ponderación
0-12%	5
12%-25%	4
25%-50%	3
50%-75%	2
Mayor a 75%	1

Fuente: *Construcción propia*

B1, Uso económico de la tierra (Tabla VIII): criterio asociado a la afectación a la economía del sector donde se ubicará el relleno debido a su construcción y operación, por tal motivo las mejores condiciones se presentan en lugares donde el suelo no tiene valor económico.

Tabla 9. *B1 Uso económico de la tierra*

Criterio	Ponderación
Agropastoril, Cuerpo de agua, Agrícola con cultivos permanentes, Conservación de recursos hídricos, Zonas urbanas	1
Agroforestal, Agrícola con cultivos transitorios, Ganadería	2
Forestal	3
Conservación de recursos hidrobiológicos	4
Conservación de suelos	5

Fuente: *Construcción propia*

B2, Depreciación (Tabla IX): corresponde a la depreciación del suelo respecto al valor económico de la misma por lo que se considera mejores condiciones en suelos deteriorados, dado su deterioro existente, la depreciación se prevé menor.

Tabla 10. *B2 Depreciación*

Criterio	Ponderación
Cuerpo de agua, Conservación de recursos hídricos e hidrobiológicos, Zonas urbanas	1
Agrícola, conservación de suelos	2
Agroforestal, agrícola	3
Forestal	4
Ganadería, silvopastoril	5

Fuente: *Construcción propia*

B3, Distancia de los asentamientos humanos: se clasifican las distancias a los asentamientos humanos (cabeceras municipales) donde entre mayor distancia hay mayor potencial social para establecer el relleno sanitario.

B4, Explotación o reserva industrial y mineral (Tabla X): se prevé mejores condiciones para la localización de rellenos sanitarios en sitios donde no existe explotación minera o de hidrocarburos.

Tabla 11. *B4 Explotación o reserva industrial y mineral*

Criterio	Ponderación
Zona urbana, Presencia Título minero y explotación de hidrocarburos	1
Títulos mineros	2
Zona de hidrocarburo en etapa de producción	3
Zona de hidrocarburo en etapa de exploración	4
Zona sin título minero ni reserva de hidrocarburos	5

Fuente: *Construcción propia*

C1, Carreteras de acceso; se clasifican las distancias a las vías donde entre menor distancia hay mayor potencial económico para establecer el relleno sanitario.

C2, Disponibilidad de material de cobertura (Tabla XI): se considera que a mayor disponibilidad de material para la construcción y operación del relleno las condiciones del terreno son mejores desde un punto de vista económico.

Tabla 12.*C2 Disponibilidad de material de cobertura*

Criterio	Ponderación
Presencia de ceniza volcánica y zonas urbanas	1
Depósitos aluviales, Cuerpos de agua y sedimentos mixtos	2
Rocas metamórficas e ígneas con presencia de granitos, cuarzodioritas, granodioritas, diabasas, basaltos, tobas, dioritas y andesitas con cobertura de ceniza volcánica	3
Sedimentos con gravillas y coluviones, Rocas sedimentarias (arcillolitas, areniscas, conglomerados), Rocas metamórficas (neisseses e inclusiones de esquistos, grabos), depósitos coluviales y aluviales	4
Rocas sedimentarias, metamórficas e ígneas con presencia de arcillolitas, areniscas, conglomerados, Miscelaneo rocoso, Depósitos mixtos (limos y arenas) mezclados con materia orgánica, Aluviones finos, medios y gruesos.	5

Fuente: *Construcción propia*

C3, Evaluación hidrológica: se clasifican las distancias a los drenajes bajo la lógica de que entre mayor distancia hay mayor potencial económico para establecer el vertedero.

C4, Distancia desde la fuente principal de consumos de residuos: se clasifican las distancias a los asentamientos humanos (cabeceras municipales) donde entre menor distancia hay mayor potencial económico para establecer el relleno sanitario.

C5, Pendientes promedio del terreno: se consideran que entre menor pendiente existe mayor potencial económico para establecer el relleno sanitario.

El Proceso Analítico Jerárquico modela el problema estableciendo prioridades mediante una consistencia lógica de juicio por pares. Para llevar a cabo el análisis de las variables aplicando el PAJ se consulta con expertos utilizando como herramienta el Software SATTY [1], el cual partiendo del método de sumatoria y probatoria estima el porcentaje de importancia para cada variable. En la Tabla XII la ponderación de las variables de acuerdo con los resultados obtenidos en la consulta SATTY-

Tabla 13. *Ponderación Variables Método SATTY*

VAR	PON	VAR	PON	VAR	PON
A1	0.15	B1	0.41	C1	0.19
A10	0.08	B2	0.20	C2	0.20
A2	0.21	B3	0.30	C3	0.25
A3	0.10	B4	0.08	C4	0.18
A4	0.07			C5	0.19
A5	0.08				
A6	0.06				
A7	0.09				
A8	0.07				
A9	0.10				

**Var: Variable; Pon: Ponderación*

Fuente: *Construcción propia*

Tales porcentajes son utilizados como factor en la suma de los mapas obtenidos del Método de Decisión Multi-criterio por cada criterio. A continuación, se presenta la estimación obtenida para los mapas: el color rojo representa las áreas con peso mayor y el color verde representa el peso menor con respecto a las características ambientales, sociales y económicas aptas para el establecimiento de rellenos sanitarios, respectivamente. En la Fig. 10, Fig. 11 y Fig. 12 se presenta el geoproceso realizado en el Proceso Analítico Jerárquico por componente y en la Fig. 13, Fig. 14 y Fig. 15 se presentan los resultados obtenidos en cada uno de los geoprocesos.

1 Herramienta diseñada por Julián D. Giraldo. Dirección de Investigaciones de la Corporación Universitaria Remington, 2014.

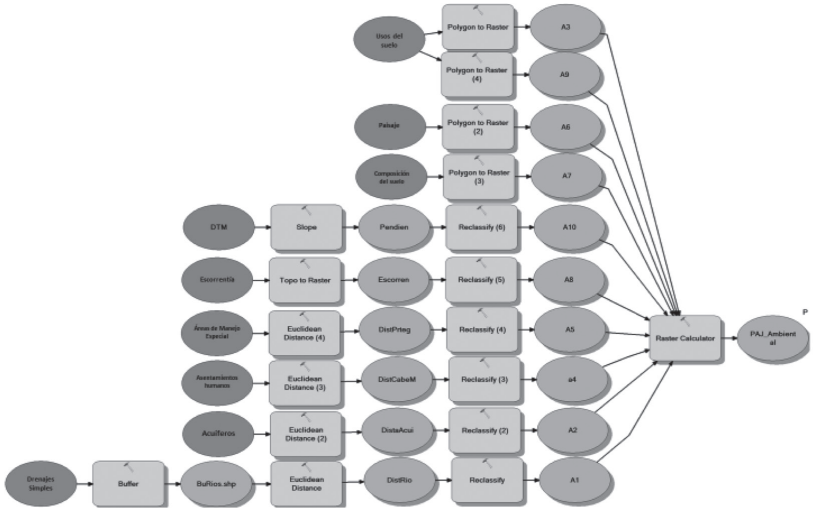


Fig. 12 *Proceso Analítico Jerárquico del Componente Ambiental*
Fuente: *Construcción propia*

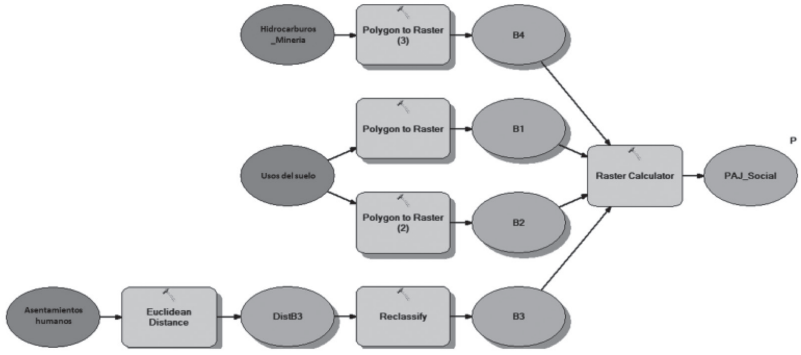


Fig. 13 *Proceso Analítico Jerárquico del Componente Social*
Fuente: *Construcción propia*

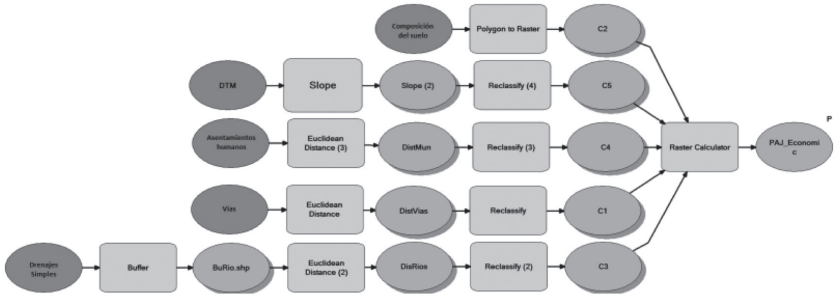


Fig. 14 *Proceso Analítico Jerárquico del Componente Económico*
Fuente: *Construcción propia*

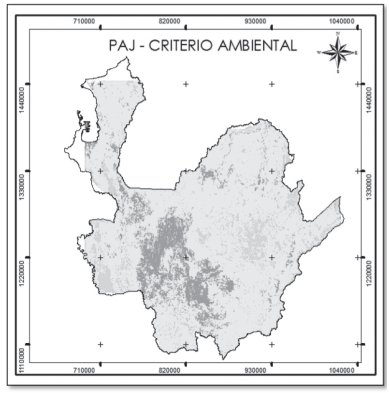


Fig. 15 *Resultado PAJ criterios Ambiental*
Fuente: *Construcción propia*

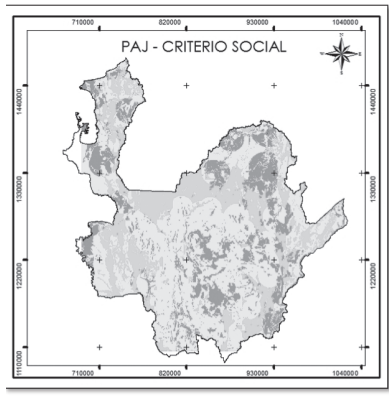


Fig. 16 *Resultado PAJ criterios Social*
Fuente: *Construcción propia*

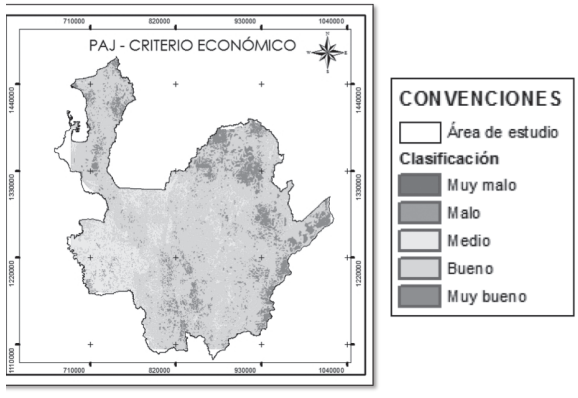


Fig. 17 *Resultado PAJ criterios Económicos*
Fuente: *Construcción propia*

Posteriormente, según el Método de Programación por Compromiso se realiza la sumatoria de los mapas resultado del Proceso Analítico Jerárquico con el fin de generar las áreas en dónde la unión de los criterios ambiental y socioeconómico tienen un peso máximo. Tras la generación del análisis anterior se seleccionaron las áreas clasificadas como "Muy bueno" (evaluación del 1-5) sobreponiéndolas con el resultado de la primera etapa (Áreas aptas) obteniendo de este modo las Áreas Óptimas, es decir, sitios en donde existen condiciones de prohibición y restricción nulas, y condiciones ambientales, sociales y económicas de máximo peso. En la Fig. 16 se detalla el geoproceso realizado para la determinación de las áreas óptimas y en la Fig. 17 el mapa resultante.

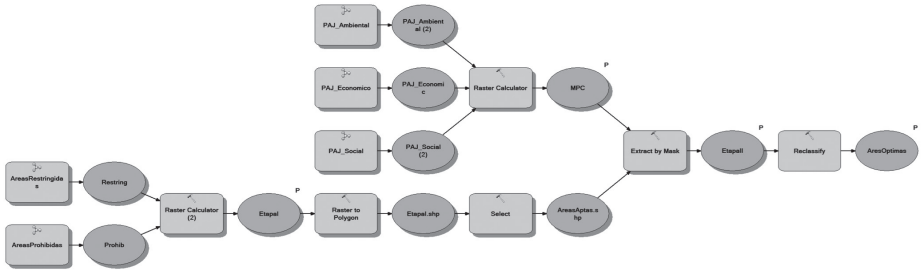


Fig. 18 *Geoproceso para determinar las Áreas óptimas*
Fuente: *Construcción propia*

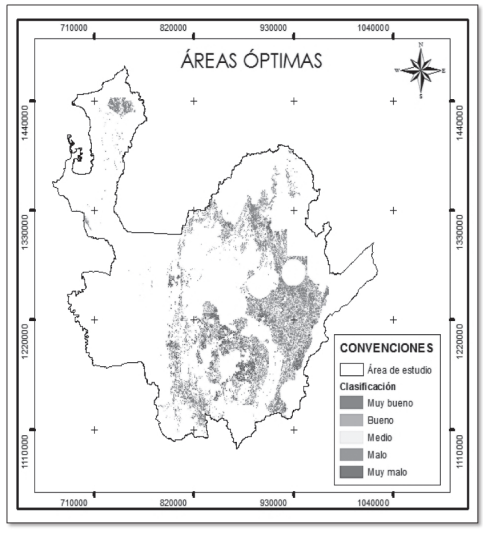


Fig. 19 *Áreas óptimas*
Fuente: *Construcción propia*

Finalmente, se estima la capacidad de los sitios generados en términos de área. Donde se identifica que el relleno sanitario la pradera, específicamente para el vaso La Música son dispuestas alrededor de 20 ha, en este sentido, se decide establecer como área mínima un valor de 30 hectáreas teniendo en cuenta el crecimiento acelerado de la población, el consumo, por tanto, el aumento de residuos sólidos respecto al tiempo. Sin embargo, se aclara que este análisis se realiza de forma cualitativa y debe ser reestructurado cuantitativamente dependiendo de lo encontrado en campo y las estimaciones realizadas en relación con el crecimiento poblacional, la generación de residuos por día y el volumen anual de residuos no aprovechables para futuros estudios. Tras restringir las áreas óptimas a 30 ha se genera el mapa de Áreas Finales (ver Fig. 18), es decir; áreas aptas y óptimas para el establecimiento de rellenos sanitarios

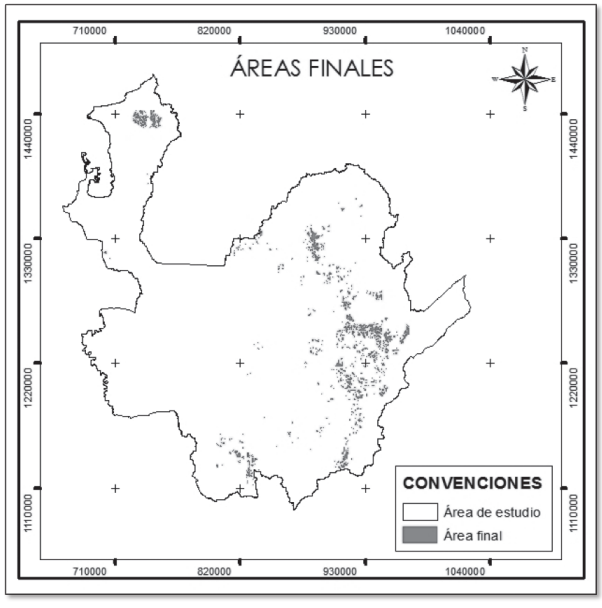


Fig. 20 *Áreas Finales*
Fuente: *Construcción propia*

3) Estructura de almacenamiento de Geodatos

El almacenamiento de los geodatos se divide en tres grupos, información primaria referente a las capas tipo vector y ráster colectadas para llevar a cabo los geoprocesos posteriores. Primera etapa que corresponde a la fase denotada con el mismo nombre y que contiene los Ráster de análisis Booleano

con sus respectivos resultados, y la capa de Áreas Aptas; resultado final de la primera etapa (Vector). Segunda etapa que corresponde a la fase denotada con el mismo nombre y que contiene los Ráster de análisis por el Método de Decisión Multi-criterio, Proceso Analítico Jerárquico y Método de Programación Por Compromiso con sus respectivos resultados, y las capas tipo vector correspondientes a las Áreas Óptimas, finalmente, el resultado de la segunda etapa: Áreas Finales, como se presenta en la Fig. 19:

C. Verificación de la calidad

Bajo el ciclo Planear, Hacer, Verificar y Actuar – PHVA (ver Fig. 20) se busca la mejora continua en el Método de Evaluación Multicriterio propuesto en el presente trabajo como herramienta de apoyo para determinar zonas óptimas para la ubicación de rellenos sanitarios, utilizando Sistemas de Información Geográfica, por lo que se deberá realizar la valoración de los procesos para confirmar que los resultados obtenidos son los esperados. Para ello se requiere realizar una validación en campo de las áreas optimas, proponer nuevos ajustes al método de acuerdo con las necesidades locales del territorio y realizar los estudios definidos en el marco legal vigente colombiano. Así mismo, se debe involucrar entidades territoriales, ambientales y sociales, como interesados en el establecimiento de un relleno sanitario, para buscar un entorno colaborativo que logre satisfacer las necesidades y expectativas. El usuario debe considerar las condiciones del Estado de Emergencia Económica, Social y Ecológica decretado por el Gobierno Colombiano, con el fin de enfrentar la pandemia del coronavirus COVID – 19 en las que se desarrollan el presente trabajo y que por obvias razones dificultan la adquisición de datos.

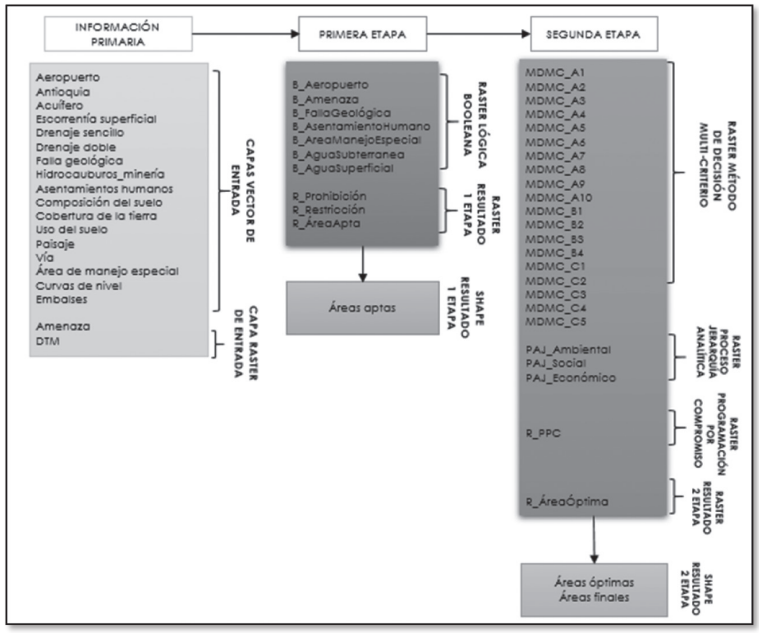


Fig. 21 *Estructura de la Geodatabase*
Fuente: *Construcción propia*

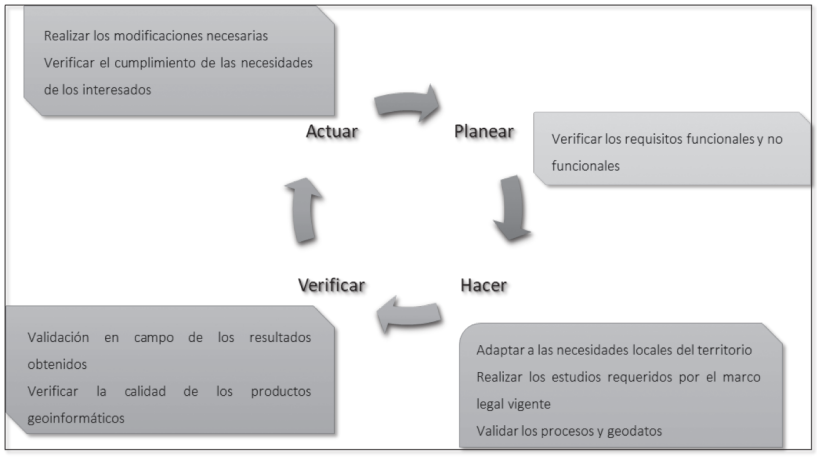


Fig. 22 *Política de calidad de los geodatos*
Fuente: *Construcción propia*

D. Mantenimiento

Con el fin de conservar el Método de Evaluación Multicriterio propuesto en el presente trabajo como herramienta de apoyo para determinar las zonas óptimas para la ubicación de rellenos sanitarios en buen estado y evitar su degradación, se debe realizar un plan de actualizaciones para responder efectivamente a los cambios de los geodatos evitando: (i) perder la validez temporal es decir, el grado de realidad en la escala del tiempo de los elementos existentes en la base de datos, y sus relaciones temporales con respecto a las especificaciones del producto, (ii) perder exactitud temática, el cual corresponde al grado de fidelidad de los valores de los atributos asignados a los elementos en la base de datos con respecto a su verdadera característica en el mundo real y la clasificación correcta de los objetos y sus relaciones de acuerdo con las especificaciones del geodato y (iii) la exactitud de posición, para describir la cercanía en posición de los objetos en el producto, con respecto a sus posiciones verdaderas (o asumidas como verdaderas) [14].

De igual manera, el mantenimiento debe incluir un plan de soporte que registre el control de las modificaciones y la documentación del proyecto donde se especifiquen las pruebas de funcionamiento y las soluciones planteadas y ejecutas para corregir los problemas hallados. También, debe incluir la gestión y adquisición de los recursos de acuerdo con las condiciones locales del territorio. La presente política de mantenimiento permite al usuario responder efectivamente a los cambios, mantener las medidas de éxito, asegurar la viabilidad del Método de Evaluación Multicriterio en el futuro y la satisfacción de los interesados

V. DISCUSIÓN

La aplicación del Método de Análisis de decisión Multicriterio propuesto es una herramienta que permiten planificar y desarrollar estratégicamente el territorio y la adecuada disposición final de los residuos sólidos municipales, previniendo, minimizando y mitigando los impactos en la salud y en el ambiente. Lo cual constituye una opción para dar solución a la problemática nacional por la inadecuada gestión de residuos sólidos en muchos municipios de Colombia.

Dadas las condiciones de análisis restrictivo que posee la lógica booleana para la interpretación de datos, se recomienda usar este método para realizar geoprocesos como la determinación de áreas aptas para la localización de los vertederos de residuos sólidos, dado que permite validar el cumplimiento de los parámetros de prohibición y restricción asignando pesos iguales a 1 en mapas binarios.

Se recomienda tener en cuenta la búsqueda de expertos para participar en el proceso llevado a cabo con el apoyo del Método SAATY en la fase de adquisición de datos, puesto que al ser método estadístico en necesaria la participación de 10 participantes con conocimientos en el área que se plantea en el presente estudio, con el fin de que los resultados arrojados sean los esperados.

La Programación por Compromiso permite evaluar la solidez de las áreas optimas obtenidas, al clasificar individualmente los sitios utilizando varios escenarios (ambiental, social y económico) con diferentes ponderaciones y puntos de equilibrio, dando como solución óptima aquella área que ocupa el primer lugar en todas las opciones de estimación.

VI. CONCLUSIONES

Se cumplió satisfactoriamente con el objetivo planteado de proponer un método para determinar zonas óptimas para la ubicación de rellenos sanitarios en el área de estudio usando herramientas SIG a partir de la propuesta de combinar tres Métodos de Análisis de decisión Multicriterio. En donde la Lógica Booleana permite establecer las restricciones y prohibiciones del territorio, el Proceso de Jerarquía Analítica (AHP) derivo la importancia comparativa de las variables involucrando múltiples criterios y la Programación por Compromiso clasificó los sitios resultantes en función de su importancia relativa en los componentes ambiental, económico y social.

A partir del geoproceso llevado a cabo en el presente estudio se obtienen dos resultados principales: el MAPA DE AREAS APTAS; que resulta de la sobreposición entre las zonas sin restricción ni prohibición ambiental y el MAPA DE ÁREAS ÓPTIMAS, en el cual se contrastan las áreas aptas y las zonas con mayor importancia ambiental y socioeconómica, generándose la espacialización de zonas que cuentan con mayor potencial para el establecimiento de rellenos sanitarios. Adicionalmente, se expone un tercer resultado que contiene la delimitación de sitios en términos de área (Ha) correspondiente al análisis cualitativo de la capacidad del territorio tomado como referencia datos encontrados en la bibliografía, con el fin de que tal criterio (capacidad) sea considerado en estudios posteriores y se recomienda se estime de forma cuantitativa.

El resultado arrojado por el Proceso Analítico Jerárquico en relación con el criterio socioeconómico tuvo incidencia significativa sobre el resultado final puesto que donde se observan áreas con pesos máximos en el PAJ-Económico, es decir al Nordeste Antioqueño; es donde se concentran finalmente las áreas con

mayor potencial para establecer rellenos sanitaros.

Como se observa en el resultado Áreas Óptimas, las zonas con mayor potencial ambiental y socioeconómico para el establecimiento de rellenos sanitarios se concentra en el Nordeste de Antioquia. Dado que la zona se caracteriza entre otras por la presencia de conflicto, se recomienda considerar dicho criterio (conflicto) dentro de las variables socioeconómicas en estudios posteriores.

De igual manera, se hace énfasis en que la calidad de geodatos adquiridos, en lo que respecta a exactitud temporal, de posición y temática no fueron las deseadas dadas las condiciones del Estado de Emergencia Económica, Social y Ecológica decretado por el Gobierno Colombiano, con el fin de enfrentar la pandemia del coronavirus COVID – 19 en las que se desarrollaron el presente trabajo y dificultan la adquisición de datos. Por lo cual es necesario llevar a cabo la verificación de calidad y la política de mantenimiento propuestas, para garantizar la mejora continua del método, conservarlo en buen estado y evitar su degradación

Hay que señalar que los resultados obtenidos no contemplan el diseño final del relleno sanitario, por lo que la capacidad deberá ser validada en función de la generación de residuos sólidos, proyección de la densidad poblacional y el periodo de diseño

VI. REFERENCIAS

[1] Grupo Sectorial de Gestión de Aseo, “Disposición Final de Residuos Sólidos Nacional Informe de Informe Disposición Final de Residuos Sólidos – 2018,” p. 97, 2019.

[2] C. Montes Cortés, *Estudio de los residuos sólidos en Colombia*, Primera Ed. Bogotá: Universidad Externado de Colombia, 2018.

[3] Presidencia de la Republica de Colombia, “Código Nacional de Recursos Naturales Renovables y de Protección al Medio Ambiente. Decreto Ley 2811 de 1974,” p. 544, 1974.

[4] Ministerio de Salud, “LEY 9 DE 1979 (enero 24) por la cual se dictan Medidas Sanitarias (Colombia),” vol. 1979, no. enero 24, 1979.

[5] Congreso de la república de Colombia, “Ley 99 de 1993, Por la cual se crea el Ministerio del medio ambiente , se reordena el sector público encargado de la gestión y conservación del medio ambiente y los recursos naturales renovables , se organiza el Sistema Nacional Ambiental -SINA- y se dict,” *D. Of. No. 41.146*, vol. 1993, pp. 1–279, 1993.

[6] K. Vatalis and O. Manoliadis, "A two - level multicriteria DSS for Landfill Site Selection Using GIS : Case Study in Western Macedonia , Greece," *J. Geogr. Inf. Decis. Anal.*, vol. 6, no. 1, pp. 49–56, 2002.

[7] C. Mena Frau, J. Gajardo Valenzuela, and Y. Ormazábal Rojas, "Modelación Espacial Mediante Geomática Y Evaluación Multicriterio Para La Ordenación Territorial," *Rev. Fac. Ing. - Univ. Tarapacá*, vol. 14, no. 1, 2006, doi: 10.4067/s0718-13372006000100009.

[8] N. Bin Chang, G. Parvathinathan, and J. B. Breeden, "Combining GIS with fuzzy multicriteria decision-making for landfill siting in a fast-growing urban region," *J. Environ. Manage.*, vol. 87, no. 1, pp. 139–153, 2008, doi: 10.1016/j.jenvman.2007.01.011.

[9] C. A. Z. Mejía, F. A. M. Castañeda, and P. A. M. Varela, "A methodology for landfill location using geographic information systems : a Colombian regional case Metodología para la localización de rellenos sanitarios mediante sistemas de información geográfica . Un caso regional colombiano," *Ing. e Investig.*, vol. 32, no. 1, pp. 64–70, 2012.

[10] M. Hanine, O. Boutkhoum, A. Tikniouine, and T. Agouti, "Comparison of fuzzy AHP and fuzzy TODIM methods for landfill location selection," *Springerplus*, vol. 5, no. 1, 2016, doi: 10.1186/s40064-016-2131-7.

[11] M. Barzehkar, N. M. Dinan, S. Mazaheri, R. M. Tayebi, and G. I. Brodie, "Landfill site selection using GIS-based multi-criteria evaluation (case study: SaharKhiz Region located in Gilan Province in Iran)," *SN Appl. Sci.*, vol. 1, no. 9, pp. 1–11, 2019, doi: 10.1007/s42452-019-1109-9.

[12] J. Manuel *et al.*, "Atlas Climátologico de Colombia," 2017.

[13] Ministerio de Ambiente Vivienda y Desarrollo Territorial, "Decreto 0838 de 2005," *45862*, no. 0838, p. 17, 2005.

[14] ISO 19517, "INTERNATIONAL STANDARD Geographic information — Data quality," vol. 2013, 2013.

Capítulo 7
Detección Automática de Eventos Generadores de Accidentes de Tránsito

Fabio Alberto Vargas Agudelo - Fvargas@tdea.edu.co
Tecnologico de Antioquia
Hugo Estrada Esquivel - hugo.estrada@infotec.com.mx
Infotec - Centro de Investigación e Innovación en Tic
Alicia Martínez Rebollar - amartinez@cenidet.edu.mx
Javier Ortiz Hernández - Javier.oh@cenidet.edu.mx
Leon Alberne Torres Restrepo - Leontorres16c@cenidet.edu.mx
Tecnológico Nacional de México CENIDET

I. INTRODUCCIÓN

La seguridad vial se centra en la implementación de medidas que permiten prevenir o mitigar los daños causados por los accidentes de tránsito, los cuales se encuentran entre las principales causas de muerte a nivel mundial [1]. A finales del año 2016, los accidentes de tránsito causaron cerca de 500.000 muertes y 14 millones de heridos en todo el mundo [2]. Asimismo, este tipo de accidentes se ubican como la principal causa de muerte en la población menor de 45 años en México [1], [3].

En este sentido, algunas organizaciones e instituciones trabajan en diferentes iniciativas enfocadas en la detección temprana de los eventos de riesgo o conductas de conducción agresivas que pueden desencadenar un accidente de tránsito. Estas iniciativas se apoyan en técnicas de procesamiento de señales y aprendizaje de máquinas [4], [5], [6], así como en diferentes herramientas tecnológicas para detectar automáticamente los tipos de eventos de riesgo que pueden presentarse al momento de conducir un automóvil [6], [7], [8], [9].

Sin embargo, el tema de la detección automática de este tipo de eventos de riesgo no está cerrado y existen áreas de oportunidad para el desarrollo de técnicas y métodos más eficaces [10].

En este trabajo de investigación se propone un método para la detección automática de eventos de riesgo al

momento de conducir, compuesto por un conjunto de algoritmos que utilizan la información obtenida del contexto y de los sensores ubicados dentro de un automóvil para detectar la ocurrencia de un evento de riesgo. El método propuesto fue utilizado en el desarrollo de un prototipo de software para detectar de manera automática algunos eventos de riesgo en la conducción de un automóvil. El prototipo utiliza los sensores de un teléfono inteligente para monitorear el contexto y enviar la información a la plataforma abierta FIWARE[1]. Esta es una plataforma en la nube para la construcción de soluciones, la cual ofrece un conjunto de herramientas y módulos genéricos que facilitan el desarrollo de aplicaciones para internet de las cosas y ciudades inteligentes [11], [12].

La evaluación del método se realizó por medio de la selección e implementación de una serie de algoritmos que permiten detectar automáticamente el exceso y los cambios de velocidad en el manejo de un automóvil.

El capítulo se encuentra organizado de la siguiente manera: en la sección 2 se muestra el estado del arte sobre el tema; en la sección 3 se describe el método para la detección de eventos generadores de accidentes de tránsito; en la sección 4 se presenta un ejemplo del método con un caso de estudio; la sección 5 muestra pruebas y resultados y, finalmente, la sección 6 presenta las conclusiones y los trabajos futuros.

II. ESTADO DEL ARTE

Los accidentes de tránsito son un grave problema para la sociedad por los daños y pérdidas que pueden causar [3], [13]. Las principales causas que provocan los accidentes están relacionadas con errores humanos, fallas del automóvil o condiciones del entorno, siendo las más frecuentes: la velocidad excesiva, no guardar distancia, invasión del carril contrario, giro indebido y distracciones [9], [14]. Sin embargo, muchas de estas causas pueden ser detectadas por medio del uso de diferentes técnicas y t ecnologías [13].

Actualmente, existe una diversidad de trabajos e implementaciones que abordan la detección automática de eventos de riesgo que pueden presentarse al momento de conducir un automóvil. Por ejemplo, los nuevos modelos de autos usan el sistema ADAS (Advanced Driver Assistance Systems) [15] para mejorar la seguridad y minimizar el riesgo de sufrir un accidente o colisionar contra otros vehículos [2], y en algunas ciudades se utilizan sistemas para supervisión y control colocados en las carreteras [16]. Asimismo, algunos trabajos se apoyan en la

1 FIWARE es una plataforma abierta para el internet de las cosas desarrollada por la Unión Europea.

utilización de dispositivos embebidos en los vehículos, tales como cámaras, sensores y radares que supervisan los comportamientos de conducción y alertan al conductor sobre posibles peligros en la vía [4], [17], mientras que otros utilizan dispositivos móviles equipados con múltiples sensores para registrar los datos provenientes del vehículo [2], [5], [6], [7], [8], [9], [10].

Sin embargo, la detección automática de eventos de riesgo que pueden desencadenar en un accidente al momento de conducir un automóvil encierra un grado de complejidad alto, y pese al esfuerzo realizado por la comunidad, aún no se ha alcanzado la madurez en cuanto a las técnicas de detección y los algoritmos desarrollados para tal fin [10]. Asimismo, hay pocos trabajos que planteen un método para el abordaje de este tipo de problema en el contexto del internet de las cosas. No obstante, al revisar diferentes trabajos se pueden observar etapas o pasos generales en común [18], [19].

III. MÉTODO PARA LA DETECCIÓN DE EVENTOS GENERADORES DE ACCIDENTES DE TRÁNSITO

El método propuesto se apoya en diversas herramientas computacionales y en los beneficios que ofrecen los paradigmas de internet de las cosas y ciudades inteligentes. Además, esta investigación ha estado soportada por técnicas y metodologías de ingeniería de software como la metodología en cascada [20], que ha sido adaptada a los nuevos requerimientos que se generan al utilizar nuevas tecnologías. La Figura 1 muestra una vista general del método propuesto, el cual está formado de los siguientes pasos: modelado del negocio, análisis de requisitos, análisis de la tecnología, diseño del sistema, construcción del sistema e implementación del sistema.

En cada uno de estos pasos se ejecutan una serie de tareas que permiten la construcción de una solución o prototipo que puede ser utilizado para la detección automática de los eventos generadores de accidentes de tránsito.

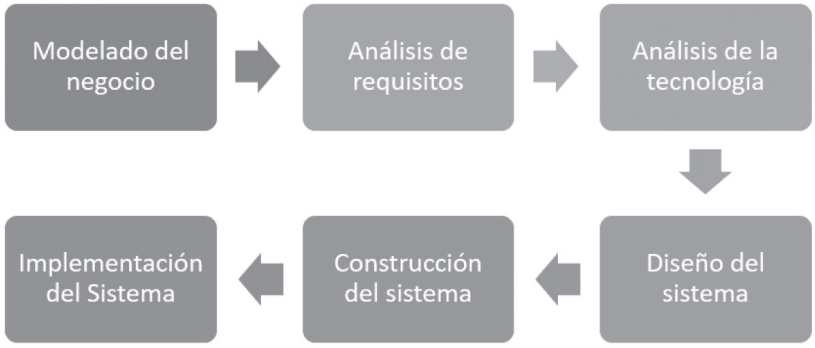


Fig. 1. Vista general del método propuesto.

A. *Paso 1: Modelado del negocio*

Se enfoca en la identificación y selección de los eventos generadores de accidentes de tránsito que se desean detectar. La selección de estos eventos se puede realizar con base en las estadísticas oficiales o a partir del criterio de un experto. Es importante tener en cuenta que los accidentes de tránsito y sus causas pueden diferir de un lugar a otro, lo cual impacta directamente en la relevancia de los eventos que se desean detectar.

B. *Paso 2: Análisis de requisitos.*

Se definen y analizan los requerimientos y las reglas de negocio que se deben tener en cuenta para la detección de los eventos seleccionados y, adicionalmente, otros aspectos tales como: el volumen de los datos, su frecuencia de adquisición, las restricciones y las características propias de cada evento que facilitan su identificación automática.

C. *Paso 3: Análisis de la tecnología.*

Este paso se propone a partir de la experiencia obtenida durante el proceso de experimentación con soluciones de internet de las cosas y ciudades inteligentes. Los principales objetivos de este paso son la identificación de las fuentes de datos y la selección de las técnicas de detección. La identificación de las fuentes de datos es necesaria porque las aplicaciones que tienen soluciones con el internet de las cosas tienen varias fuentes de datos (aplicaciones, servicios, sensores o dispositivos). Incluso se pueden tener diferentes tipos de datos (texto, audio o video), los cuales pueden estar estructurados o no estructurados, lo que impacta directamente en la complejidad de la solución que se desea construir.

Además, es importante determinar las técnicas que pueden ser utilizadas para la detección automática de cada uno de los eventos seleccionados, dado que estas deben estar alineadas con las

fuentes de datos seleccionadas para que la solución sea factible. En este paso es muy importante tener presente la capacidad instalada, las características propias de cada escenario y la disponibilidad de la información.

D. *Paso 4: Diseño del sistema.*

Se realiza el diseño de la arquitectura del prototipo. Este modela la estructura final del sistema que se desea obtener. En este diseño deben estar claramente identificados cada uno de los subsistemas y módulos que se espera que tenga el prototipo o solución final.

E. *Paso 5: Construcción del sistema.*

Se enfoca en la construcción de la solución que permite la detección automática de los eventos seleccionados. Está compuesto por cuatro tareas consecutivas: carga de valores de referencia, adquisición de datos, detección de eventos y envío de notificaciones.

Carga de valores de referencia. En esta tarea se alimenta el sistema con los valores de referencia que se utilizarán para determinar la ocurrencia de un evento. Un ejemplo de un valor de referencia puede ser el límite de la velocidad permitida, la distancia mínima de frenado o una serie de tiempo que representa un tope. Una característica importante de los valores de referencia es que estos generalmente son de tipo estático y son tomados de la norma o de estudios anteriores.

Adquisición de datos. En esta tarea se toman los datos de las fuentes de información seleccionadas. Este paso se centra en la obtención de información en tiempo real. Se realizan tres procesos: adquisición, tratamiento y distribución de los datos.

Detección de eventos. En esta tarea se realiza la detección y/o clasificación de los eventos. Su objetivo principal es identificar cuándo se presenta la ocurrencia de un evento para generar una notificación. Esta tarea tiene tres procesos básicos: recepción de datos, identificación de los eventos e identificación del área de ocurrencia del evento.

Envío de notificaciones. El objetivo de esta tarea es notificar a una entidad o aplicación sobre la ocurrencia de un determinado evento de riesgo. Esta tarea está compuesta por dos procesos: construcción del mensaje y distribución del mensaje.

F. *Paso 6: Implementación del sistema*

Se implementa el sistema para la detección de los eventos. El método está diseñado para que pueda ser aplicado, independientemente de las herramientas tecnológicas que se deseen utilizar. En la siguiente sección se presenta un ejemplo de la aplicación del método.

IV. EJEMPLO DE APLICACIÓN DEL MÉTODO A UN CASO DE ESTUDIO

En esta sección se presenta un ejemplo que describe la aplicación del método. Los eventos se seleccionaron buscando cumplir con los siguientes dos criterios: Mantener el ejemplo lo más sencillo posible y cubrir los aspectos más relevantes.

Los primeros pasos del método se enfocan en la identificación general de los eventos que se van a detectar y las técnicas que se han de utilizar. El objetivo es diseñar una solución que esté alineada con los recursos disponibles y las limitantes de cada escenario.

A. *Paso 1. Modelado del negocio*

En este paso se identifican y seleccionan los eventos que se desea detectar. Para este ejemplo se seleccionaron los eventos con base en datos de fuentes oficiales. Los eventos de riesgo se tomaron con base en la información presentada en los estudios realizados por el CESVI en 2015 [14] y el INEGI en 2017 [21]. En estos estudios se observa que el exceso de velocidad es el principal causante de accidentes automovilísticos (Figura 2). Con esta información se decide detectar automáticamente dos eventos específicos: velocidad no permitida y cambios de velocidad.

B. *Paso 2. Análisis de requisitos*

Una vez seleccionados los eventos, es importante identificar las características y reglas de negocio de cada uno de ellos. En este caso la velocidad no permitida está asociada con un concepto que se conoce como límite de velocidad, el cual se utiliza para determinar si un automóvil se desplaza a una velocidad autorizada de acuerdo con el lugar por el que se desplaza, y está compuesto por un límite inferior y un límite superior que representan la velocidad mínima y la máxima permitida, respectivamente.

Cualquier valor de velocidad que se encuentre fuera de estos límites se toma como una velocidad no permitida. Asimismo, durante el desplazamiento de un vehículo, éste experimenta diversos cambios de velocidad de acuerdo con el contexto. Algunos de estos cambios pueden ser paradas, inicios de marcha, aumento/disminución de velocidad, cambios bruscos de velocidad, etc. En este caso lo que se desea observar es el comportamiento del desplazamiento en un periodo de tiempo, por lo que para su detección puede ser más conveniente utilizar series de tiempos.

En este paso también se seleccionan las fuentes de datos de donde se obtendrá la información necesaria para detectar cada evento. En el caso de la velocidad no permitida se requieren datos estáticos y dinámicos, los cuales

provienen de diferentes fuentes. Los datos dinámicos corresponden a la velocidad de desplazamiento y la ubicación del vehículo.

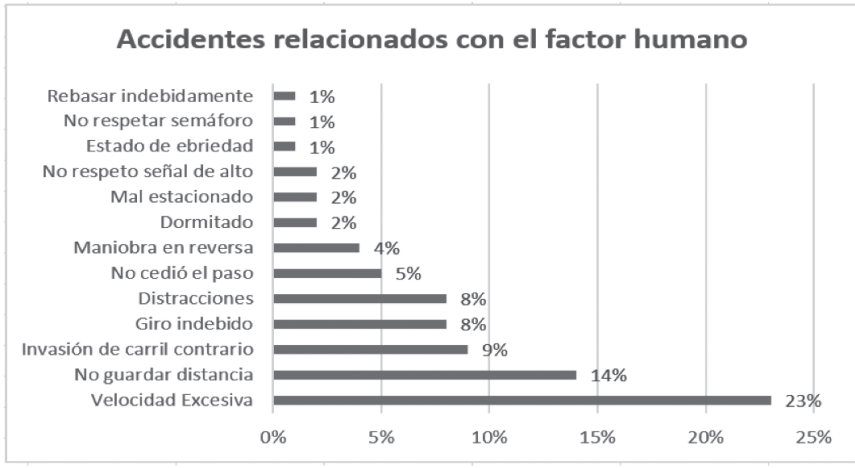


Fig. 2. Estadísticas de los resultados obtenidos en la investigación realizada por el CESVI México (tomado de [14] y parcialmente modificada)

Los datos estáticos corresponden a la delimitación de las áreas y los límites de velocidad establecidos para cada área. Los datos dinámicos se pueden obtener de diferentes fuentes (dispositivos móviles, dispositivos embebidos o dispositivos externos como una cámara). Los datos estáticos se toman directamente de la norma o el reglamento de tránsito. En el caso del evento de cambio de velocidad, solo se requiere el histórico de la velocidad de desplazamiento, la cual se puede obtener utilizando algunos de los medios mencionados anteriormente. Para este ejemplo, los datos dinámicos son obtenidos utilizando el GPS de un teléfono inteligente.

C. *Paso 3. Análisis de la tecnología*

El objetivo de este paso es analizar y seleccionar las tecnologías que se utilizarán para la detección automática de cada evento. Esto debido a que, en una solución para ciudades inteligentes o internet de las cosas, es posible detectar un mismo evento de múltiples formas. La Tabla 1 muestra las técnicas utilizadas para detectar la velocidad no permitida y los cambios de velocidad.

Tabla 1. Técnicas utilizadas para la identificación de los eventos seleccionados

N.°	Causa	Fuente de datos	Técnicas de detección
1	Velocidad no permitida	GPS y normativa de tránsito	Operaciones matemáticas
2	Cambios de velocidad	GPS	Reconocimiento de patrones

En el caso de la velocidad no permitida, el proceso es menos complicado porque se trata de hacer una comparación entre la velocidad actual y los límites de velocidad del área en la que se encuentra el vehículo. En este sentido, conviene la utilización de operaciones matemáticas básicas, con lo cual se gana en costo computacional y velocidad de procesamiento. En el caso de los cambios de velocidad, se desea analizar el comportamiento del desplazamiento del vehículo en un periodo de tiempo determinado, durante el cual se realizan varias lecturas de velocidad. Con este evento se busca ejemplificar un escenario bastante común en la detección automática de eventos generadores de accidentes de tránsito, en el cual se requiere utilizar técnicas de inteligencia artificial para obtener resultados más acertados. Para este evento, se propone la utilización de una técnica de reconocimiento de patrones y el uso de un vector de características compuesto por siete datos dinámicos distribuidos de la siguiente forma: las últimas cinco lecturas de la velocidad de desplazamiento del vehículo con un traslape de dos lecturas, la media y la varianza estadística de las cinco lecturas obtenidas.

D. *Paso 4. Diseño del sistema*

La arquitectura propuesta para la detección de los eventos seleccionados está compuesta por dos subsistemas: 1) detección de eventos y 2) monitoreo y visualización.

El subsistema de detección de eventos se ejecuta del lado del teléfono inteligente. Éste está compuesto por dos módulos. Uno cubre las tareas de interacción con el usuario y el otro las tareas que se encargan de la adquisición de datos y detección de eventos. En la figura 3 se muestra la arquitectura.

El subsistema de monitoreo y visualización permite la visualización de los datos provenientes del subsistema de detección de eventos. Al igual que el subsistema de detección, este cuenta con dos módulos: interfaz de usuario y librería para la detección de eventos. La razón para tener un bloque de detección de eventos en ambos subsistemas se

debe a que algunas técnicas requieren una capacidad computacional mayor, la cual puede resultar costosa para algunos dispositivos de adquisición de señal.

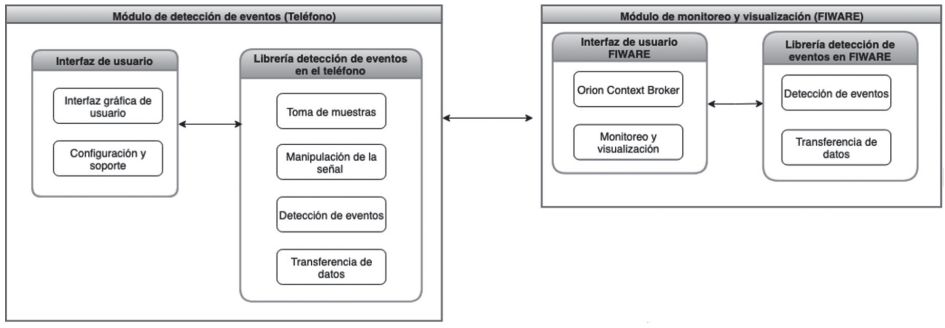


Fig. 3. Arquitectura general del prototipo

Para el módulo de interfaz de usuario se propone utilizar dos componentes genéricos de la plataforma de *FIWARE*: el gestionador de contexto central (llamado: *Orion Context Broker* [22]) y el habilitador genérico *QuantumLeap* [23] para conservar datos de contexto en una base de datos. Para el módulo de detección de eventos se propone el uso de Python, por el conjunto de herramientas que posee para el trabajo con reconocimiento de patrones.

E. *Paso 5 Construcción del sistema*

Este paso tiene como objetivo construir la solución que permitirá detectar automáticamente los eventos seleccionados en el paso anterior, en este caso, la velocidad no permitida y los cambios de velocidad.

La primera tarea es definir los valores de referencia o parámetros de configuración que deben ser ingresados en el sistema y que son requeridos por las técnicas de detección para determinar si ocurre un evento. Estos valores generalmente vienen dados por el reglamento de tránsito y la dinámica del sistema y pueden ser cuantitativos o cualitativos, como es el caso del nombre de la zona que puede tomar valores como: norte, sur, metropolitana, etc. Generalmente, estos datos son de tipo estático.

Los valores de referencia que se utilizan en este ejemplo para la determinación del exceso de velocidad son los siguientes:

Velocidad mínima. Su valor base se establece como cero. La razón para establecerla en cero inicialmente es que el velocímetro no marca velocidades negativas y que, a excepción de las vías de alta velocidad, generalmente no se establece un valor de velocidad mínima permitida explícitamente.

Velocidad máxima. Esta se toma de acuerdo con el reglamento de tránsito vigente para México [24], en el cual se establecen los siguientes valores de velocidad máxima: 110 km/h para autopistas, 80 km/h para vías primarias, 50 km/h par vías terciarias y 10 km/h vías ubicadas en el interior de conjuntos residenciales.

Zona (Área). Se refiere a una región geo-referenciada la cual cuenta con límites de velocidad previamente definidos.

Los valores de referencia para la determinación de los cambios de velocidad no corresponden a un valor específico, sino a un conjunto de ejemplos o muestras que permiten entrenar los algoritmos de clasificación. Este conjunto de muestras representa paradas, inicios de marcha, aumento o disminución de velocidad, cambios bruscos de velocidad, etc. Para este ejemplo, se determinó que una muestra va a estar compuesta por un total de 7 datos, de los cuales los primeros 5 corresponden a las últimas lecturas de velocidad obtenidas, y los siguientes dos corresponden a la media y la varianza de las últimas cinco lecturas obtenidas.

Una vez cargados los valores de referencia, se procede con la segunda tarea de este paso que corresponde a la adquisición de los datos, los cuales corresponden al monitoreo del desplazamiento del vehículo. Estos datos son el principal insumo para la detección automática de los eventos y generalmente provienen de sensores, aplicaciones o servicios que obtienen la información del entorno en tiempo real. Los datos que se deben recolectar son la fecha y la hora, la longitud, la latitud y la velocidad. Estos datos son tomados del sistema y del GPS, tal como se definió en los pasos anteriores. En esta tarea se llevan a cabo dos procesos simultáneos: toma de datos y distribución de datos.

El proceso de toma de datos se realiza de forma continua. En este proceso se adquieren los datos, se elimina el ruido en los datos, se llena el vector de características y se distribuyen los datos.

En la actividad de adquisición de datos se toman los datos de los sensores y de las diferentes fuentes de información y se lleva a cabo un proceso de filtrado de datos para la eliminación de ruido utilizando la técnica de medias móviles. En el Algoritmo I se presenta el pseudocódigo para la adquisición de datos.

Cada dato filtrado se pasa en simultáneo al bloque de distribución de información y al bloque que construye el vector de características. La razón de pasar directamente el dato al bloque de distribución es que algunas técnicas, como la utilizada para velocidad no permitida, no requieren un vector de características. Una vez que el bloque que construye

el vector de características recibe el dato filtrado, lo empaqueta de acuerdo con los valores de referencia establecidos y lo pasa al bloque de detección de eventos. El pseudocódigo del proceso de construcción del vector de características se muestra en el Algoritmo II. El funcionamiento general es el siguiente:

- El valor filtrado se adiciona a un arreglo que almacena los valores del vector de características.
- Se verifica que el tamaño del vector de características no sea mayor o igual al tamaño establecido en los valores de referencia, dado que un tamaño diferente afectaría negativamente el resultado obtenido al momento de la clasificación de los eventos.
- Cuando el tamaño del vector de características es igual al valor de referencia, se pasa el vector a la etapa de detección de eventos.
- Una vez que se pasan los datos del vector de características a la etapa de detección de eventos, se borran todos sus datos y se le adicionan los datos de traslape de acuerdo con la ventana de tiempo establecida en los valores de referencia.

ALGORITMO I

Algoritmo 1. Pseudocódigo para adquisición de datos

```
entero p // tamaño de la ventana para suavizar el ruido.
cola A // almacena los últimos p elementos.
double suma // suma total de los elementos de A.
entero n // número de elementos de traslape
cola B // almacena los últimos n elementos de traslape.
double dato // variable para almacenar el último dato adquirido del sensor.
double valorFiltrado // Valor que será pasado al vector de características.
 // Ciclo  para recolectar los datos provenientes del sensor.
mientras(verdadero) {
dato = leerDatosSensor ()
A.adicionar(dato)
si (A.tamaño<p)
suma+=dato
sino
valorFiltrado =suma/n
construirVectCaracter(valorFiltrado, sensor)
distribuirInformacion(valorFiltrado, sensor)
suma -= A.removerPrimero()
A.adicionar(dato)
 suma += dato
si (B.tamaño<n)
B.adicionar(valorFiltrado)
   sino
      B.removerPrimero()
      B.adicionar(valorFiltrado)
}
```

ALGORITMO II

Algoritmo 2. Pseudocódigo del proceso de construcción del vector de características

```
entero k // tamaño total del vector de características
array vectorCaracteristicas
// estructura de tamaño k para almacenar los elementos del vector de características.
double dato // Valor que es recibido del proceso de toma de muestras
construirVectCaracter (dato){
vector Características. adicionar(dato)
si vector Características. tamaño>=k
distribuir Informacion (vector Caracteristicas,
sensor )
vector Caracteristicas. vaciar()
si n > 0
entero i = 0
mientras (B.tamanio > 0)
 C[i] = B.removerPrimero()
i += 1
}
```

El bloque de distribución toma la información recibida y le agrega algunos datos adicionales. La información recibida puede ser un solo dato o un vector de características; a esta información se le agrega información adicional, como: fecha, hora y geolocalización de los datos obtenidos. Una vez que se tiene el conjunto de datos, se transfieren al bloque de detección de eventos (ver Algoritmo III).

ALGORITMO III

Algoritmo 3. Pseudocódigo del proceso de distribución de información

```
lista info Contexto //estructura que contiene información adicional relacionada con los datos adquiridos.
objeto info Data  // objeto que encapsula la de contexto y la obtenida de los sensores.
distribuir Informacion (vector Caracteristicas, sensor) {
info Data. Adicionar (vector Caracteristicas)
info Data. Adicionar (info Contexto)
info Data. Adicionar (sensor)
selecionar Modelo Deteccion Evento (info Data)
}
```

La tercera tarea de este paso es la detección de los eventos. En ella se reciben los datos adquiridos en el paso anterior y se distribuyen a los diferentes algoritmos de reconocimiento de eventos, que para este caso serían los encargados de detectar los eventos de velocidad no permitida y cambios de velocidad. La primera actividad es seleccionar el modelo que se utilizará para la detección de los eventos de acuerdo con el sensor del cual provienen los datos. Es importante tener en cuenta que los datos provenientes de un mismo sensor pueden ser utilizados para la detección de eventos de distinta naturaleza. En el Algoritmo

IV se muestra el pseudocódigo de la selección del modelo de detección.

ALGORITMO IV

Algoritmo 4. Pseudocódigo selección del modelo de detección

lista modelos // variable que contiene la relación de los modelos de detección aplicables de acuerdo con la fuente de datos.

selecionarModeloDeteccionEvento (vectorCaracteristicas){
mientras (modelo.siguiente()!= null){
if(modelo.sensor =
vectorCaracteristicas.sensor){
analizarDatos(modelo,
vectorCaracteristicas)
}
}
}

Una vez que un modelo es seleccionado, se invoca una función para el análisis del conjunto de datos recibido. Ésta se encarga de llamar a los algoritmos que comparan los datos recibidos contra los valores de referencia del modelo indicado. Cada algoritmo retorna una de dos respuestas posibles: tipo de evento detectado o ninguno (ver Algoritmo V).

ALGORITMO V

Algoritmo 5. Pseudocódigo para el análisis de datos

Cadena respuestaAnalisis //variable para obtener la respuesta retornada por el algoritmo de detección.
analizar Datos (modelo, vectorCaracteristicas) {
algoritmo
=obtenerAlgoritmo(modelo.id)
respuestaAnalisis = algoritmo.
analizar (vectorCaracteristica.
vectorCaracteristica, modelo)
si respuestaAnalisis != 'Ninguno'
enviarNotificacion (respuesta),
vectorCaracteristica.infoContexto)
}

A continuación, se presenta el pseudocódigo base para la detección automática de los eventos tomados como ejemplo para la explicación del método propuesto (ver Algoritmo VI).

ALGORITMO VI

Algoritmo 6. Pseudocódigo para detección de velocidad no permitida

objeto valores Referencia // Variable que contiene los valores de referencia establecidos en la etapa uno
deteccion Vel No Autorizada (vel Actual Automovil, [modelo]) {
vel Min Aut= valores Referencia.vel Min Aut
vel Max Aut= valores Referencia.vel Max Aut
tolerancia= valores Referencia. tolerancia
si (velMinAut>0 && velMinAut>0))
si (vel Min Aut- tolerancia >vel Actual Automovil ||
vel Max Aut< vel Actual Automovil+ tolerancia)

Retornar “Vel. No Autorizada”
sino
Retornar “Ninguno”
sino
si (velMaxAut > 0)
si(vel Max Aut+ tolerancia< vel Actual Automovil)
Retornar “Vel. No Autorizada”
sino
Retornar “Ninguno”
sino
si (vel Min Aut-tolerancia > velActualAutomovil)
Retornar “Vel. No Autorizada”
sino
Retornar falso
}

Es importante recordar que para la detección de los eventos del ejemplo propuesto se seleccionaron dos técnicas diferentes. Para el evento de velocidad no permitida se utilizan operaciones matemáticas básicas. Se compara el valor actual de la velocidad de desplazamiento con los límites establecidos por el reglamento de tránsito. Mientras que para la detección de los cambios de velocidad se utiliza el reconocimiento de patrones, ya que este evento se detecta a partir de una serie de tiempo que contiene el histórico de las últimas velocidades registradas. El primer algoritmo está enfocado en la detección de desplazamiento a una velocidad no permitida.

La detección de cambios de velocidad se realiza por medio de una función que invoca un módulo que está alojado en el servidor y no en el dispositivo. La razón para contemplar la utilización de algoritmos que se encuentran alojados en un módulo al lado del servidor, es que algunos algoritmos están basados en técnicas de reconocimiento de patrones o de redes neuronales que requieren una alta capacidad de cómputo. Sin embargo, la capacidad de cómputo de algunos dispositivos utilizados para la adquisición de los datos es limitada, como es el caso de los teléfonos inteligentes o de algunos dispositivos embebidos (ver Algoritmo VII).

ALGORITMO VII

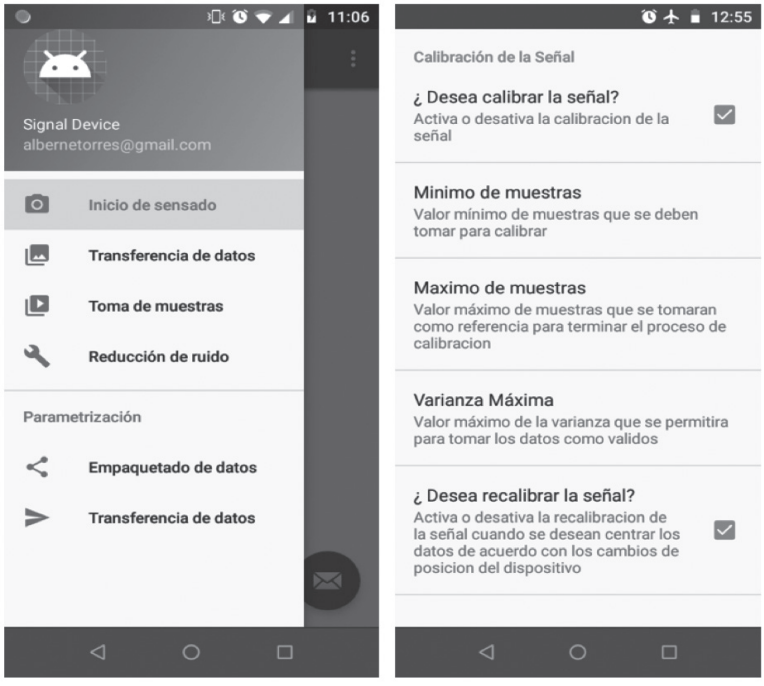
Algoritmo 7. Pseudocódigo para invocación servicio externo

detec Cambios Vel (vector Caracteristicas, modelo) {
respuesta Evento= llamar Punto AccesoExterno(vectorCaracteristicas, modelo)
retorna respuestaEvento
}

La última tarea de este paso corresponde al envío de las notificaciones. En ella se desarrollan dos procesos: Emisión de notificaciones locales y distribución de notificaciones. La emisión de notificaciones locales tiene el objetivo de generar una alerta en el dispositivo, para que el usuario tome las medidas preventivas necesarias.

Las notificaciones externas, por su parte, tienen el objetivo de informar a las autoridades competentes y a otros dispositivos el hecho de que se ha detectado un evento de riesgo. La información de las notificaciones está empaquetada en el modelo de datos Alerta que hace parte del catálogo de modelos de datos de la plataforma FIWARE [25].

Paso 6. Implementación del sistema. Finalmente, el último paso es la construcción e implementación del prototipo de acuerdo con la arquitectura propuesta. En la Figura 4 se muestran algunas pantallas del módulo de detección de eventos implementado en un teléfono inteligente. En la imagen se presentan la vista principal y las vistas de transferencia de datos y toma de muestras.



a) Pantalla principal *b) Parametrización toma de muestras*

Fig. 4. Opciones de parametrización que permite l módulo del teléfono inteligente

V. PRUEBAS Y RESULTADOS

En esta sección se presenta una descripción de las pruebas realizadas.

La fase de pruebas se realizó por un periodo de 14 meses. Durante este periodo se realizaron más de 100 recorridos en zonas rurales y urbanas y se recolectaron más de 7 millones de registros provenientes del GPS. Los recursos utilizados para la evaluación del método fueron los siguientes: tres teléfonos inteligentes (ZTE Blade V6, un Samsung Galaxy J4 y un Motorola Moto G5), una computadora portátil Gamer ASUS (Core i7, 8 GB, 1 TB, NVIDIA GeForce GTX 1050) y dos máquinas virtuales alojadas en la plataforma *FIWARE Lab* nodo México. Además, para la toma de muestras se utilizaron diferentes medios de transporte público y particular. El procedimiento para la realización de las pruebas fue el siguiente:

1. Ubicar el teléfono en una posición fija dentro del vehículo de tal forma que los ejes del teléfono coincidan en lo posible con los ejes de referencia del vehículo (ver la Figura 5).

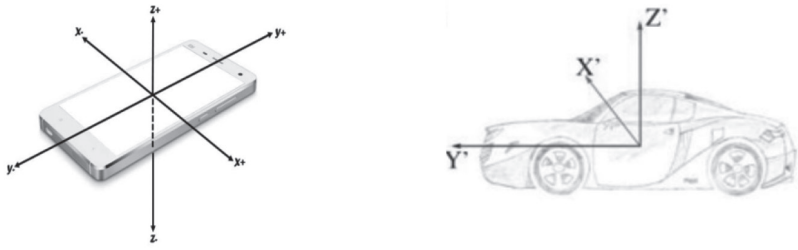


Fig. 5. Sistema de referencia utilizado con respecto al vehículo y el teléfono [11]

2. Seleccionar los sensores que se utilizaran para la toma de datos (ver Figura 6).
3. Cuando el vehículo esté en marcha se presiona el botón *iniciar* para que se inicie el proceso de toma de datos para la detección de eventos.
4. Realizar el procesamiento de los datos. Después de la toma de la señal se procede con la limpieza, transformación y análisis de los datos.

En la Figura 6 se muestra la vista de la interfaz para toma de datos y un archivo plano representativo de los registros adquiridos.

id	Longitud	Latitud	Altitud	Velocidad	Precisión Horizontal
354	-99.2402	18.93679	1570.9	6.503000736	5.4000001
355	-99.2402	18.93685	1571	6.354245663	5.4000001
356	-99.2402	18.9369	1571.1	6.32078886	5.4000001
357	-99.2403	18.93697	1571.2	6.32336235	5.4000001
358	-99.2403	18.93702	1571.2	6.327994823	5.4000001
359	-99.2403	18.93708	1571.3	6.329024315	5.3000002
360	-99.2403	18.93714	1571.5	6.346525192	5.3000002
361	-99.2403	18.9372	1571.6	6.348069191	5.3000002
362	-99.2404	18.93725	1571.8	6.332112789	5.3000002
363	-99.2404	18.93731	1571.9	6.318729877	5.3000002
364	-99.2404	18.93737	1572.1	6.209609032	5.3000002
365	-99.2404	18.93743	1572.2	6.122106075	5.4000001
366	-99.2404	18.93748	1572.4	6.111297131	5.5
367	-99.2405	18.93753	1572.5	6.197255611	5.5999999
368	-99.2405	18.93759	1572.7	6.202917576	5.5999999
369	-99.2405	18.93764	1572.8	6.29093504	5.5999999
370	-99.2405	18.93768	1573	6.326450825	5.5999999
371	-99.2405	18.93775	1573.1	6.464910984	5.5
372	-99.2405	18.9378	1573.2	6.463366985	5.5
373	-99.2405	18.93786	1573.3	6.599253654	5.5

Fig. 6. Vista para la toma de datos y del archivo de registros generado

Para detectar el evento de cambios de velocidad se monitorea el comportamiento del desplazamiento del vehículo a partir de la velocidad registrada por el GPS. Para este caso, se seleccionó un recorrido en particular y se construyó un conjunto de datos de entrenamiento el cual está compuesto por 391 ejemplos debidamente etiquetados. Los eventos registrados en el conjunto de datos son: inicio de marcha, en reposo, parada, incremento de velocidad, decremento de velocidad y velocidad constante. Este conjunto de datos fue utilizado como insumo para el entrenamiento de los algoritmos de detección, los cuales utilizan la librería Scikit-learn. Una vez realizado el entrenamiento, se construyó un servicio web que es utilizado para determinar si se presentó un cambio de velocidad. Este servicio es consumido desde el teléfono inteligente que pasa un paquete de datos con un vector de características y recibe una respuesta, la cual indica el resultado del análisis. El algoritmo con el que se obtuvieron mejores resultados fue con RandomForest, lo cual es coherente con lo reportado en la literatura [2]. Los resultados obtenidos durante las pruebas con los diferentes algoritmos se muestran en la Tabla 2.

Tabla 2. Resumen de resultados obtenidos al pasar los datos por los clasificadores

Técnica	Clasificados correctamente		Clasificados incorrectamente	
	Cantidad	%	Cantidad	%
Naive Bayes	281	71,81	110	28,13
RandomForest	368	94,12	23	5,88
J48	367	93,86	24	6,14

Para detectar el evento de exceso de velocidad se efectuó un recorrido formado por 5.090 registros (tramo A). Cada registro está formado por el valor de la velocidad registrada en ese momento y las coordenadas GPS del lugar en el que se hizo la lectura. De este conjunto de datos se extrajo un conjunto de 490 registros (tramo B) que cumplían con la característica de estar en un segmento de la vía homogéneo con muy pocas curvas, donde el tráfico vehicular no es tan alto (ver Figura 7). El objetivo de extraer este segmento era observar el comportamiento del algoritmo y la influencia de la definición de los tramos de la ruta en la precisión de los resultados obtenidos. El desplazamiento se realizó siempre en sentido correcto y sin tener segmentaciones en ninguno de los tramos. En la Figura 8 se muestran algunos de los valores obtenidos para el tramo A. El porcentaje de detección no es muy alentador, ya que casi un 40 % de las lecturas obtenidas se clasificaron como desplazamiento a una velocidad no permitida.

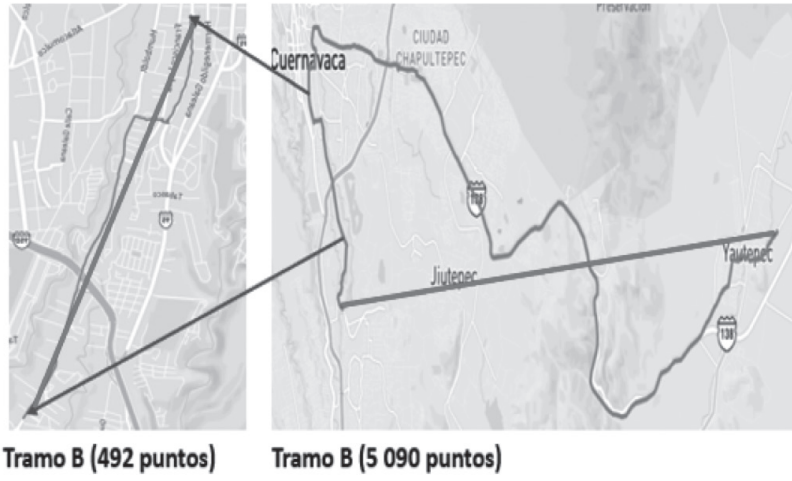
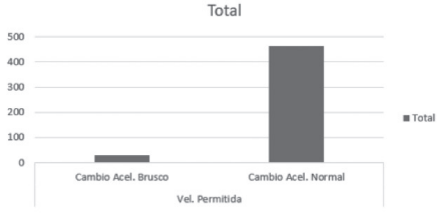


Fig. 7. Línea entre los puntos inicial y final del recorrido

Luego se ejecutaron los algoritmos con los mismos parámetros, pero sobre el tramo B. Los resultados obtenidos mejoran con respecto a los obtenidos en el tramo A, se obtuvieron valores por encima del 93 % y no se detectó ningún exceso de velocidad. Es importante recordar que el tramo B es un segmento del tramo A, por lo tanto los valores del tramo B son exactamente iguales a los contenidos para ese mismo segmento en el tramo A.

Los datos más cercanos a la realidad del recorrido realizado se presentan en el tramo B. Una de las razones por las que esto ocurre está relacionada con los algoritmos seleccionados.



Vel. Permitida	492
Cambio Acel. Brusco	28
Cambio Acel. Normal	464
Total general	**492**

Fig. 8. Resultados obtenidos para el tramo B

Los valores relacionados con la detección de velocidad no permitida se deben a que, en términos de geometría y ubicación espacial, el tramo B es más homogéneo que el tramo A, el tramo B se encuentra en una zona urbana, por lo que el límite de velocidad de 50 k/h se encuentra dentro del rango normal. El tramo A no corresponde a un tramo homogéneo en cuanto a su geometría y su ubicación espacial, ya que algunas partes corresponden a zonas urbanas y otras, a zonas rurales. Algunas partes corresponden a vías secundarias o terciarias, mientras que otras corresponden a vías primarias con velocidades permitidas de 80 k/h, velocidad que está muy por encima de la establecida en los parámetros de referencia.

VI. CONCLUSIONES Y TRABAJOS FUTUROS

En este trabajo se presenta un método para la construcción de soluciones enfocadas en la detección

automática de eventos generadores de accidentes de tránsito. El método desarrollado se utilizó para crear un prototipo que permite la detección automática de exceso y cambios de velocidad enfocado en la utilización de soluciones dirigidas al internet de las cosas y ciudades inteligentes. Los resultados obtenidos en las pruebas del prototipo permiten inferir que es viable utilizar el método para el desarrollo de este tipo de soluciones. Como aportes del presente trabajo se encuentra el método desarrollado, el ejemplo de aplicación del método y el pseudocódigo de los algoritmos de adquisición de datos y detección de eventos. En este trabajo se presenta un ejemplo que abarca dos escenarios que son recurrentes en la implementación de este tipo de solución: procesamiento local y procesamiento externo. Como trabajo futuro se plantea el refinamiento de los algoritmos y la prueba del método en otros dominios.

Agradecimientos

Esta investigación ha sido parcialmente financiada por los proyectos TecNM. 9264.20-P y 8819.20-P

VII. REFERENCIAS

[1] Organización Mundial de la Salud, "Informe Sobre La Situación Mundial de la Seguridad Vial 2015", pp. 1-12, 2015.

[2] P. Chaovalit, C. Saiprasert, and T. Pholprasit, "A method for driving event detection using SAX with resource usage exploration on smartphone platform", *Eurasip J. Wirel. Commun. Netw.*, vol. 2014, n.º 1, 2014. doi: 10.1186/1687-1499-2014-135.

[3] Organización Mundial de la Salud, "Global Status Report on Road Safety 2018: Summary", Tech. Rep. (WHO/NMH/NVI/18.20). Licence: CC BY-NC-SA 3.0 IGO). 2018.

[4] B. Fernandes, M. Alam, V. Gomes, J. Ferreira, and A. Oliveira, "Automatic accident detection with multi-modal alert system implementation for ITS", *Veh. Commun.*, 2016. doi: 10.1016/j.vehcom.2015.11.001

[5] H. M. Ali and Z. S. Alwan, "Car Accident Detection and Notification System Using Smartphone", *Int. J. Comput. Sci. Mob. Comput.*, vol. 4, n.º 4, pp. 620-635, 2015.

[6] A. Bin Faiz, A. Imteaj, and M. Chowdhury, "Smart

vehicle accident detection and alarming system using a smartphone", in *1st International Conference on Computer and Information Engineering, ICCIE 2015*, 2016. doi: 10.1109/CCIE.2015.7399319

[7] N. Kattukkaran, A. George, and T. P. M. Haridas, "Intelligent accident detection and alert system for emergency medical assistance", in *2017 International Conference on Computer Communication and Informatics, ICCCI 2017*. doi: 10.1109/ICCCI.2017.8117791

[8] L.-J. Kau and C.-S. Chen, "A Smart Phone-Based Pocket Fall Accident Detection, Positioning, and Rescue System", *IEEE J. Biomed. Heal. Informatics*, vol. 19, n.º 1, pp. 44-56, 2015, doi: 10.1109/JBHI.2014.2328593.

[9] C. Saiprasert, T. Pholprasit, and S. Thajchayapong, "Detection of Driving Events using Sensory Data on Smartphone", *Int. J. Intell. Transp. Syst. Res.*, vol. 15, n.º 1, pp. 1-12, July 2015. doi: 10.1007/s13177-015-0116-5

[10] L. C. González, R. Moreno, H. J. Escalante, F. Martínez, and M. R. Carlos, "Learning Roadway Surface Disruption Patterns Using the Bag of Words Representation", *IEEE Trans. Intell. Transp. Syst.*, pp. 1-13, 2017. doi: 10.1109/TITS.2017.2662483.

[11] M. Muñoz, J. D. Gil, L. Roca, F. Rodríguez, and M. Berenguel, "An iot architecture for water resource management in agroindustrial environments: A case study in almería (Spain)", *Sensors (Switzerland)*, vol. 20, n.º 3, 2020. doi: 10.3390/s20030596.

[12] A. M. Alberti, M. A. S. Santos, R. Souza, and H. D. L. Da Silva, "Platforms for Smart Environments and Future Internet Design: A Survey", *IEEE Access*, vol. 7, pp. 165748-165778, 2019. doi: 10.1109/ACCESS.2019.2950656.

[13] Organización Mundial de la Salud, *Salve VIDAS. Paquete de medidas técnicas sobre seguridad vial.* Ginebra, Organización Mundial de la Salud, 2017. Licencia: CC-BY-SA 3.0 IGO.

[14] CESVI, "Factores que generan los accidentes de tránsito", 2015.

[15] N. O. Alsrehin, A. F. Klaib, and A. Magableh, "Intelligent Transportation and Control Systems Using Data Mining and Machine Learning Techniques: A Comprehensive Study", *IEEE Access*, vol. 7, pp. 49830-

49857, 2019. doi: 10.1109/ACCESS.2019.2909114

[16] K. Rostamzadeh, H. Nicanfar, N. Torabi, S. Gopalakrishnan, and V. C. M. Leung, "A context-aware trust-based information dissemination framework for vehicular networks", *IEEE Internet Things J.*, vol. 2, n.º 2, pp. 121-132, 2015. doi: 10.1109/JIOT.2015.2388581

[17] Z. Liu, M. Wu, K. Zhu, and L. Zhang, "SenSafe: A smartphone-based traffic safety framework by sensing vehicle and pedestrian behaviors", *Mob. Inf. Syst.*, vol. 2016, Article ID 7967249, 2016. doi: 10.1155/2016/7967249

[18] M. Karaduman and H. Eren, "Smart driving in smart city", in *ICSG 2017 - 5th Int. Istanbul Smart Grids Cities Congr. Fair*, pp. 115-119, 2017. doi: 10.1109/SGCF.2017.7947614

[19] O. Avatefipour and F. Sadry, "Traffic Management System Using IoT Technology - A Comparative Review", *IEEE Int. Conf. Electro Inf. Technol.*, vol. 2018-May, pp. 1041-1047, 2018. doi: 10.1109/EIT.2018.8500246

[20] J. B. de Areba, "Metodología del análisis estructurado de sistemas", 2001. [En línea]. Disponible en: https://books.google.com/book s?id =PU qxs NVa QC8 C& pgis=1 [Accedido: 08-Jun-2020].

[21] Dirección General de Servicios Técnicos, "Estadística accidentes de tránsito 2017, república mexicana", 2018.

[22] FIWARE, "FIWARE -Orion." [En línea]. Disponible en: https://fiware-orion.readthedocs.io/en/master/. [Accedido: 05-Oct-2020].

[23] FIWARE, "FIWARE - QuantumLeap." [En línea]. Disponible en: https://quantumleap.readthedocs.io/en/latest/. [Accessed: 05-Oct-2020].

[24] Segunda Sección, Poder Ejecutivo, Secretaría de Economía, "Reglamento de Tránsito en Carreteras y Puentes de Jurisdicción Federal", Diario oficial, 5-di.-2014.

[25] FIWARE, "Alert - Fiware-DataModels." [En línea]. Disponible en: https://fiware-datamodels.readthedocs.io/en/latest/Alert/doc/spec/index.html [Accedido: 05-Oct-2020].

Capítulo 8
Un Método de Etiquetado Semiautomático de Corpus Lingüísticos

Maria Isabel Marín Morales - mmarinm2@tdea.edu.co
Tecnologico de Antioquia
Jorge Molina - Jorge.molina@udea.edu.co
Universidad de Antioquia

I. INTRODUCCIÓN

La lingüística computacional es un campo interdisciplinar cuyo interés principal es la creación de modelos formales que permitan el procesamiento automático de las lenguas naturales. A partir de estos modelos se procesa el lenguaje natural emulando, desde la inteligencia artificial, la capacidad humana de comprender o reproducir el lenguaje dentro de un entorno comunicativo (producción escrita u oral de datos lingüísticos) [1], [2].

Uno de los mecanismos o enfoques metodológicos de los cuales se sirve la lingüística computacional es la lingüística de corpus. Esta se encarga de estudiar la manera más adecuada de gestionar producciones reales de la lengua, almacenadas en textos u otros formatos (audio, video, imágenes, etc.) con el fin de analizar la propia lengua desde sus distintos niveles: fonético, morfológico, sintáctico, lexical, semántico y pragmático [3]. De ahí que la lingüística de corpus se aplique a lo que tiene que ver con la recolección, procesamiento y análisis de grandes muestras de una determinada lengua [4]. Dichas muestras son definidas como corpus, los cuales son conjuntos de textos digitalizados, reunidos y seleccionados teniendo en cuenta unos criterios lingüísticos, a saber: a) que sean recolectados en entornos naturales; b) que posean rasgos definitorios explícitos; c) con una similar extensión o tamaño; d) que se encuentren disponibles; e) que sean accesibles; f) que sean representativos de una determinada lengua; g) que cuenten con metadatos descriptores, y h) que se encuentren compilados bajo criterios o parámetros de organización [3].

Muchos de los algoritmos para el procesamiento del lenguaje natural no

permiten etiquetar, de manera precisa, fenómenos importantes en el nivel textual como las anáforas, las catáforas o los marcadores lógico-temporales y discursivos, entre muchos otros inherentes al lenguaje escrito u oral [5]. Si bien se han encontrado algunas propuestas, como lo es el caso de "AnCora corpora" [6], o de los modelos de resolución anafórica de Palomar et al. [7] para el español, o de Kundu et al. [8] para el inglés, no ha sido posible, hasta el momento, encontrar un modelo específicamente destinado a los textos para la enseñanza del español como lengua extranjera (ELE), los cuales contienen algunas especificidades que no poseen los textos académicos o literarios. Esto tiene que ver con el carácter referencial de algunos elementos que demandan al lector unos vastos conocimientos del mundo [9]. Por otra parte, es necesario contar con corpus para la enseñanza de la lengua española que se encuentren perfectamente analizados y etiquetados, ya que no es lo ideal enseñar errores a los estudiantes extranjeros de dicha lengua [9].

Las razones antes esbozadas son las que se han tenido en cuenta a lo largo del presente capítulo, en el que se presenta un método que permite etiquetar de manera semiautomática un corpus de documentos textuales dedicados a la enseñanza de ELE. El método cubre las etapas de etiquetado morfosintáctico automáticamente, con la aplicación FreeLing [10], y el etiquetado, de manera manual, de correferencias textuales como las anáforas y catáforas. Las etiquetas finales se expresan sobre un documento XML (eXtensible Markup Language o Lenguaje de Marcado Extensible, en español) extraído de un corpus para la enseñanza de la lengua española y dirigido a un público de estudiantes extranjeros [11].

Si bien el trabajo aquí presentado se encuentra, sobre todo, destinado a la aplicación de la enseñanza del español como lengua extranjera (ELE), como se expone en la discusión, el método puede utilizarse en otros contextos de aplicación como, por ejemplo, el aprendizaje de máquina (Machine Learning).

Este capítulo se distribuye como sigue: en el apartado II se presentan los conceptos y la revisión de literatura con relación al etiquetado de corpus y, en especial, se expone una síntesis de la necesidad de efectuar una aplicación en la enseñanza del español como lengua extranjera; en el apartado III se exhiben los diferentes pasos que constituyen el método propuesto; en el apartado IV se aborda la discusión de la aplicabilidad del método y cómo aquel puede extenderse a otros casos de aplicación y, finalmente, en la sección V se enuncian las conclusiones referentes al presente trabajo.

II. MARCO TEÓRICO Y REVISIÓN DE LITERATURA

A. Sistemas de etiquetado

Un sistema de etiquetado, o analizador morfosintáctico, es un programa informático que tiene como entrada un texto plano (generalmente en formato *.txt), a partir del cual se crea una versión con etiquetas (*tags*) que dan cuenta de unas características, a la vez, morfológicas y sintácticas que son inherentes al lenguaje natural. En este sentido, el sistema tiene en cuenta la posición de la palabra dentro de la oración al momento de generar las etiquetas que marcan, por una parte, las nociones morfológicas, como el género y el número, o las funciones sintácticas como, por ejemplo, que una palabra actúe como determinante, nombre propio o común, verbo (con sus marcas de modo, tiempo, persona), etc. También son importantes las fases de desambiguado de las unidades etiquetadas (Figura 1), lo que permite que cada elemento etiquetado no se confunda con otro. En este último proceso, la mayor parte de los sistemas se fundamentan en métodos estadísticos, principalmente, en cadenas de Markov que proveen información estadística acerca de la posibilidad de que luego de una determinada unidad gramatical vaya otra íntimamente relacionada con aquella otra; por ejemplo, luego de un determinante es más probable que le siga un sustantivo (DET + NC).

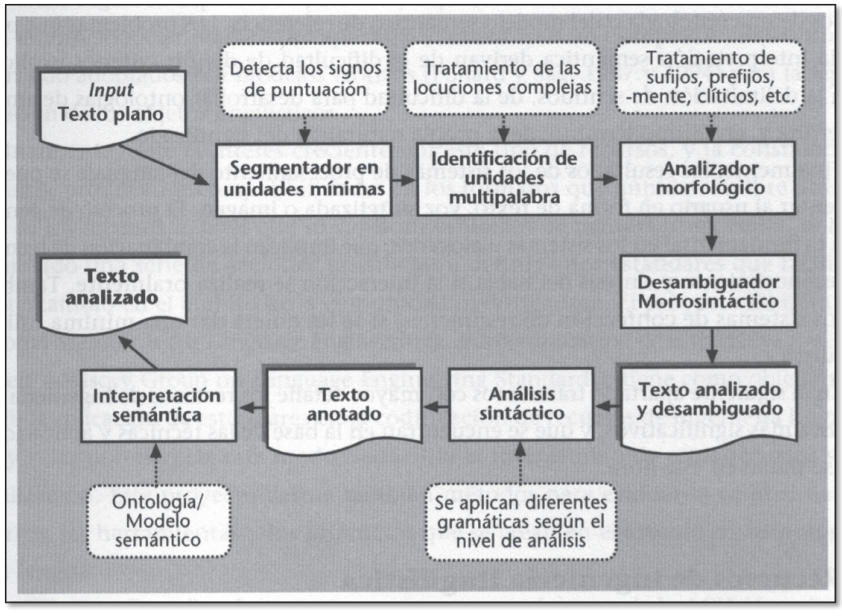


Fig. 1. Diferentes etapas en un análisis textual morfosintáctico y semántico [12].

Desde la entrada del texto plano, hasta la extracción del texto completamente analizado, el sistema de etiquetado realiza diversos procedimientos: 1) un tokenizado (segmentación en palabras y signos de puntuación); 2) la identificación de unidades multipalabra o de frasemas (diferentes tipos de locuciones, por ejemplo); 3) el desambiguado de tipo morfosintáctico, etc. Es así como muchos sistemas, en función de su nivel de granularidad y de fineza en el análisis, realizan el etiquetado con un mayor o con un menor grado de efectividad.

En etiquetadores como *FreeLing* [10], las etiquetas contienen rasgos importantes como lo son: si el género de un sustantivo o de un adjetivo es masculino o femenino, o si el número de alguno de estos es singular o plural. Dicha información permite destacar elementos, como es el caso de las entidades nombradas, las cuales permiten, entre otras funciones, el reconocimiento de referentes y correferentes textuales, por ejemplo.

La mayor parte de las herramientas suelen tener un conjunto de etiquetas que proveen la información gramatical más pertinente en lo que tiene que ver con los elementos que constituyen las oraciones. En el caso de *FreeLing* [10], las etiquetas propuestas dan el género y número de elementos o información acerca de los verbos (modo, tiempo, persona), etc.

B. Corpus en la enseñanza del español como lengua extranjera

Con respecto al uso de corpus en la Enseñanza del Español como lengua extranjera, la idea, según Boulton y Tyne [13], es que los aprendientes de lenguas adquieran las estructuras lingüísticas de la lengua objeto gracias al empleo de documentos auténticos. Esto ayudaría no solamente a entender mejor la lengua, sino, además, a expresarse en la lengua estudiada de una forma correcta y adecuada, teniendo para ello el modelo lingüístico de primer orden: los textos producidos por hablantes nativos y cuyo objetivo no es la enseñanza sino informar o ser leídos por otros hablantes nativos. La enseñanza a partir de la lingüística computacional y de corpus, y que en lenguas como el inglés y francés tiene un gran desarrollo en países como Estados Unidos, Inglaterra y Francia, ha sido poco desarrollada en el área del español [5], [11].

Por todo ello, se tiene en cuenta que la lingüística computacional proporciona las herramientas que permiten realizar un tratamiento informático de las lenguas naturales de una manera más rápida y con un nivel de precisión altamente aceptable. A partir de dicho tratamiento, se pueden obtener corpus anotados y etiquetados de manera morfológica y sintáctica. En este trabajo inicial se proporciona la base de conocimiento de sistemas

para la enseñanza y aprendizaje de lenguas que funcionan a partir de corpus lingüísticos, como lo señala Antoniadis [14]. Los dispositivos informáticos que resultan de estos procedimientos aplicados sobre corpus se implementan con tres finalidades: ayudar a los docentes en su proceso de planificación de actividades de una forma fácil y automática; hacer que el aprendizaje de la lengua sea más interesante y participativo por parte de los aprendientes, y permitir un aprendizaje a partir de datos auténticos, es decir, de textos que dan cuenta de la lengua real en la que se expresan sus hablantes [11], [14], [15].

C. Correferencia textual

La correferencia textual es una categoría problemática en la gramática de las lenguas [16], debido al amplio espectro de mecanismos que intervienen en ella, como la repetición de palabras, la sustitución por pronombres y la elipsis. El manejo correcto de estos mecanismos incluye habilidades de naturaleza tanto gramatical como cognitiva, por lo que se dificulta su sistematización en el procesamiento del lenguaje natural [17]. Existen diversos mecanismos de correferencia, como la repetición de unidades léxicas, su sustitución por medio de elementos fóricos (anáforas y catáforas) y la elipsis.

D. Revisión de la literatura

Varios autores que tratan las lenguas naturales, entre otras, el español, han abordado la creación de una metodología específica para el etiquetado morfosintáctico de textos escritos: *TreeTagger* [18]; *Stanford Parser* [19]; *FreeLing* [20]; *TagAnt* [21], entre otras.

En primer lugar, Castellón *et al*. [22] presentan: 1) el etiquetado morfosintáctico automático de un corpus, utilizando para ello la herramienta *FreeLing*; 2) la aplicación de una interfaz de anotación que permite refinar el etiquetado; 3) la realización de pruebas de anotación que permitan definir los criterios de desambiguación, y 4) la extensión de operadores, entre los cuales destacan usos metafóricos, metonimias, multipalabras y partitivos.

Sánchez [23] introduce las siguientes etapas: 1) la compilación semiautomática del corpus con las herramientas *BootCaT* y *WebBootCaT,* y 2) el análisis del corpus con *Sketch Engine*, etapa que se encarga del proceso de etiquetado y que presenta, además, una interfaz de explotación del corpus.

Los autores Martí *et al*. [24] presentan una metodología dividida en cinco etapas: 1) el etiquetado morfosintáctico; 2) el etiquetado sintáctico superficial, es decir,

automático; 3) el etiquetado sintáctico profundo, el cual corresponde a una revisión manual del etiquetado automático; 4) la asignación de unos papeles temáticos de los predicados verbales, y 5) el análisis de sentidos de nombres más frecuentes a partir de *WordNet*.

Anxo, Portela y Xavier [25] especifican el proceso de creación de un corpus paralelo inglés-gallego en las etapas: 1) la adaptación automática al *WordNet 3.0* de las etiquetas semánticas del corpus en inglés; 2) la traducción manual al gallego de los textos; 3) la proyección en los textos en gallego de las etiquetas semánticas del inglés, y 4) la consolidación del corpus paralelo inglés-gallego con el resultado de la anotación semántica de la lengua gallega.

Zafra, Gómez y Navarro [26] presentan la anotación de un corpus en las siguientes etapas: 1) la búsqueda y selección de textos; 2) la anotación manual de metadatos; 3) la anotación automática del idioma con la herramienta *Gate Developer,* y 4) la revisión manual del etiquetado.

Rioja [27] y Vila [28] presentan una metodología que, si bien en algunas etapas coincide con la propuesta que se presentará en la sección siguiente, no ahonda en el detalle referente al proceso de etiquetado. El primero especifica las etapas: 1) la definición de fuentes de información; 2) los criterios de aceptación o rechazo de los documentos; 3) el filtro por tipología textual: novelas o relatos cortos, y 4) el filtro por lengua, únicamente en inglés. El segundo trabajo detalla el carácter de los textos y sus fuentes de consecución, pero deja de lado los elementos puntuales del etiquetado.

III. RESULTADOS

Con el fin de construir corpus etiquetados que contengan suficiente información sobre las unidades correferenciales (anáforas y catáforas) al interior de un conjunto de textos y que este pueda ser usado como base de conocimiento para sistemas informáticos como los que tienen por objeto la enseñanza del español como lengua extranjera o para el entrenamiento de algoritmos de aprendizaje de máquina, se establece un método que define los pasos y las consideraciones que se deben tener en su ejecución, y además un sistema de etiquetas que reglan la estructura de los archivos trabajados.

Según lo anterior, es menester decidir cuál es la información más relevante que debe acompañar a cada una de las unidades etiquetables y qué factores son los más apropiados para tener en cuenta en el proceso de etiquetado. Es decir, qué datos se deberían priorizar. Al respecto, se piensa en tres factores principales, según la bibliografía antes señalada: los elementos deberían marcarse con

etiquetas que contengan información sobre 1) Estructuras sintácticas. 2) Información léxico-semántica. 3) Correferencias textuales.

Por otro lado, en cuanto a las fases y etapas, en la revisión realizada se encontró que los trabajos a menudo abordaron en las propuestas los aspectos en relación con la obtención de los textos, el etiquetado morfosintáctico y la corrección manual. No obstante, se dejan de lado consideraciones adicionales como la limpieza de los textos una vez obtenidos, y el almacenamiento y el etiquetado de correferencias textuales. Estos últimos se abordan en el método que se presenta en esta sección y que se aplicó en la consolidación de un corpus que sirve como base del conocimiento en una aplicación para la enseñanza del español como lengua extranjera.

El método que se presenta se divide en dos fases principales: la conformación del corpus, y su etiquetado, cada una dividida, a su vez, en tres etapas (ver Figura 2).

A. Conformación del corpus

Como acción previa al etiquetado, se hace necesario pasar por una fase de conformación del corpus, la cual consta de las etapas que se presentan a continuación:

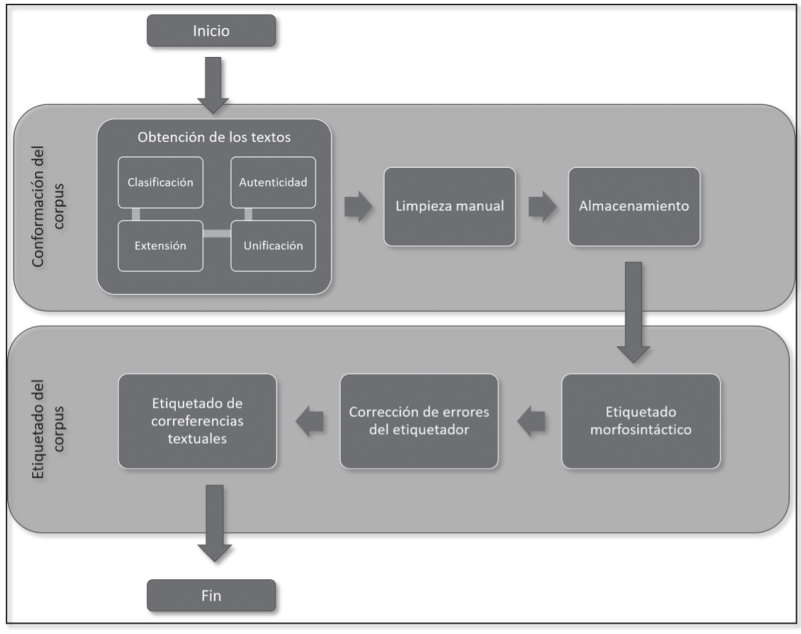


Fig. 2. Fases y etapas del método propuesto

A.1. Obtención de los textos

La primera etapa corresponde a la definición de las diferentes características que permitirán tener un conjunto de textos homogéneos, a saber:

a) *Clasificación de los textos según el nivel de dificultad y los parámetros planteados por el Marco Común Europeo de Referencia (MCER)*: Conviene, en primer lugar, obtener textos según determinados niveles del MCER (B1, B2 y C1, por ejemplo), puesto que estos serán necesarios para realizar actividades de forma automática en una aplicación basada en corpus. La clasificación de los textos en cada uno de estos niveles se hizo con la asistencia del programa informático presentado por Lamprinos [29], que permite calcular el nivel de dificultad de textos en español dentro del Marco Común Europeo de Referencia para las lenguas (MCER), y se buscó que incorporaran variedad con relación a la tipología: textos argumentativos, explicativos, informativos, etc.; y a las fuentes: periódicos, revistas de divulgación científica, extractos de novelas, ensayos breves, etc.

b) *Definición de la extensión de los textos*: Es importante tener un corpus lo más homogéneo posible en lo que respecta al número de *tokens* (palabras y signos de puntuación). Por ello se estableció un umbral recomendado de entre 500 y 1.000 palabras (alrededor de 800 y 1.200 tokens) para cada uno. Esto se estableció con el fin de que exista un equilibrio de contenido en los tres grupos (B1, B2 y C1), además de que los textos cortos facilitan la ejecución de actividades de enseñanza por la agilidad que permiten en el tiempo de las sesiones de estudio. A partir de la Tabla 1 es posible apreciar los valores aproximados para la cantidad de *types* (entradas únicas de cada palabra en el grupo de textos) y para los *tokens* (cantidad total de palabras de cada subcorpus).

Tabla 1. Cantidad de *types* y de *tokens* del corpus DICEELE.

Nivel MCER	Cantidad *types*	Cantidad *tokens*
B1	5.844	30.326
B2	7.518	37.621
C1	7.794	43.356
Total corpus	21.156	111.303

c) *Autenticidad de los textos*: En cuanto a la naturaleza de los textos, es condición que estos sean realizaciones auténticas del español. Se considera importante que los textos no hayan sido previamente didactizados, es decir, construidos artificialmente con

fines de enseñanza u otros fines, ya que esto permite el aprendizaje sobre un entorno más cercano a la realidad. Para ello, es posible hacer un rastreo en Internet y efectuar una selección y extracción de los textos de forma manual.

d) *Unificación del número de textos por nivel a trabajar*: En este caso se hace necesario establecer un número igual por nivel, que, como punto de partida, se recomienda sean 50 documentos por cada uno. No obstante, el corpus debe permitir ser incrementado, de forma igualmente homogénea a futuro, con la consecución de nuevos documentos que sigan los parámetros establecidos en los puntos antes expuestos.

A.2 Limpieza manual

Una vez que se tiene todo el corpus recolectado, se hace necesario pasar a la segunda etapa dentro de la primera fase: la limpieza manual. En este sentido, es importante realizar una revisión cruzada que permita validar que los textos cumplen con el grupo de características previamente definidas y que, además, están correctamente categorizados dentro de su tipología. También, que correspondan a realizaciones ejemplares del uso de la lengua en cuanto a los niveles ortográfico, sintáctico, lexical, etc. Es por eso necesario que se defina un número mínimo de errores corregibles, a partir del cual el texto no podría ser usado en el corpus, de tal manera que se conserve la naturaleza del primero. El problema de una intervención demasiado grande de los textos es que estos podrían perder su carácter de auténticos.

A.3 Almacenamiento

El almacenamiento es la etapa final de la primera fase, y una condición previa frente al proceso de etiquetado. Los recursos se deben unificar a partir de un formato de texto plano (formato *.txt). Esta unificación se hace necesaria toda vez que en la siguiente etapa se empleará el componente automático del método que consiste en etiquetar morfosintácticamente los recursos, con la herramienta *FreeLing* [10]. Por otra parte, es indispensable que los documentos cumplan con una codificación estandarizada de caracteres; en este caso en particular, se recomienda la codificación UTF-8, la cual permite la impresión de letras del alfabeto español como es el caso de la ñ, de las vocales tildadas y de algunos signos de puntuación.

B. Etiquetado del corpus

Una vez que se ha llevado a cabo todo el proceso de recolección de los textos, su limpieza y la codificación, se procede al etiquetado. Para ello, se tienen en cuenta dos niveles de etiquetado: el etiquetado morfosintáctico y el etiquetado de

las correferencias en el nivel de las unidades textuales y discursivas; este último basado en la lingüística textual.

B.1 Etiquetado morfosintáctico

Con el fin de realizar el proceso de etiquetado morfosintáctico, se recomienda usar la herramienta *FreeLing*. Esta es una herramienta libre para el procesamiento del lenguaje natural. *FreeLing* puede ser utilizado de dos formas: a) directamente sobre la página web del proyecto, en su versión demo, o b) por consola, es decir, descargando las librerías y ejecutando el etiquetado de manera local. Este analizador morfosintáctico presenta una documentación amplia, con una descripción generosa de cada *token*. El programa ofrece, además, la posibilidad de utilizar el formato XML en el etiquetado, lo que proporciona la estructura principal sobre la cual se puede hacer, *a posteriori*, todo el etiquetado manual textual y discursivo que se realizará en la etapa final.

En el nivel de etiquetado morfosintáctico se obtienen etiquetas como las de nombre, verbo, adjetivo, etc., para las categorías gramaticales, y las de género, número, persona, tiempo, modo y aspecto, etc., que acompañan a las mencionadas categorías gramaticales. Por otra parte, el sistema permite la segmentación en párrafos y en oraciones, anotados en el mencionado formato XML (ver Figura 3).

Fig. 3. Salida de texto etiquetado en formato XML utilizando *FreeLing* [10].

B.2 Corrección de errores del etiquetador

Por lo general todo sistema para el análisis y etiquetado de las lenguas naturales tiene un margen de error en la asignación de determinadas etiquetas. Esto se debe, fundamentalmente, a la ambigüedad que generan las lenguas naturales. Por esta razón también *FreeLing* incurre en errores de ambigüedad en el etiquetado, aunque con menos frecuencia que otros etiquetadores como *TreeTagger* o *TagAnt*. El nivel de precisión de *FreeLing* en el etiquetado es de 97 a 98 %, gracias a los modelos ocultos de Markov y al *relaxation labelling,* lo que permite la combinación estadística con reglas elaboradas de forma manual [30]. Debido a los márgenes de errores mencionados, se hace necesario que, una vez etiquetados los textos en una estructura XML, se realice una revisión manual de los resultados. Algunos de los errores que este sistema presenta son, *grosso modo,* la confusión entre el uso de la tercera persona y el registro formal de la segunda persona en el singular, el desconocimiento de palabras propias de un dialecto, o la marcación errónea de palabras que comienzan con mayúscula como nombres propios, entre otros.

B.3 Etiquetado de correferencias textuales

En virtud de los propósitos de uso del corpus para la generación automática de actividades que permitan la enseñanza del español, se hace necesario etiquetar diversos fenómenos textuales como la correferencia textual (uso de anáforas y catáforas), la progresión temática (tema y rema), los marcadores y conectores lógico-temporales y discursivos, etc. Hasta el día de hoy no existe una aplicación que lo haga de manera automática para el español [11], por lo tanto, esta etapa debe hacerse de forma manual sobre los XML obtenidos en la etapa anterior. Para este fin se hace uso de los atributos ID e IDREF que arroja el etiquetador. Anotar estas nociones permite llevar a cabo todo un trabajo en la lengua española alrededor de las nociones de coherencia y de cohesión textuales.

Finalmente, una vez obtenidos todos los documentos XML de los recursos que conformarán el corpus, se recomienda su validación a partir de la definición del tipo de documento que especifica todos los documentos en su versión XML (DTD: *Document Type Definition*).

C. Conformación de la DTD

La DTD es un documento que permite, por una parte, la estructuración, lógica y gramática de todo un grupo de textos o de documentos que previamente han sido etiquetados o anotados bajo el formato XML. Por otra parte, la DTD permite la verificación y la validación de dichos documentos con el fin de

que contengan, efectivamente, todos los elementos que se necesitan para la realización de actividades, o que no contengan, si tal es el caso, elementos innecesarios. Además, una DTD puede mapearse informáticamente en el momento en que se quieren realizar acciones como, en este caso, la generación automática de actividades.

La DTD contendrá, además de los elementos etiquetados de manera morfosintáctica y textual, los metadatos necesarios para la realización de los ejercicios. Esto permite que se puedan buscar los textos por nivel, tipo de texto, autor, variante del español, etc. En la Figura 4 se presenta una parte de la DTD recomendada.[1]

La revisión de literatura permitió conocer otras formas de abordar el etiquetado de un corpus. En el caso del uso para la enseñanza del español como lengua extranjera, surgieron particularidades que demandaron una propuesta, fundamentalmente el dominio característico de los textos elegidos, evidenciado en la segunda fase del etiquetado en donde estos se validan: el almacenamiento y el etiquetado manual de las correferencias textuales.

IV. DISCUSIÓN

El método presentado facilita la creación de corpus con las condiciones necesarias para ser utilizados en la enseñanza de lenguas. Es conveniente, siempre, que los sistemas para la enseñanza de ELE sean parametrizables, es decir que puedan dinamizarse sus usos con tipos de textos y ejercicios variados y dinámicos.

El método fue aplicado en la construcción del corpus DICEELE: Dispositivo Informático basado en Corpus para la Enseñanza del Español Lengua Extranjera [11]; se trabajaron 150 textos recolectados de internet, originarios de Colombia, España, Venezuela, México, Argentina, Panamá, Honduras, Nicaragua, Cuba, Ecuador, Uruguay y Perú, y han sido organizados de acuerdo con los niveles B1, B2 y C1 del MCER. Las tipologías textuales recuperadas fueron informativa, narrativa, argumentativa y explicativa, y entre sus diferentes microtipologías se destacan: noticias, cartas, relatos cotidianos, columnas de opinión, reseñas, instrucciones y crónicas. Cada texto cuenta con metadatos que dan cuenta de informaciones tales como el título, el autor, el tipo de publicación, la ubicación, el género textual, el país de procedencia y el investigador encargado de su recolección y sistematización. En la Figura 5 se presenta un ejemplo de cómo se verían etiquetados los elementos correferenciales (unidades anafóricas): *veían, la policía, se, el patrullero, esos chapas*, que retoman al referente de tipo GN (grupo nominal) *un carro de la policía*. Adicionalmente,

1 La DTD completa puede ser solicitada a los autores.

se puede observar el marcado morfológico que del referente principal se hace, por ejemplo, la palabra *carro*, se presenta como un *nombre* (sustantivo), *masculino*, *singular*; esta información puede extraerse de la codificación *NCMS0000* realizada por *FreeLing*.

```
<!ATTLIST sentence id CDATA #IMPLIED>
    <!ELEMENT token (morpho)>
        <!ATTLIST token begin CDATA #IMPLIED
                        ctag (Fia | Fx | Fd | Fz | AQ | AO | AP | Fra | Frc | Fh | DE | VSG | PR | DP | VAS | VAN | VSN |
                        VSP | VSS | VAI | W | VMS | VMP | VMM | Fs | CC | PP | PD | Faa | Fat | PT | I | VMG | VMN | DI |
                        P0 | VSI | Fp | DT | DD | VMI | NC | CS | DA | Fit | RG | Fc | SP | PI | RN) #IMPLIED
                        end CDATA #IMPLIED
                        form CDATA #IMPLIED
                        id CDATA #REQUIRED
                        lemma CDATA #REQUIRED
                        pos (punctuation | adjective | pronoun | date |determiner | interjection | adposition | verb |
                        noun | conjunction | adverb) #IMPLIED
                        tag (AQ0FP00 | AQ0MS000 | AQ0FS000 | AQVMS00 | AQVFS00 | AQVMP00 | AQ0CN00 | AQ0MS00 | AQ0FP00 |
                        AQ0FS00 | AQ0MP00 | AQ0CS000 | AQ0CP00 | AO0MP00 | AO0FS00 | AO0MS00 | RG | RN | DD0FP0 | DD0CS0 |
                        DD0FS0 | DD0MS0 | DD0MP0 | DP2CSS | DP1MPP | DP1FPP | DP1MSP | DP1CPS | DP3CPN | DP3CSN | DP1CSS |
                        DT0CS0 | DT0MP0 | DT0CN0 | DE0CS0 | DE0CN0 | DI0CP0 | DI0CS0 | DI0FP0 | DI0FS0 | DI0MP0 | DI0MS0 |
                        DA00S0 | DA0FS0 | DA0MP0 | DA0FP0 | DA0MS0 | NCFN000 | NCMP00V | NCFP00V | NCCN000 | NCMS00V |
                        NCFS00V | NCCP000 | NCCS000 | NCFP000 | NCFS000 | NCMN000 | NCMP000 | NCMS000 | VAIP2S0 | VAII3P0 |
                        VASI3P0 | VAII1P0 | VAN0000 | VAIC1S0 | VASP1S0 | VAIF3S0 | VAIP1S0 | VAII3S0 | VASI1S0 | VAIP1P0 |
                        VAIP3P0 | VAIS1S0 | VAIP3S0 | VMIS2S0 | VMM02P0 | VMIF1S0 | VMIP2S0 | VMSF2S0 | VMIC3P0 | VMM03S0 |
                        VMSI1P0 | VMIC1S0 | VMIF3P0 | VMSI1S0 | VMP00PF | VMIS1P0 | VMIF3S0 | VMIS3P0 | VMM02S0 | VMSP3P0 |
                        VMSI3S0 | VMSP1P0 | VMSI3P0 | VMII1P0 | VMM03P0 | VMIC3S0 | VMP00PM | VMIS1S0 | VMIP1P0 | VMSP3S0 |
                        VMII3S0 | VMII3P0 | VMIP3P0 | VMM01P0 | VMIS3S0 | VMG0000 | VMP0000 | VMP00SM | VMP00SF | VMN0000 |
                        VMIP1S0 | VMIP3S0 | VSSP3P0 | VSIS1P0 | VSG0000 | VSII1P0 | VSIS3P0 | VSIF3S0 | VSII3S0 | VSIP1S0 |
                        VSN0000 | VSSI3S0 | VSII3P0 | VSIC1S0 | VSIP1P0 | VSIS3S0 | VSSP1S0 | VSP00SM | VSIP3P0 | VSIP3S0 |
                        PP2CP00 | PP1FP00 | PP2CS00 | PP3FP00 | PP3MS00 | PP3MPA0 | PP2CP0P | PP3CPD0 | PP3CN00 | PP3FPA0 |
                        PP3CN00 | PP3MP00 | PP3FS00 | PP3FSA0 | PP1MP00 | PP1CP00 | PP3MSA0 | PP3CSD0 | PP1CSN0 | PP2CS0P |
                        PP1CS00 | PP1CS00 | PD0MP00 | PD0CS00 | PD0FS00 | PD0FP00 | PD0MS00 | PD00S00 | PI0CP00 | PI0FP00 |
                        PI0FS00 | PI0CS00 | PI0MP00 | PI0MS00 | PT0CP00 | PT00000 | PT0CS00 | PR0CS00 | PR00000 | PR0CN00 |
                        CC | CS | I | P02CS00 | P01CP00 | P00CN00 | P01CS00 | AP0MS1S | AP0MS2S | SP | Fia | Fx | Fd | Fra |
                        Frc | Fz | Fh | W | Fs | Faa | Fat | Fp | Fit | Fc) #IMPLIED
                        type (questionmark | qualificative | slash | quotation | exclamative | ordinal | semicolon |
                        other | colon | possessive | relative | etc | exclamationmark | comma | personal | period |
                        interrogative | preposition | demonstrative | coordinating | semiauxiliary | negative |
                        indefinite | auxiliary | main | common | subordinating | article | general) #IMPLIED
```

Fig. 4. Fragmento de la DTD en la que se aprecian parte de las etiquetas contenidas en los documentos XML

```
<REF ANT="nom" type="SN" id="50" from="268" to="272" chain="8">

    <element morpho="DI0MS0" id="268">un</element>
    <element morpho="NCMS000" id="269">carro</element>
    <element morpho="SP" id="270">de</element>
    <element morpho="DA0FS0" id="271">la</element>
    <element morpho="NCCS000" id="272">policía</element>
</REF>
<chain id="8">
<REF ANT="nom" type="SN" id="50" words="un carro de la policía "/>
<COREF ANA="nom" type="SN" id="53" words="la patrulla policial "/>
<COREF ANA="pron" type="imp" id="54" words="veían "/>
<COREF ANA="nom" type="SN" id="62" words="la policía "/>
<COREF ANA="nom" type="SN" id="85" words="la policía "/>
<COREF ANA="pron" type="PP" id="87" words="se "/>
<COREF ANA="nom" type="SN" id="97" words="esos chapas "/>
<COREF ANA="nom" type="SN" id="98" words="el patrullero "/>
</chain>
```

Fig. 5. Etiquetas de correferencia textual: anáforas

Si bien el método se puso en marcha en el contexto mencionado, es decir para un contexto de aplicación de ELE, puede ser extendido a otros contextos como el entrenamiento de algoritmos de aprendizaje de máquina.

En los modelos de aprendizaje de máquina, a menudo implementados a través de redes neuronales, la capacidad del aprendizaje hacia uno profundo (*deep learning*) es posible gracias a la obtención de suficientes datos para

etapas de entrenamiento, validación y verificación. Por eso se hace necesaria la creación de corpus en distintas lenguas, con objetivos de marcaje de la información variados, y que funcionen como bases de conocimiento para la algoritmia tras la inteligencia artificial. El tamaño de las redes neuronales y la cantidad suficiente de datos es, por lo tanto, fundamental para impulsar el avance del aprendizaje profundo y alcanzar, así, un nivel más preciso de la extracción de sentido de las distintas maneras que tienen los humanos de producirlo. En esta línea es esencial avanzar más allá del análisis sintáctico hacia una integración con modelos de conocimiento que permitan establecer una relación lenguaje-realidad que se aproxime más a las características y capacidades humanas, características que se consiguen, en un grado inicial, a partir del marcado de correferencias.

V. CONCLUSIONES

En este trabajo se presentó un método que conduce a la consolidación de un corpus para su uso en sistemas informáticos que medien la enseñanza del español como lengua extranjera (ELE), que, sin embargo, se puede extender a otros casos de aplicación. Se introduce la herramienta *FreeLing*, que asiste en el etiquetado morfosintáctico en la etapa automatizada. Sobre esta última también se expusieron sus principales limitantes, las cuales se deben tener en consideración siempre que se utilice, y de esta forma se podrán realizar los ajustes, de manera manual, a los que haya lugar.

El método presentado incorpora las fases de consolidación de los textos, así como el proceso de etiquetado como tal, y se sugiere un conjunto de pasos que, llevados a cabo, permitan garantizan la idoneidad del corpus. Se trata de un método semiautomático, puesto que el etiquetado morfosintáctico se efectúa de manera automática, mientras que el etiquetado de correferencias textuales debe realizarse de manera manual o, en su defecto, a partir de algunos procedimientos que permitan agilizar ciertos pasos del proceso de consolidación de las anotaciones, por medio de *scripts* programados, por ejemplo, en *Python* o en *Perl*.

VI. REFERENCIAS

[1] S. A. Moreno, *Lingüística computacional: Introducción a los modelos simbólicos, estadísticos y biológicos*. Madrid: Editorial Síntesis, 1998.

[2] R. Hausser, *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language*. New York: Springer Science & Business Media, 2013.

[3] G. Parodi, *Lingüística de Corpus: de la teoría a la empiria*. Madrid: Iberoamericana Vervuert, 2010.

[4] J. Bernal y D. Hincapié, *Lingüística de corpus*. Colombia: Instituto Caro y Cuervo. 2018.

[5] J. M. Molina, "Lingüística de corpus utilizada en la concepción de plataformas de formación de docentes de lenguas", *CHIMERA: Romance Corpora and Linguistic Studies,* vol. 3, n.° 1, pp. 57-85, 2016.

[6] M. Recasens and M. A. Martí, "AnCora-CO: Coreferentially annotated corpora for Spanish and Catalan", Lang Resources & Evaluation, n.° 44, 2010.

[7] M. Palomar, L. Moreno, J. Peral, R. Muñoz, A. Fernández, P. Martínez-Barco, and M. Saiz-Noeda, "An Algorithm for Anaphora Resolution in Spanish Texts", *Association for Computational Linguistics,* vol. 27, n.°, 4, pp. 545-567, 2001.

[8] G. Kundu, A. Sil, R. Florian, and W. Hamza. "Neural Cross-Lingual Coreference Resolution and its Application to Entity Linking", in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (Short Papers), Melbourne, Australia, July 15-20, 2018, pp. 395-400.

[9] A. F. Grajales Ramírez y J. M. Molina Mejía, "Problemática actual del procesamiento computacional anafórico: el caso de FreeLing 4.1", *Lenguaje,* vol. 47, n.° 2S, pp. 537-568, 2019. doi: 10.25100/lenguaje.v47i3.8356

[10] L. Padró and E. Stanilovsky, "FreeLing 3.0: Towards Wider Multilinguality", in *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, ELRA. Istanbul, Turkey. May 2012.

[11] J. M. Molina, A. F. Grajales, and J. L. Pemberty, "Hacia un dispositivo informático basado en Corpus para la Enseñanza del Español Lengua Extranjera (DICEELE)", *Revista Internacional de Tecnología, Conocimiento y Sociedad*. vol. 7, n.° 1, pp 1-13, 2019. doi: 10.18848/2474-588X/CGP/v07i01/1-13

[12] M. A. Martí, *Introducción Tecnologías del lenguaje*, Barcelona: Editorial UOC, 2003.

[13] A. Boulton et H. Tyne, *Des documents authentiques aux corpus: démarches pour l'apprentissage des langues*. Paris. Didier Coll Langues & didactique, 2014.

[14] G. Antoniadis. "De l'apport pertinent du TAL pour les systèmes d'ALAO. L'exemple

du projet MIRTO", en 2eme Congrès Mondial de Linguistique Française (CMLF-2010). La Nouvelle Orléans, USA, 12-15 juillet 2010.

[15] J. M. Molina, "ELiTe-[FLE]: Un environnement d'ALAO fondé sur la linguistique textuelle, pour la formation linguistique des futurs enseignants de FLE en Colombie », Tesis doctoral, Université Grenoble Alpes, 2015.

[16] R. Mitkov, "Outstanding Issues in Anaphora Resolution", In *Computational Linguistics and Intelligent Text Processing,* A. Gelbukh, Ed., Springer, 2001, pp. 110-125.

[17] F. Landragin, "Anaphores et coréférences : analyse assistée par ordinateur", en *Nouvelles perspectives sur l'anaphore: Points de vue linguistique, psycholinguistique et acquisitionnel,* M. Fossard & M.-J. Béguelin, Eds., Peter Lang. 2014, pp. 29-54.

[18] H. Schmid. "Probabilistic Part-of-Speech Tagging Using Decision Trees", in Proceedings of International Conference on New Methods in Language Processing, Manchester, UK, 1994.

[19] D. Klein and C. D. Manning. "Accurate Unlexicalized Parsing", in *Proceedings of the 41st Meeting of the Association for Computational Linguistics,* 2003, pp. 423-430.

[20] X. Carreras, I. Chao, L. Padró and M. Padró. "FreeLing: An Open-Source Suite of Language Analyzers", in Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04), 2004.

[21] L. Anthony. "TagAnt (Version 1.1.0) [Computer Software]". Tokyo, Japan: Waseda University, 2015. Available from http://www.laurenceanthony.net/software

[22] I. Castellón; S. Climent; M. Coll-Florit; M. Lloberes and G. Rigau. "Constitución de un corpus de semántica verbal del español: Metodología de anotación de núcleos argumentales", *RLA. Revista de lingüística teórica y aplicada,* vol. 50, n.° 1, 2012, pp. 13-38.

[23] M. del M. Sánchez, "Metodología de corpus y formación en la traducción especializada (inglés-español): una propuesta para la mejora de la adquisición de vocabulario especializado", *Revista de Lingüística y Lenguas Aplicadas,* vol. 12, n.° 1, 2017, pp. 137.

[24] M. A. Martí, M. Taulé, L. Màrquez, and M. Bertran. "Anotación semiautomática con papeles temáticos de los corpus CESS-ECE", Procesamiento del Lenguaje Natural, n.° 38, 2017, pp. 67-76.

[25] M. Anxo; S. Portela and G. Xavier. "Diseño y elaboración del corpus SemCor del gallego anotado semánticamente con WordNet 3.0", *Procesamiento del Lenguaje Natural,* N.° 59, 2017, pp. 137-140.

[26] S. Zafra Cremades, J. M. Gómez Soriano, and B. Navarro-Colorado. "Diseño, compilación y anotación de un corpus para la detección de mensajes suicidas en redes sociales", *Procesamiento del Lenguaje Natural,* vol. 59, n.° 0, 2017, pp. 65-72.

[27] M. Rioja Barrocal. "Metodología para la narrativa de traducciones censuradas inglés-español: análisis del Corpus 0 TRACEni", in Estudios Humanísticos. Filología. 317. 10.18002/ehf. v0i29.2821, 2007.

[28] A. R. Vila. "Canon y corpus literario latinoamericano y caribeño. Una metodología de construcción del corpus", in Palabra Clave (La Plata), vol. 7, N.° 2, 2018, pp. 0-12.

[29] D. Lamprinos, "Clasificación automática de textos españoles en los niveles del Marco Común Europeo de Referencia para las Lenguas", Tesis de maestría, Universidad Nacional y Kapodistriaca de Atenas y la Universidad Politécnica Nacional de Atenas, 2014.

[30] L. Padró, "Analizadores Multilingües en FreeLing", in Linguamatica, vol. 3, N.° 2, 2011, pp. 13–20.

Capítulo 9
Modelo de Transformación Digital para la Creación de Destinos Turísticos Inteligentes Caso de Estudio: Támesis

Sebastian Gomez Jaramillo - Sgomezja@tdea.edu.co
Juan David Aranzazu - David3356@gmail.com
Jorge Mario Cadavid - Jcadavi7@correo.tdea.edu.co
Tecnologico de Antioquia

I. INTRODUCCION

En todo el mundo, la migración de población a las grandes capitales ha provocado, por un lado, su sobrepoblación y, por el otro, que las zonas rurales, y en el caso de Colombia sus municipios, disminuyan su población y crecimiento económico, lo cual ha generado impactos negativos en sus habitantes y territorios aumentado la brecha digital y el desempleo.

Por eso se plantea implementar y apropiar el uso de las TIC y de datos abiertos para transformar los municipios de Colombia en destinos turísticos inteligentes e incluyentes, con el fin de disminuir la migración de sus habitantes, mejorar la economía y atraer a visitantes a lo que se conoce como turismo rural o de naturaleza, ya que por sus riquezas naturales, biodiversidad y cultura, Colombia está considerada como una de las potencias mundiales en materia de turismo sostenible, eficiente y competitivo.

Sumado a esto, en Colombia se ha firmado el acuerdo de paz con la guerrilla de las FARC que pone punto final a una guerra de más de 50 años que ha dejado miles de víctimas, el acuerdo ha propiciado una época de transición hacia la paz y ha generado oportunidades para sectores afectados directamente por el conflicto, relacionados con el turismo y la movilidad en las carreteras del país, entre otros aspectos. Es así que la comunidad internacional se ha puesto los ojos sobre el país y ha aumentado la cantidad de visitantes extranjeros, lo que perfila a Colombia como potencia turística mundial gracias a su riqueza cultural y biodiversidad.

Se generan, entonces, nuevas oportunidades en el sector turismo, como la necesidad de documentar toda la oferta turística, el patrimonio cultural, natural y social de cada región rural de Colombia, poniendo esa oferta accesible en internet e internacionalizando territorios rurales y municipios con fin de promover un turismo eficiente para las subregiones de este país, estimulando el crecimiento social y económico de los municipios mediante el aprovechamiento sostenible de sus recursos naturales, convirtiéndolos en destinos inteligentes.

Lo que hace un destino turístico inteligente es el trabajo conjunto entre los sectores público, privado y mixto, donde se generan propuestas innovadoras y tecnológicas positivas por medio del uso de las TIC para diversos sectores, como los de movilidad, cultura, seguridad, salud, economía y servicios públicos, propiciando crecimiento económico competitivo, con gestión eficiente, mejorando la calidad de vida del residente y la experiencia del visitante.

Este trabajo de investigación diseña y desarrolla un modelo de ciudad inteligente que se adapta para ser aplicado a los municipios de Colombia. Este modelo basado en TIC y datos abiertos fue implementado con éxito en el municipio de Támesis, Antioquia, que se convirtió en el primer destino turístico inteligente, haciendo énfasis en puntos tan importantes como: la oficina turística del siglo XXI, la seguridad y accesibilidad para las personas con movilidad reducida y la centralización de la información. Además, genera interacción entre residentes, turistas y gobierno y muestra toda la oferta turística y de territorio a los visitantes quienes desean mejorar sus expectativas.

II. ANTECEDENTES

Los desarrollos recientes han culminado en la noción del destino inteligente, en el cual las TIC se consolidan como el motor y la base para la innovación y competitividad del destino.

En algunas economías emergentes, el concepto de destino inteligente se está convirtiendo en una parte integral de la política nacional para el desarrollo económico y la mejora general de la calidad del servicio con la ayuda de las TIC. Está, por ejemplo, la Iniciativa de Destino Inteligente propuesta por el Consejo de Estado del Gobierno Central de China, cuyo objetivo es construir una plataforma en la que las TIC se puedan integrar sin problemas en las actividades turísticas, el consumo de productos turísticos y la gestión de los recursos turísticos a través de una variedad de dispositivos de usuario final [1]. Por lo tanto, el concepto de destino inteligente se ha adoptado no solo como un marco

conceptual, sino también como una herramienta estratégica para destinos en todo el mundo con el fin de apoyar la creación de experiencia turística, comunicarse con los consumidores de manera efectiva y medir y evaluar la competitividad del destino.

Recientemente, ha habido un creciente interés de la academia de turismo para conceptualizar el destino inteligente. En general, el destino inteligente se ha basado en la noción de ciudad inteligente que prevé que la nueva estrategia de desarrollo urbano se base en el uso de las TIC en varias áreas claves como la economía, el medioambiente, la movilidad y la gobernanza para transformar la infraestructura de la ciudad y los servicios [2].

La evolución de una ciudad inteligente converge en el concepto de "destino inteligente", sin embargo, una serie de transformaciones digitales y estructurales en el sector turístico justifican la necesidad de un cambio en la gestión y la gobernanza. En [3] se define el concepto de una *Smart destination*:

> Un destino turístico innovador, consolidado sobre una infraestructura tecnológica de vanguardia, que garantiza el desarrollo sostenible del territorio turístico, que promueve la accesibilidad universal, que facilita la interacción e integración del visitante con el entorno e incrementa la calidad de su experiencia en el destino y mejora la calidad de vida del residente.

Para López de Ávila [4], el destino inteligente es un destino turístico innovador construido sobre una infraestructura de tecnología de punta, que garantiza el desarrollo sostenible de las zonas turísticas, le facilita al visitante que pueda interactuar e integrarse con el entorno, aumenta la calidad de la experiencia en el destino y mejora la calidad de vida de los residentes.

Respecto al destino turístico inteligente, Buhalis [5] sugiere:

> Basado en la investigación y las metodologías de Smart Cities, un Smart Tourism Destination implementa con éxito la inteligencia fomentada por la innovación abierta, respaldada por inversiones en capital humano y social, y sustentada por la gobernanza participativa para desarrollar la competitividad colectiva de los destinos turísticos para mejorar la prosperidad social, económica y ambiental de todos los interesados. La interoperabilidad y la computación ubicua aseguran que todos estén interconectados y que los procesos se integren para generar valor, a través de la co-creación dinámica, los recursos sostenibles y la personalización dinámica y la adaptación al

contexto. Todos los proveedores y los intermediarios, el sector público, así como los consumidores y varias partes interesadas están en red, coproduciendo dinámicamente valor para todos los interconectados en el ecosistema.

Desde la perspectiva del consumidor, se dice que un destino es inteligente cuando hace un uso intensivo de la infraestructura tecnológica con el fin de mejorar la experiencia turística de los visitantes al personalizarlos y darles a conocer los servicios y productos locales y turísticos disponibles en el destino.

Desde la perspectiva gerencial y de gobernanza, un destino inteligente puede empoderar a las organizaciones de gestión de destinos, las instituciones locales y las empresas de turismo para tomar sus decisiones y emprender acciones basadas en los datos producidos en el destino, recopilados, gestionados y procesados mediante la infraestructura de las TIC [6]. Dentro de este entorno, las partes interesadas del turismo se deben interconectar dinámicamente a través de plataformas tecnológicas para recopilar, crear e intercambiar información que pueda usarse para enriquecer las experiencias turísticas en tiempo real [7].

Recientemente, se visualiza el destino inteligente como parte integral del ecosistema de turismo inteligente en el que la tecnología genera nuevos modelos de negocios, paradigmas de interacción e incluso nuevas especies de negocios turísticos. En general, el destino inteligente sigue siendo un tema emergente en la investigación turística que requiere la integración del conocimiento de una serie de campos relevantes, como los sistemas de información, el comportamiento de los viajes, el *marketing,* la planificación urbana, la gestión del destino y la gobernanza, así como el análisis de datos importantes y ciencias de datos [8].

En esta perspectiva, la mera adquisición y aplicación de tecnología no hace un *Smart Destination* ni asegura la inteligencia. Es fundamental que la tecnología vaya acompañada de un proceso de cambio a todos los niveles (estrategia turística, mentalidad y cultura empresarial, capacidad de absorber innovación, y generación de nuevos modelos de negocio) [9](Castro 2018). Por tanto, no se trata simplemente de introducir las nuevas aplicaciones tecnológicas, sino de revolucionar la gestión turística según las posibilidades tecnológicas y la capacidad de acción local. Para lograrlo se requiere entonces, según Fabronni (2017), que diferentes actores del sector turístico lideren la misión de cooperar y coordinar acciones encaminadas a estructurar la oferta turística y el valor de sus recursos. Las partes interesadas incluyen la academia y la ciencia, así como representantes de los sectores

público y privado y la sociedad civil en general [10].

III. INVENTARIO TURÍSTICO EN COLOMBIA

Un inventario turístico es un proceso por el cual se registran, de forma ordenada, los factores físicos y culturales como un conjunto de atractivos que sirven de base para elaborar productos turísticos para una región. Los recursos turísticos son aquellas posesiones que tiene una población, que poseen características que implican un atractivo para el turismo, bien sea de carácter natural de esparcimiento y recreación histórico-cultural, para utilizarlos de la mejor manera y poder ofrecer un producto turístico.

En Colombia, el Ministerio de Industria y Comercio desarrolló y publicó su metodología para el levantamiento de inventarios turísticos en el territorio colombiano y teniendo en cuenta la información suministrada no se encuentra actualizada de manera uniforme en las diferentes regiones del país y considerando la importancia del inventario turístico, como instrumento informativo para las diferentes entidades, el Ministerio de Desarrollo, por intermedio del viceministro de Turismo, construyó en 1996 una nueva guía para la elaboración del inventario turístico nacional, integrado a la metodología del Programa de Asistencia Técnica en Planificación Turística. Se deben evaluar tres categorías:

1. Sitios naturales
2. Recursos culturales
3. Festividades y eventos.

El levantamiento del inventario de recursos turísticos comprende dos fases de trabajo como son ordenar y clasificar los lugares, objetos, acontecimientos y otros de interés turístico de los recursos turísticos del país, la región o área determinada, proceso que se refiere a la recopilación de información, ordenamiento y clasificación de los datos.

A continuación, se presentan tres tablas con la clasificación del inventario turístico en Colombia que fue utilizado para aplicarla al municipio de Tamesis en el caso de estudio.

Tabla 1. Clasificación del inventario turístico – Sitios Naturales

Sitios naturales	Subtipos
Montañas	Nudo o macizo
	Loma o colina
	Nevado Volcán
	Cordillera
	Sierra
	Serranía
	Monte

Sitios naturales	Subtipos
Altiplanicies	Meseta
	Valle
	Depresión
	Cañón
Llanuras	Sabana o llano
	Pradera
	Llanura selvática
	Desierto
	Depresión
Lagunas	
Ríos	
Cascadas	
Costas o litorales	Playa
	Acantilado
	Golfo
	Bahía
	Ensenada
	Península
	Istmo
	Estrecho
Islas	Archipiélago
	Islote
	Cayo
	Arrecifes
	Otros
Lugares de caza	
Lugares de observación de fauna y flora	
Parques nacionales naturales	
Parque nacional natural	
Reserva natural	
Área natural náutica	
Santuario de fauna y flora	
Vía parque	
Otros	

Tabla 2. Clasificación del inventario turístico – Recursos culturales

Recursos culturales	Subtipos
Arqueológicos	Zona
	Conjunto
Arquitectónicos	Civil
	Religioso
	Militar
	Funerario
Históricos	Ciudad
	Sector
	Conjunto
	Sitio
Obras de arte	Pintura
	Escultura
	Otra
Expresiones religiosas	Peregrinación
	Ofrenda
	Promesa
	Ceremonia
Manifestaciones folclóricas	Comida
	Bebida
	Música
	Danza
	Trajes
	Artesanías
	Literatura
Realizaciones técnico-científicas	Minería
	Agropecuaria
	Industrial
	Técnica
Grupos de interés especial	Comunidad indígena
	Comunidad negra
	Comunidad raizal
	Otra

Tabla 3. Clasificación del inventario turístico – Festividades y eventos

Festividades y eventos	Subtipos
Ferias y exposiciones	
Eventos:	Artísticos
	Culturales
	Deportivos
Aniversarios	
Reinados	
Encuentros	
Día del mercado	
Otros	

IV. CREACIÓN DE DESTINOS TURÍSTICOS INTELIGENTES

Para crear un destino turístico inteligente (o *Smart Destination*) es necesario que ese espacio turístico sea innovador, accesible para todos, consolidado sobre una infraestructura tecnológica de vanguardia que pueda garantizar el desarrollo sostenible del territorio, facilitar la interacción e integración del visitante y el morador con el entorno e incrementa la calidad de su experiencia en el destino y la calidad de vida de los residentes.

Este modelo se plantea con el fin de contribuir al desarrollo económico de los municipios para mejorar su competitividad en factores tangibles como los atractivos turísticos y servicios que ofrecen: la calidad integral del destino (calidad del medioambiente, paisajismo del entorno, facilidades y servicios turísticos); innovación en sus atractivos, actitud competitiva.

No se puede hablar de competitividad sin tener en cuenta la sostenibilidad, la cual es fundamental para preservar y cuidar los recursos de los municipios. En el caso particular de Támesis, Antioquia, la gestión de todos los recursos debe satisfacer los aspectos económico, social, estético, integral, cultural, ecológico, la diversidad biológica y los sistemas que apoyan la vida. En otras palabras, se trata de un modelo que responde a las necesidades de los turistas y de la región anfitriona, a la vez que protege y mejora las oportunidades futuras sin afectar sus valores sociales o su ambiente.

Es importante mencionar el papel que cumple la gobernanza, ya que le compete al sector público determinar la forma en que se ejerce el poder para desplegar los recursos económicos y sociales, y que se caracteriza por la participación, la legalidad, la transparencia, el consenso, la equidad, la eficacia, la eficiencia y la sensibilidad. Para lograr la competitividad, la gobernanza interactúa con el sector privado en el área de turismo según sus objetivos, intereses y estrategias.

En busca de un modelo que cumpliera con las características antes mencionadas, y teniendo en cuenta el potencial y los servicios turísticos que se ofrecen en el municipio de

Támesis, Antioquia, se creó un modelo por medio del uso de las TIC y se generaron datos abiertos donde toda la información del destino turístico se encuentra centralizada y accesible para todos. Por ello se diseñó un aplicativo móvil que centra toda la oferta de cultura, noticias, atractivos, restaurantes, hoteles, comercio y accesibilidad para personas con algún grado de discapacidad, y une al sector público, es decir sus gobernantes, con la población. De este modo, se logra que el visitante encuentre a un clic todo lo que el municipio le ofrece, además, se acerca a los gobernantes a su población y se genera interacción entre turistas, anfitriones, sector público y privado, con el objetivo de producir competitividad y desarrollo mediante un modelo de negocio fundamentado en el uso de las TIC para el impulso de los destinos turísticos inteligentes.

Con la generación de datos abiertos mediante el uso de las TIC para diseñar una aplicación móvil, surge una idea de negocio que pudiera ser usada y replicada en los municipios de Colombia como una solución a la migración de la población, buscando que los ojos de las personas nuevamente se enfocaran en el turismo rural y el deseo de volver a tierras que en épocas pasadas fueron abandonadas por el conflicto armado.

El modelo de turismo se desarrolla a partir de los conceptos extraídos del libro *Bussiness Model Generation* de Alex Osterwalder e Yves Pigneur [11], como se describe a continuación:

A. Segmento de clientes

Son los clientes a los que va dirigida la aplicación de destino de turismo inteligente. Los usuarios del sistema son todas las personas que la descargan y navegan por nuestros municipios inteligentes usando sus servicios y conociendo el territorio.

Los clientes son las personas que pagan por un sitio dentro de nuestra plataforma, el cual sirve como vitrina virtual para la promoción de sus productos.

B. Propuesta de valor

Indica el factor diferencial de la aplicación, la cual permite atender una necesidad del cliente y solucionarle un problema. En el caso turístico, la necesidad es relativa al número de visitantes, los clientes potenciales y el impacto económico que pueda tener sobre el municipio.

La propuesta descrita en el presente trabajo es una plataforma web/app con un sistema de información basado en una arquitectura de datos que centraliza la oferta turística, cultural y comercial de los municipios, con el fin de hacerlos visibles, integrarlos e internacionalizarlos, buscando su alianza para el aprovechamiento de sus territorios y poder crecer de manera

social, económica y sostenible, todo con base en arquitectura de datos abiertos para que cualquier persona, en cualquier parte del mundo a cualquier hora, pueda navegar por los municipios digitales de Colombia y acceder a sus servicios y oferta comercial.

C. Canales

Los canales indican el medio de comunicación para llegar a los clientes y a los usuarios finales que van a interactuar con la plataforma.

- Canales directos
  - Sitio web
  - App móvil
- Canales indirectos
  - Oficinas de turismo
  - Alianzas con terceros

D. Relación con clientes

Identifica los medios y los tipos de relación que se mantendrán con los clientes y usuarios de la aplicación del destino turístico.

Para atraer, mantener y generar nuevos clientes y alianzas, es vital para la Alianza colombia.travel generar procesos de acompañamiento personalizado en temas de tecnología y de emprendimiento a los negocios y clientes de la app.

El contacto y el acompañamiento deben ser constantes, ya sea virtual, personalizado o telefónico, se mantienen comunicaciones todo el tiempo con el cliente para generar confianza inversionista.

E. Fuentes de ingresos

Los ingresos están determinados por varias fuentes; venta de la app a:

*Alcaldías municipales, hoteles, bares, negocios y operadores turísticos, también ganamos dinero por concepto de publicidad y creación de contenidos audiovisuales para nuestros clientes. también nos contratan los clientes para que hagamos administración de sus redes sociales o *community manager*.

F. Recursos claves

Infraestructura de sistema de información: *Hosting*, dominio, almacenamiento y cuentas en línea, terminales móviles android y iphone, licencias de software.

Recurso humano especializado: desarrolladores, animadores, guionista, mercadólogo, soporte técnico, georreferenciación, logística y herramientas de *marketing* tradicional y digital.

G. Actividades clave

- Infraestructura tecnológica: es necesario un equipo de trabajo interdisciplinario para completar las tareas técnicas propias de nuestra app, es necesario diseñador, que puedan ayudar con labores audiovisuales, desarrollador de software, quien mantenga el producto.

- Relacionamiento público: es vital el cabildeo y los contactos que podamos aliar a la causa de la promoción de los destinos turísticos rurales, se necesitan contactos con los líderes de diferentes gremios con el fin de promover alianzas y el trabajo en equipo en los territorios.
- Mercadeo: conocer mantener y generar nuevos clientes con ayudad de la publicidad.
- Producción de contenidos: crear producciones audiovisuales para vender fácilmente las actividades y experiencias que se pueden vivir en un destino turístico.
- Documentación del patrimonio rural turístico y cultural de cada territorio mediante contenidos digitales.

H. Socios clave

Permiten generar ayudas para poder responder a las necesidades de los clientes:

- Ministerio de las Tecnologías de la Información y las Comunicaciones (MinTIC)
- Ministerio de Comercio, Industria y Turismo
- Provincia del Cartama
- Operadores turísticos
- Equipo de trabajo interno
- Gremios y grupos de empresarios

I. Costos de operación

La mayor parte de los costos de la operación se enfocan al pago del capital humano empleado, como son los programadores, diseñadores, animadores y productores audiovisuales. También se necesita el recurso técnico para que ese personal ejecute sus tareas: cámaras y equipos de video, infraestructura en la nube, servicios continuos de seguridad de la información. En general sería

- Recurso humano
- Infraestructura tecnológica
- Logística
- Publicidad
- Costos indirectos
- Georreferenciación

V. CASO DE ESTUDIO

El municipio de Támesis, ubicado en la subregión Suroeste del departamento de Antioquia, en Colombia, limita por el norte con los municipios de Jericó y Fredonia, por el este con el municipio de Valparaíso, por el sur con los municipios de Caramanta y Jardín y además con el departamento de Caldas, y por el oeste con los municipios de Jardín y Jericó [12].

Fig. 1. Ubicación geográfica del unicipio de Támesis

El turismo en el municipio cuenta con una excelente infraestructura hotelera que va desde hoteles spa hasta pequeñas residencias, además en reiteradas ocasiones ha sido sede del encuentro nacional de caminantes y practicantes del *rappel* y otros deportes extremos, es una potencia turística por su riqueza hídrica, boscosa y paisajística y cuenta con un punto de información turística. Támesis se encuentra en una zona rica en petroglifos que datan de la época prehispánica, se pueden visitar aproximadamente 93 de las más de 300 piedras talladas por indígenas embera chamí.

Una vez documentados y clasificados todos los sitios naturales, recursos culturales, festividades y eventos en Támesis - Antioquia , también se documenta toda la oferta territorial y comercial que al turista y al morador les interesa acceder, no solo basta con el inventario turístico, el concepto de destino turístico está asociado también a toda la oferta de territorio; ese sitio donde se realizan las actividades que desempeña el turista se llama espacio turístico, y su caracterización o distinción define el paisaje, el cual puede ser clasificado como natural, cultural o urbano.

Para el municipio de Támesis se hizo una clasificación territorial siguiendo los lineamientos del inventario turístico de Colombia, la cual se detalla en la Tabla 4.

Tabla 4. Elementos clasificados en el municipio de Támesis

Ítem	Descripción
Perfil	Descripción detallada sobre el lugar o negocio, incluir elementos diferenciadores que hacen ese lugar único.
Oferta	Portafolio de productos o de servicios del negocio.
Promociones	Descuentos, promociones y beneficios que ofrece ese lugar a sus visitantes.
Contacto	Información de contacto y canal de comunicación directo entre el cliente y el proveedor o administrador del negocio
Ubicación	Información georrefeneciada sobre la ubicación mediante sistema de GPS y dirección del lugar o negocio.
Galería digital	Fotos, videos y contenido transmedia que permitan dar a conocer y promocionar el lugar, negocio o servicio.
Fuentes primarias	– Observación Directa Mesas de trabajo con la oficina encargada de promocionar el turismo. – Información recolectada por presidentes de juntas comunales – Mesas de trabajo con personas involucradas en el sector turismo
Fuentes secundarias	– Documento del EOT (Esquema de Ordenamiento Territorial) – Programas de gobierno del actual alcalde – Plan de desarrollo municipal – Plan de desarrollo departamental – Plan nacional de desarrollo – DANE – Ministerio de Industria Comercio y Turismo – Cámara de Comercio

La recapitulación de todo el proceso de documentación, no solo turística, sino también comercial y territorial comprendió las siguientes etapas:

Etapa 1. Recopilación de información: se recopila la información existente del inventario turístico de cada una de las veredas por medio de planes, esquemas de ordenamiento territorial, entrevistas no estructuradas a los habitantes de cada una de las veredas para la concertación de los atractivos existentes, trabajos de tesis de grado y otros documentos que se utilizarán respecto de tales atractivos.

Etapa 2. Verificación, complementación de la información y reconocimiento de condiciones: se realizan visitas a los recursos naturales, facilidades y servicios para confrontar, complementar y registrar la información inicial en los formularios.

Etapa 3. Registro audiovisual: se elaboran contenidos audiovisuales que muestran cada atractivo y facilidades

de acceso, recomendaciones y promoción turística del sitio.

Etapa 4. Elaboración de bases de datos abiertas: en esta etapa se lleva a cabo el respectivo registro de la información oficial que será reportada al MinTIC sobre la oferta disponible en el municipio y sus territorios rurales.

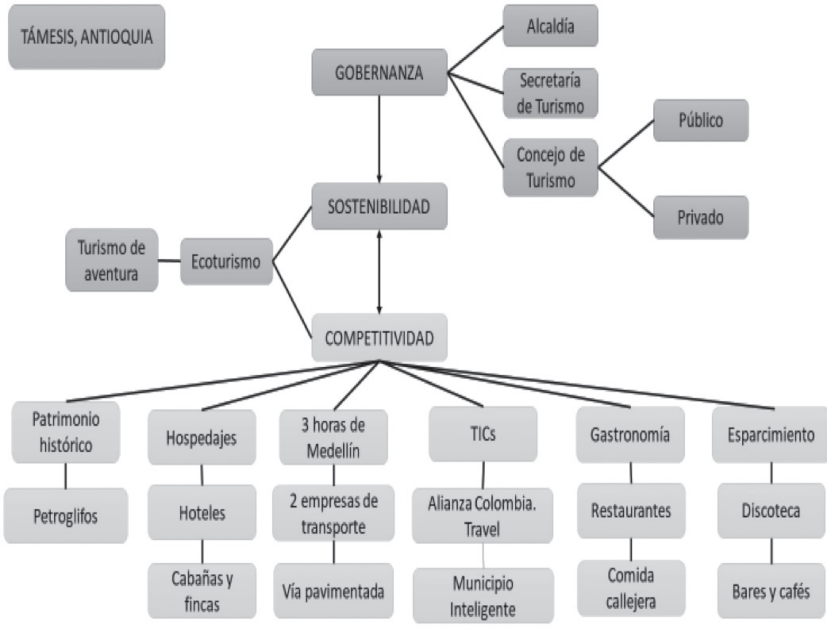


Fig. 2. Modelo de gobernanza, sostenibilidad y competitividad en Támesis, Antioquia

Como resultado, se obtuvo un inventario de los sitios turísticos que se resume en la Figura 3, que abarca no solo el casco urbano, sino también las veredas y asentamientos de interés. Se agregaron 22 sitios turísticos cercanos, correspondientes al Valle de Cartama, que generan un valor agregado para los visitantes al municipio de Támesis.

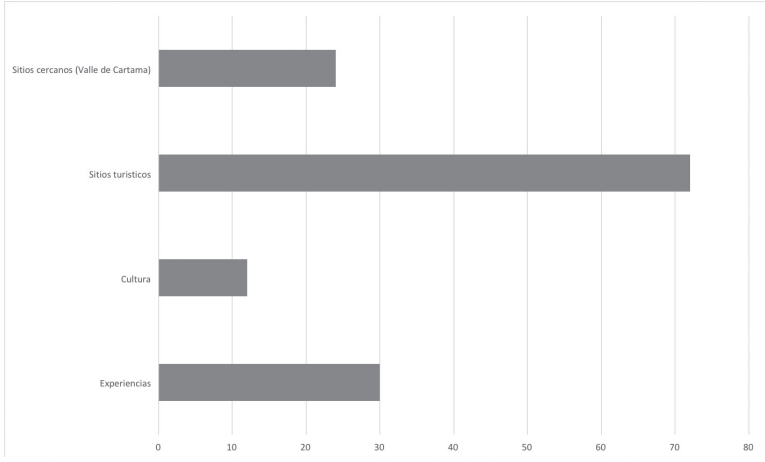


Fig. 3. Inventario turístico del municipio de Támesis

VI. RESULTADOS

La generación de bases de datos abiertas con la oferta turística del municipio, como hoteles, restaurantes, bares, comercio, cultura e información municipal, y la creación de la alianza turística entre diversos operadores y prestadores turísticos, permitieron el desarrollo de la APP Alianza Colombia Travel, cuyo propósito es organizar y centralizar la información obtenida, mostrando de forma más amplia y mediante el uso de las TIC toda la oferta turística, con el fin de mejorar la experiencia del visitante y ayudar a la generación de crecimiento económico sostenible para los residentes que dependen económicamente del turismo. En la Figura 4 se presentan algunas imágenes de la aplicación desarrollada.

Fig. 4. Aplicación móvil para el municipio de Támesis como destino inteligente

Según datos del Centro de Innovación Pública Digital del MinTIC, durante la prueba piloto de implementación en el municipio de Támesis, se registró un incremento de los turistas en un 70 % [13], lo que mejoró la economía local y visibilizó e internacionalizó a Támesis como destino turístico inteligente. Asimismo, el proyecto tuvo los siguientes impactos positivos:

Impacto en la economía local:

- Generación de más de 100 negocios virtuales mediante la comercialización de productos o servicios a través de aplicativo móvil desde el municipio de Támesis a la ciudad de Medellín.
- Desarrollo de nuevos productos y servicios turísticos en el municipio.
- Promoción, posicionamiento y ventas de los productos autóctonos y de origen tamesino.
- Bases de datos con el directorio comercial y turístico del municipio.

Impacto social:

- Creación de puestos de empleo.
- Incremento del atractivo regional.
- Reducción de la población local que migra en busca de trabajo a la ciudad.
- Agremiación de los sectores productivos, hoteles, restaurantes, bares, cafés especiales para trabajo conjunto y networking.
- Apropiación de la cultura y la arqueología locales
- Producción de contenidos audiovisuales educativos que promueven el arte y la cultura locales, con más de 100.000 visitas en plataformas y redes sociales.

Impacto ambiental:

- Creación de campañas educativas para la presentación y conservación de los recursos naturales.
- Desarrollo de circuitos turísticos comunitarios, con familias campesinas productoras de café, cacao, cítricos y panela.
- Documentación de la actividad económica y productiva de las 37 veredas del municipio de Támesis.
- Autoabastecimiento local.
- Equidad de género: el 59 % del total de personas beneficiadas con la implementación del proyecto en el municipio de Támesis fueron mujeres entre 20 y 50 años de edad, y el 41 % fueron hombres en el mismo rango de edad.

VII. CONCLUSIONES

El desarrollo y la implementación de un modelo de transformación digital para convertir al municipio de Támesis, Antioquia, en un destino turístico inteligente han tenido un gran impacto local, tanto que incluso a nivel nacional ya es considerado por el Centro de Innovación Pública Digital de Colombia como un referente por ser el primero de su tipo. Se trata de una solución que centraliza datos abiertos de la oferta turística, cultural y territorial del municipio por medio de una aplicación móvil que resuelve los retos que plantea ser un destino turístico inteligente, para aumentar el desarrollo sostenible, eco-eficiente, competitivo y accesible del municipio en un trabajo conjunto entre los sectores público y privado, que mejora notablemente la experiencia de los turistas y la calidad de vida de los habitantes del territorio.

VIII. REFERENCIAS

[1] D. Wang, X. Li, and Y. Li, "China´s 'smart tourism destination' initiative: a taste of the service-dominant logic", *J. Destination, Market. Manage.,* vol. 2, n.° 2, pp. 59-61, 2013.

[2] T. Bakici, E. Almirall, and J. Wareham, "A smart city initiative: The case of Barcelona", *J. Knowledge Econ.,* vol. 4, n.° 2, pp. 135-148, 2013.

[3] L. Rojas Pérez, "Destinos turísticos inteligentes. Una aproximación a Castellón como destino inteligente", Tesis de grado, Universitat Jaume, Catellón de la Plana, España, 2019.

[4] A. Lopez de Avila, "Smart destinations: XXI century tourism". *ENTER2015 Conference on Information and Communication Technologies in Tourism,* Lugano, Switzerland, February 4-6, 2015.

[5] D. Buhalis, "Working Definitions of Smartness and Smart Tourism

Destination", Buhalis Blog, Feb. 2015. [En línea]. Disponible en: http://t.co/xrLRpGipvu, 10

[6] C. Lamsfus, D. Martín, A. Alzua-Sorzabal, and E. Torres-Manzanera, "Smart Tourism Destinations: An Extended Conception of Smart Cities Focusing on Human Mobility". In *Information and Communication Technologies in Tourism,* I. Tussyadiah and A. Inversini (Eds.), Heidelberg, Germany: Springer, 2015, pp. 363-375.

[7] D. Buhalis and A. Amaranggana, "Smart tourism destinations". In *Information and Communication Technologies in Tourism,* Z. Xiang and I. Tussyadiah (Eds.), Heidelberg: Springer, 2014, pp. 553-564.

[8] U. Gretzel, H. Werthner, C. Koo, and C. Lamsfus, "Conceptual foundations for understanding smart tourism ecosystems". *Computers in Human Behavior,* vol. 50, Issue C, 2015. [En línea]. https://doi.org/10.1016/j.chb.2015.03.043

[9] P. Castro, "Proposal of a management model for the smart tourism destination (STD) of the city of Cuenca, Ecuador", PhD. Thesis, The University of Havana, Havana, Cuba, 2018.

[10] M. Fabronni, *Fundamentos del turismo. Nuevo enfoque en el Siglo XXI.* Buenos Aires: Editorial Ugerman, 2017,

[11] A. Osterwalder and Y. Pigneur, *Business model generation. A handbook for visionaries, games changers, and challengers*. Chichester, UK: John Wiley and Sons Limited, 2010.

[12] Corporación Encuentro de Dirigentes del Suroeste Antioqueño, "Administraciones Municipales", s. a. [En línea]. Disponible en: http://www.encuentrodelsuroeste.org/administraciones-municipales/

[13] Centro de Innovación Pública Digital, "Alianza Colombia. Travel", nov. 2018 [En línea]. Disponible en: https://centrodeinnovacion.mintic.gov.co/es/experiencias/alianza-colombiatravel

Capítulo 10
Gestión del Conocimiento Aplicada al Proceso de Autoevaluación

Dario Enrique Soto Duran - Dsoto@tdea.edu.co
Fabio Alberto Vargas Agudelo - Fvargas@tdea.edu.co
Tecnologico de Antioquia
Adriana Xiomara Reyes Gamboa - Axreyes@elpoli.edu.co
Politécnico Colombiano Jaime Isaza Cadavid

I. INTRODUCCIÓN

La Gestión de conocimiento define el proceso mediante el cual se desarrolla, estructura y mantiene la información, con el objetivo de transformarla en un activo crítico y ponerla a disposición de una comunidad de usuarios; la gestión incluye: el aprendizaje, la información, las aptitudes y la experiencia desarrollada durante la historia de la organización [1]. Desde sus inicios, las instituciones universitarias han sido las principales instancias encargadas de la generación, acumulación y transmisión del conocimiento, y han tenido como principal aporte, o al menos el más reconocido, la formación de profesionales competentes y comprometidos. Hoy la universidad sigue siendo una organización cuyo recurso más valioso es el conocimiento que atesora y gestiona diariamente. Las universidades son uno de los principales motores de la sociedad actual [2](Muñoz, 2016) donde se forman los futuros profesionales, mediante el desarrollo en ellos de capacidades de adaptación a nuevas situaciones en las que deben ser capaces de generar un conocimiento específico para cada momento y situación. Para lograr esto, la universidad debe realmente ser una organización intensiva en conocimiento [3] y en constante evolución, buscando el mejoramiento continuo de cada uno de sus ejes misionales, por ello debe tener un proceso de autoevaluación que le permita identificar esas fortalezas y debilidades.

La autoevaluación es un proceso a través del cual se recoge y se interpreta, formal y sistemáticamente, información pertinente sobre un programa educativo, se producen juicios de valor a partir de esa información y se toman decisiones conducentes a mantener, proyectar,

mejorar, revisar, ajustar, reformar o eliminar elementos del programa o de su totalidad [3], por esta razón, este proceso es clave para el mejoramiento continuo de los programas. Lo anterior evidencia la importancia de la definición de estrategias de gestión de conocimiento para la adquisición, conservación y transferencia del conocimiento relacionado con el proceso de autoevaluación de los programas en las universidades, lo que permite a las personas encargadas gestionar todo el conocimiento relevante que ayude a su desarrollo y mejora, manteniendo y preservando una efectiva autoevaluación de los programas.

Este capítulo presenta el diseño de estrategias de gestión del conocimiento aplicadas a los procesos de autoevaluación, con el fin de convertir el conocimiento tácito en explícito y que de esta forma se pueda exteriorizar, combinar, interiorizar y socializar buscando el mejoramiento del proceso.

El documento se encuentra organizado en tres secciones: la primera corresponde a los antecedentes de la Gestión de conocimiento en diversos contextos, la segunda sección presenta los resultados que corresponden al diseño y desarrollo de las estrategias y la última sección presenta las conclusiones del trabajo realizado.

II. ANTECEDENTES

Son diversos los trabajos que hacen evidente la importancia de la implementación de gestión de conocimiento para la mejora de las organizaciones. Dentro de los trabajos analizados en el contexto de gestión de conocimiento se encuentran:

En el estudio sobre la aplicación de la Gestión del conocimiento en contextos sociales y tecnológicos de países de Latinoamérica y Europa [4], se expone como pilar de la gestión del conocimiento a las "personas", ya que son ellas quienes por medio de interacciones, investigación y comunicación logran un conocimiento que luego se gestiona para darlo a conocer. Al referirse a la gestión del conocimiento se hace énfasis en tratar de detectar quién sabe, qué sabe, qué es utilizable de lo que sabe y determinar cómo ese conjunto de saberes puede ser incorporado de forma eficaz para estar disponible para todos de una manera sencilla, es decir, identificar: qué es lo que se quiere saber, quién es la persona que posee dicho conocimiento, qué de ese conocimiento sirve o es relevante para lo que se necesita y cómo se puede trasladar el conocimiento que tiene esa persona para lograr hacerlo entendible a cualquier persona que esté interesada.

En [5], los autores presentan un Modelo de evaluación de gestión del conocimiento para las pymes del sector

de tecnologías de la información. Este modelo está integrado con los procesos de la organización, por lo cual se enmarca desde una perspectiva de gestión de procesos. Este enfoque hacia la gestión de procesos consiste en la identificación y el análisis de los procesos desarrollados en la organización y específicamente sus interacciones, lo cual facilita estructurar las relaciones que se dan entre personas, grupos de trabajo y dependencias organizacionales.

El trabajo realizado por [6] presenta un modelo de autoaprendizaje con integración de las TIC y los métodos de gestión del conocimiento, en donde se permite al estudiante acceder a la información, tanto del profesor como de otros profesores y otros estudiantes. Las TIC pueden ser explotadas óptimamente en la medida en que se adapte el modelo de enseñanza/aprendizaje en contexto, así como los roles del mismo.

El estudio exploratorio acerca del impacto de la gestión del conocimiento en la calidad de las universidades que se presenta en [7] pretende determinar si existe una relación entre el proceso de gestión del conocimiento que se lleva a cabo dentro de las unidades académicas universitarias y la calidad de las mismas. Efectivamente, la aplicación del conocimiento es directamente proporcional al nivel de calidad de la institución: entre más se desarrollen y apliquen los conocimientos, más alto será el nivel de calidad de la institución. Por medio del estudio se concluye que la gestión del conocimiento debe mirarse desde una perspectiva integradora de sus dimensiones: crear, compartir y aplicar. El proceso de gestión del conocimiento al interior de los equipos dependerá justamente de la forma en que sus miembros en particular crean, compartan y apliquen el conocimiento para tomar decisiones.

En [8] se analiza la motivación como elemento clave para el desarrollo de la creatividad en estudiantes, usando ésta como estrategia principal en la gestión del conocimiento de las instituciones; la motivación y la creatividad van de la mano. Se pretende identificar y socializar los factores que constituyen el principal impulsor de la creatividad, tanto en docentes como alumnos en formación, por medio de la gestión del conocimiento. Allí proponen una interfaz desarrollada para web 2.0 que permite gestionar simultáneamente contenidos y usuarios para que tanto docentes como alumnos participen de manera colaborativa en la creación y difusión de contenidos relevantes a los factores que influyen en su motivación. Se analiza qué piensan los estudiantes acerca de las estrategias que usan sus docentes en el aula para fomentar la creatividad. ¿Los estudiantes se sienten motivados? Esta información es muy útil a la hora de brindar a los

docentes retroalimentación sobre sus fortalezas y debilidades, lo que les permite reflexionar en sus procesos clave, así como también socializar las experiencias de éxito con otros docentes interesados en la temática.

En [9], el autor analiza la relación entre políticas nacionales de innovación y la gestión del conocimiento en las organizaciones para darles valor a la economía y la sociedad. Adicionalmente, resalta la importancia de valorar el capital intelectual fomentando su crecimiento y aplicación para poder llegar a innovar. La gestión del conocimiento se aplica tanto en organizaciones privadas como públicas con el fin de crecer dentro de un entorno que promueva la innovación y la eficiencia tanto en el ámbito de las organizaciones individuales como de la nación.

En el artículo "Sistemas curriculares para la gestión de información y conocimiento institucional" [10] se analizan los sistemas de gestión de información y conocimiento institucional, en especial los que tienen la hoja de vida del profesor como única fuente de información, medición y representación del conocimiento dentro de una institución. Por supuesto es importante analizar y mantener actualizado el *curriculum vitae* del personal y de los investigadores, pero más importante aún es que los conocimientos que tengan sean clasificados, compartidos y aplicados a través de diferentes métodos en los sistemas curriculares. Por último, se apoya completamente la propuesta del uso de las ontologías como principal herramienta para la organización semántica de la información en un sistema curricular de gestión de información y conocimiento.

Vargas y Rodríguez [11] presentan una propuesta orientada al mejoramiento continuo de los procesos de gestión del conocimiento en la universidad para docentes, personal administrativo y estudiantes. Se plantean cuatro objetivos estratégicos: crear mecanismos para la generación y adquisición del conocimiento; favorecer el almacenamiento y protección del conocimiento; mejorar e incrementar la trasmisión del conocimiento, y fomentar el adecuado y eficaz uso del conocimiento. De igual forma, para cada objetivo se determinaron una serie de acciones que permitan dar cumplimiento a lo expuesto, destacando de modo general la importancia de recibir una mayor información al ingresar a la institución, al igual que la creación de alianzas con otras entidades para la generación de nuevo conocimiento, la divulgación periódica de la información sobre los avances en la institución y la disminución de la resistencia a compartir información entre áreas, dependencias y personal, con el fin de incrementar la información a favor de la toma de decisiones futuras.

En otro artículo [12] los autores sostienen que las relaciones, reconocimientos, satisfacciones y demás de un empleado repercuten en el trabajo que desarrolla, de manera que si está motivado llevará a cabo su trabajo de una manera más productiva que si no lo está, lo cual se ve reflejado en su rendimiento y bienestar. Existen modelos, como el que se presenta en este artículo, que generan de manera científica y sistemática diagnósticos motivacionales y recomendaciones de mejora de especial utilidad para administradores y departamentos de recursos humanos.

Siemens [13] precisa que, tradicionalmente, se ha tratado de desarrollar el conocimiento ya existente o añadir nuevas partes, pero las condiciones tecnológicas actuales hacen necesario cambiar los enfoques por completo. Más importante aún es crear conexiones: como no podemos experimentarlo todo, podemos apoyarnos en las experiencias de los demás. "Almacenar mi conocimiento en mis amigos" es recolectar conocimiento a través de la recolección de personas. Una red es una conexión entre entidades, si una se afecta repercute en todas las demás: "los nodos siempre compiten por las conexiones porque los enlaces representan la supervivencia en un mundo interconectado". La probabilidad de que el aprendizaje sea exitoso actualmente depende de lo bien que esté vinculado. Los vínculos débiles no son inútiles, permiten conexiones breves entre la información: las conexiones entre ideas y campos dispares pueden crear nuevas innovaciones. La tubería es más importante que el contenido dentro de ella.

Los autores Keishing y Renukadevi [14] investigan en su artículo de qué manera la gestión del conocimiento puede influir positivamente en la educación superior para que realmente sea útil en la vida laboral. Nos proponen, entonces, combinar las técnicas de exploración de carreras con las técnicas de gestión del conocimiento para que la fusión de ambas pueda brindar mejores resultados a los estudiantes en el momento de seleccionar la carrera más adecuada para su perfil, esto también involucra a las diferentes empresas que hacen parte del medio para construir un buen conocimiento y poder aplicarlo correctamente.

En [15] los autores aportan una comprensión básica de los procesos de gestión del conocimiento internos de las instituciones educativas y, adicionalmente, buscan llenar el vacío de conocimiento en los procesos que apoyan a los docentes en sus diversas tareas de compartirlo: reunir, crear, compartir y explicar el conocimiento.

En su artículo [16], Rampai busca explorar cuál sería el modelo

apropiado de gestión del conocimiento combinado con redes sociales, para mejorar la enseñanza y el aprendizaje autodirigido de los estudiantes, ayudándolos a aprender a usar las tecnologías como herramientas de instrucción con el fin de mejorar sus conocimientos a través de un enfoque sistemático.

Mussenova y Abdirova plantean en su artículo [17] que se debe contribuir a crear las condiciones necesarias para que las tecnologías educativas disponibles generen la posibilidad de modelar y llevar a cabo actividades de comunicación en un entorno profesional, en donde la comunicación entre profesionales sea el canal de aprendizaje más importante. La presencia de un componente implícito del conocimiento permite definir el proceso educativo como la interacción no lineal de los participantes. Los términos de la sociedad del conocimiento dictan requisitos específicos para la calidad de la educación, lo que incluye una lista de competencias clave del componente de información del graduado.

Los autores Torres y Lamenta [18] buscan en su artículo demostrar por medio del método hermenéutico (interpretar, aclarar, traducir) que la gestión del conocimiento es una fuente positivamente transformadora para las organizaciones que usan sistemas de información. En un entorno tan cambiante como el actual hay que adaptarse para sobrevivir, y estos mismos cambios aceleran la generación y adquisición de diferentes conocimientos que necesitan ser tratados con el valor que se merecen para alcanzar las ventajas competitivas que se buscan con la ayuda de los sistemas de información. Por lo anterior, constantemente emergen situaciones en las que hay que tomar decisiones sobre problemas, para lo cual, si se tiene bien estructurada la gestión del conocimiento y los sistemas de información, se puede llegar a la mejor decisión utilizando una cantidad mínima de recursos. La información se debe analizar, desarrollar y compartir de manera constante para generar nuevas ideas, mejorar los procesos existentes y poder llegar a innovar.

III. MATERIALES Y MÉTODOS

La gestión del conocimiento es clave en el contexto universitario, y más aún en los procesos de autoevaluación de los programas de las universidades, dado que es un proceso que involucra diferentes fuentes, información y conocimiento que son relevantes para el mejoramiento continuo de los programas académicos. Por esta razón, la metodología propuesta plantea tres fases: la primera se enfoca en seleccionar técnicas y herramientas adecuadas para la gestión del conocimiento, la segunda fase se enfoca en tomar un modelo

de gestión del conocimiento y definir las actividades que se deben aplicar para poder implementar el modelo, la tercera fase consiste en integrar a las etapas de autoevaluación, el modelo de gestión del conocimiento con sus respectivas actividades e incorporar las técnicas y herramientas que faciliten el desarrollo de cada actividad propuesta, y de esta forma realizar una adecuada gestión del conocimiento.

A. Técnicas y/o herramientas de gestión del conocimiento

Para realizar una adecuada gestión del conocimiento se hace necesario definir técnicas o herramientas que faciliten este proceso. Existe gran variedad de estas técnicas y herramientas, por tanto, teniendo en cuenta que el proceso que se desea abordar en este artículo es el de la autoevaluación de programas académicos, se hace necesario seleccionar las técnicas y/o herramientas de acuerdo con las etapas que tiene definida la autoevaluación, las cuales corresponden a:

E1: Planeación y organización

E2: Socialización y sensibilización

E3: Recolección y selección de información

E4: Análisis de resultados

E5: Elaboración del Plan de mejoramiento

E6: Socialización de Resultados

Entre algunas de las técnicas o herramientas más utilizadas y que se pueden aplicar en el proceso se encuentran las siguientes:

Mapas de conocimiento: los mapas de conocimiento facilitan la identificación y visualización de aquellas cosas "que sabemos" o que "necesitamos saber" para llevar adelante nuestra actividad, y por ende nos permiten llevar a cabo acciones para desarrollar y sacarle el mejor provecho a dichos recursos. Los mapas sirven para:

- Identificar: los conocimientos clave y sus cadenas de valor.
- Orientar: sobre acciones de gestión del conocimiento.

Foros de debate: los foros de debate tienen por objetivo promover el intercambio y la reflexión colectiva del conocimiento que permitan proponer, discutir y aportar nuevas ideas o información útil para generar nuevos conocimientos.

Ciclos de intercambio: se trata de una técnica que nos permite intercambiar o rotar las funciones de los miembros del proceso para que se involucren en áreas cuyo conocimiento es poco o nulo, y con el acompañamiento de personas expertas en el tema puedan generar nuevos conocimientos y facilitar el desarrollo del proceso.

Encuestas: son procedimientos que permiten recopilar datos con base en un cuestionario. Las encuestas nos permiten identificar internamente cómo se está realizando el proceso de gestión del conocimiento y qué tan importante es para el proceso de autoevaluación.

Histogramas: son las representaciones gráficas de diferentes tipos de valores, resultados y/o estadísticas. Sirven también como método para analizar de manera más fácil y rápida los valores obtenidos de algún otro proceso.

Rúbrica: las rúbricas son guías precisas que valoran los aprendizajes y productos realizados [19]. Son tablas que permiten desglosar los puntos a valorar o calificar dentro del aprendizaje.

Repositorio de información: es una herramienta web que permite gestionar la información obtenida de un proceso.

En la Tabla 1 se presenta un cuadro relacional de las diferentes técnicas y herramientas con las etapas del proceso de autoevaluación.

Tabla 1. Etapas autoevaluación vs técnicas y/o herramientas

Etapas Proceso ---------- Técnicas	E1	E2	E3	E4	E5	E6
Mapas de conocimiento	X				X	
Foros de debate	X			X	X	
Ciclos de intercambio	X		X	X	X	
Encuestas		X				X
Histogramas				X		
Rúbrica Genérica	X		X			
Repositorio de información	X	X	X	X	X	X

Se concluye que las técnicas y/o herramientas antes mencionadas son las que mejor se adaptan a la estrategia y a las diferentes etapas del proceso de autoevaluación.

B. Modelo de Gestión del Conocimiento

Existen diversos modelos de gestión del conocimiento. Como se evidenció en los antecedentes, uno de los más utilizados es el definido por Nonaka y Takeuchi [20], en el que se plantean cuatro modos de conversión del conocimiento, lo cual se da a través de una conversión entre el conocimiento tácito y el explícito que permite postular diferentes modos de conversión del conocimiento: 1) del conocimiento tácito al conocimiento tácito; 2) del conocimiento explícito al conocimiento explícito; 3) del conocimiento tácito al conocimiento explícito, y 4) del conocimiento explícito al conocimiento tácito.

En el primer caso, existe un modo de conversión del conocimiento que nos permite transformar el conocimiento tácito a través de la interacción entre los individuos, este proceso de creación de conocimiento tácito a través de compartir experiencia se conoce como *socialización.* El segundo modo de conocimiento, *la conversión,* implica la utilización de procesos sociales para combinar distintos análisis de conocimiento explícito celebrado por los individuos, lo que se busca en este proceso es crear conocimiento explícito del conocimiento y este se transforma en *combinación*. El tercero y cuarto modos de conversión de conocimiento se refieren a patrones de conversión que involucran tanto el conocimiento tácito como el explícito. Estos modos de conversión buscan captar la idea de que el conocimiento tácito y el explícito se complementan y pueden expandirse a lo largo del tiempo a través de un proceso de interacción mutua, la cual involucra dos operaciones diferentes. Uno de ellos es la conversión del conocimiento tácito en conocimiento explícito, que se llamará *la externalización,* y la otra es la conversión de conocimiento explícito en conocimiento tácito, que tiene cierta similitud con el concepto tradicional de aprendizaje y que se conoce como *internalización.*

Lo anterior lleva a definir una serie de actividades que faciliten esa conversión de conocimiento. Tales actividades son: filtrar el conocimiento tácito, seleccionar conocimiento tácito, generar conocimiento explícito, presentar conocimiento, combinar conocimientos, capturar conocimiento explícito, procesar conocimiento explícito, generar conocimiento tácito, seleccionar conocimiento tácito y explícito, y presentar conocimiento tácito y explícito.

IV. RESULTADOS

La estrategia propuesta está formada por cuatro procesos que se ejecutan cíclicamente para aprovechar todos los instantes del conocimiento. Coincidiendo con [20], se determinó que estos cuatro procesos son:

Exteriorización, Combinación, Interiorización y Socialización del conocimiento. Transversales a estos procesos, se encuentran las técnicas y herramientas que facilitan la ejecución de cada uno de los procesos de la estrategia.

A. Procesos

A continuación, se definen los procesos adaptados de [20] que se aplican en la presente estrategia para el proceso de autoevaluación.

Proceso de Exteriorización (tácito a explícito): se encarga de convertir el conocimiento tácito en conocimiento explícito, haciéndolo comprensible para otras personas interesadas [20]. Es apropiado para los procesos de autoevaluación, pues permite que el conocimiento tácito sea mucho más comprensible utilizando principalmente el idioma como fuente para expresarlo.

Proceso de Combinación (explícito a explícito): busca, en primera instancia, sintetizar y formalizar el conocimiento explícito de manera que cualquier persona pueda acceder a él y, en segunda instancia, busca generar más de este conocimiento a partir de la interacción entre los miembros del proceso [20].

Proceso de Interiorización (explícito a tácito): técnicamente, es entendido como la conversión del conocimiento explícito en conocimiento tácito. En este proceso cada persona capta, analiza e integra el volumen de información adquirida del proceso anterior y lo convierte en nuevos conocimientos y experiencias [20] fundamentales para el proceso de autoevaluación.

Proceso de Socialización (tácito a tácito): este proceso busca el intercambio de experiencias y la creación de nuevo conocimiento tácito.

Este proceso suele confundirse con el de combinación, sin embargo, su diferencia radica en que éste se enfoca a compartir el conocimiento tácito y generar otro nuevo a través de la experiencia, mientras que la combinación si bien busca fusionar y generar nuevo conocimiento, también busca formalizarlo para que sea accesible [20].

B. Técnicas y Herramientas.

A continuación, se presentan las diferentes técnicas y herramientas útiles en la aplicación de la presente estrategia de gestión del conocimiento.

Mapas de conocimiento: analizando las diferentes etapas del proceso de autoevaluación, se utilizaron los mapas del conocimiento en las etapas de Planeación y organización y Elaboración [del] plan de mejoramiento, ya que proporcionan una representación de los puntos

claves de las etapas y de las fuentes de conocimiento que se deben mantener para darles apoyo y facilitar su desarrollo.

Foros de debate: Para la estrategia se utilizó esta técnica en las etapas de Planeación y organización, Análisis de resultados y Elaboración [del] plan de mejoramiento, ya que son un pilar fundamental en el que el debate de ideas y la consolidación de nuevas ideas son muy importantes para tomar decisiones.

Ciclos de intercambio: esta técnica es muy transversal y podría aplicarse en cualquier instante. Se decidió aplicarla en las etapas de Planeación y organización, Recolección de información y Elaboración [del] plan de mejoramiento.

Encuestas: esta herramienta se utilizó principalmente en las etapas de Socialización y sensibilización y Socialización de resultados.

Histogramas: para la estrategia planteada se decidió utilizar esta herramienta en la etapa de Análisis de resultados, que es donde más influencia logra tener.

Rúbrica: esta herramienta se usó dentro de la estrategia como un instrumento para tener unas guías y puntos claros para evaluar en el proceso y se aplicó específicamente a las etapas de Planeación y organización y Recolección y selección de información.

Repositorio de información: el repositorio consiste en un portal web que contiene diferentes ítems que permiten realizar la gestión de las etapas del proceso.

C. Actividades.

Se identificaron varias actividades que sirven como pilar fundamental para facilitar la implementación de cada uno de los procesos planteados en la estrategia y se asociaron las diferentes técnicas y herramientas seleccionadas para su incorporación en la estrategia de gestión del conocimiento. En la Tabla 2 se presentan las relaciones de las diferentes actividades con los procesos en los cuales son aplicables.

Tabla 2. Actividades vs. procesos de estrategia de gestión del conocimiento

Actividades	Procesos
Filtrar conocimiento tácito	Exteriorización Combinación
Seleccionar conocimiento tácito	Exteriorización Combinación
Generar conocimiento explícito	Exteriorización Combinación
Presentar conocimiento	Combinación
Combinar conocimientos	Combinación
Capturar conocimiento explícito	Interiorización
Procesar conocimiento explícito	Interiorización
Generar conocimiento tácito	Interiorización
Seleccionar conocimiento tácito y explícito	Socialización
Presentar conocimiento tácito y explícito	Socialización

En la Tabla 3 se presentan las estrategias que se fundamentan en la relación de cada uno de los procesos de la estrategia planteada con sus respectivas actividades y herramientas, y asociadas a cada etapa del proceso de autoevaluación.

Tabla 3. Relación procesos vs. actividades y herramientas

Proceso de estrategia	Actividades	Técnicas y/o herramientas	Etapas autoevaluación
Exteriorización	-Capturar CT. -Seleccionar CT. -Generar CE.	-Mapas de conocimiento -Rubrica -Portal	-Planeación -Recolección información
Combinación	-Filtrar CT. -Seleccionar CT. -Generar CE.	-Rubrica -Foros de debate -Histogramas -Portal	-Planeación -Análisis de resultados -Recolección información
Interiorización	-Capturar CE. -Procesar CE. -Presentar CT.	-Ciclos de intercambio -Foros de debate -Portal	-Análisis de resultados -Elaboración plan de mejoramiento
Socialización	-Seleccionar CT. -Seleccionar CE. -Presentar CT. -Presentar CE.	-Encuestas -Portal	-Socialización y sensibilización -Socialización de resultados

Lo mencionado hasta aquí da a conocer la estrategia diseñada de gestión del conocimiento para el proceso de autoevaluación.

V. CONCLUSIONES

Con el pasar del tiempo, se ha evidenciado la importancia y el crecimiento exponencial de la información. Este crecimiento hace necesario también un hacer un control, llevar un registro y darle un orden para poder acceder a ella de manera fácil y rápida cuando se necesite posteriormente. De ahí nace la gestión del conocimiento, de la necesidad de llevar un control sobre la información existente para garantizar el futuro acceso a ella con el fin de generar nuevos conocimientos de manera más rápida y eficiente. De igual forma, se hacen importantes las conexiones que permiten llevar a cabo una comunicación entre los diferentes tipos y fuentes de información para facilitar el acceso a ella.

Las Instituciones de Educación Superior vienen realizando procesos de autoevaluación de sus programas, que involucran la gestión de mucha información y conocimiento que se encuentra muy dispersos, lo cual

genera un problema evidente al momento de realizar este proceso, ya que se presenta pérdida de información, reproceso y falta de gestión de conocimiento sobre los diferentes aspectos misionales de la institución, los cuales son claves para el mejoramiento de los programas.

Este trabajo busca dar solución al problema que se hace evidente cada vez que se inicia un proceso de autoevaluación, por tal motivo, se presenta el diseño de una estrategia de gestión del conocimiento aplicada al proceso de autoevaluación con el fin de convertir el conocimiento tácito en explícito, y que de esta forma se pueda exteriorizar, combinar, interiorizar y socializar con el fin de mejorar el proceso.

En el artículo se presenta una identificación de las técnicas y/o herramientas más apropiadas para llevar a cabo una adecuada gestión del conocimiento, las cuales se utilizaron para diseñar la estrategia de gestión del conocimiento, haciendo uso de los conceptos analizados para aplicarlos al proceso de autoevaluación. Las herramientas aplicadas corresponden a Mapas de conocimiento, Foros de debate, CiclBos de intercambio, Encuestas, Histogramas, Rúbrica y un Repositorio de información.

El balance final de la investigación ha conducido a la aplicación del conocimiento tratado, donde la estrategia propuesta permitirá facilitar el proceso de autoevaluación de los programas de educación superior.

VI. REFERENCIAS

[1] Y. Pérez Rodríguez y A. Coutín Domínguez, “La gestión del conocimiento: un nuevo enfoque en la gestión empresarial”, *Acimed,* vol. 13, n.° 6, 2005.

[2] S. P. Muñoz, “Funciones de la Universidad en el siglo XXI: humanística, básica e integral”, *Revista electrónica interuniversitaria de formación del profesorado,* vol. 19, n.° 1, pp. 191-199, 2016.

[3] O. L. López Abreu, J. J. García Muñoz, I. Batte Monter, y M. E. Cobas Vilches, “La mejora continua: objetivo determinante para alcanzar la excelencia en instituciones de educación superior”, *Edumecentro,* vol. 7, n.° 14, pp. 96-215, 2015.

[4] C. B. Valbuena, *Evidencias de la gestión de conocimiento en contextos sociales y tecnológicos de países de Latinoamérica y Europa.* OmniaScience, 2015.

[5] C. E., Echeverry, J. A. García, y H. M. Gómez, “Modelo de evaluación de gestión del conocimiento para las pymes del sector de tecnologías de la Información”, *AD-minister*, 26,

2015.

[6] R. R. Aedo, E. C. Ramos, y M. E. Romero, "Un modelo de autoaprendizaje con integración de las TIC y los métodos de gestión del conocimiento", *RIED. Revista Iberoamericana de Educación a Distancia,* pp. 137-149, 2015.

[7] E. Rodríguez Ponce, "Estudio exploratorio del impacto de la gestión del conocimiento en la calidad de las universidades", *Interciencia,* vol. 41, n.° 4, pp. 228-234, 2016.

[8] R. I. Carrasco, "Gestión del conocimiento en la educación en management: Entorno 2.0 Como espacio para promover la motivación y desarrollo de la creatividad/Knowledge Management Education: 2.0 Environment to promote motivation and creativity development", *Palermo Business Review,* n.° 13, 33, 2016.

[9] J. A. Cadavid, "Gestión del conocimiento y políticas de innovación", *Revista Propiedad Inmaterial,* n.° 22, 19, 2016.

[10] M. D. Pérez, "Sistemas curriculares para la gestión de información y conocimiento institucional. Estudio de caso", *Revista General de Información y Documentación,* vol. 26, n.° 1, pp. 11-24, 2016.

[11] S. M. Vargas y L. R. Rodríguez, "Análisis de la gestión del conocimiento en una institución de educación superior", *Criterio Libre,* vol. 13, n.° 22, pp. 279-297, 2015.

[12] D. López Fernández, P. P. Alarcón, E. Tovar, y S. Alonso, "El Modelo de Gestión del Conocimiento Motivacional: propuesta de aplicación en el sector bibliotecario", *Revista Española de Documentación Científica,* vol. 39, n.° 4, 157, 2016.

[13] G. Siemens, "Connectivism: a learning theory for the digital age", *e-Learning Library*, 1-8, 2014. Disponible en: http://www.itdl.org/Journal/Jan_05/article01.htm

[14] V. Keishing y S. Renukadevi, "A review of knowledge management based career exploration system in engineering education", *International Journal of Modern Education and Computer Science,* vol. 8, n.° 8, 2016.

[15] P. Syysnummi y H. Laihonen, "Top management's perception of knowledge management in a vocational education and

training organization in Finland", *International Journal of Educational Management,* vol. 28, n.° 1, pp. 53-65, 2014.

[16] N. Rampai, "Model of Knowledge Management via Social Media to Enhance Graduated Student's Self-Directed Learning Skill", *International Journal of Information and Education Technology,* vol. 5, n.° 10, 799, 2015.

[17] E. K. Mussenova and N. T. Abdirova, "Knowledge Management in the Educational Process of the University in Modern Conditions", *Education & Science Without Borders,* vol. 7, n.° 13, 2016.

[18] K. Torres y P. Lamenta, "La gestión del conocimiento y los sistemas de información en las organizaciones", *Negotium,* vol. 11, n.° 32, pp. 3-20, 2015.

[19] F. Gatica-Lara y T. Uribarren-Berrueta, "Cómo elaborar una rúbrica", *Investigación en educación médica,* vol. 2, n.° 1, pp. 61-65, 2012.

[19] I. Nonaka and H. Takeuchi, *The knowledge-creating company: How Japanese companies create the dynamics of innovation.* Oxford University Press, 1995.

Capítulo 11
Monitoreo de Indicadores de Rendimiento en Cadenas *Blockchain Ethereum*

Juan Camilo Giraldo Mejia - Jgiraldo1@tdea.edu.co
Julian Grisales Z. - Juliangris@gmail.com
Mauricio Amariles Camacho - Mauricio.amariles@tdea.edu.co
Tecnologico de Antioquia

I. INTRODUCCIÓN

Blockchain es una tecnología que cambia el concepto de la confianza entre las personas, ya que resuelve el problema de validar la autenticidad de los datos sin requerir un intermediario de confianza, lo que permite que cualquier miembro de la red verifique la validez de los datos de forma independiente, por medio de algoritmos criptográficos que autorizan y registran una serie de transacciones que se ejecutan y registran sobre todo el sistema [1]. Esta tecnología permite a gobiernos, tecnologías financieras (*Fintech*), empresas emergentes e individuos agregar a sus procesos: registro, trazabilidad, seguridad y almacenamiento distribuido de sus datos [2].

Una de las tecnologías más destacadas de *Blockchain* es Ethereum, ya que desde noviembre de 2018 sus aplicaciones contaban con una cuota de mercado estimada en US$21.447.799.407, ubicándose en segundo lugar por debajo de Bitcoin [3].

Lo que hace tan atractiva esta tecnología es que su base es de código abierto, en donde una gran cantidad de personas o empresas podrían beneficiarse aumentando la seguridad, mejorando sus procesos y reduciendo costos de manera considerable. Sin embargo, *Ethereum* todavía tiene una serie de desafíos técnicos referentes a la privacidad, escalabilidad y rendimiento [4].

Habilitar *Ethereum* puede llevar tiempo por la falta de conocimiento en todos los elementos que hacen parte del sistema, y por la ausencia de herramientas y modelos confiables y de fácil acceso que permitan identificar anomalías que afecten el rendimiento.

En investigaciones como la que se presenta en [5], se afirma que los modelos actuales de monitoreo generan sobrecarga en el sistema *Blockchain,* ofrecen pocos parámetros en el monitoreo y carecen de capacidad de escalabilidad.

En [6] se ilustra cuáles son los componentes y variables que generan los mayores cuellos de botella en el sistema, como son la velocidad de procesamiento, la memoria de uso, la velocidad de respuesta entre nodos y el consumo de energía. En [7], por su parte, se demuestra que el rendimiento también puede verse afectado por ataques informáticos que puede sufrir el sistema como está construido actualmente.

El presente trabajo busca identificar y caracterizar las métricas del sistema *Blockchain Ethereum,* y con base en esta información se propone un modelo de monitoreo en relación con el rendimiento en un caso aplicado.

En la actualidad, la adopción de sistemas que se ejecutan sobre redes privadas *Blockchain Ethereum* viene en aumento, debido a que es una plataforma de software abierta con características de descentralización, inmutabilidad y trazabilidad, que permite la creación de aplicaciones en diferentes ámbitos, sin embargo, al igual que otras arquitecturas, este sistema está soportado por múltiples componentes de hardware como: CPU, memoria RAM, memoria ROM y la tarjeta de red, los cuales, al presentar saturación o fallas, provocan cuellos de botella en el sistema *Blockchain*. Estas saturaciones pueden ser provocadas por la variación de parámetros como: tamaño de bloque, política de respaldo, latencia en la red, asignación de recursos, etc. Estos eventos ocasionan ineficiencia y conflictos en el sistema [8].

Se ha evidenciado que a medida que el sistema crece, adicionando nuevos nodos a la red, el algoritmo de consenso genera también cuellos de botella, ya que todos los nodos de la red deben dar el consenso para aprobar las transacciones que se ejecutan [9].

II. MATERIALES Y MÉTODOS

Blockchain fue presentado por primera vez en el año 2008 por el autor con pseudónimo Satoshi Nakamoto, en el tratado llamado *Bitcoin* [10], en el cual propone un sistema de dinero electrónico definido como criptomoneda, conformado por múltiples nodos, soportados en una red punto a punto, donde cada uno de los nodos se encuentra ubicado en diferentes sitios [11].

En su propuesta, Nakamoto describe los inconvenientes de un sistema de pago centralizado donde se depende de una única entidad de confianza, la cual se encarga de procesar y almacenar las transacciones en el sistema de pagos,

y por la condición de estar soportado sobre un esquema centralizado, puede presentar problemas como costo de transacción, eficiencia y seguridad [10]. Lo anterior lleva a que la tecnología *Blockchain* pueda ser utilizada como base de datos, contratos inteligentes y transacciones condicionales [12].

En enero de 2018, se realizó una medición en la velocidad de procesamiento de transacciones por sistemas de pago tradicionales, en comparación con los sistemas de pagos no tradicionales soportados sobre redes *Blockchain* como *bitcoin* y *Ethereum.* Se concluyó que los sistemas de pagos tradicionales como Visa son más eficientes que los sistemas de pago no tradicionales, en lo referente a la velocidad en la que se procesan las transacciones.

Al evidenciar esta problemática, diferentes autores han profundizado en el análisis de las métricas que inciden directamente en el rendimiento del procesamiento de transacciones en las redes *Blockchain*, y han identificado cuellos de botella en diferentes elementos.

La red, conformada por múltiples nodos, generalmente localizados en diferentes ubicaciones geográficas, tiene una métrica llamada latencia, la cual incide directamente en el rendimiento de las transacciones, razón por la cual se han propuesto simuladores que modelen los componentes de bloque, nodo y características de la red para predecir el comportamiento relacionado con el rendimiento [13].

Adicionalmente, algunas investigaciones han determinado que la base de datos debe ser modelada a nivel local, nivel remoto y nivel distribuido, ya que en cada punto tiene mayor o menor incidencia en el rendimiento del sistema en métricas, como latencia de escritura y latencia de lectura [14].

El algoritmo de consenso en redes *Blockchain* cumple un papel fundamental en el correcto funcionamiento del protocolo, porque es el que permite aprobar transacciones con el consenso de los nodos. Precisamente, el tiempo que tardan los nodos en dar el consenso para aprobar las transacciones incide en el rendimiento del sistema, donde los componentes de CPU, memoria y red desempeñan un papel fundamental en la respuesta efectiva del consenso de cada nodo [15].

La memoria RAM en *Blockchain* es la encargada de almacenar las transacciones que posteriormente serán procesadas en el sistema. Este componente funciona con esquemas de encolamiento, por lo tanto, su desempeño incide considerablemente en el rendimiento del nodo.

Las métricas que inciden en el rendimiento son las transacciones por bloque, el tiempo de minado en cada bloque, transacciones por segundo, *pool* de memoria, tiempo de espera en memoria, número de transacciones sin confirmar, número total de transacciones y número de bloques generados [16].

En los sistemas *Blockchain*, la CPU procesa las transacciones en los nodos, y es la encargada del proceso de minado en redes *Ethereum,* ya que realiza una de las tareas más críticas en el rendimiento de transacciones del sistema. Algunos autores proponen evaluar su desempeño de procesamiento por medio de la ejecución de bloques consecutivos de transacciones sintéticas, que permitan ajustar los valores para alcanzar el máximo desempeño [12].

En algunas aplicaciones, como los contratos inteligentes en *Blockchain Ethereum,* es necesario firmar digitalmente las transacciones para garantizar la seguridad entre el remitente y el destinatario de la transacción. Esta propiedad afecta el rendimiento de las transacciones del sistema, ya que incide en la métrica de tiempo de propagación de bloque [17].

Al identificar los elementos que generan los principales cuellos de botella en el sistema en cuanto al rendimiento de transacciones, nace la necesidad de monitorear diferentes métricas y establecer un modelo que permita medir y analizar detalladamente las diferentes etapas del proceso de generación de la cadena de bloques.

Para el monitoreo de los indicadores, se utilizaron dos equipos con sistema operativo Linux, se instala cliente *Geth Ethereum* (Aplicación oficial *Ethereum*), *Json* con los datos del bloque, y el administrador de redes privadas de *Ethereum Puppeth.*

III. RESULTADOS

Se presentan los siguientes elementos para el monitoreo de indicadores de rendimiento *Blockchain* de *Ethereum*:

Colector de *logs*: Corresponde a los registros requeridos en los diferentes componentes que se están generando correctamente, este elemento tiene la función de recolectarlos: colecta los logs en Red, los logs de Blockhain de Ethereum, y los registros del rendimiento del nodo.

Colector de Registros en Capa de Red y nodo: La forma más simple de recolectar los registros en la capa de red y los componentes del nodo es por medio de subprogramas, APIS o sistemas de registros (*syslogs*), que consuman directamente los datos de los archivos de *logs* del sistema operativo o por medio de la invocación de comandos nativos que permitan extraer de forma segura los datos

almacenados. Se propone extraer las métricas en los componentes de Red, CPU, Memoria RAM, Disco duro.

Colector de Registros en Capa *Blockchain Etherum:* Para recolectar los registros del nodo *Blockchain Ethereum*, también existen APIS que nos facilitan el trabajo de recolección de estos registros, los cuales se conectan directamente sobre el nodo y ejecutan operaciones para extraer las métricas.

Almacenamiento de *logs* (normalizados y convertidos): Este componente lo que busca es almacenar únicamente los registros relacionados con métricas de rendimiento de una forma estructurada y organizada, con el fin de que pueda ser analizada y cargada en componentes externos al sistema en donde, por ejemplo, se le haga una minería de datos a esta información.

Presentación WEB: Una vez se obtienen los datos de rendimiento de métricas *Blockchain* de *Ethereum,* se requiere un componente web con gráficas enriquecidas que permita visualizar la variación en el tiempo de los diferentes indicadores, con el fin de entender el comportamiento del sistema y poder tomar decisiones cuando se presenten demoras o fallas en el mismo.

Analítica de datos para anomalías: Contar con un repositorio de *logs* referentes a métricas de rendimiento para analizarlas en modelos de minería de datos, como las redes neuronales, y de árboles de decisión. Por lo tanto, se debe construir una secuencia de extracción, transformación y carga, conocida como ETL, en donde el destino sea una base de datos independiente del nodo *Ethereum*, para ejecutar modelos de entrenamiento con predicción y experimentos determinando valores de referencia en rendimiento, y que con base en la comparación de dos sets de datos el modelo tenga la capacidad de determinar anomalías y causas y muestre en tiempo real el desfase de rendimiento del sistema.

IV. PRUEBAS – FUNCIONALIDAD

A. Generador y colector de métricas de rendimiento

Paso 1: Recolectar métricas de red, del procesamiento de la CPU, memoria RAM y ROM.

El ambiente de pruebas corresponde a dos Nodos *blockchain* sobre máquinas virtuales Linux Ubuntu. En cada nodo está instalado el cliente *blockchain* (*geth*). Los nodos se unen en una red *blockchain* por medio de un archivo Génesis (*static-nodes.json)*.

Para recolectar las métricas de los componentes de red, CPU, memoria y disco se hace uso del agente *open source telegraf,* el cual se instala como

un plugin. Este tiene la capacidad de recolectar métricas y eventos de aplicaciones en diferentes sistemas operativos.

Este colector de registros se instala en los dos nodos *Blockchain Ethereum* y se habilita el servicio: Una vez se tiene el agente instalado y se está ejecutando como servicio, se indican en el archivo de configuración las métricas requeridas para recolectar.

```
1794 ###############################################################################
1795 #                            INPUT PLUGINS                                    #
1796 ###############################################################################
1797
1798
1799 # Read metrics about cpu usage
1800 [[inputs.cpu]]
1801   ## Whether to report per-cpu stats or not
1802   percpu = true
1803   ## Whether to report total system cpu stats or not
1804   totalcpu = true
1805   ## If true, collect raw CPU time metrics.
1806   collect_cpu_time = false
1807   ## If true, compute and report the sum of all non-idle CPU states.
1808   report_active = false
```

Fig. 1. *Configuración de las métricas en CPU*

```
1821 # Read metrics about disk IO by device
1822 [[inputs.diskio]]
1823   ## By default, telegraf will gather stats for all devices including
1824   ## disk partitions.
1825   ## Setting devices will restrict the stats to the specified devices.
1826   devices = ["sda", "sdb", "vd*"]
1827   ## Uncomment the following line if you need disk serial numbers.
1828   # skip_serial_number = false
```

Fig. 2. *Configuración de las métricas en disco*

```
1855 # Read metrics about memory usage
1856 [[inputs.mem]]
1857   # no configuration
1858
1859
1860 # Get the number of processes and group them by status
1861 [[inputs.processes]]
1862   # no configuration
1863
1864
1865 # Read metrics about swap memory usage
1866 [[inputs.swap]]
1867   # no configuration
```

Fig. 3. *Recolección de las métricas en memoria*

```
3762 # # Read metrics about network interface usage
3763  [[inputs.net]]
3764 #   ## By default, telegraf gathers stats from any up interface (excluding loopback)
3765 #   ## Setting interfaces will tell it to gather these explicit interfaces,
3766 #   ## regardless of status.
3767 #   ##
3768     interfaces = ["ens33", "lo"]
3769 #   ##
3770 #   ## On linux systems telegraf also collects protocol stats.
3771 #   ## Setting ignore_protocol_stats to true will skip reporting of protocol metrics.
3772 #   ##
3773 #   # ignore_protocol_stats = false
3774 #   ##
```

Fig. 4. *Recolección de métricas en Red*

Paso 2: Se recolectan las métricas *Ethereum* vía API.

La recolección de las métricas del nodo *Ethereum* se obtiene por tres métodos: API, *Syslog,* Aplicación *Ethereum.* La red está conformada por dos nodos debido a las limitaciones de recursos en la máquina física.

Vía API: Ejecutar servicio que extraiga métricas de monitoreo. Aquí hacemos uso de un API llamada *eth-net-intelligence-api,* donde debemos indicar los datos del nodo en el archivo *app.json* para recolectar los registros generados por el nodo *Ethereum.* Los parámetros que se deben diligenciar son:

– RPC_HOST (Dirección IP del nodo donde se van a extraer los datos).

– RPC_PORT (Puerto de conexión RPC para ejecutar comandos).

– LISTENING_PORT (Puerto del nodo donde se van a extraer los datos).

– INSTANCE_NAME (Identificador de la instancia cuando inicia el nodo).

– WS_SECRET (Contraseña que se define para conectarse a un *front end* para graficar información).

– WS_SERVER (Servidor donde se van a enviar los datos recolectados).

Al ejecutar el API podemos visualizar las métricas que extrae el nodo *Blockchain Ethereum:*

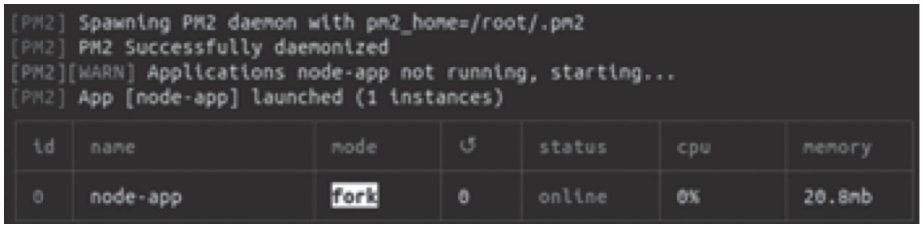


Fig. 5. *Métricas Blockchain Ethereum*

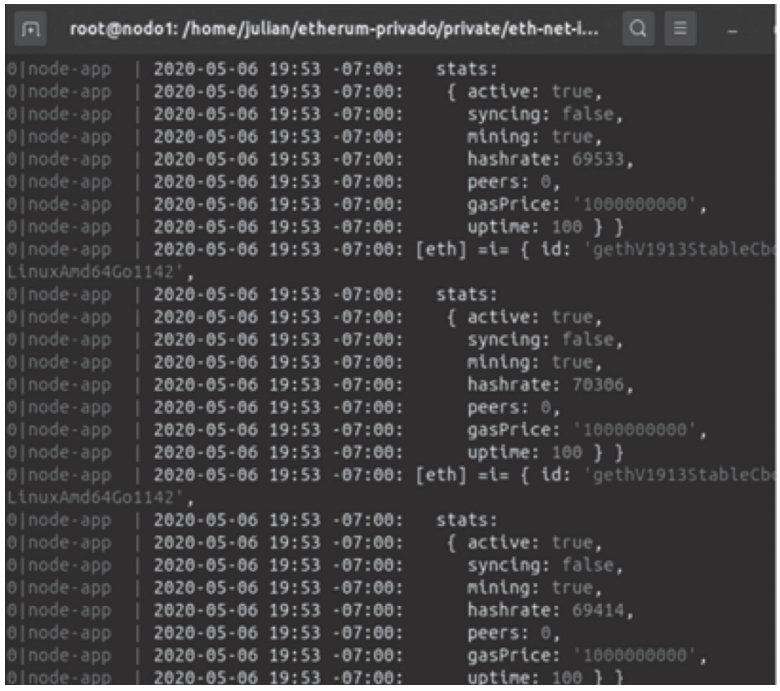


Fig. 6. *Extracción de métricas Ethereum vía API*

En la Figura 6 se presentan las métricas del gas consumido, el tiempo de validación de la transacción y el id del bloque.

Almacenamiento y visualización de métricas de rendimiento Blockchain Ethereum:

En el paso anterior, el API *eth-net-intelligence-api* extraía métricas de *Blockchain ethereum*. Existe un componente de capa de presentación (*front-end*) llamado *eth-netsats,* que consume esos registros y los grafica en una interfaz web: Número de bloque, Número de nodos (en este punto solo es un nodo), tiempo de último bloque, Promedio de tiempo de bloque, Promedio de latencia de la red, Tiempo de bloque, Dificultad, Precio de GAS, Límite de GAS, Propagación del bloque, Últimos bloques minados, Velocidad de minado, Transacciones pendientes, Bloque actual.

Almacenar métricas Blockchain Ethereum en bases de datos:

Para almacenar las métricas de Blockchain Ethereum en base de datos existen dos métodos:

- Vía *Syslog*
- Vía Aplicación *Ethereum*

Vía *Syslog:* El servicio *syslog* permite visualizar los registros del nodo en tiempo real, se puede

visualizar con el comando *tail-f/var/log/syslog|rep eth_blocknumber*: Este registro puede almacenarse en una base de datos *open source* llamada *influxdb*, por medio del servicio *telegraf:*

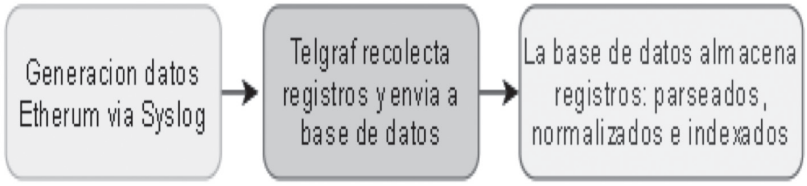
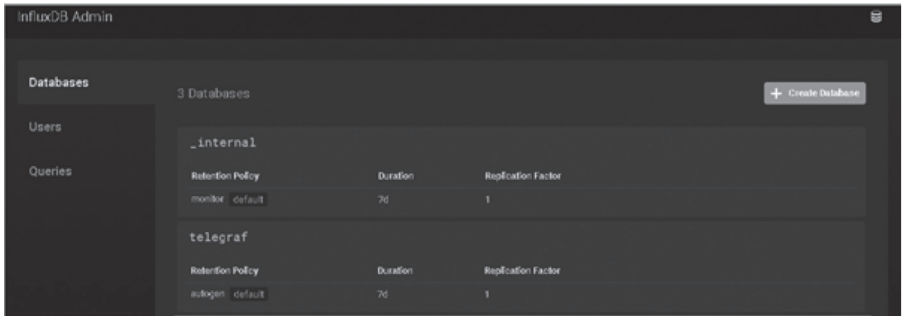


Fig. 7. *Flujo de datos Syslog hacia Base de Datos*

Para habilitar el almacenamiento en la base de datos, es necesario: Instalar un software llamado *influxdb,* crear una base de datos llamada *telegraf* con una tabla llamada *syslog,* crear un usuario para la aplicación con permisos de escritura y habilitar envío de registros desde *telegraf:*




```
104 # Configuration for sending metrics to InfluxDB
105 [[outputs.influxdb]]
106   ## The full HTTP or UDP URL for your InfluxDB instance.
107   ##
108   ## Multiple URLs can be specified for a single cluster, only ONE of the
109   ## urls will be written to each interval.
110   # urls = ["unix:///var/run/influxdb.sock"]
111   # urls = ["udp://127.0.0.1:8089"]
112   urls = ["http://127.0.0.1:8086"]
113   ## HTTP Basic Auth
114   username = "test"
115   password = "[redacted]"
```

Fig. 8. *Configuración telegraf para escritura de registros en influxDB*

Paso 3: Visualización de métricas de red, CPU, memoria y disco

La aplicación *Influxdb* es compatible con la aplicación *Chronograf,* la cual tiene la capacidad de consumir las métricas de la base de datos y graficarlas en una interfaz web:

La aplicación permite ejecutar *querys* a la base de datos vía web y extraer cualquier registro que esté almacenado en esta. También es posible consultar vía web los eventos *syslog* y datos en la base de datos en tiempo real o tiempo pasado.

Paso 4: Almacenar métricas de red, CPU, memoria y disco en la base de datos:

El almacenamiento se realiza desde el cliente *telegraf* con la configuración que anteriormente se había aplicado.

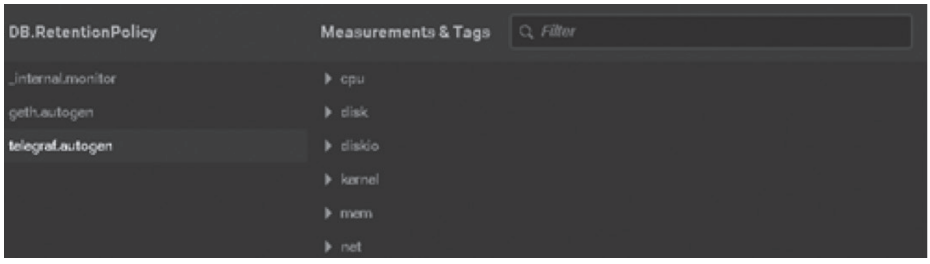


Fig. 9. Almacenamiento de métricas en red, CPU, memoria y disco en base de datos Influxdb

Paso 5: Alertar eventos críticos y modelar detección de anomalías

Para alertar en tiempo real, hacemos uso de la herramienta *Kapacitor,* la cual es un analizador de *logs* en tiempo real y genera alertas preestablecidas de acuerdo con la parametrización. A continuación, se presenta alerta cuando la CPU está por encima del 80 % en relación con el consumo total.

V. CONCLUSIONES

En el nivel de monitoreo no se recomienda hacer uso del método RPC, ya que este induce en el sistema una sobrecarga adicional de procesamiento de los datos que puede impactar el rendimiento. Se propone monitorear el sistema por medio de la extracción de métricas en el nivel de aplicaciones colectoras de métricas y emplear el método *Syslog* para su análisis.

Almacenar todas las métricas en una base de datos centralizada permite tener trazabilidad sobre los indicadores de rendimiento y ayuda a entrenar el sistema bajo algoritmos de *machine learning* para detectar anomalías.

La validación con herramientas de código abierto permite la integración de nuevos componentes en el sistema para resolver necesidades específicas de rendimiento y monitoreo en un sistema *Blockchain Ethereum.*

Se proponen pruebas de evaluación de rendimiento del sistema, con el fin de que las organizaciones e investigadores pongan a prueba sus redes *Blockchain Ethereum,* ante eventos relacionados con degradación del servicio por la afectación del rendimiento en algunos componentes, y tengan la posibilidad de monitorearlos y predecirlos bajo el modelo propuesto.

En un servidor *Ethereum,* los elementos de CPU, memoria, disco y red tienen un alto impacto en el rendimiento de las transacciones.

VI. REFERENCIAS

[1] J. Mattila, "The Blockchain Phenomenon. The Disruptive Potential of Distributed Consensus Architectures. *ETLA Working Papers,* n.° 38, 2016.

[2] B. Jepkemei and A. Kipkebut, "Blockchain - A Disruptive Technology in Financial Assets", *IRE Journals,* vol. 2, n.° 9, pp. 38-47, 2019.

[3] A. S. Deshpande, *Design and Implementation of an Ethereum-like Blockchain Simulation Framework.* 2018.

[4] H. F. Leppelsack, "Experimental Performance Evaluation of Private Distributed Ledger Implementations" (Experimentelle Performanz Evaluierung von privaten Distributed Ledger Implementierungen), Master's thesis in Informatics, Technical University of Munich, April 2018.

[5] P. Zheng, Z. Zheng, X. Luo, X. Chen, and X. A. Liu, "A detailed and real-time performance monitoring framework for blockchain systems. In *Proceedings of International Conference on Software Engineering,* 2018, pp. 134-143. [On line]. https://doi.org/10.1145/3183519.3183546.

[6] H. Sukhwani, J. M. Martínez, X. Chang, K. S. Trivedi, and A. Rindos, "Performance modeling of PBFT consensus process for permissioned blockchain network (hyperledger fabric)". In *Proceedings of the IEEE Symposium on Reliable Distributed Systems,* Sept. 2017, pp. 253-255. [On line]. https://doi.org/10.1109/SRDS.2017.36

[7] F. Schussler, P. Nasirifard, H.-A. Jacobsen, "Attack and Vulnerability Simulation Framework for Bitcoin-like Blockchain Technologies". In *Middleware '18: Proceedings of the 19th International Middleware Conference* (Posters), December 2018 pp. 5-6. [On line]. https://doi.org/10.1145/3284014.3284017

[8] P. Thakkar, S. Nathan, and B. Viswanathan, "Performance benchmarking and optimizing hyperledger fabric blockchain platform". In *Proceedings of 26th IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems,* MASCOTS 2018, pp. 264-276. [On line]. https://doi.org/10.1109/MASCOTS.2018.00034

[9] D. Mingxiao, M. Xiaofeng, Z. Zhe, W. Xiangwei, and C. Qijun, "A review on consensus algorithm of blockchain". In *IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017,* 2017-Janua, pp. 2567-2572. [On line]. https:/ /doi. org/ 10.1109/SMC.2017.8123011

[10] S. Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System,* 2008.

[11] T. T. A. Dinh, R. Liu, M. Zhang, G. Chen, B. C. Ooi, and J. Wang, "Untangling Blockchain: A Data Processing View of Blockchain Systems". In *IEEE Transactions on Knowledge and Data Engineering,* vol. 30, n.° 7, pp. 1366-1385, 2018. [On line]. https://doi.org/10.1109/TKDE.2017.2781227

[12] R. Yasaweerasinghelage, M. Staples, and I. Weber, "Predicting Latency of Blockchain-Based Systems Using Architectural Modelling and Simulation". In *Proceedings of IEEE International Conference on Software Architecture, ICSA 2017,* pp. 253-256. [On line]. https://doi.org/10.1109/ICSA.2017.22

[13] Y. Aoki, K. Otsuki, T. Kaneko, R. Banno, and K. Shudo, "SimBlock: A Blockchain Network Simulator". In *INFOCOM 2019 - IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS 2019,* pp. 325-329. [On line]. https://doi.org/10.1109/INFCOMW.2019.8845253

[14] X. Xu, Q. Lu, Y. Liu, L. Zhu, H. Yao, and A. V. Vasilakos, "Designing blockchain-based applications a case study for imported product traceability", *Future Generation Computer Systems,* vol. 92, pp. 399-406. https://doi.org/10.1016/j.future.2018.10.010

[15] H. Sukhwani, J. M. Martínez, X. Chang, K. S. Trivedi, A. Rindos, "Performance modeling of PBFT consensus process for permissioned blockchain network (hyperledger fabric)". In *Proceedings of the IEEE Symposium on Reliable Distributed Systems,* Sept. 2017, pp. 253-255. [On line]. https://doi.org/10.1109/SRDS.2017.36

[16] R. A., Memon, J. P., Li, and J. Ahmed, "Simulation model for blockchain systems using queuing theory", *Electronics (Switzerland),* vol. 8, n.° 2, pp. 1-19. [On line]. https://doi.org/10.3390/electronics8020234

[17] C. Faria and M. Correia, "BlockSim: Blockchain simulator". In *Proceedings of 2nd IEEE International Conference on Blockchain,* Blockchain 2019, pp. 439-446. [On line]. https://doi.org/10.1109/Blockchain.2019.00067

Este libro se terminó de imprimir
en los talleres de Divegráficas S.A.S
en el mes de diciembre de 2020

El libro Investigación e Innovación en Ingeniería de Software, en su cuarto volumen, compila once artículos que relacionan las disciplinas y tópicos de la industria 4.0, con referencia a la especificación de requisitos de software, la inteligencia de negocios, la gestión del conocimiento, los sistemas de información geográfica, la seguridad de la información, la lingüística computacional y la inteligencia artificial.

Tecnológico de Antioquia
Institución Universitaria
VIGILADA MINEDUCACIÓN

SELLO EDITORIAL TdeA

GIISTA
Grupo de Investigación en Ingeniería de Software del Tecnológico de Antioquia