

ENCUENTRO
NACIONAL DE
COMPUTACION
2021

Presentación

Actualmente la ciencia de la computación tiene un papel fundamental en la innovación y el desarrollo tecnológico, su aplicabilidad en todos los ámbitos permite coadyuvar en la resolución de problemas que enfrenta la sociedad en diversas áreas del conocimiento con el fin de mejorar la condición de vida de los seres humanos.

Este libro presenta una pequeña muestra de los trabajos en progreso de estudiantes de licenciatura y maestría de distintas instituciones educativas. Esta selección, donde participaron 18 instituciones diferentes a nivel nacional, muestra investigaciones recientes que ofrecen un panorama de los avances en distintas áreas donde la ciencia de la computación tiene una incidencia transversal en temas de educación, seguridad, salud, aplicaciones móviles en telemedicina, conteo de células, y tráfico vehicular, entre otros.

Este libro de memorias del ENC 2021 contiene 12 capítulos cada uno de los cuales es el trabajo de investigación donde al menos un estudiante participa y los productos generados servirán como evidencia de las competencias desarrolladas y en algunos casos, cumplir con requisitos para obtener el certificado de formación en sus estudios. El objetivo es que cada estudiante comparta el trabajo de investigación desarrollado y obtenga retroalimentación por parte de otros estudiantes y un grupo de expertos. Además de fomentar la colaboración y participación en proyectos de investigación en el área de ciencia de la computación.

La Sociedad Mexicana de Ciencia de la Computación se complace en presentar estas memorias del Encuentro Nacional de Computación 2021, las cuales constituyen un producto sometido a un arduo trabajo de revisión en un proceso doble ciego. El comité técnico estuvo integrado por 12 profesores e investigadores con reconocida trayectoria académica, que se

Instituciones

Amphora Health

CICESE-UT3

Cinvestav Tamaulipas

CONACyT-Universidad Michoacana de San Nicolás de Hidalgo

Instituto de energías renovables, UNAM

LANIA

TecNM - Instituto Tecnológico de Chihuahua

TecNM - Instituto Tecnológico de Ciudad Guzmán

TecNM - Instituto Tecnológico de Culiacán

TecNM - Instituto Tecnológico de Morelia

TecNM - Instituto Tecnológico Superior de Martínez de la Torre

Universidad Autónoma de Tamaulipas

Universidad Autónoma del Carmen

Universidad Autónoma Metropolitana

Universidad Nacional Autónoma de México

Universidad Michoacana de San Nicolás de Hidalgo

Comité Técnico

Dr. Noel Enrique Rodríguez Maya
Tecnológico Nacional de México - Instituto Tecnológico de Zitácuaro

Luis Alberto Morales Rosales
CONACYT - Universidad Michoacana de San Nicolás de Hidalgo

Juan J. Flores Romero
Universidad Michoacana de San Nicolás de Hidalgo

Mariana Lobato Báez
Tecnológico Nacional de México - Instituto Tecnológico Superior de Libres

Gloria Ekaterine Peralta Peñuñuri
Tecnológico Nacional de México - Instituto Tecnológico de Culiacán

Ignacio Algreto Badillo
Instituto Nacional de Astrofísica, óptica y electrónica

Abraham Rodríguez Mata
Tecnológico Nacional de México - Instituto Tecnológico de Culiacán

Víctor Alejandro González Huitrón
CONACYT- Tecnológico Nacional de México - Instituto Tecnológico de Culiacán

Carlos Lara Álvarez
Centro de Investigación en Matemáticas (CIMAT)

Ansel Y. Rodríguez González
Centro de Investigación Científica y de Educación Superior de Ensenada
Unidad de Transferencia Tecnológica Tepic

Carlos Arturo Hernandez Gracida
CONACyT

Sofia Isabel Fernández Gregorio
Tecnológico Nacional de México - Instituto Tecnológico Superior de Martínez
de la Torre

Índice general

- 1. *Ejecución y Descubrimiento de IOBPs mediante Blockchain* 1**
José Antonio Molina De la Fuente y Miguel Morales Sandoval
- 2. *Diseño de un marco de evaluación para sistemas de in-formación de soporte a servicios de salud: Caso SirVi* 7**
Victor Manuel Gonzalez Bello y María de Lourdes Hernández Rodríguez
- 3. *Marco de Referencia para trazabilidad y almacenamiento de activos digitales* 14**
José Antonio Jiménez Miramontes y Rocío Aldeco-Pérez
- 4. *Una nueva herramienta educativa basada en tecnología de videojuegos para fortalecer habilidades blandas y cognitivas de los infantes* 19**
Daniela Barradas, Emily Zavala, Flor Radilla, Eduardo Vazquez-Santacruz y Elvia Morales Turrubiates
- 5. *Identificación y Conteo de células *Chlorella sp.* por medio de inteligencia artificial* 25**
Luis Alberto Miranda Torres, Héctor Rodríguez Rangel, Víctor Alejandro González Huitrón, Abraham Efraim Rodríguez Mata y Dulce María Arias Lizarraga
- 6. *Sistema de Recolección y Procesamiento de Datos para Estudios de Tránsito* 31**
Héctor Rodríguez Rangel, Rafael Imperial Rojo, Luis Alberto Morales

Rosales, Sofia Isabel Fernández Gregorio y Abraham Efraim Rodríguez Mata

7. *Implementación de un Servicio REST-API para el uso de la firma electrónica para instituciones públicas* 37

Jesus Alfredo Bravo Méndez, Anastacio Antolino Hernández y Heberto Ferreira Medina

8. *Metodología para obtención de Zonas Homogéneas en pavimentos flexibles* 43

Eduardo Daniel Raya Gamiño, Jorge Alarcón Ibarra, Luis Alberto Morales Rosales, Jaime Saavedra Rosales y Juan Carlos Gallegos Cornejo

9. *Estrategia de Verificación de la Conformidad entre Modelos de Procesos y Bitácoras de Eventos.* 49

Arely Moreno y Heidy M. Marin-Castro

10. *Desarrollo e implementación de un sistema proactivo de iluminación* 55

Rosario de La Luz Cantero Ramírez, Marco Antonio Meza Aguilar y Ansel Yoan Rodríguez González

11. *Algoritmo de Anonimización de Notas Médicas para Proteger la Identidad de los Pacientes* 61

Jesús Mercado, Karina Figueroa, Arturo López, Cleto Álvarez y Anel Gómez

12. *Modelo difuso para la evaluación de cortes carreteros* 67

Juan Carlos Gallegos Cornejo, Luis Alberto Morales Rosales, Jaime Saavedra Rosales, Daniel Raya Gamiño y Virgilio López Morales

Índice de autores 73

CAPÍTULO 1

Ejecución y Descubrimiento de IOBPs mediante Blockchain

JOSÉ ANTONIO MOLINA DE LA FUENTE Y MIGUEL MORALES SANDOVAL

Ejecución y Descubrimiento de procesos de negocio interorganizacionales mediante Blockchain

J. A. Molina-De la Fuente and M. Morales-Sandoval

Cinvestav Tamaulipas, Cd. Victoria 87138, Mexico
jose.antonio.molina@cinvestav.mx

Resumen En este trabajo se presenta el desarrollo de una herramienta basada en la cadena de bloques que permita realizar 1) la ejecución y monitoreo de procesos de negocio interorganizacionales (IOBPs), 2) captura y conformación de la bitácora de eventos durante la ejecución del IOBP y 3) preparación de datos en el formato requerido para descubrimiento de IOBP en el contexto de minería de procesos.

Palabras clave: Procesos de negocio interorganizacionales · Cadena de bloques · Minería de procesos · Contratos inteligentes.

1. Introducción

Un proceso de negocio interorganizacional [2] (*IOBP*, por sus siglas en inglés) es un grupo de actividades conjuntas ejecutadas por dos o más organizaciones para lograr un objetivo en común. Los IOBPs son generalmente representados mediante la notación *Business Process Model and Notation (BPMN*, por sus siglas en inglés). La Fig. 1 muestra el IOBP del proceso de cadena de suministro en BPMN, en el cual participan cinco organizaciones, tiene un evento de inicio, un evento de fin, 10 tareas y dos compuertas paralelas.

La ejecución de IOBPs es apoyada por sistemas de software distribuidos. Cada organización realiza una parte del proceso de forma interna, mediante sus propios sistemas de información. Ejecutar una instancia de un IOBP genera una serie de eventos llamada traza que corresponde a un caso del IOBP. Un evento representa una actividad en algún proceso y tiene las siguientes propiedades: *id del caso*, *id del evento*, *tiempo* en el cual ocurrió el evento, *actividad* acción realizada en el proceso, *recurso* nombre de la entidad que realizó la actividad y *costo*. Comúnmente, las trazas son distintas entre sí debido a situaciones que se presentan durante la ejecución del proceso. El conjunto de trazas registradas por los sistemas de información forman la bitácora de eventos. El objetivo de la minería de procesos [1] es crear modelos que representan el comportamiento real de un proceso a través de la extracción de conocimiento de una bitácora de eventos del proceso. El modelo descubierto permite identificar problemas, áreas de mejora, así como diferencias y similitudes entre el modelo descubierto y su bitácora de eventos.

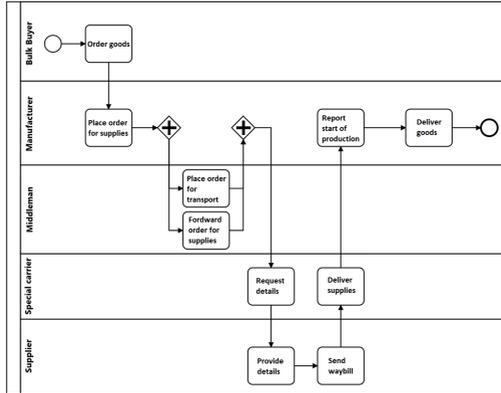


Figura 1: IOBP del proceso de cadena de suministro [7].

En la ejecución de un IOBP las organizaciones participantes son independientes y se rigen por normas y políticas distintas, por lo que es necesario un mecanismo que asegure la confianza entre ellas. La cadena de bloques [6] es una tecnología que permite a usuarios desconocidos realizar y dar seguimiento a transacciones de forma segura y transparente sin necesidad de un intermediario confiable. La cadena de bloques podría utilizarse para establecer confianza entre los participantes de un IOBP y como una fuente de datos confiable para almacenar la bitácora de eventos, dado que es una estructura inmutable.

2. Antecedentes

En la literatura se han presentado indicios de cómo incluir una cadena de bloques para resolver el problema de confianza en IOBPs [6,7]. Sin embargo, estas soluciones solo se enfocan en el monitoreo y ejecución, sin abordar la minería de IOBPs. La cadena de bloques podría utilizarse para que al mismo tiempo que se usa para asegurar la fiabilidad de la ejecución del IOBP, se use para almacenar la bitácora de eventos requerida para la minería de IOBPs. Algunos trabajos [5] parten del hecho de que existen transacciones de un IOBP ya registradas en la cadena de bloques, y con esos datos se realiza un proceso de extracción y transformación para conformar la bitácora de eventos asociada, que después se usa para la minería del IOBP. Los trabajos que intentan resolver el problema de confianza entre las organizaciones, no incluyen realizar minería de procesos. Para aplicar minería de procesos a una bitácora de eventos existe una serie de problemas asociados que van desde recabar la bitácora de eventos durante la ejecución

del IOBP, extraer la bitácora de eventos que fue recopilada y preprocesarla, la obtención del modelo y su evaluación.

El Cuadro 1 muestra las principales diferencias de la propuesta presentada en este trabajo en relación con otros enfoques, el cual es el primero que permite ejecutar un IOBP de forma segura, corrigiendo errores en los eventos y recopilándolos de forma dinámica para formar la bitácora de eventos. Todo lo anterior utilizando la cadena de bloques.

Ref.	Ejec. y Monit.	Alg. extracc.	Limp. datos	Validación MP
[4]	✗	✓	✗	✓
[5]	✗	✓	✗	✓
[3]	✗	✗	✓	✗
Este trabajo	✓	✗	✓	✓

Cuadro 1: Uso de blockchain para ejecución y monitoreo de IOBPs, y para la creación de la bitácora de eventos requerida en minería de procesos (MP).

Por tanto, el objetivo de este trabajo es crear un método basado en el uso de la cadena de bloques que permita 1) el despliegue y monitoreo de IOBPs para resolver el problema de confianza entre los participantes y 2) desplegar una estrategia para la creación de la bitácora de eventos y limpieza de datos que impacte positivamente en la tarea de descubrimiento de procesos en el contexto de minería de IOBPs.

3. Metodología

A partir de un IOBP, se utiliza un parser que recibe un IOBP en BPMN y extrae la lógica para configurar el contrato inteligente que permitirá la ejecución y monitoreo (EM) del IOBP en la cadena de bloques. Cualquier entidad que interactúa con la cadena de bloques se implementa como un *Trigger*. El encargado de realizar la configuración del contrato inteligente es el trigger supervisor. Después, se despliegan los triggers organizacionales del IOBP que realizan la ejecución del IOBP. Cuando EM recibe un evento, verifica los permisos y la conformidad, si el evento es correcto es enviado a un contrato inteligente de limpieza y preprocesamiento de eventos (PDC), el cual consulta información en un contrato inteligente dedicado a la conformación de la bitácora de eventos (ELC), corrige errores en los atributos del evento y lo registra en ELC. Después de ejecutar el IOBP, cualquier usuario puede extraer la bitácora de eventos de ELC. La bitácora extraída es transformada a formatos CSV y XES para después pasar a un algoritmo de descubrimiento de minería de procesos para obtener el modelo real. El modelo descubierto y la bitácora son la entrada de algoritmos de verificación de conformidad para evaluar el impacto de la estrategia en PDC.

4. Avances y resultados

Los contratos inteligentes propuestos en este trabajo (EM, PDC y ELC) fueron implementados en *Solidity* y desplegados en Ganache, una cadena de

bloques de Ethereum privada. Los *Triggers* desarrollados utilizan la biblioteca *Web3j*. El método propuesto se implementó como una herramienta, en Java. La herramienta recibe como entrada un IOBP en BPMN 2.0, las billeteras digitales de las organizaciones participantes, un número n de instancias del proceso a ejecutar, y un porcentaje de ruido p . Con estos datos, realiza la ejecución de de las n instancias y genera una bitácora de eventos en la cadena de bloques, con limpieza de datos on-chain incluida. Durante la ejecución de tareas, se inyecta ruido en $(p/100,0)*n$ instancias hasta en un 50 % de los eventos de una instancia. La herramienta extrae y transforma la bitácora de eventos a formato XES y se utilizan las herramientas ProM y P-miner para las tareas de minería de procesos. La Fig. 3 muestra el diagrama del método implementado.

Mediante la herramienta, se implementó el caso de estudio del IOBP de cadena de suministro de la Fig. 1. Se generaron 1200 instancias, inyectando los porcentajes de ruido: $p = \{10, 20, 30\}$. Las bitácoras generadas se utilizaron para las tareas de minería de procesos utilizando las herramientas ProM y P-miner. Los modelos descubiertos se evaluaron mediante las métricas fitness y precisión. La Fig. 2 muestra la gráfica de resultados de fitness y precisión obtenidos por P-miner. Con esta prueba, se valida el correcto funcionamiento del enfoque propuesto. Como se observa, el mecanismo de limpieza de datos on-chain mejora de manera significativa la precisión de los modelos descubiertos.



Figura 2: Fitness y precisión obtenidos por P-miner del IOBP *Supply chain*.

5. Conclusiones

Se ha presentado el desarrollo de un método que permite ejecutar y monitorear IOBPs, recabar la bitácora de eventos y limpiar eventos en la cadena de bloques mediante tres contratos inteligentes: uno para la ejecución y monitoreo del IOBP, otro para la limpieza de los eventos recabados durante la ejecución del IOBP y otro más para la conformación de la bitácora de eventos. Además, el método contempla la configuración automática del IOBP a partir de un modelo documentado en notación BPMN. Como prueba de concepto, se implementó una herramienta en Java que implementa el método completo. La herramienta permite simular la ejecución de un IOBP de manera real y crear una bitácora de

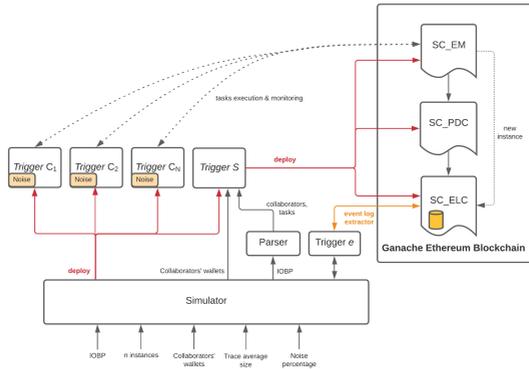


Figura 3: Diagrama del simulador implementado en Java.

eventos de n trazas en formatos CSV y XES. El desarrollo realizado demuestra la factibilidad de contar con una herramienta que no solo apoye la ejecución confiable de un IOBP en la cadena de bloques sino contar con datos confiables (la bitácora de eventos) para las tareas de minería de procesos.

Referencias

1. van der Aalst, W.M.P.: Process Mining: Data Science in Action. Springer, Berlin, Heidelberg, 2 edn. (2016)
2. Bouchbout, K., Alimazighi, Z.: Inter-organizational business processes modelling framework. vol. 789, pp. 45–54 (2011)
3. Ekici, B., Tarhan, A., Ozsoy, A.: Data cleaning for process mining with smart contract. In: 4th International Conference on Computer Science and Engineering (UBMK). pp. 1–6 (2019). <https://doi.org/10.1109/UBMK.2019.8907140>
4. Klinkmüller, C., Ponomarev, A., Tran, A.B., Weber, I.: Mining Blockchain Processes: Extracting Process Mining Data from Blockchain Applications, pp. 71–86 (08 2019). https://doi.org/10.1007/978-3-030-30429-4_6
5. Mühlberger, R., Bachhofner, S., Di Ciccio, C., García-Bañuelos, L., Pintado, O.: Extracting event logs for process mining from data stored on the blockchain (2019)
6. Sturm, C., Szalanczi, J., Schönig, S., Jablonski, S.: A lean architecture for blockchain based decentralized process execution. In: Business Process Management Workshops. pp. 361–373. Springer, Cham (2019)
7. Weber, I., Xu, X., Riveret, R., Governatori, G., Ponomarev, A., Mendling, J.: Untrusted business process monitoring and execution using blockchain (2016). https://doi.org/10.1007/978-3-319-45348-4_19

CAPÍTULO 2

*Diseño de un marco de evaluación para
sistemas de in-formación de soporte a
servicios de salud: Caso SirVi*

VICTOR MANUEL GONZALEZ BELLO Y MARÍA DE LOURDES HERNÁNDEZ
RODRÍGUEZ

Diseño de un marco de evaluación para sistemas de información de soporte a servicios de salud: Caso SirVi

Víctor Gonzalez Bello¹[0000-0002-4538-9369]; Lourdes Hernández R²[0000-0003-0069-3639]

^{1,2} Laboratorio Nacional de Informática Avanzada, A. C.,
Rébsamen 80, CP 91000, Xalapa, Ver., México

¹vgonzalez.mca20@lania.edu.mx; ²lourdes.hernandez@lania.edu.mx

Resumen. El desarrollo o adopción de Sistemas de información para Soporte a servicios de salud requiere una evaluación holística para toma de decisiones de implantación o mejora lo que implica un proceso complejo. Investigaciones en el área reportan diferentes estrategias de evaluación dirigidas a dominios o problemáticas específicas en contextos internacionales sin identificarse un marco estándar de referencia para salud. Esta investigación propone un marco de evaluación basado en la familia de normas ISO/IEC 25000 y la adaptación del modelo de evaluación de impacto de un sistema, IS-Impact, aplicado en al menos un Sistema de Salud Pública. El marco de evaluación se aplicará al Sistema SirVi, diseñado para el área de servicios de terapia fonológica del Centro de Rehabilitación e Inclusión Social (CRISVER), ubicado en Xalapa, Ver. México. Los resultados parciales incluyen características e indicadores de mayor relevancia a considerar en la evaluación, el informe final incluirá el porcentaje de cumplimiento de tres aspectos, a saber, calidad del Sistema, calidad de la información y calidad del servicio. Reportando indicadores de soporte a la toma de decisiones en su implantación.

Keywords: Sistemas de salud, evaluación de software, marco de evaluación

1 Introducción

La operación exitosa de una organización se sustenta en las decisiones sobre la tecnología que apoya sus actividades sustantivas y que le permitirá mantener o mejorar sus servicios en entornos dinámicos de la población que atiende [1]. Específicamente para organizaciones en el área de servicios de salud es de particular relevancia evaluar la calidad de los Sistemas informáticos que utiliza considerando un modelo holístico sustentado en estándares internacionales, incluyendo aspectos organizacionales, sociales y técnicos [2] [3]. Las perspectivas de interés son complejas de evaluar y requieren de un análisis cuidadoso observando la interrelación de las variables bajo análisis.

En este contexto el Centro de Rehabilitación e Inclusión Social de Veracruz (CRISVER), se encarga de brindar servicios de salud con respecto a rehabilitación integral de calidad y calidez a personas con discapacidad, así como contribuir en la eliminación de las barreras físicas, sociales y culturales, que permitan su inserción

social [4], estas actividades enmarcan un contexto de percepción individual en referencia a quienes reciben el servicio además organizacional en cuanto al cumplimiento de la misión de la institución. Adicionalmente, entidades de gobierno estatal y federal están interesadas en alcanzar metas de brindar servicios de salud de calidad a la población en general y a la población con alguna discapacidad en particular.

El Laboratorio Nacional de Informática Avanzada y el grupo de médicos del área fonológica del CRIVER realizaron un sistema de registro estandarizado para el diagnóstico y seguimiento en terapia fonológica (SirVi) [5]. Sirvi está conformado por los módulos de registro y de visualización de datos, orientado a optimizar el tiempo de atención a pacientes, la comunicación entre terapeutas y médicos así como presentar de forma gráfica progreso de terapia y monitoreo de la población atendida en el área de fonología [4][5]. Los aspectos funcionales y de cobertura son de relevancia para la aplicación del modelo de evaluación propuesto en este trabajo.

2 Trabajo relacionado

El trabajo Initial Framework for Software Quality Evaluation based on ISO/IEC 25022 and ISO/IEC 25023 [6], cuyo propósito es desarrollar un marco basado en un estándar internacional propone un marco integral para todas las características de calidad de ISO / IEC 25022 e ISO / IEC 25023, reducir métricas, medidas ambiguas y definir entradas y las salidas para mediciones de calidad con claridad. La metodología de los autores y contribuciones se resumen en (a) la definición de un marco para las mediciones y evaluaciones de calidad, (b) establecimiento de un procedimiento de uso del marco para evaluar la calidad del software, (c) incorporar métricas factibles y mediciones en el marco propuesto, y (d) demostrar la efectividad del marco para las partes interesadas del proyecto a través de un estudio de caso.

La mantenibilidad de software es una característica importante y en el trabajo Applying the ISO/IEC 25010 Quality Models to Software Product [7] plantea como las aplicaciones de valor significativo pasarán por una serie de versiones de mantenimiento para cubrir el soporte de versiones futuras. Su objetivo es explorar el alcance y la interpretación de los modelos de calidad ISO / IEC 25010, identificando aspectos significativos del producto que conciernen a los adquirentes de software.

En el trabajo “Software Quality Evaluation of the Laboratory Information System used in the Santa Catarina’s State wide Integrated Telemedicine and Telehealth System” [8] se presenta un caso de estudio aplicando un modelo de evaluación personalizado para validar la calidad del software de un sistema de información de laboratorio. Los autores analizaron las características del sistema LACEN, personalizando un modelo para la evaluación de la calidad de los sistemas de telemedicina basado en la norma ISO / IEC 25010 mediante el uso de un cuestionario. Concluyen que, la calidad del sistema LACEN es buena, específicamente por características como eficacia, eficiencia, satisfacción, ausencia de riesgo, idoneidad funcional y seguridad. Por otro lado, pocas características se han considerado de baja calidad, incluidas la eficiencia y el rendimiento.

Davidson, et al [2] presentan la adaptación de un modelo de evaluación de impacto de un sistema, IS -impact, en un sistema de salud de amplia cobertura en Canadá. El trabajo reporta 48 aspectos evaluados integrados en dos dimensiones calidad e impacto a nivel individual, organizacional y sectorial. Los resultados exhiben valores promedios 3.4 de 7 del sistema evaluado. No obstante, el modelo es de particular relevancia para el trabajo presentado dado que amplía el ámbito de evaluación.

La evaluación de calidad de productos dirigidos a la salud es fundamental para asegurar la digitalización de servicios, el establecimiento de un marco de referencia basado en estándares internacionales dará soporte al proceso de evaluación de calidad de software como SIRVI. La Evaluación debe incluir tanto aspectos de calidad externa como interna, así como calidad en uso.

3 Objetivos

El objetivo general del trabajo presentado es establecer un marco holístico de evaluación de sistemas de información orientados a dar soporte a servicios de salud con base en la familia de normas internacionales ISO/IEC 25000 y modelos de referencia en el área de salud.

Los objetivos específicos asociados son: analizar la familia de normas ISO/IEC 25000, así como modelos de aplicados con éxito en el dominio de salud tanto pública como privada; diseñar el marco de evaluación considerando aspectos sociales, organizacionales y técnicos; aplicar el marco de evaluación al sistema SirVi y realizar el análisis de resultados correspondiente.

4 Metodología

La revisión de trabajos relacionados incluye tanto el estándar ISO/IEC 25000 [3] cómo modelos reportados en McCrorie et al[1], B. Davidson et al [2], Nakai et al [6], Estdale & Georgiadou [7] considerados como más apropiados para evaluar un Sistema que proporciona Servicios de Salud. El estándar ISO/IEC 25000 provee un conjunto general de características a considerar en calidad externa e interna de los cuales se han seleccionado los de mayor relevancia para el dominio de salud, a saber: seguridad, integridad, escalabilidad y facilidad de mantenimiento, en cuanto a calidad externa: usabilidad y fiabilidad.

En cuanto a aspectos organizacionales se han revisado con especial detalle, la modificación al modelo IS-Impact publicada por Davidson [2], un caso de aplicación que cubre los aspectos de interés a evaluar, particularmente impactos a nivel individual, organizacional y a nivel de sector gubernamental.

Los elementos analizados cómo referentes dan lugar a un marco específico de evaluación, que incluirá características e indicadores. Una vez delimitado el marco de evaluación y los sujetos participantes se procederá a la elaboración y aplicación de instrumentos.

Los datos obtenidos se analizarán a través de estadística descriptiva para obtener una calificación global de calidad del producto así como impactos en al menos dos dimensiones, a saber, individual y organizacional.

5 Avances

A partir del análisis realizado con respecto a características de calidad de particular relevancia en sistemas de soporte a servicios de salud se realizó la conceptualización inicial del modelo representado en la Figura 1 y la selección características presentadas en la Tabla 1 y 2, correspondiendo a atributos de calidad externa e interna.

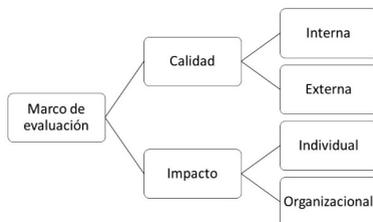


Fig. 1. Esquema general inicial del modelo de evaluación propuesto

Tabla 1. Características de calidad externa sujetas a evaluación en el modelo propuesto.

Característica	Descripción
Disponibilidad	Identificación puntual de los módulos que requieren mantenerse en operación indicando periodos de tiempo críticos así como temporalidad para mantenimiento
Integridad	Asegurar ejecución de operaciones de persistencia y actualización de datos, considerando operaciones incompletas.
Seguridad	Protección de acceso al sistema y a datos considerados sensibles
Usabilidad	Indicadores de tiempo invertido en la realización de operaciones con el sistema, cuantificación de tareas realizadas, operación autónoma del usuario sin requerir ayuda

Tabla 2. Características de calidad interna sujetas a evaluación en el modelo propuesto.

Característica	Descripción
Eficiencia	Utilización óptima de recursos en funcionalidades con mayor demanda, manteniendo un nivel aceptable de

	operación
Facilidad de modificación	Tiempo requerido para atender realizar una mejora o arreglar incidentes.
Escalabilidad	Mantenimiento de niveles de rendimiento en aumento de demanda de uso asignando los recursos requeridos
Facilidad de verificación	Especificidad de características que deben exhibir los resultados generados por la operación del sistema en los escenarios establecidos para cada funcionalidad

6 Conclusiones

La evaluación de criterios de calidad de software alineada con estándares internacionales que permitan garantizar un nivel óptimo de operación de los servicios que soportan es fundamental para la determinación de puesta en operación de un sistema en el dominio de salud pública. Los trabajos inicialmente revisados exhiben marcos de evaluación propios y basados en estándares. Orientando el establecimiento de un marco de evaluación multidimensional con características de calidad interrelacionadas con dimensiones de impacto de uso de la población objetivo a nivel individual y organizacional.

El marco de evaluación que se tiene como meta pretende ser un referente para evaluación de calidad específicamente en el dominio de salud. El caso de estudio que se aborda es un sistema de alta demanda y la determinación de calidad de este permitirá continuar el proceso de despliegue y adopción. La toma de decisiones del CRISVER en referencia a tecnología apoyará la continuación de sus servicios de alto impacto social y la evaluación propuesta se orienta a proveer de indicadores que soporten la toma de decisiones.

Referencias

1. C. McCrorie, J. Benn, O. A. Johnson, and A. Scantlebury, "Staff expectations for the implementation of an electronic health record system: A qualitative study using normalisation process theory," *BMC Med. Informat. Decis. Making*, vol. 19, no. 1, pp. 1–14 (2019)
2. B. Davidson, M. A. A. Dewan, V. S. Kumar, M. Chang and B. Liggett, "Visualizing Benefits: Evaluating Healthcare Information System Using IS-Impact Model," in *IEEE Access*, vol. 8, pp. 148052-148065, doi: 10.1109/ACCESS.2020.3015467. (2020)
3. ISO (International Organization for Standardization), ISO/IEC 25022 Software Engineering. Software Product Quality Requirements and Evaluation (SQuaRE). Measurement of Internal Quality. ISO, (2005)
4. I. Zurita Barbosa, F., Hernandez Rodríguez, Visualización De Datos Médicos De Terapias Fonológicas Para Soporte A Toma De Decisiones. Laboratorio Nacional de Informática Avanzada. (2020)
5. Bardales, A; Hernández Rodríguez.; Marin de la Luz; Diseño de software orientado a visualización y registro de datos médicos en terapia de lenguaje. 5th International Conference in Software Engineering Research and Innovation.(2015)

6. Nakai, H., Tsuda, N., Honda, K., Washizaki, H., & Fukazawa, Y. Initial framework for software quality evaluation based on iso/iec 25022 and iso/iec 25023. In 2016 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C) (pp. 410-411). IEEE. (2016).
7. Estdale, J., & Georgiadou, E. Applying the ISO/IEC 25010 quality models to software product. In European Conference on Software Process Improvement (pp. 492-503). Springer, Cham.(2018)
8. Alves, J. M., Savaris, A., Von Wangenheim, C. G., & Von Wangenheim, A. Software quality evaluation of the laboratory information system used in the santa catarina state integrated telemedicine and telehealth system. In 2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS) (pp. 76-81). IEEE. (2016)

CAPÍTULO 3

*Marco de Referencia para trazabilidad y
almacenamiento de activos digitales*

JOSÉ ANTONIO JIMÉNEZ MIRAMONTES Y ROCÍO ALDECO-PÉREZ

Marco de referencia para trazabilidad y almacenamiento de activos digitales

José Antonio Jiménez Miramontes¹[0000-0001-9917-1149] and Rocío Aldeco-Pérez²[0000-0002-7003-2724]

¹ UNAM, IIMAS, Posgrado en Ciencias e Ingeniería de la Computación, CU, México
ja.jimenez.mi@gmail.com

² UNAM, Facultad de Ingeniería, Departamento de Computación, CU, México
raldeco@unam.mx

Abstract. El control que la gente tiene sobre sus activos digitales una vez que están en internet es debatible. Se propone un marco de referencia que haga uso del linaje electrónico para satisfacer la necesidad de controlar lo que sucede con los activos digitales y saber qué es lo que está pasando con estos. Se propone que dicho marco haga uso de la tecnología *Blockchain* para proporcionar la integridad y el no repudio del linaje electrónico. A su vez, se plantea el uso de un protocolo distribuido para almacenar los activos digitales proporcionando descentralización.

Keywords: Blockchain · Trazabilidad · Activos digitales · IPFS.

1 Introducción

El control que la gente tiene sobre sus activos digitales una vez que están en internet es debatible. Estos activos pueden ser duplicados con facilidad o caer en manos de un tercero que no necesariamente es de confianza. Con esto surge la necesidad de controlar lo que sucede con los activos digitales. El *linaje electrónico* nos permite conocer el origen y los procesos por los que pasa un activo digital, con esto conocemos toda su historia [6]. Este registro de procesos le da la posibilidad al dueño del activo de saber qué es lo que ocurre con su propiedad. Para lograr esto, las aplicaciones deben generar documentación de los procesos que le están ocurriendo a los activos digitales y almacenarlos de forma segura para que no sean alterados y sean confiables. Debido al papel que esta información juega, asegurar su integrar se vuelve clave. La primeras propuestas para almacenar el *linaje electrónico* involucran un sistema centralizado donde se encuentran todos los registros de procesos [6]. Esto nos obliga a tener un tercero de confianza sin garantía alguna de su comportamiento. Para resolver este problema es necesario el uso esquemas distribuidos y descentralizados, una opción es *Blockchain*.

Blockchain, al ser una base de datos distribuida, puede ser una alternativa al almacenamiento centralizado del *linaje electrónico* aunque presenta ciertas problemáticas como su lentitud, alto consumo eléctrico y baja eficiencia en el almacenamiento de archivos [3]. En vista de esto, proponemos el uso de una *Blockchain* basada en un protocolo de consenso eficiente y sustentable descrita

en el trabajo Yakovenko [9]. Para el caso del almacenamiento del activo digital en sí mismo, proponemos el uso de otro protocolo diseñado para este fin y descrito en [2].

2 Antecedentes

El *linaje* puede ser conocido como la derivación desde un origen particular hasta un estado específico de un elemento [6]. El *linaje electrónico* apoya la información y la integridad del proceso documentando las entidades, sistemas y procesos que operan y contribuyen a los datos de interés, sirviendo como un registro histórico inalterable de la duración de los datos y sus orígenes [5].

Por otro lado tenemos *Blockchain* que es una lista creciente de registros llamados bloques, unidos entre sí a través del uso de *hashes* criptográficos. Mediante el uso de redes punto a punto y protocolos de consenso, *Blockchain* hace que la historia de activos digitales sea inalterable y transparente [8].

Yakovenko [9] propone el uso del protocolo de consenso llamado Proof of Stake (PoS) con la adición de un segundo protocolo llamado Proof of History (PoH). Este es una secuencia de hashes que mantienen un registro del tiempo en la *Blockchain* creando un orden para las transacciones y proporcionando rápida sincronización entre los nodos de la red.

Un sistema distribuido para almacenamiento de archivos muy prometedor en la actualidad es el InterPlanetary File System (IPFS), el cual está diseñado como infraestructura para construirse sobre él y proporciona un modelo de almacenamiento en bloques con dirección a contenido de alto rendimiento con hiperenlaces con dirección de contenido, formando un árbol de Merkle. Además, IPFS combina una tabla hash distribuida, un intercambio de bloques incentivado y un espacio de nombres autocertificable [2].

Existen propuestas de usar *Blockchain* para el almacenamiento del *linaje electrónico*. Azaria et al. [1] proponen un esquema descentralizado para el manejo de registros médicos con el uso de *Blockchain*. Sifahet al. [7], por otro lado, presentan el uso de *Blockchain* para el *linaje electrónico* archivos de en la nube. Khatal et al. [4] proponen de igual forma un marco de referencia para el *linaje electrónico* de archivos con el uso de *Blockchain* y de IPFS. Sin embargo, las primeras dos proponen almacenamiento centralizado mientras que en [4] se propone un esquema descentralizado enfocándose únicamente en el *linaje electrónico* de archivos de texto. Estos trabajos también proponen usar la red de Ethereum para la implementación de *Blockchain*, la cual se basa en el algoritmo de consenso *Proof of Work* (PoW) generando una solución lenta y costosa. El uso del protocolo de consenso Proof of Stake con Proof of History permite mejores tiempos de respuesta y un menor costo de transacción.

3 Objetivo y Metodología

El objetivo de este trabajo es diseñar un marco de referencia para el manejo de activos digitales que capture el *linaje electrónico* de estos haciendo uso de

una *Blockchain* pública basada en PoS y PoH y un protocolo distribuido de almacenamiento de archivos, para brindar control sobre un activo digital dado y monitorear su historia. La metodología que se propone seguir para cumplir con el objetivo es la siguiente: (1) Analizar el funcionamiento del protocolo de consenso y del protocolo distribuido, así como sus componentes. Analizar la integración de estos protocolos así como la información a incluir en las transacciones de cada protocolo basado en el tipo de activo digital. Esto haciendo uso de contratos inteligentes. (2) Diseñar y modelar los componentes del marco de referencia y sus procesos para ofrecer trazabilidad y almacenamiento distribuido. (3) Verificar que es posible tener trazabilidad de activos digitales de manera eficiente usando el modelo de los protocolos correspondientes y casos de uso. Hasta ahora se ha realizado la investigación de las propuestas actuales de *linaje electrónico*, los trabajos de *Blockchain* y sistemas distribuidos para almacenamiento de archivos así como de los protocolos de consenso. Se modeló la arquitectura preliminar de marco de referencia propuesto integrando PoS, PoH y el protocolo IPFS. Como se muestra en la Fig. 1, se proponen dos capas, una que gestiona el *linaje electrónico* a través de *Blockchain* y otra para el almacenamiento del mismo archivo y sus versiones modificadas. Esta última utiliza IPFS.

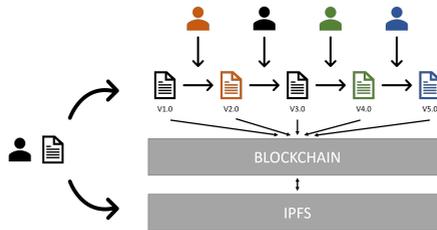


Fig. 1. Arquitectura General

También se modeló la capa *Blockchain* donde el protocolo de consenso PoS interactúa con PoH.

4 Resultados esperados y Conclusiones

Como resultado de este trabajo se tendrá un marco de referencia que integre una *Blockchain* que utilice el protocolo de consenso Proof of Stake con Proof of History para almacenar el *linaje electrónico* de archivos y un protocolo para almacenar dichos activos de forma distribuida. Esto dará como resultado la base para la creación de aplicaciones seguras que den el control a los usuarios sobre sus activos digitales y que no violen su derecho a la privacidad. A su vez, se

generarán modelos para el marco de referencia con los cuales se puede observar su funcionamiento, incluyendo los contratos digitales y los procesos que se llevarán a cabo dentro de éste. Para evaluar los modelos durante esta investigación se realizará su formalización, lo que permitirá probar tanto la eficiencia de esta propuesta como la trazabilidad que brinda. Si el tiempo lo permite, se creará una prueba de concepto usando la plataforma de Solana.

Es clara la necesidad que existe de que las personas tengan control sobre la información que comparten en internet. La utilización de *Blockchain* para almacenar el *linaje electrónico* de activos digitales brinda integridad y no repudio a la información referente a la historia de dicho activo. Sin embargo, esto requiere de una tecnología eficiente, sustentable y que permita el almacenamiento de archivos. En vista de esto, el marco de referencia propuesto incluye el uso del protocolo de consenso PoS con PoH y el protocolo de almacenamiento IPFS. Esto nos permitirá observar los cambios que sufren sus activos digitales usando contratos inteligentes con tiempos de transacción cortos y un almacenamiento distribuido. Finalmente, el soportar este sistema con una arquitectura distribuida, permite confiar en toda la información generada, además de garantizar que las medidas de control y trazabilidad implementadas serán cumplidas.

References

1. Azaria, A., Ekblaw, A., Vieira, T., Lippman, A.: Medrec: Using blockchain for medical data access and permission management. In: 2016 2nd International Conference on Open and Big Data (OBD). pp. 25–30 (2016). <https://doi.org/10.1109/OBD.2016.11>
2. Benet, J.: Ipfs - content addressed, versioned, p2p file system (2014)
3. Joshi, A., Han, M., Wang, Y.: A survey on security and privacy issues of blockchain technology. *Mathematical Foundations of Computing* **1**, 121–147 (01 2018). <https://doi.org/10.3934/mfc.2018007>
4. Khatal, S., Rane, J., Patel, D., Patel, P., Busnel, Y.: Fileshare: A blockchain and ipfs framework for secure file sharing and data provenance. In: Patnaik, S., Yang, X.S., Sethi, I.K. (eds.) *Advances in Machine Learning and Computational Intelligence*. pp. 825–833. Springer Singapore, Singapore (2021)
5. McDaniel, P.: Data provenance and security. *IEEE Security Privacy* **9**(2), 83–85 (2011). <https://doi.org/10.1109/MSP.2011.27>
6. Moreau, L., Groth, P., Miles, S., Vázquez-Salceda, J., Ibbotson, J., Sheng, J., Munroe, S., Rana, O., Schreiber, A., Tan, V., Varga, L.: The provenance of electronic data. *Commun. ACM* **51**, 52–58 (04 2008). <https://doi.org/10.1145/1330311.1330323>
7. Sifah, E.B., Xia, Q., Agyekum, K.O.B.O., Xia, H., Smahi, A., Gao, J.: A blockchain approach to ensuring provenance to outsourced cloud data in a sharing ecosystem. *IEEE Systems Journal* pp. 1–12 (2021). <https://doi.org/10.1109/JSYST.2021.3068224>
8. Sivleen, K., Sheetal, C., Aabha, S., Jayaprakash, K.: A research survey on applications of consensus protocols in blockchain. *Security and Communication Networks* **2021** (01 2021). <https://doi.org/10.1155/2021/6693731>
9. Yakovenko, A.: Solana: A new architecture for a high performance blockchain v0.8.13. Tech. rep.

CAPÍTULO 4

*Una nueva herramienta educativa basada en
tecnología de videojuegos para fortalecer
habilidades blandas y cognitivas de los
infantes*

DANIELA BARRADAS, EMILY ZAVALA, FLOR RADILLA, EDUARDO VAZQUEZ-
SANTACRUZ Y ELVIA MORALES TURRUBIATES

Una nueva herramienta educativa basada en tecnología de videojuegos para fortalecer habilidades blandas y cognitivas de los infantes

Naisha Daniela Barradas Martínez¹, Emily Monserrath Zavala Mendoza¹, Flor Radilla López¹, Eduardo Vazquez Santacruz² y Elvia Morales Turrubiates¹

¹Universidad Autónoma del Carmen. Ciudad del Carmen Campeche, México
Facultad de Ciencias de la Información

152126@mail.unacar.mx, 141830@mail.unacar.mx
fradilla@pampano.unacar.mx, emorales@delfin.unacar.mx

²Universidad Autónoma Metropolitana, Ciudad de México, México
Ciencias Básicas e Ingeniería. Departamento de Ingeniería Eléctrica. Computación y Sistemas.
Ingeniería de Software
evazquez.santacruz@izt.uam.mx

Resumen. The boundaries of the videogame industry expand from the entertainment to become an educational tool. In recent years, videogames represent an educational potential as teaching and learning resource. Important contributions of videogames can cover each of the three main fields of psychology: the affective (awakening feelings); the connate (aggressive or impulsive behavior); and cognitive (learning-related skills). Abilities and skills like problem solving, critical thinking, logic memory, attention, creativity and perseverance can be acquired through tools based on video games technology. The present work proposes the development of a software prototype throughout a videogame to encourage perseverance and goal achievement. Our tool is currently a functional prototype that allows greater usability and playability than other similar games. In addition, since each game is a challenge to the user to achieve a goal avoiding obstacles and avoiding leaving the game, it represents an ideal scenario to develop skills such as reasoning, analytical thinking, complex problem solving, creativity, resilience, generation of ideas, among other. The software will be mainly aimed at children.

Keywords: Videogames, education, software, teaching, learning, software development. Unity.

1 Introducción

En la actualidad, gracias a los avances de la tecnología, el entretenimiento está presente en el día a día de una gran parte de la humanidad y, por lo tanto, llega a ser influyente en la vida de cada persona, lo que puede ser irrelevante o, todo lo contrario. Uno de los casos especiales en el que el entretenimiento llega a tener una gran responsabilidad es

el de los videojuegos, específicamente en los usuarios identificados como infantes. Si bien, no es estrictamente necesario, es muy importante que este grupo de usuarios obtenga enseñanzas importantes a través del entretenimiento y éste impacte de manera positiva en su vida [1]. En este contexto se desarrolló un prototipo de software que, a través de una dinámica divertida y lúdica de juego, se le plantea al usuario enseñanzas que le serán útiles a lo largo de la vida, en particular, la perseverancia, estrategia, dedicación, constancia y deseo por cumplir sus metas. Lo que la herramienta busca es incentivar al usuario infante a través de diversos retos representados con elementos gráficos animados que le resulten atractivos.

2 Antecedentes

Existen similitudes entre el proyecto realizado y otros videojuegos con el mismo enfoque, es decir transmite diversión, es colorido, la mecánica de juego similar, se juegan usando una computadora personal y lo más importante pone a prueba el objetivo que queremos transmitir “no darse por vencido, lograr llegar a la meta”. Un ejemplo es el videojuego “Happy Hop” [2], a la cual puede accederse mediante la tienda Microsoft Store. Esta herramienta consiste en un arcade con gráficos en dos dimensiones en la que el jugador tendrá el reto de llegar lo más alto posible saltando de plataforma en plataforma con sumo cuidado, un paso en falso y nuestro simpático protagonista caerá al vacío. Y dado que los monos son todos los personajes del juego, nadie quiere que se caigan. Otro ejemplo es el videojuego “Tap Tap Dash” [3], a la que puede accederse a través de una descarga en la tienda *Microsoft Store*. Ésta es un arcade frenético en el que, mediante una serie de laberintos muy estrechos, los jugadores tendrán que intentar llegar tan lejos como les sea posible. Un toque a destiempo en estos laberintos, y nuestro personaje se precipitará al vacío. A diferencia de nuestra propuesta, los videojuegos referidos requieren dispositivos móviles, manejo de comandos, instrucciones, tienen interfaces más austeras (se necesita más esfuerzo de los dedos), sólo usan el mouse, únicamente tienen 1 nivel y la oportunidad de juego para cada usuario es muy corta. Así, nuestra propuesta ofrece mejor usabilidad, mayor jugabilidad y busca estar al alcance de cualquier persona que tenga una computadora personal [2],[3].

3 Objetivo

Desarrollar un prototipo de software de apoyo pedagógico basado en tecnología de videojuegos de tipo plataforma. El contenido de este videojuego será interactivo para desarrollar destrezas psicomotoras tales como reflejos y coordinación óculo-manual, el videojuego es elaborado para usuarios infantiles. Además, motivará el desarrollo de habilidades blandas como la perseverancia y el logro de metas enfrentando retos de una forma divertida y lúdica. El software permitirá incrementar destrezas psicomotoras y habilidades blandas de los usuarios por medio de los tres niveles de juego que se proporcionan en la herramienta.

4 Metodología

La metodología usada es la siguiente: 1) Definición del problema, que consiste en que el personaje, una “versión animada de un gatito”.2) Especificaciones de los requerimientos. 3) Diseño definido con base en los requerimientos, utilizando programas de diseño y funciones de programación. 4) Implementar el funcionamiento diseñado para su fluidez y dinamismo. 5) Pruebas de verificación de la funcionalidad de juego.

4.1 Herramientas utilizadas

Para la creación del videojuego usamos 3 programas, dos de ellas son Unity y Visual Studio. Unity es el escenario donde colocamos todo lo que se creó aun sean bocetos, el diseño de los personajes que fueron creados en el programa de Photoshop, igualmente para que el personaje funcionara se necesitaba órdenes lo cual fue implementado con Visual Studio donde se programa para la funcionalidad del juego.

5 Avances

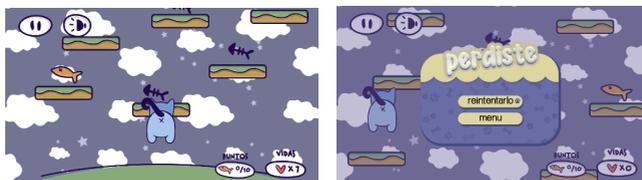


Fig. 1. Pantalla nivel 3 y pantalla de Game Over. Fuente: elaboración propia

Como se puede observar a través de la Fig. 1, nuestro trabajo se encuentra en un nivel de desarrollo tecnológico de prototipo funcional ya que se puede interactuar con la herramienta atendiendo las características básicas ya mencionadas. Se atendieron los siguientes requerimientos de manera fundamental dado el escenario de su uso potencial en un futuro cercano: que el videojuego sea ejecutado en una PC, que se desarrollará en 2D, el diseño de la interfaz deberá ser atractivo para el usuario, debe indicar instrucciones de cómo debe empezar a jugarse, el personaje principal tendrá 3 oportunidades (vidas) al inicio del juego y se ejecutará de manera local. Se hizo un trabajo de profundo detalle en la creación de diseños de los personajes y elementos tales como: corazón, pez, esqueleto, e iconos diversos. Además, se trabajó en la ejecución de movimientos del personaje en saltos y esquivo de obstáculos. Se trabajó además en el “Robo y recuperación de vidas” y se realizó investigación respecto al tipo de videojuego que se realizaría, diseño de menús, diseño de escenarios y colocación de

audio. Se atendió un conjunto de historias de usuario para el desarrollo de este videojuego referido en este trabajo. Estos requerimientos fueron establecidos a partir de una iterada retroalimentación con expertos usuarios de videojuegos diversos, la experiencia de los desarrolladores y profesores expertos en la temática de multimedia. Se tiene considerada la posibilidad de realizar ajustes a nuestro videojuego a partir de la retroalimentación de nuevos usuarios en las etapas de pruebas e implantación. Este software permite al usuario enfrentar retos y diseñar estrategias de solución para evitar perder la oportunidad de seguir jugando y lograr llegar a la meta indicada, en este sentido se ve fortalecida la experiencia educativa y pedagógica de fortalecer las habilidades blandas y cognitivas dado que cada escenario de juego es diferente y siempre hay nuevos objetivos que lograr. Cada vez que se usa el prototipo propuesta, el usuario tiene que esforzarse para diseñar su estrategia y lograr el reto que le propone cada oportunidad de juego lo que requiere habilidades metodológicas para construir soluciones, generar conocimiento a partir de las experiencias, autoaprendizaje, diseño de nuevas herramientas lógicas, resiliencia, pensamiento analítico, aprendizaje activo, capacidad de resolución de problemas, pensamiento y análisis crítico, creatividad, originalidad, iniciativa, pensamiento analítico, uso y control de la tecnología, diseño y programación tecnológica, tolerancia al estrés y flexibilidad, razonamiento, resolución de problemas y generación de ideas nuevas para enfrentar retos que son coherentes con las habilidades identificadas por el Foro Económico Mundial que serán demandadas por la sociedad para el año 2025 [5],[6],[7],[8].

6 Resultados esperados

Actualmente nuestro videojuego se encuentra en un nivel de madurez tecnológica que se puede considerar como prototipo funcional. Es un desarrollo que contiene todos los elementos básicos de un videojuego educativo sin embargo podemos incrementar los niveles de complejidad y mejorar el detalle gráfico y estético. Se pretende también aplicar este videojuego en áreas infantiles como guarderías, preescolares y primarias. En este sentido, el prototipo tendrá que ser analizado y mejorado considerando los comentarios de los usuarios finales como son profesores, padres de familia e incluso los mismos infantes. En general, nuestro software tiene las funciones básicas necesarias pero su mejora dependerá mucho de la retroalimentación de los actores del ambiente productivo en el que se puede implantar. Los elementos dentro de la escena del videojuego serán ampliados y rediseñados, para permitir al usuario elegir entre diferentes escenarios y personajes. La herramienta deberá ser analizada por los usuarios finales, que incluirá a los profesores, en la facilidad para la evaluación del progreso académico de los alumnos; y a los alumnos, en cuanto a la jugabilidad. Se prevé también un proceso de implementación de las mejoras posibles, identificadas durante el análisis.

7 Conclusiones

La realización de este software implicó conocimientos en programación y diseño gráfico y de videojuegos [1],[4]. Asimismo, dada su finalidad de brindar enseñanzas y

el impacto positivo que se busca que tenga en la vida de aquellos que la emplean, se considera una herramienta de gran importancia, ya sea sólo por su contenido en sí o cual sea su aportación como instrumento educativo. Se trabajó cuidadosa y enfáticamente en el diseño integral de la herramienta y desde una perspectiva de diseño del software. Se pretende realmente implantar este software en ambientes educativos en beneficio de los niños mexicanos primeramente y quizás de otros lugares del mundo. Actualmente atenderemos diversos comentarios de usuarios potenciales para realizar las mejoras y modificaciones correspondientes con el objetivo de tener versiones más robustas y completas.

Referencias

1. De Aguilera, Miguel & Mendiz Noguero, Alfonso. (2003). Video games and education: (Education in the face of a “parallel school”). *Computers in Entertainment*. 1. 10.
2. Happy Hop: Kawaii Jump <https://sites.google.com/platonicgames.com/platonicgames/happy-hop?authuser=0>
3. Tap Tap Dash <https://www.minijuegos.mx/juego/tap-tap-dash>
4. Widitiarsa, Arsa & Sn, M. (2018). Video Games as Tools for Education. 10.5281/zenodo.2669725
5. World Economic Forum. (2021). Competitiveness Report 2020-2021. Génova, Suiza
6. Pitt, Laura. El Secreto de uno de los mejores sistemas educativos del mundo. [En Línea]. Disponible en: http://www.bbc.co.uk/mundo/noticias/2013/06/130604_educacion_finlandia_lp.shtml; accesado el 30 de enero de 2014
7. Harvard T. Wagner. El Sistema Educativo Finlandés. [En Línea]. Disponible en: <http://redesoei.ning.com/video/el-sistema-educativo-finland-s-realizado-por-el-profesor-de>; accesado el 30 de enero de 2014
8. Muro M. Finlandia: las claves del sistema educativo más exitoso de Europa. [En Línea]. Disponible en: <http://www.libertaddigital.com/internacional/europa/2013-10-04/sabado-que-tiene-el-modelo-educativo-finlandes-que-no-tiene-el-espanol-1276501044/>; accesado el 30 de enero de 2014.

CAPÍTULO 5

*Identificación y Conteo de células Chlorella
sp. por medio de inteligencia artificial*

LUIS ALBERTO MIRANDA TORRES, HÉCTOR RODRÍGUEZ RANGEL,
VÍCTOR ALEJANDRO GONZÁLEZ HUITRÓN, ABRAHAM EFRAIM RODRÍGUEZ
MATA Y DULCE MARÍA ARIAS LIZARRAGA

Identificación y Conteo de células *Chlorella sp.* por medio de inteligencia artificial

Luis Alberto Miranda ¹, Héctor Rodríguez Rangel¹, Víctor Alejandro González Huitrón¹, and Abraham Efraím Rodríguez-Mata² y Dulce María Arias Lizárraga³

¹ Tecnológico Nacional de México/IT Culiacán, Juan de Dios Batiz No. 310 pte Culiacán, Sinaloa hector.rr@culiacan.tecnm.mx

² Tecnológico Nacional de México/IT de Chihuahua, Av. Tecnológico 2909, Tecnológico, 31200 Chihuahua, Chihuahua, México

³ Instituto de Energías Renovables, Universidad Nacional Autónoma de México, Priv. Xochicalco s/n, Col. Centro, Temixco, Morelos

Resumen Artificial intelligence has had an impressive impact on industrial processes, but this time we focus especially on the production of *chlorella vulgaris*, which has many applications in the food and pharmaceutical industry. For this reason it is essential to experiment with these cells. In this work we propose the development of a software capable of speeding up and optimizing the results, through artificial intelligence and its pattern detection methods based on autoencoders to count the cells present in a certain number of samples, achieving an identification process of more than 95% efficiency in prediction and identification.

Keywords: Autoencoder · Pattern detection · Microalgae Detection.

1. Introducción

En la actualidad la tecnología tiene un gran impacto en el desarrollo del área de la biología, ayudando a estudiar los seres vivos mas de cerca, creando instrumentos y herramientas que facilitan la obtención de resultados más eficaces.

El conteo de microalgas se usa para calcular la cantidad de biomasa, de forma manual, usando una cámara de Neubauer y el juicio de un experto. Actualmente la microalga tiene muchos beneficios para los humanos, algunos de ellos son propiedades en la alimentación, tratamiento de aguas residuales, producción de biocombustibles y farmacología de alto nivel, por lo que su estudio, producción a gran escala, así como también la investigación de su optimización sigue abierta [6].

El desarrollo de este proyecto tiene el fin de automatizar las tareas de conteo de células para la optimizar tiempos. Para ello se desarrolló un software capaz de detectar células y contabilizarlas, utilizando técnicas como la transformada de Hough para la detección de rectas y circunferencias, detección de bordes con Canny, así como también la inteligencia artificial usando los autoencoders. Los

cuales son capaces de aprovechar la detección de patrones para reconocer si los datos que se analizan son o no son el objeto que se necesita detectar.

Los autoencoders son un tipo de inteligencia artificial, la cual actúa como un filtro que aprende a reconstruir la imagen que se quiere clasificar, para poder obtener un porcentaje de semejanza con el objeto a detectar, aprovechando su capacidad de almacenar características únicas del mismo durante el entrenamiento.

2. Antecedentes

Anteriormente, diversos equipos han realizado investigaciones acerca de estas metodologías de trabajo, para poder ayudar a hacer más efectivas las actividades en sus áreas. Uno de estos trabajos es el “Sistema de conteo de colonias de células tumorales” [5], quienes por medio de la detección de bordes con Camy, lograron detectar y contabilizar colonias de células tumorales.

3. Objetivos

- Objetivo general
 1. Diseñar, desarrollar e implementar una herramienta capaz de agilizar la detección y conteo de células en imágenes tomadas a muestras biológicas de células *Chlorella* sp. por medio de inteligencia artificial.
- Objetivos específicos
 1. Diseñar y desarrollar un algoritmo para la detección de la región de interés por medio de procesamiento de imágenes.
 2. Diseñar y crear interfaz gráfica para la mejor manipulación del usuario.
 3. Modelar sistema de inteligencia artificial para detección automática y conteo de las células *Chlorella* sp.
 4. Lograr significativamente mejorar la rapidez en el conteo de células.
 5. Obtener una mayor efectividad en el conteo de células.

4. Metodología

El proyecto se desarrolla en dos lenguajes de programación, Python, con el que entrenamos nuestro auto-codificador e hicimos la clasificación de verdaderas células positivas, y MATLAB con el cual se hace la detección de posibles células dentro de las imágenes, así como también el desarrollo de la interfaz de usuario.

4.1. Transformada de Hough[2]

La transformada Hough es un algoritmo que se utiliza en la detección de patrones de imágenes, con el cual es posible detectar formas como círculos, líneas, entre otras.

Detección de líneas La detección de líneas rectas es representada por los parámetros ρ y θ , donde ρ es la distancia entre la línea y el origen, y θ es el ángulo del vector desde el origen al punto más cercano. Por lo tanto el punto se representa como una función sinusoidal. En caso de tener dos puntos, estos se representan por dos sinusoidales que se desfazarán por alfa grados de acuerdo a las coordenadas de los puntos.

$$\rho = x \cos \theta + y \sin \theta \quad (1)$$

Detección de circunferencias[4] Para encontrar una circunferencia son necesarios tres parámetros: los ejes del punto central de la circunferencia (a,b) y el radio (r).

$$(x - a)^2 + (y - b)^2 = r^2 \quad (2)$$

El algoritmo usa un acumulador de 3 dimensiones (a, b, r) que guarda los valores de cada variable, después los puntos se agrupan con cada una de las circunferencias con las que podría encajar. Los valores más altos en el acumulador determinarán el valor del radio y el centro de la circunferencia.

4.2. Detección de bordes por el método de Canny[3]

El algoritmo de Canny es un método que utiliza el cálculo de variaciones que encuentra la función óptima. La función esta compuesta por cuatro términos exponenciales, que también a su vez es posible acercarse por medio de la primer derivada de una matriz gaussiana.

El algoritmo es óptimo si éste cumple con los siguientes objetivos:

- Buena detección: se debe marcar la mayor cantidad real posible de bordes en la imagen.
- Buena localización: los bordes marcados deben estar lo más cerca al borde de la imagen real.
- Respuesta mínima: El borde de los objetos solo deben marcarse una vez, y el ruido de la imagen no debe afectar creando bordes falsos.

Reducción de ruido El algoritmo de Canny utiliza un filtro mediante la primera derivada de una matriz gaussiana. Dado que es sensible al ruido en los datos de la imagen original, se utiliza un filtro gaussiano para transformar la imagen original. Comparando la versión original con el resultado, ésta es una imagen ligeramente borrosa.

$$B = \frac{1}{159} \left(\begin{bmatrix} 2 & 4 & 8 & 2 & 3 \\ 4 & 9 & 7 & 7 & 9 \\ 5 & 5 & 6 & 5 & 3 \\ 4 & 1 & 6 & 1 & 7 \\ 2 & 3 & 8 & 4 & 7 \end{bmatrix} \right) * A \quad (3)$$

Encontrar la intensidad del gradiente de la imagen El algoritmo de Canny hace uso de cuatro filtros para detectar de manera horizontal, vertical y diagonal en los bordes de la imagen borrosa. La función de detección de bordes devuelve un valor para la primera derivada en la dirección horizontal (G_y) y la dirección vertical (G_x).

$$G = \sqrt{G_x^2 + G_y^2} \quad (4)$$

$$\theta = \arctan \frac{G_y}{G_x} \quad (5)$$

5. Resultados preliminares

El proyecto consta de tres partes principales, detección de células, desarrollo y entrenamiento del auto-codificador que procesará las imágenes para determinar si estas son o no una célula, y el desarrollo de la interfaz de usuario.

5.1. Segmentación de imagen y detección de posibles células

El primer paso fue recortar y delimitar la imagen a nuestra área de interés, utilizando los métodos de procesamiento de imágenes como Canny[3], así como también la t. de Hough [2].

Generamos 4 máscaras binarias, a partir de la detección de las líneas al límite del área de interés. Las máscaras binarias se sumarán, creando una sola que multiplicaremos por la imagen original, dando como resultado una imagen con el área de interés recortada. Posteriormente aplicamos el método de Canny y la t. de Hough para circunferencias[4], para detectar posibles células y recortar pequeñas imágenes de estas. Y finalmente pasarlas a nuestra red neuronal para así determinar si es una célula o no.

5.2. Desarrollo y entrenamiento del auto-codificador

El desarrollo de nuestro auto-codificador se basó en uno hecho por Keras [1], modificando y ajustando los parámetros de retención de información, número de épocas de entrenamiento, etc. Con la ayuda del experto, obtuvimos las características de una célula para entrenar el auto-codificador. En el entrenamiento se usó una base de datos de imágenes verdaderas positivas las cuales nos ayudaron a realizar el entrenamiento, y de verdaderas negativas nos sirvieron para realizar pruebas de la red. La red neuronal pudo reconocer el 100% de imágenes negativas como verdaderas negativas y mas de 95% de las imágenes positivas como verdaderas positivas, concluyendo que la red es altamente efectiva en la detección. Actualmente seguimos desarrollando y diseñando la interfaz de usuario en AppDesigner de MATLAB en conjunto con el experto quien nos dará la pauta del manejo y el diseño de la misma.

6. Conclusiones

El desarrollo de esta plataforma hace posible la automatización y optimización, ya que no es necesario que se pierda demasiado tiempo contando y/o verificando si existen células o no.

Referencias

1. Chollet, F. (2016). Building Autoencoders in Keras. The Keras Blog, 1–14. <https://blog.keras.io/building-autoencoders-in-keras.html>
2. Fokkinga, M. (2011). The Hough transform. Journal of Functional Programming. <https://doi.org/10.1017/S0956796810000341>
3. Canny, J. (1986). A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. <https://doi.org/10.1109/TPAMI.1986.4767851>
4. Find circles using circular Hough transform - MATLAB imfindcircles. (n.d.). Retrieved January 25, 2021 <https://www.mathworks.com/help/images/ref/imfindcircles.html>
5. Ramos, L. R., Fernanda, M., Castro, R., Fajardo, R., Forero, K. C., Ondo Méndez, A., Manuel López López, J., & Forero, M. G. (n.d.). SISTEMA DE CONTEO AUTOMÁTICO DE COLONIAS DE CÉLULAS TUMORALES <https://antiguo.acofipapers.org/index.php/eiei2018/2018/paper/viewFile/2845/1083>
6. Shah, M., Mahfuzur, R., Liang, Y., Cheng, J. J., & Daroch, M. (2016). Astaxanthin-producing green microalga *Haematococcus pluvialis*: from single cell to high value commercial products. *Frontiers in plant science*, 7, 531. <https://www.redalyc.org/jatsRepo/5122/512253717006/html/index.html>
7. Schulze, K., Tillich, U. M., Dandekar, T., & Frohme, M. (2013). PlanktoVision-an automated analysis system for the identification of phytoplankton. *BMC bioinformatics*, 14(1), 1-10.
8. Shukla, A. K., Janmajaya, M., Abraham, A., & Muhuri, P. K. (2019). Engineering applications of artificial intelligence: A bibliometric analysis of 30 years (1988–2018). *Engineering Applications of Artificial Intelligence*, 85, 517-532.

CAPÍTULO 6

Sistema de Recolección y Procesamiento de Datos para Estudios de Tránsito

HÉCTOR RODRÍGUEZ RANGEL, RAFAEL IMPERIAL ROJO, LUIS ALBERTO MORALES ROSALES, SOFIA ISABEL FERNÁNDEZ GREGORIO Y ABRAHAM EFRAIM RODRÍGUEZ MATA

Sistema de Recolección y Procesamiento de Datos para Estudios de Tránsito

Héctor Rodríguez Rangel¹, Rafael Imperial Rojo¹ Luis Alberto Morales Rosales², Sofía Isabel Fernández Gregorio³ y Abraham Efraím Rodríguez Mata⁴

¹Tecnológico Nacional de México, Campus Culiacán, hector.rr@culiacan.tecnm.mx

²Conacyt-Universidad Michoacana de San Nicolás de Hidalgo

³Instituto Tecnológico Superior de Martínez de la Torre

⁴Tecnológico Nacional de México, Campus Chihuahua

Resumen La movilidad urbana se incremento con el uso de automóviles, lo que origino también un aumento de accidentes de tránsito. Para estudiar este fenómeno se requiere de estudios (ET) de tránsito. Con el uso de inteligencia artificial (IA) es posible realizarlos sin modificar en gran medida la infraestructura urbana. Para el diseño de soluciones basadas en IA es necesario generar bases de datos que proporcionen información confiable para la calibración y desarrollo de soluciones que permitan realizar ET de manera automatizada. En este trabajo se presenta un sistema para generar un conjunto de datos a partir de vídeos tomados en un punto de observación.

Keywords: Base de Datos · Velocidad vehicular · Visión Artificial.

1. Introducción

El uso de automóviles es esencial en nuestra vida cotidiana. Según la organización Association for Safe International Road Travel (ASIRT), cada año mueren más de 1,3 millones de personas en accidentes de tráfico, y entre 20 y 50 millones más resultan heridos o incapacitados ([5]). En los accidentes viales, el exceso de velocidad es el factor más mortífero por lo que los estudios de tránsito proporcionan el principal suministro de información para detectar las posibles causas. El enfoque tradicional para realizar estudios de tránsito está influenciado por el uso de equipo especial. Uno de los sensores que no requiere modificaciones en la infraestructura de una ciudad son las cámaras de videovigilancia proporcionando información en tiempo real. Al analizar los datos provenientes de cámaras es posible estimar atributos del flujo de tráfico, detectar congestión, accidentes y observar el comportamiento de los conductores entre otros factores ([4]). La inteligencia artificial (IA) está revolucionando la sociedad moderna. Los investigadores y desarrolladores tanto de la industria automotriz como de seguridad vial están desarrollando activamente enfoques de conducción autónoma y monitoreo basados en el aprendizaje profundo. Sin embargo, antes de que la red neuronal llegue al vehículo de producción, primero requiere una evaluación de seguridad

funcional rigurosa y la construcción de base de datos para poder entrenar dichas redes neuronales. En [1] se describe las posibilidades y desafíos de integrar el aprendizaje profundo en vehículos autónomos y se estudian la creación de base de datos para el entrenamiento de dichas redes neuronales.

En este trabajo se presenta un sistema para generar un conjunto de datos a partir de vídeos tomados en un punto de observación de una vía carretera. La principal característica de estos vídeos es que no son tomados paralelamente a la vista de la cámara. Con la creación de este sistema podemos ayudar a desarrollar diferentes conjuntos de datos con el propósito de inferir la velocidad de objetos y extraer las características que el usuario considere como las más importantes. Por otro lado, el desarrollo de soluciones para la generación automática de estudios de tránsito requieren de bases de datos que recopilen diferentes condiciones para observar comportamientos y calibrar sus resultados. En el caso del estudio del flujo vehicular mexicano, no existen bases de datos públicas confiables que ayuden a esta labor.

2. Metodología

Este trabajo está enfocado en la generación de la información del flujo vehicular de una vialidad a partir de secuencias de vídeo. Al investigar, no se encontró con un dataset (conjunto de datos) acorde a las necesidades del proyecto, por lo cual se decidió crear uno que se adapte. La metodología para la extracción de estas características se describe en el diagrama de flujo de la Figura 1 la cual cuenta con 4 etapas: 1) Toma de muestras, 2) Limpieza de muestras, 3) Extracción de información, 4) Validación de la información, estas se describen a continuación:



Figura 1. Diagrama de flujo metodología implementada.

- *Toma de muestras.* Se requirieron dos dispositivos para extraer el conjunto de datos. Los dispositivos son el radar Bushnell con una precisión de +/- 1.6 kilómetros por hora y una cámara de vídeo, para este caso se utilizó la cámara de teléfono inteligente configurados a 60 FPS y una resolución de Full HD (1920 x 1080 píxeles). Se posiciona la cámara de forma que los vehículos no pasen paralelamente a la vista de la cámara para que el cálculo no sea considerado como tomado en un plano 2D.
- *Limpieza de las muestras.* El radar no está integrado con el sistema de seguimiento. Se realiza la unión manual de la velocidad proporcionada por el

radar y la información obtenida. Se requiere una inspección visual de las muestras tomadas, lo primero es identificar los puntos en x que nos servirán como referencia para tomarlos como entrada y salida de los vehículos. Usando como referencia nuestro punto de salida, identificaremos los vehículos a los que se les tomo la velocidad y se anota el segundo en que pasa. Para esto se utiliza un archivo CSV, el cual cuenta con 2 campos:

- El segundo en que el objeto pasa por el punto de salida.
 - La velocidad del objeto.
- *Extracción de información.* Para la generación del conjunto de datos se extraen las características principales del vídeo. Para el sistema presentado estas características son las siguientes:
- Ángulo de salida: A partir del punto de entrada hasta el punto de salida.
 - Distancia de salida: Distancia del punto de entrada hasta el de salida.
 - Área de entrada: Área del vehículo en píxeles en el punto de entrada.
 - Área de salida: Área del vehículo en píxeles en el punto de salida.
 - FPS: Fotogramas por segundo del vídeo.
 - Tiempo: Tiempo del punto entrada al de salida.
 - Velocidad: Velocidad detectada por el radar.
 - Identificador: Correspondiente a una imagen generada.

El proceso completo realizado por el sistema, comienza ejecutándolo dando como entrada el archivo CSV con las velocidades y los límites configurados previamente. Se decodifica el vídeo utilizando la biblioteca OpenCV la cual examina fotograma por fotograma, e identifica los vehículos con la red neuronal YOLO ([2]) esta se encarga de reconocer múltiples objetos en una sola predicción. Detectados todos los objetos dibuja la caja correspondiente a cada uno de ellos. El seguimiento se realiza por medio del filtro Kalman ([3]) el cual determina su ubicación en el próximo fotograma. El sistema se encarga de guardar todas las ubicaciones de los vehículos en el transcurso del tiempo, con lo cual se traza todo el trayecto que han tenido cada uno de ellos, así como calcular la recta que corresponde a cada trayectoria. La Figura 2 muestra un vehículo detectado en un recuadro blanco, y un par de líneas amarilla y roja que corresponden al seguimiento y a la recta calculada a partir del seguimiento.



Figura 2. Vehículo identificado con su respectivo seguimiento.

Cuando se detecta que un vehículo pasa por el primer limite se guarda el fotograma para después unirlo con el fotograma pasando por el punto de

salida, cabe mencionar que esto se realiza para todos los vehículos detectados. Sin embargo, solo se guardan aquellos que correspondan al segundo de salida que se identificó en el archivo CSV. Una vez se detecta que un vehículo sale del límite, se guardan los datos y la velocidad en una línea del archivo CSV, al mismo tiempo se guarda una imagen con el vehículo entrando y saliendo por los límites, y un identificador para la imagen.

- *Limpeza de los datos.* El sistema genera una imagen por cada línea del CSV resultante, esta imagen es la combinación de 2 imágenes, cuando el vehículo pasa por el punto de entrada y cuando pasa por el punto de salida, con ayuda de esta imagen validamos que el vehículo haya sido tomado correctamente. Una imagen valida es aquella que no se aleja demasiado del punto de salida y el vehículo se detecta completo o gran parte del mismo. Para las imágenes invalidas tenemos un vehículo incompleto. La Figura 3 muestra un ejemplo de un vehículo capturado parcialmente.



Figura 3. Detección invalida del sistema.

3. Avances

Hasta el momento de redactar este artículo contamos con un total de 97 videos de entre 1 y 2 minutos de duración, de los cuales solo 51 videos han sido procesados. Para finalmente obtener 408 muestras. Ejemplo del resultado obtenido es la Tabla 1 que muestra 3 muestras obtenidas durante el proceso de extracción de características.

Tabla 1: Ejemplo de muestras resultantes.

Angulo Salida	Distancia Salida	Area Entrada	Area Salida	FPS	Tiempo	Velocidad	Identificador
78.65	711	26335	145475	60.01	0.866	37	117
82.71	693	23421	122661	60.01	0.866	32	358
78.97	700	55815	350730	60.01	0.633	45	407

La Figura 4 muestra en un recuadro blanco el vehículo que el sistema esta realizando el seguimiento para generar su información, mientras que en amarillo el resto de los vehículos detectados a los que no se les generara una línea para el CSV resultante.



Figura 4. Imagen generada por el sistema.

4. Conclusiones

El sistema es capaz de generar un conjunto de datos de calidad, con este podemos realizar experimentos con diferentes métodos ya sean estadísticos o de inteligencia artificial para predecir la velocidad de objetos que aparecen en una secuencia de vídeo. Además de darnos la libertad de integrar sensores para generar más parámetros que pueden ser de valor para inferir la velocidad de objetos. Por otra parte, el conjunto de datos generado requiere la eliminación de datos que puedan afectar las predicciones, lo cual implica un trabajo extra por parte de quien genera el conjunto de datos, además de tomar una suma considerable de vídeos para generar un conjunto de datos aceptable. Una actividad de mejora importante para el sistema es eliminar la necesidad de una persona que genere el archivo CSV con las velocidades y también quien valida que la información generada es correcta, esto ayudara a reducir el tiempo necesario para generar el conjunto de datos.

Referencias

1. Rao, Q., Frtunikj, J.: Deep learning for self-driving cars: Chances and challenges. In: Proceedings of the 1st International Workshop on Software Engineering for AI in Autonomous Systems. pp. 35–38 (2018)
2. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 779–788 (2016). <https://doi.org/10.1109/CVPR.2016.91>
3. Welch, G., Bishop, G., et al.: An introduction to the kalman filter (1995)
4. Yang, L., Li, M., Song, X., Xiong, Z., Hou, C., Qu, B.: Vehicle speed measurement based on binocular stereovision system. *IEEE Access* **7**, 106628–106641 (2019). <https://doi.org/10.1109/ACCESS.2019.2932120>
5. Zaki, P.S., William, M.M., Soliman, B.K., Alexsan, K.G., Khalil, K., El-Moursy, M.: Traffic signs detection and recognition system using deep learning. *arXiv preprint arXiv:2003.03256* (2020)

CAPÍTULO 7

*Implementación de un Servicio REST-API
para el uso de la firma electrónica para
instituciones públicas*

JESUS ALFREDO BRAVO MÉNDEZ, ANASTACIO ANTOLINO HERNÁNDEZ
Y HEBERTO FERREIRA MEDINA

Implementación de un Servicio REST-API para el Uso de la Firma Electrónica en Instituciones Públicas

Jesús Alfredo Bravo Méndez¹, Anastacio Antolino Hernández¹, and Heberto Ferreira Medina¹

TecNM/Instituto Tecnológico de Morelia, Michoacán 58120, México {1m19123005, anastacio.ah, heberto.fm}@morelia.tecnm.mx

Resumen En la Fiscalía General del Estado de Michoacán (FGE) surge la necesidad de implementar y de brindar una mayor seguridad en la documentación emitida, permitiendo la implementación de la Firma Electrónica Avanzada, donde los emisores firmaran los datos correspondientes al documento, demostrando su autenticidad, confiabilidad y no repudio. Se ha realizado la investigación necesaria para definir las tecnologías que se usarán y lograr generar las firmas electrónicas, permitiendo la gestión de los certificados, y la conexión con otros sistemas para emitir y verificar las firmas. Se implementó un servicio en una aplicación web de gestión administrativa para fungir como una autoridad certificadora siguiendo el protocolo x509 versión 3 y permitiendo conexiones a través de una arquitectura RESTful, para que los mecanismos tradicionales y digitales puedan hacer uso de éste y realizar un cambio disruptivo en el ambiente informático de la institución y sus sistemas ya implementados en producción.

Keywords: CLQ · PKI · x509 · RSA · Firma Electrónica.

1. Introducción

En las instituciones gubernamentales se emite documentación y se les provee de seguridad a través de mecanismos físicos, los cuales pueden ser una firma, un sello o un troquelado, sin embargo, estos mecanismos no son suficientemente seguros, llevando a las instituciones a padecer de falsificación a través de la réplica de estos mecanismos. Es por ello que en la Fiscalía General del Estado de Michoacán (FGE) se propuso otorgar a los documentos una firma electrónica que permita darles mayor seguridad y a su información, requiriendo el uso de un estándar basado en Public Key Infrastructure (PKI), con un algoritmo de cifrado que sea resistente a los ataques modernos y un mecanismo para permitir la interacción con otros sistemas de la institución. Con la finalidad de establecer estos mecanismos la propuesta implementa el uso de la firma electrónica a través de un servicio REST-API, que permite una fácil integración y extiende su uso a otras aplicaciones. Contando con un sitio, gestionando la información correspondiente a las personas que emitirán firmas electrónicas, basada en el estándar

x509 y las especificaciones de la Ley de Firma Electrónica Avanzada en México (LFEAM).

2. Antecedentes

La necesidad de implementar mecanismos que protejan la información para la sociedad fue primera prioridad. El gobierno de Guanajuato logró la implementación de la firma digital en diferentes servicios públicos, demostrando buenos resultados. El gobierno federal observó dicha práctica y con interés comenzó a impulsar la implementación de la firma digital en el 2004, reconociendo las ventajas y extendiéndolos al alcance de la sociedad. La firma digital identifica al emisor de un correo electrónico como autor legítimo en el proyecto “Tu Firma”, a cargo del Sistema de Administración Tributaria (SAT) y usándose como base para la implementación de la Firma Electrónica Avanzada (FIA) [4], definiendo a la tecnología PKI para la criptografía de claves, la firma electrónica y el uso de certificados electrónicos [1], así mismo el método de cifrado asimétrico RSA que hace uso del problema no computable de búsqueda de factores primos [6], el estándar X.509 permite el nivel de autenticación necesario para las claves emitidas, siendo que las contraseñas reforzarán la seguridad de su ejecución [5].

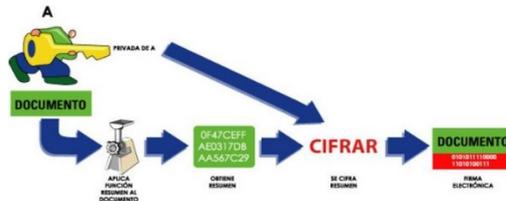


Figura 1: Procedimiento de firma electrónica realizado por el SAT [1].

3. Objetivos

Diseñar y desarrollar un software REST-API para el uso de la firma electrónica en documentos emitidos por instituciones gubernamentales.

- Fungir como una autoridad certificadora al uso interno de la institución.
- Gestionar los mecanismos para permitir la autenticación, verificación y firma de información con el estándar x509.
- Generar, verificar y validar firmas electrónicas para cualquier documento.
- Ofrecer seguridad en los procesos de emisión de las firmas por parte de la institución.

4. Metodología

- *Investigación de Antecedentes, Mecanismos y Requisitos.* Se realizó la investigación de los requisitos de las leyes para la implementación de la Firma Electrónica, requiriendo la gestión de los certificados, así como ser la fuente que permita emitir y verificar las firmas emitidas [3], el uso del estándar X.509, el uso de RSA para definir un par de claves que permita cifrar o descifrar datos a través de una contraseña [2], la implementación de un mecanismo que permita a otras aplicaciones comunicarse con el servicio, este es requerimiento de la FGE, por lo que se identificaron tres servicios: SOAP, REST y GRAPHQL, usando diferentes estructuras de archivos para interactuar, como XML para SOAP, HTTP para REST y GRAPHQL [5].
- *Análisis de las Tecnologías Para la Implementación.* Entre los requerimientos de la FGE se solicitaba usar el lenguaje PHP, sin embargo, se comparó con otros lenguajes como Python y Ruby, y se observaron distintos frameworks que ya cuentan con mecanismos contra ataques, a diferencia de las presentes vulnerabilidades de PHP, se definió a Python como lenguaje de programación junto a su framework Django. Comparando los tipos de servicios para la comunicación se observó a SOAP, a REST y a GRAPHQL donde ambos usan el protocolo HTTP y pueden enmascarse con SSL/TLS, requiriendo menos datos para su funcionamiento que SOAP. Como arquitectura final se seleccionó a RESTFUL por la compatibilidad con Django-rest-framework. La librería Cryptography fue seleccionada por el uso del estándar X.509 y el cifrado RSA, logrando con éste establecer que el servicio sea una Autoridad Certificadora (AC) y pueda gestionar los certificados de los clientes.
- *Métodos de Cifrado.* Se evaluaron los métodos simétricos y asimétricos. De los métodos asimétricos RSA y Diffie-Hellman son los más importantes, donde RSA se basa en el problema de encontrar la factorización de dos números primos grandes de forma aleatoria. El protocolo Diffie-Hellman es un sistema de intercambio de claves a través de un canal inseguro y sin alguna autenticación, siendo menos costoso de procesar que un cifrado asimétrico [3]. La Ley de Firma Electrónica indica al algoritmo RSA como válido, en comparación con Diffie-Hellman, cumpliendo con el requisito de ser asimétrico.
- *Desarrollo e Implementación.* Para la implementación se usó el servidor web Apache para desplegar la aplicación, creando un entorno virtual para el manejo de librerías. Se usa a Django v3.0.10, Bootstrap 4, Pyodbc para el manejo de la conexión al motor MS SQL Server, la librería Pyzipper para la compresión de archivos. Para la generación de certificados y llaves se usó la librería Cryptography, basado en el estándar X.509 que usa el método RSA. Para integrar la arquitectura RESTful se usó la librería Django-rest-framework logrando la creación de rutas API y usando el protocolo HTTP para la generación y verificación de las firmas. Se definieron los módulos que integran la aplicación, con un inicio de sesión, un gestor de la AC generando el certificado raíz, la gestión de certificados generados para los usuarios solicitando información de la persona y emitiendo su par de llaves para cifrar. La visualización de firmas generadas, el listado de revocación y la bitácora de

movimientos que sucedan en el uso del servicio. La API permite el consumo por los otros sistemas, logrando generar firmas electrónicas y entregándolas en formato JSON.

- **Pruebas y Resultados.** El desarrollo de la aplicación integra todos los mecanismos y requisitos solicitados por las leyes y la FGE. Se realizaron pruebas de calidad para verificar su funcionamiento con certificados emitidos por la AC y para los usuarios. A través de la aplicación Postman se probaron de las rutas API, generando firmas electrónicas a partir de los datos brindados con longitudes de 4096 bytes y 2048 bytes, Fig. 2a.

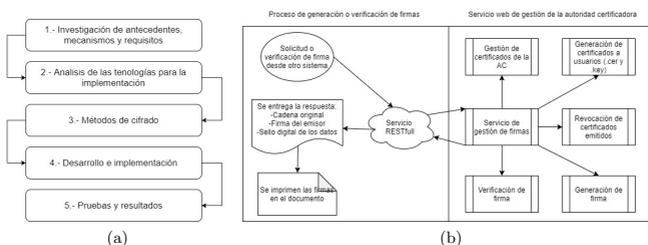


Figura 2: Diagrama de metodología, y Diagrama del servicio y el servicio RESTful.

5. Avances

Los requerimientos analizados para la implementación del servicio, cumple con la Ley de Firma Electrónica, permitiendo el manejo de los certificados y firmas emiti-das, y la posibilidad de contactarse con otros sistemas, almacenando la información de los certificados y firmas generadas y gestionando los certificados raíz para la AC, así como la lista de revocación y la bitácora de la actividad. El servicio está compuesto por dos partes. 1) La interfaz web, la cual hace gestión de certificados para la AC y los usuarios, la visualización de información correspondiente a las firmas, el listado de certificados revocados y la bitácora de las actividades de la aplicación web. 2) El servicio de arquitectura RESTful, Fig. 2b, el cual permite a otras aplicaciones solicitar y verificar firmas, verificando la vigencia del certificado usado y si éste no se encuentra en la lista de revocación. Se han realizado pruebas de calidad en un entorno local comprobando su funcionalidad, a la espera de la autorización de implementarlo para las pruebas de calidad en un entorno productivo, con el servicio de gestión documental de la FGE.

6. Resultados Esperados

Con la implementación del servicio y el fácil acceso para otros sistemas, estampar una firma electrónica en un documento dará seguridad a la institución y a los usuarios de los servicios. Permitirá colocar a los mecanismos usados como elementos secundarios y opcionales, otorgando certeza y confianza a la información contenida en los documentos, permitiendo nuevas formas de emisión totalmente electrónica, identificando al documento como único y asegurando que los datos sean firmados por el emisor.

7. Conclusiones

RSA como método de cifrado brinda seguridad a partir de claves de 2048 bytes. La posibilidad de conexión se extiende, permitiendo hacer uso en otras aplicaciones y retirando la carga de la generación de las firmas, mencionando la importancia de la escalabilidad y compatibilidad que permite el framework Django y el lenguaje de programación Python al ser de los lenguajes más seguros. El proyecto es el inicio de un cambio disruptivo para la institución, asegurando que la información es confiable, íntegra y no repudiable, permitiendo extender sus bondades al uso de mecanismos de consumo REST a sistemas externos si así se desea y como Autoridad Certificadora pueda representar validez ante cualquier entidad de gobierno. Las emisiones de documentos con firma electrónica mejorarán los documentos y brindarán un mayor índice de seguridad. Agradecemos al TecNM por el apoyo al proyecto número 10614.21-P.

Referencias

1. Firma de documentos electrónicos, http://m.sat.gob.mx/fichas_tematicas/fiel/Paginas/firma_de_documentos_electronicos.aspx
2. Pki – definición y despliegue. Instituto Nacional de Administración pública
3. Héctor Corrales Sánchez, Carlos Cilleruelo Rodríguez, A.C.N.: Criptografía y métodos de cifrado. clasespersonales.com <http://clasespersonales.com/sisdis/paratranscribir1.pdf>
4. Izquierdo, E.L.: La implementación de la firma electrónica en México. *Economía Informa* pp. 97–103 (8 2011), <http://www.economia.unam.mx/publicaciones/econinforma/369/08leonizquierdo.pdf>
5. Jaime P. Sayago H., Evelin L. Flores C., A.R.: Análisis comparativo entre los estándares orientados a servicios web soap, rest y graphql. *Pontificia Universidad Católica del Ecuador* pp. 10–22 (2019), <https://zenodo.org/record/3592004/files/n17a2.pdf>
6. Ken Beer, R.H.: Seguridad de datos en reposo con cifrado. *Amazon Web Services* pp. 1–16 (11 2013), https://d1.awsstatic.com/whitepapers/es_ES/AWS_Securing_Data_at_Rest_with_Encryption.pdf

CAPÍTULO 8

Metodología para obtención de Zonas Homogéneas en pavimentos flexibles

EDUARDO DANIEL RAYA GAMIÑO, JORGE ALARCÓN IBARRA, LUIS ALBERTO MORALES ROSALES, JAIME SAAVEDRA ROSALES Y JUAN CARLOS GALLEGOS CORNEJO

Metodología para la Obtención de Zonas Homogéneas en Pavimentos Flexibles

Eduardo Daniel Raya Gamiño¹, Jorge Alarcón Ibarra¹, Luis Alberto Morales Rosales², Jaime Saavedra Rosales¹, Juan Carlos Gallegos¹

¹ Universidad Michoacana de San Nicolás de Hidalgo

² Conacyt-Universidad Michoacana de San Nicolás de Hidalgo

Abstract. Las zonas homogéneas son espacios de la superficie terrestre que presentan condiciones, características o cualidades similares. En el campo de la ingeniería civil, la determinación de zonas homogéneas es aplicable en el área de la conservación de carreteras, permitiendo delimitar tramos que presenten similitudes respecto a un factor de desempeño obtenido de su auscultación. El método estadístico de las diferencias acumuladas en conjunto con la experiencia de las personas encargadas de la programación de trabajos de conservación de una carretera, son los principales elementos utilizados para delimitar las zonas homogéneas, obteniéndose resultados parciales y confusos. El método de diferencias acumuladas, sólo puede seccionar el tramo en estudio mediante un factor de desempeño por lo que el seccionamiento no refleja una completa evaluación del estado del pavimento. La Secretaría de Comunicaciones y Transportes (SCT) en México, sugiere la auscultación de un pavimento con varios factores de desempeño identificando así diferentes condiciones de la estructura y el desempeño del pavimento. Por esta razón, este artículo presenta una metodología integral para la obtención de zonas homogéneas de tramos carreteros basada en la evaluación de la energía de los factores de desempeño propuestos por la SCT (Deflexiones, Índice de Regularidad Internacional, Profundidad de roderas, Coeficiente de fricción y Macrotextura). La principal aportación radica en fortalecer las características de juicio durante el proceso de seccionamiento al emplear de manera conjunta y jerárquica los factores de desempeño para representar una mejor descripción de los deterioros presentados por zonas homogéneas.

Keywords: zonas homogéneas · factores de desempeño · transformadas Wavelet.

1 Introducción

Es importante tener presente que la estructura que conforma a un pavimento flexible, con el paso del tiempo sufrirá daño y deterioros, aun cuando su diseño y construcción hayan sido acorde a las normas de calidad y especificaciones requeridas; razón por la cual es importante dar seguimiento a la infraestructura vial, realizando intervenciones de mantenimiento y conservación, seleccionando

el tratamiento adecuado y el tiempo preciso en que debería aplicarse[1], extendiendo así la vida útil del pavimento.

Los tramos que resultan de seccionar una carretera de acuerdo a características o condiciones del estado en que se encuentran, son denominados “Zonas Homogéneas” y resulta importante recalcar que en muchas ocasiones, el éxito de la adecuada implementación y ejecución de los trabajos de conservación de una carretera, dependen en gran medida de la determinación adecuada de los tramos homogéneos.

En este artículo se presenta una metodología integral para la obtención de zonas homogéneas de tramos carreteros, permitiendo así fortalecer las características de juicio durante el proceso de seccionamiento, incluyendo una evaluación de la energía de los factores de desempeño propuestos por la SCT, jerarquizándolos para que representen una mejor descripción de los deterioros presentados por zonas homogéneas.

2 Antecedentes

Las zonas homogéneas están definidas como espacios de la superficie terrestre que presentan condiciones, características o cualidades similares en toda su longitud; concepto que puede aplicarse en diferentes temas de análisis; como estudios catastrales, climáticos, litológicos, tipos de suelo, etcétera.

Xu Yang et. al [2] propuso un método para seccionar una región en Michigan (USA) basándose en las condiciones de temperatura del aire y temperatura sobre la superficie del pavimento apoyándose de metodologías como isoyetas y polígonos de Thiessen, con la finalidad de definir el desempeño del pavimento para las temporadas más frías, más calientes y la temperatura media anual.

Pashant et. al [3] desarrolló un proyecto ad-hoc que considera las necesidades del seccionamiento para la conservación de un pavimento. Para dicho proyecto se toma como tramo de estudio un pavimento en la India y se busca delimitar zonas homogéneas desde su aspecto estructural, utilizando la metodología de las diferencias acumuladas y el factor de desempeño de las Deflexiones.

Los anteriores proyectos de investigación cumplen con su propósito y obtienen resultados aplicables a su respectivo caso, pero para el caso del proyecto que se desarrolla en este artículo, las isoyetas, polígonos de Thiessen o las diferencias acumuladas no resultan del todo adecuadas ya que solo consideran un solo factor de desempeño. En este artículo se propone caracterizar y evaluar al pavimento tanto en su aspecto estructural como funcional y de seguridad de manera integral.

3 Objetivos

El objetivo de este artículo es presentar una metodología que permita seccionar a un pavimento basándose en los factores propuestos por la SCT de manera conjunta y utilizando herramientas que den sustento a la toma de decisión sobre la condición estructural, funcional o de seguridad, mayormente afectada en un pavimento flexible.

4 Metodología

La propuesta metodológica esta enfocada en generar zonas homogéneas poniendo el siguiente sentido de revisión: Estructural →Funcional →Seguridad.

La metodología toma como base la jerarquización de los factores de desempeño propuestos por la SCT: siendo las deflexiones la deformación vertical del pavimento en respuesta a una carga externa, el índice de regularidad internacional (IRI) como la medida de referencia de la regularidad de la superficie, la profundidad de roderas como los surcos formados por neumáticos de los vehículos, la macrotextura que representa el relieve de la capa de rodadura y el coeficiente de fricción que valora la resistencia al movimiento relativo neumático - superficie del pavimento.

La jerarquización revisa la estructura del pavimento de manera ascendente, primero con las capas inferiores hasta llegar a la superficie de rodadura, acorde a los siguientes pasos:

- Primero, se revisa el comportamiento de las capas estructurales del pavimento mediante el análisis de las deflexiones.
- En seguida, se revisa al pavimento de manera longitudinal y transversal, considerando el comportamiento estructural y funcional de las capas superiores con el análisis del IRI y la profundidad de roderas.
- Finalmente, se revisa la seguridad sobre la capa de rodadura con el análisis del coeficiente de fricción y la macrotextura.

La metodología propuesta para el cálculo se muestra en la figura 1 y se describen a continuación:

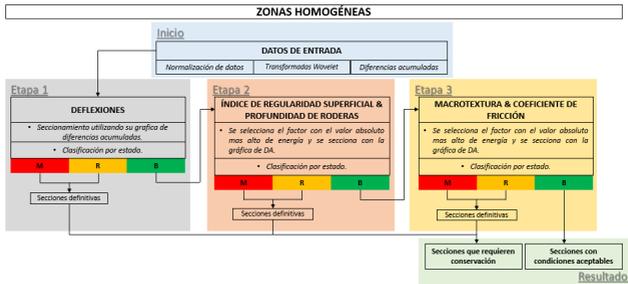


Fig. 1. Metodología para cálculo de zonas homogéneas.

- Inicio: hace referencia al proceso de normalización de los datos, cálculo de las diferencias acumuladas para cada uno de los factores de desempeño y la

obtención de las transformadas Wavelet que toman un papel importante en la propuesta al ser el valor de energía o desgaste que resulten para los factores de desempeño, lo que dicte cual es el que presenta un estado mayormente crítico para ser seleccionado con su gráfica de diferencias acumuladas y seccionar el tramo carretero.

- Etapa 1: se realiza un primer seccionamiento, basándose en la gráfica de diferencias acumuladas de las deflexiones, considerando que este factor de desempeño nos permite hacer la evaluación estructural de un pavimento, lo cual se alinea con la propuesta revisando primero la parte estructural y ascender en las capas del mismo hasta llegar a la parte funcional. Los tramos resultantes se clasifican asignando etiquetas de estado, bueno, regular o malo, considerando las especificaciones que se tienen para dicho factor de desempeño. Lo que resulte en estado regular o malo, se deberá definir como sección definitiva, pasando a una segunda etapa de seccionamiento, los tramos clasificados como en estado bueno.
- Etapa 2: se revisan los tramos que en la etapa anterior tengan clasificación de estado bueno, esto con el IRI y la profundidad de roderas en conjunto, caracterizando así el pavimento en el sentido longitudinal y transversal. Ya que se están tratando en conjunto los dos factores de desempeño, se debe elegir a uno de ellos como el más crítico, para esto comparamos los valores absolutos de energía obtenidos para ambos factores de la transformada wavelet y se elige al que resulte mayor para este segundo seccionamiento. De la misma manera que para la etapa 1, los tramos obtenidos se clasifican por su nivel de estado según las especificaciones del proyecto de estudio como regular o malo siendo secciones definitivas y los de estado bueno pasan a una tercera etapa.
- Etapa 3: para el aspecto funcional del pavimento, se analizan en conjunto la macrotextura y el coeficiente de fricción. El proceso en esta etapa es similar a la etapa 2, revisando la energía entre ambos factores de desempeño y eligiendo el que presente un estado más crítico de desgaste (transformadas Wavelet), para ser utilizado en este último seccionamiento con su gráfica de diferencias acumuladas.

El seccionamiento del tramo carretero será la división obtenida en las secciones definitivas obtenidas en cada etapa, las cuales pueden tener una clasificación de estado bueno, regular o malo.

5 Resultados

La metodología propuesta se ha aplicado en un ejemplo práctico y considerando que actualmente el seccionamiento se basa en análisis empiristas, el realizar comparaciones no resulta posible; con esta consideración el resultado del caso ejemplo es de 14 zonas homogéneas con variadas combinaciones en los valores de factores de desempeño y por consiguiente, diferentes condiciones de estado, como se muestra en la figura 2:

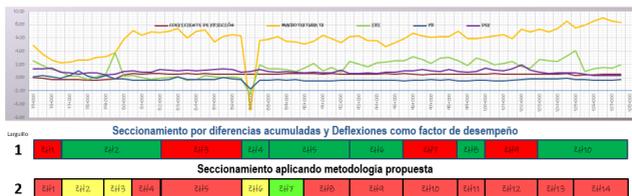


Fig. 2. Zonas Homogéneas (ejemplo práctico)

En el larguillo 1, de la figura 1, se utilizó la metodología tradicional de las diferencias acumuladas, y de las 10 zonas homogéneas que resultan únicamente 4 destacan con necesidades de intervención, pero se discrimina el resto de características que presentan los tramos en sus factores de desempeño. Caso contrario, en el larguillo 2, que expone la aplicación de la metodología propuesta, se muestra que en su mayoría, las zonas homogéneas presentan algún tipo de daño pudiendo ser estructural, funcional o de seguridad, considerando cada factor de desempeño.

6 Conclusiones

La correcta delimitación de zonas homogéneas en un pavimento flexible es de vital importancia ya que de este aspecto puede depender el correcto desempeño de la vialidad a lo largo del tiempo al aplicar conservación carretera. Además, contribuye a proteger la inversión de recursos, ya que si el seccionamiento es correcto el tratamiento de conservación será el adecuado para detener o reparar la evolución del deterioro identificado en cada tramo. La metodología presentada es una herramienta de apoyo para los encargados de la conservación de carreteras, al ofrecer bases para la toma de decisiones referente a la obtención de zonas homogéneas contribuyendo a disminuir subjetividades en la apreciación de los daños del pavimento por tramo.

References

1. Delmar Salomón. (2009). Conservación de pavimentos: Metodología y estrategias. EUA.
2. Xu Yang, Zhanping You, Jacob Hiller y David Watkins (2018): Pavement performance zone based on mechanistic-empirical design and temperature indices, Transportmetrica A: Transport Science, DOI: 10.1080/23249935.2018.1457734
3. Pashant R. Pambhar, Dr. L.B: Zala, Amit A. Amin (2018). Flexible pavement overlay design cumulative difference approach of homogeneous section.
4. AASHTO. (1993). Método aashto 93 para el diseño de pavimentos. Apéndice A
5. Normativa para conservación: Libro CVS. Conservación, Parte: Evaluación. (2016 – 2017). Secretaría de Comunicaciones y Transportes

CAPÍTULO 9

*Estrategia de Verificación de la Conformidad
entre Modelos de Procesos y Bitácoras de
Eventos.*

ARELY MORENO Y HEIDY M. MARIN-CASTRO

Estrategia de Verificación de la Conformidad entre Modelos de Procesos y Bitácoras de Eventos

Arely Moreno Bautista¹ and Heidy M. Marin-Castro²

¹ Universidad Autónoma de Tamaulipas, Cd. Victoria Tamaulipas, México
a2133080093@alumnos.uat.edu.mx

² Cátedras Conacyt - Universidad Autónoma de Tamaulipas, Cd Victoria
Tamaulipas, México
hmarisol@docentes.uat.edu.mx

Resumen Un modelo de proceso es una representación gráfica y analítica utilizado para capturar el comportamiento del proceso de negocio de una organización. La generación de modelos de procesos robustos permite mejorar la toma de decisiones dentro de las organizaciones, lo que implica explotar el conocimiento adquirido a partir de las bitácoras de eventos disponibles. Para ello, es necesario utilizar técnicas de minería de procesos, con el objetivo de descubrir, monitorear y mejorar los procesos de negocio en ejecución. Una de las principales tareas de la minería de procesos es la verificación de la conformidad encargada de comparar un modelo de proceso documentado o descubierto con registros almacenados en la bitácora de eventos, con el fin de identificar los puntos en común o desviaciones en el modelo. Este trabajo de investigación propone el desarrollo de una nueva estrategia de verificación de la conformidad, basada en el uso de técnicas de preprocesamiento de eventos, considerando como caso de estudio el dominio de salud.

Keywords: CLQ · Minería de Procesos · Bitácora de Eventos · Modelo de proceso · Verificación de Conformidad · Descubrimiento de modelos

1. Introducción

Las organizaciones a menudo especifican el comportamiento de sus procesos de negocio por medio de un modelo de proceso, el cual representa el flujo de ejecución de las tareas que se llevan a cabo. La minería de procesos es un área de estudio relativamente nueva que provee un enlace importante entre la computación, el modelado y análisis de procesos de negocio. Esta área de estudio permite realizar un análisis a partir de la extracción de conocimiento de las bitácoras de eventos. De esta forma, las organizaciones pueden saber realmente de qué manera se están llevando a cabo sus procesos e identificar problemas no visibles como cuellos de botella, decisiones, violaciones de políticas, gestión de recursos, entre otros, y dar solución a estos problemas. Una de las tareas más importantes de la minería de procesos es la verificación de conformidad, cuyo

objetivo es cuantificar la calidad de un modelo de proceso de negocio en relación con los datos que se registraron durante la ejecución del proceso, es decir, nos permiten monitorear las desviaciones al comparar un modelo de proceso dado con el registro de eventos real y establecer en qué nivel corresponde el uno al otro. Esta propuesta de investigación se desarrolla en el contexto de la minería de procesos [7], particularmente en la verificación de la conformidad de modelos de procesos y su bitácora de eventos. Una parte importante de la propuesta consiste en mejorar la calidad de la bitácora de eventos mediante la aplicación de técnicas de preprocesamiento y limpieza de los registros almacenados. Una vez mejorada la calidad de los eventos el siguiente paso consiste en realizar el descubrimiento del modelo de proceso a partir de la bitácora preprocesada. Posteriormente realizar la tarea de verificación de conformidad a fin de identificar desviaciones o anomalías relacionadas con el seguimiento o cumplimiento del proceso de la organización. Finalmente, como parte de este trabajo se plantea la validación y evaluación de la estrategia de conformidad propuesta a través de métricas reportadas en la literatura.

2. Antecedentes

En los últimos años el mercado analítico de procesos globales presenta un alto crecimiento. Esto sugiere que el estudio, análisis y desarrollo de modelos de procesos de negocio a través de técnicas de minería de procesos será uno de los temas más interesantes para el análisis de procesos en los años siguientes. El tener modelos de proceso eficientes y de calidad es una tarea ardua que requiere constantemente de aplicar técnicas diseñadas para mejorar la calidad de las bitácora de eventos, y así reducir la complejidad de los modelos de procesos construidos. En los trabajos [5] y [4] se presentan dos enfoques de preprocesamiento en los registros de eventos. En [5] se propone un enfoque para la selección de instancias la cual consiste en extraer un subconjunto de registros a través de una medida umbral y así establecer que el subconjunto seleccionado es realmente representativo del conjunto total de datos. En el trabajo presentado en [4] se utiliza el algoritmo CloFast, y técnicas de clustering para extraer patrones de secuencia frecuentes en los registros y posteriormente agruparlos de acuerdo a su similitud. Todo esto con el fin de obtener una subselección o agrupación de datos reducida y representativa que ayude a reducir la complejidad de los modelos de proceso, y así permita una fácil interpretación de estos. Los trabajos presentados en [1] y [6] están enfocados en resolver la tarea de la verificación de la conformidad utilizando métricas basadas en la alineación como el Fitness, el cual, es una métrica que permite calcular la diferencia entre las trazas en una bitácora de eventos con respecto a un modelo de proceso. A diferencia de estos dos trabajos, Este trabajo propone abordar la misma tarea usando previamente técnicas de preprocesamiento para mejorar la calidad de la bitácora de eventos, así como diseñar una estrategia de conformidad que no requiera hacer una comparación completa entre el modelo de proceso y la bitácora de eventos (alineamiento parcial).

3. Objetivos

3.1. Objetivo General

Construir una estrategia de verificación de la conformidad entre el modelo de proceso y la bitácora de eventos basada en la limpieza de registros de eventos con el fin de mejorar la calidad de los modelos de procesos, usando como un caso de estudio el dominio de salud.

3.2. Objetivos Específicos

- Implementar un algoritmo de descubrimiento de modelos de procesos a partir de una bitácora de eventos que incorpore la limpieza de los registros de eventos.
- Desarrollar un algoritmo de verificación de la conformidad entre modelos de procesos y bitácoras de eventos tomando como caso de estudio el dominio de salud.
- Evaluar el impacto logrado por la estrategia desarrollada mediante el uso de métricas de calidad.
- Crear una herramienta de despliegue y evaluación de la estrategia de verificación de conformidad propuesta.

4. Metodología

La metodología que se propone en este proyecto consta de tres etapas, dando como resultado la realización de una herramienta de despliegue del modelo de proceso, y la evaluación de la conformidad propuesta. La primera etapa consiste en el estudio e implementación de técnicas de descubrimiento de modelos de procesos. En la segunda etapa, se enfoca en la verificación de la conformidad. En esta etapa se desarrollará una nueva técnica para obtener la conformidad entre un modelo de proceso y una bitácora de evento. En la última etapa de esta investigación se estudiarán las métricas de la evaluación de la conformidad, y finalmente se construirá una herramienta de evaluación de la conformidad resultante de la segunda etapa de esta metodología propuesta.

5. Avances

Para el desarrollo del proyecto de investigación se realizó la búsqueda de 3 bitácoras de eventos, las cuales se describen a continuación: 1. Bitácora de Eventos Sepsis Cases. Esta bitácora de eventos contiene registros de pacientes que han contraído la enfermedad de sepsis, una enfermedad potencialmente mortal debido a un virus. Esta bitácora cuenta con 1000 casos, con un total de 15,000 eventos para 16 actividades. 2. Bitácora de eventos Real-life event logs-Hospital log. Esta bitácora de registros de la vida real de un hospital académico holandés originalmente diseñado para su uso en el primer Concurso de Inteligencia de

Procesos de Negocios. 3. Bitácora de eventos Hospital Billing- Event Log. Esta bitácora contiene registros de eventos de facturación de un hospital. Esta se obtuvo de los módulos financieros del sistema ERP de un hospital regional con alrededor de 100000 trazas.

Como parte del desarrollo de este proyecto se implementaron 3 algoritmos de descubrimiento de modelos de proceso disponibles en el área de estudio, considerando los factores de tiempo, diseño, estructura y popularidad en el área. Los algoritmos implementados para el descubrimiento de modelos de procesos son: Heuristic Miner [8], InductiveMiner [3] y SplitMiner [2]. Para el preprocesamiento de las bitácoras de eventos se implementaron 2 técnicas de limpieza de datos, una de ellas es la selección de instancias [5], y la otra es la agrupación a través de patrones de secuencia frecuentes [4]. La selección de instancias consiste en analizar la bitácora de eventos calculando las probabilidades de cada par de secuencias dentro del total de registros, utilizando una medida umbral para seleccionar solo el subconjunto de trazas más representativo del total de registros. En la técnica de agrupación se utiliza el algoritmo cloFast, algoritmo capaz de detectar patrones de secuencia similares y seleccionar únicamente las secuencias de mayor frecuencia para después aplicar un algoritmo de clustering y obtener agrupaciones que cuenten con patrones de secuencia frecuentes. Como parte de los resultados parciales obtenidos, en las Figuras 1 y 2 se muestran dos modelos de proceso descubiertos utilizando el algoritmo Heuristic Miner con y sin el preprocesamiento de la bitácora de eventos. En la Fig. 2 se aprecia que el uso de técnicas de preprocesamiento reduce la complejidad del modelo y facilita su comprensión.

6. Resultados Esperados

En continuidad con la investigación, se espera diseñar una estrategia para la verificación de la conformidad basada en alineación utilizando bitácoras previamente preprocesadas. Además, como parte de los resultados esperados se diseñará e implementará una herramienta de despliegue y evaluación de la estrategia propuesta.

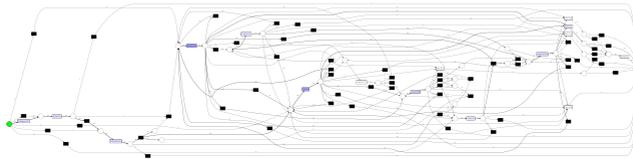


Figura 1. Modelo de proceso generado utilizando el algoritmo Heuristic Miner y la bitácora Sepsis sin preprocesamiento de datos.

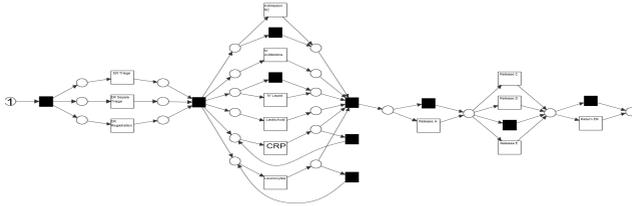


Figura 2. Modelo de proceso generado utilizando el algoritmo Heuristic Miner y la bitácora Sepsis con preprocesamiento de datos.

7. Conclusiones

A lo largo de esta investigación se ha podido observar la importancia del uso de técnicas de preprocesamiento de datos para la limpieza de los registros de eventos. Al generar modelos de proceso con bitácoras de eventos preprocesadas, utilizando algoritmos como inductiveMiner o HeuristicMiner, los modelos resultantes son menos complejos y requieren de menor el tiempo de ejecución. Por lo tanto, estos modelos más simples de interpretar pueden ser de gran ayuda dentro de las organizaciones en la toma de decisiones.

Referencias

1. E. Asare, L. Wang, and X. Fang. *Conformance Checking: Workflow of Hospitals and Workflow of Open-Source EMRs*. IEEE, 2010.
2. Adriano Augusto, Raffaele Conforti, Marlon Dumas, and Marcello La Rosa. Split miner: Discovering accurate and simple business process models from event logs. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 1–10. IEEE, 2017.
3. Sander JJ Leemans, Dirk Fahland, and Wil MP van der Aalst. Scalable process discovery with guarantees. In *Enterprise, Business-Process and Information Systems Modeling*, pages 85–101. Springer, 2015.
4. S. Lu, X. and Tabatabae, M. Hoogendoorn, and H. Reijers. *Trace Clustering on Very Large Event Data in Healthcare Using Frequent Sequence Patterns*. Springer, 2019.
5. F. Mohamadreza, S. van Zelst, and W. van der Aalst. *The Impact of Event Log Subset Selection on the Discovery of Process Models*. 2019.
6. V. Naderifar, S. Sahran, and Z. Shukur. *A Review on Conformance Checking Technique for the Evaluation*. TEM JOURNAL, 2019.
7. W. Van der Aalst. *Process Mining*. Springer, 2016.
8. AJMM Weijters, Wil MP van Der Aalst, and AK Alves De Medeiros. Process mining with the heuristics miner-algorithm. *Technische Universiteit Eindhoven, Tech. Rep. WP*, 166:1–34, 2006.

CAPÍTULO 10

Desarrollo e implementación de un sistema proactivo de iluminación

ROSARIO DE LA LUZ CANTERO RAMÍREZ, MARCO ANTONIO MEZA AGUILAR Y ANSEL YOAN RODRÍGUEZ GONZÁLEZ

Desarrollo e implementación de un sistema proactivo de iluminación

Rosario de la Luz Cantero Ramírez¹, Marco Antonio Meza Aguilar¹ y Ansel Y. Rodríguez González,²

¹ ITCG - Instituto Tecnológico de Ciudad Guzman, Jalisco, México

² CICESE-UT3, Nayarit, México

Resumen El desperdicio de energía eléctrica utilizada en iluminación, provocado por el control manual de las luminarias es un problema que no solo afecta en el aspecto económico pues también afecta al consumidor, aunque es un problema estudiado alrededor del mundo, en México se ha tratado poco. Por ello, el presente trabajo tiene como objetivo desarrollar e implementar un sistema de iluminación modular y escalable que aprenda de su entorno y proporcione iluminación adecuada de forma proactiva y sin disminuir el confort de los consumidores en un conjunto de oficinas.

Keywords: Iluminación · Sistema · Control · Aprendizaje · Proactivo.

1 Introducción

Una buena iluminación se ve reflejada en el desempeño laboral dado que puede afectar la manera en que el ser humano se desenvuelve [5]. Los edificios residenciales y no residenciales son considerados como el sector de mayor intensidad eléctrica [1]. En [4] se señala que el 40% del total de las facturas de electricidad de una empresa se deben a la iluminación. Se estima que mejorar las prácticas en el diseño de la envolvente y la iluminación puede ahorrar, al menos, 40% del uso total de energía de un edificio [3]. El desperdicio de energía eléctrica ocurre incluso con el uso de un control automático de las luminarias. El presente trabajo muestra los avances en el diseño de un sistema inteligente proactivo que pueda proporcionar la iluminación adecuada a los usuarios para el desarrollo de actividades, dentro de un conjunto de oficinas.

2 Antecedentes

Con el uso de la tecnología se puede mejorar el control de las luminarias, ofreciendo al usuario mayor comodidad y ahorro de energía. En el mercado existe una gran variedad de componentes que ofrecen esta posibilidad mediante el uso de la domótica. Sin embargo, el control se encuentra centralizado en un solo usuario, lo cual puede generar conflictos de preferencias. Ante ello se han desarrollado sistemas reactivos que son más independientes del usuario, cuyo funcionamiento

está dirigido por eventos y es determinista. En su tesis de maestría, [6] presenta un control reactivo de luminarias en áreas de un edificio. El control se basa en datos sensados, sin embargo, su nula tolerancia a cambios puede causar malas experiencias.

Aunado a ello, se sabe que los sistemas inteligentes son capaces de aprender y adaptarse a su entorno, [2] presenta un método, que permite hacer predicciones con el uso de conjuntos predefinidos de características que influyen en las preferencias de iluminación de los usuarios. Mediante la simulación se demostró que las características más influyentes en la toma de decisiones del sistema son: luminosidad exterior, área de actividad y el tipo de actividad realizada por el usuario. En [7] se propone descubrir los patrones secuenciales temporales de ocupación, como complemento de los datos sensados. Los resultados obtenidos mostraron que la predicción de ocupación se correlaciona con el índice de ocupación histórico y la característica secuencial de tiempo, logrando un 97.4 % de precisión y un 79.5 % de menos salidas falsas.

3 Objetivo

Desarrollar e implementar un sistema de iluminación, modular y escalable, que aprenda de su entorno y proporcione iluminación adecuada de forma proactiva sin disminuir el confort de los consumidores en un conjunto de oficinas.

4 Metodología

A través de un análisis experimental, se seguirá la siguiente metodología: como primera fase se realizarán consultas del estado del arte, con el fin de conocer más sobre el área y sus avances, haciendo énfasis en aquellas soluciones propuestas para conjuntos de oficinas. En la segunda etapa se diseñarán los módulos del sistema, así como la arquitectura de comunicación, la cual permita mantener una comunicación, constante y estable, entre los módulos. En la tercera fase el módulo inteligente será dotado de un algoritmo de aprendizaje reforzado, que le permitirá percibir el entorno y los cambios que suceden en éste, así como reaccionar y reajustar sus acciones sobre los cambios percibidos. La integración del sistema, las pruebas de campo y el análisis de su comportamiento, serán parte de la cuarta fase.

5 Avances

Durante los primeros cinco meses se realizó una investigación del estado del arte, encontrando que los principales sistemas de iluminación en el mercado son domóticos o reactivos, además se encontró con trabajos de investigación que analizan la relación entre las variables del entorno y la iluminación, así como propuestas de sistemas de iluminación inteligente probados mediante simulaciones. Una vez realizada la investigación se inició el estudio del entorno, donde se

implementará, posteriormente se realizó la selección de componentes para cada módulo. A continuación, se muestran los principales componentes que integran los módulos:

- Iluminación/Sensado: NodeMCU, HC-SR501, SCT-013-030, bombilla
- Sensado: NodeMCU, HC-SR04, SCT-013-030 ,KY-018
- Inteligente: Raspberry Pi, SCT-013-030
- Manual: Esp-01s, botón, ACS712

Se realizaron los diseños electrónicos, así como el diseño de las carcasas protectoras, para cada uno de los módulos, en las cuales serán montados una vez que las pruebas de campo sean iniciadas para evitar daños en el circuito. En la Figura 1 se muestra el diseño electrónico, así como la carcasa del módulo de iluminación/Sensado. El sistema cuenta con cuatro tipos de módulos diferentes, cada habitación del lugar estudiado contendrá los módulos necesarios para obtener las mediciones del ambiente. Estos módulos realizarán el intercambio de paquetes entre sí, a través del diseño de comunicación que se muestra en la Figura 1. Se cuenta con un programa estructurado para el establecimiento de la comunicación, por medio del protocolo MQTT. Actualmente se está trabajando con la programación orientada a objetos, que permita el encapsulamiento y la reutilización de código, lo cual hará más sencillo el mantenimiento y la integración de nuevos módulos al sistema para una mayor escalabilidad.

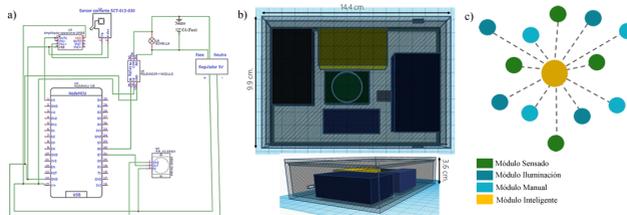


Figura 1. En la figura a) y b) se muestra el esquema eléctrico y la carcasa diseñada para el módulo de iluminación/sensado respectivamente. La figura c) muestra el diseño de la arquitectura de comunicación del sistema.

Aún queda pendiente el diseño y la programación de la base de datos, así como el diseño y programación del algoritmo inteligente para la toma de decisiones, su entrenamiento y pruebas.

6 Resultados Esperados

En este trabajo se pretende desarrollar e implementar un sistema de iluminación, modular y escalable, en un conjunto de oficinas, el cual, por medio de un

algoritmo inteligente, pueda aprender de su entorno y hacer predicciones. La finalidad del sistema es que sea capaz de proporcionar la iluminación adecuada para el usuario de forma proactiva, sin disminuir el confort. Al mismo tiempo se busca reducir el desperdicio de energía eléctrica, ocasionado por el control deficiente del sistema de iluminación tradicional.

Las variables que se utilizarán para las predicciones del sistema serán: presencia, luminosidad, día y hora, los cuales serán recopilados por el sistema y posteriormente almacenados en la base de datos que alimentará al algoritmo inteligente para la toma de decisiones. Se hará uso del algoritmo de aprendizaje reforzado Q-Learning, de esta forma el sistema será capaz de aprender y mejorar sus decisiones de acuerdo a las recompensas obtenidas por las acciones que se realizaron.

El sistema inicialmente no sabrá cómo comportarse, por lo cual tomará una de las posibles acciones a realizar de forma aleatoria recibiendo una recompensa positiva o negativa, conforme vaya interactuando con el ambiente y las recompensas obtenidas el sistema sabrá qué debe hacer y qué no, de esta forma creará políticas con base en el conocimiento adquirido para ofrecer al usuario una buena experiencia en cualquier momento de acuerdo a los datos sensados en su entorno.

7 Conclusiones

En este trabajo se logró crear diseños y montar prototipos de cada uno de los módulos que integran el sistema, así como el diseño de carcasas protectoras para los módulos. Se logró además establecer la comunicación entre dichos módulos, a través de la programación estructurada, con el uso del protocolo MQTT. Actualmente el sistema es capaz de realizar tareas programadas, simulando el funcionamiento del sistema manual.

Como trabajo futuro se diseñará la base de datos, así como el algoritmo inteligente, con el cual será dotado el módulo inteligente y le permitirá aprender de su entorno, de esta forma los datos sensados se almacenarán en la base de datos y permitirán entrenar el algoritmo. Posteriormente se implementará el sistema en un entorno real, lo que permitirá obtener datos empíricos de su funcionamiento y de la satisfacción del consumidor.

Referencias

1. Chatellier, D., McNeil, M.: Consumo de electricidad de edificios no residenciales en México: La importancia del sector de servicios. https://www.conuee.gob.mx/transparencia/boletines/Cuadernos/cuaderno3nvo ciclo_2.pdf. Último acceso 3 Jun 2021
2. Gopalakrishna, A. K., Özçelebi, T., Liotta, A., Lukkien, J.J.: Exploiting machine learning for intelligent room lighting applications. 2012 6th IEEE International Conference Intelligent Systems, pp. 406–411. <https://doi.org/10.1109/IS.2012.6335169>
3. De Buen, O.: Las Acciones de la Eficiencia Energética de la Conuee en la Edificación,

- url[https://www.gob.mx/cms/uploads/attachment/file/84265/](https://www.gob.mx/cms/uploads/attachment/file/84265/CONUEEEEdificio-Agosto2013.pdf) CONUEEEEdificio-Agosto2013.pdf. Último acceso 3 Jun 2021
4. Ramirez, A.: Recomendaciones y guías de eficiencia energética: control de alumbrado, submedición y verificación energética, <https://imej.org.mx/2016/07/04/iluminacion-eficiente-en-edificios/>. Último acceso 3 Jun 2021
 5. Talero, C.: Efectos y mejoras del proceso de iluminación en las áreas de trabajo, <https://repositorio.ecci.edu.co/handle/001/531>. Último acceso 3 Jun 2021
 6. Montaguano, C.: Diseño e implementación de un sistema automatizado de control de iluminación de las áreas comunales del edificio Torre de Suite, mediante el uso del PLC Siemens (logo 230 RCE), <https://repositoriolatinoamericanos.uchile.cl/handle/2250/2789555>. Último acceso 3 Jun 2021
 7. Jin, Y., Da, Y., Zhang, X., An, J., Han, M.: A data-driven model predictive control for lighting system based on historical occupancy in an office building: Methodology development. *Building Simulation*, volumen 14, 219–235 (2021)

CAPÍTULO 11

*Algoritmo de Anonimización de Notas
Médicas para Proteger la Identidad de los
Pacientes*

JESÚS MERCADO, KARINA FIGUEROA, ARTURO LÓPEZ, CLETO ÁLVAREZ Y ANEL GÓMEZ

Algoritmo de Anonimización de Notas Médicas para Proteger la Identidad de los Pacientes

Jesús Mercado Ríos¹, Karina Figueroa¹, Arturo López², Cleto Álvarez² y Anel Gómez³

¹ Universidad Michoacana de San Nicolás de Hidalgo
{1548272a,karina.figueroa}@umich.mx
² Amphora Health
³ IMSS

Keywords: Procesamiento de Lenguaje Natural, recuperación de información, anonimización de datos personales

Resumen La digitalización en expedientes médicos ha sido pieza fundamental para el estudio de éstos. Los centros de salud capturan valiosa información de los pacientes en su expediente. Cada expediente debería poder ser estudiado con técnicas automatizadas para descubrir patrones, comportamientos, etc, que puedan apoyar el bienestar de la población. Sin embargo, para que esto pueda ser una realidad, las notas médicas deben anonimizar los datos personales de los pacientes. Una tarea que necesariamente se debe automatizar pues la cantidad de datos podría llegar a ser enorme. En este trabajo se muestra un algoritmo para anonimizar las notas de expedientes clínicos y así eliminar datos personales de las descripciones capturadas. Los resultados obtenidos alcanzan hasta un 85% de la identificación y eliminación de éstos.

1. Introducción

Hoy en día es natural que tanto hospitales como clínicas mantengan los registros de los pacientes de manera digital. Esto es muy importante pues se podrían analizar los datos ahí reflejados para detectar patrones, sin embargo, en muchos de estos expedientes existe información sobre los datos personales, v.g. nombre del paciente, dirección, teléfono y ciudad. Lo cual, de acuerdo con la ley de privacidad de los datos, éstos deben ser eliminados.

En este trabajo, se presenta un algoritmo que permite anonimizar notas de expedientes clínicos obtenidas de las bases de datos del IMSS (Instituto Mexicano de Seguro Social). Los expedientes clínicos anonimizados permitirán proteger la información personal de los pacientes y que se puedan hacer estudios de la información contentida, por ejemplo, modelos de predicción de pacientes con alguna condición específica, sistemas de recuperación de información, análisis estadísticos, algoritmos de agrupamiento, etc.

Este trabajo está organizado de la siguiente manera: en la sección 2 se muestra el trabajo relacionado; posteriormente en la sección 3 se muestra la metodología

usada para la organización y diseño del algoritmo planteado. En la sección 4 se presenta el resultado obtenido en una muestra de la base de datos. Finalmente, en la sección 5 se darán las conclusiones y trabajo futuro.

2. Antecedentes

Dada la importancia del estudio de los expedientes clínicos (ExCl) y lo indicado por la ley en cada país, existen diversos artículos que hablan sobre la anonimización de los ExCl [4,5]; sin embargo, todos estos trabajos son para ExCl en inglés, no así para el español.

Por otro lado, dado que no se ha trabajado con Bases de datos de ExCl en español, tampoco existen bases de datos para entrenar sistemas como los que utilizan algunas técnicas de procesamiento de cadenas que requieren entrenamiento con muchos datos.

3. Metodología

En esta sección se describirán los pasos seguidos para el desarrollo del trabajo, y más adelante se explicarán detalles técnicos de cada uno.

3.1. Base de datos

Los datos médicos utilizados, son parte de la base de datos (BD) del IMSS. Dicha BD es de tipo relacional y los datos con acceso ya no cuentan con columnas obvias de información personal.

Las únicas columnas con posibles datos personas son las notas capturadas por un médico o un trabajador social. En la notas médicas no se encontraron datos personales, sin embargo, en las descripciones realizadas por el personal de trabajo social (NTS) se puede encontrar, en algunas, datos personales, como nombre, dirección, ciudad y teléfono de los pacientes o familiares directos.

Para este trabajo las NTS usadas fueron 450 como prueba y 3000 con las que actualmente se trabaja. A continuación se muestra un ejemplo de una NTS:

“Se recibe de jefatura de servicio reporte de asistente medica de paciente remisa en control por diabetes e hipertensión con inasistencia del dia 17 de agosto; se verifica en SIMF no habiendo registro de reintegracion a control medico, se realiza llamada telefónica al número {TEL} contestando su hija {NOMBRE-FAMILIAR} se le explica el motivo de la llamada, refiere de momento la sra {NOMBRE-PACIENTE} no se encuentra.”

3.2. Datos personales

Al leer un conjunto de notas, se observó que los datos personales, o identificadores, sólo serán: Nombre, domicilio, ciudad y teléfono.

Las NTS analizadas tienen como patrones *palabras disparadoras*, es decir, si se va a escribir el nombre del paciente generalmente existe la palabra *paciente* antes del nombre; lo mismo sucede cuando se comienza a describir el domicilio o los números telefónicos.

3.3. Palabras disparadoras

Las palabras disparadoras, son palabras que pueden indicar que habrá un dato personal; es decir, anteceden los datos identificadores.

- Para el caso de los nombres, le anteceden palabras como: “paciente”, “hijo”, “hija”, “padre”, “madre”, etc.
- Para el caso de los domicilios, podemos encontrarlos con las palabras disparadoras: “domicilio particular”, “Domicilio calle”, entre otros.
- Para los números de teléfono, las palabras disparadoras pueden ser: “tel.”, “tel”, “teléfono”, “cel”, “celular”, “número”.

3.4. Algoritmo

Para la anonimización de los datos personales se utilizaron técnicas de recuperación de información como las descritas en [1,3]. Al algoritmo aquí planteado lo llamaremos *Anonimización de Datos Personales en Notas Médicas (ADPNM)*.

El algoritmo realizará su tarea de anonimizar las NTS de la siguiente manera: se considerará cada narrativa como un documento independiente; se construye un índice que permita localizar todas las palabras; se identificarán los datos personales, los eliminará y en su lugar pondrá una etiqueta. Las etiquetas serán: {DOM} y {TEL}, para los domicilios y los teléfonos respectivamente. Para el caso de los nombres, la etiqueta corresponderá a la palabra disparadora que anteceda al dato personal, {HIJO}, {PADRE}, {MADRE}, etc.

El primer paso para llevar a cabo la anonimización, es un procesamiento al texto, es decir, se realizará un parser lingüístico (preprocesamiento) a cada documento; éste consistirá en dividir en palabras (tokens), sin tomar en cuenta las palabras vacías o (stop words) (v.g. artículos, preposiciones, etc.). Con las palabras conseguidas se forma un diccionario de los tokens válidos.

Con este diccionario se crea un índice invertido que contiene una lista posicional del documento y la palabra encontrada [2]. Si consideramos nuestro ejemplo presentado en la sección 3.1, y tomamos algunas palabras, por ejemplo: paciente, esta palabra se encuentra en el documento 1 siendo la 7a palabra. Lo mismo sucede para las siguientes 2 palabras presentadas.

```
paciente -> [1, [7]], [2, [4]]
diabetes -> [1, [10]]
remisa -> [1, [8]], [2, [5]]
```

El diccionario creado será la llave para acceder a la lista invertida de cada término (de ahora en adelante podrá referirse a un token como término).

Diccionarios Para poder identificar los nombres, apellidos y municipios, se utilizaron bases de datos contextuales con 4400 elementos de nombres, 7456 apellidos y 145 municipios.

Búsqueda En este punto del trabajo se tiene: un índice invertido, un diccionario con las palabras encontradas en las NTS, varios diccionarios contextuales y un conjunto de palabras disparadoras. El algoritmo en general es:

- Dada una lista de palabras disparadoras Lp
- Se usa cada palabra disparadora como llave del índice invertido para obtener la lista de documentos donde se encuentra.
- En el caso de más de una palabra disparadora se verifica que sean contiguas (con ayuda de la lista posicional).
- Se toman las palabras siguientes y se verifican si son parte de algún diccionario contextual.

Dada la naturaleza de las NTS existen algunos detalles específicos para cada tipo de dato a remover, como se muestra a continuación.

- Para los nombres, se verifica si la palabra está contenida en el diccionario de los nombres o de los apellidos, si está dentro, se revisa la siguiente palabra. Hasta 4 veces se realiza este paso
- Para el caso de los domicilios, se utilizan dos palabras disparadoras, la primera podrá ser: “domicilio”, “particular” o “calle”; la segunda podrá ser: “col”, “colonia”. Una vez identificadas, se eliminarán las palabras contenidas entre ellas, así como las siguiente dos palabras después de la segunda palabra disparadora.
- Para los teléfonos, se verifica que los caracteres siguientes a la palabra disparadora sean números, y se pide que sean mínimo 7 para poder ser reemplazados por la etiqueta.

4. Resultados

Recordemos que el tamaño del conjunto de validación es de 450 documentos. En la tabla 1 se mostrará la comparación entre los resultados de identificar manualmente los datos personales, y los obtenidos por el algoritmo ADPNM. En particular se muestra la precisión obtenida. En el caso de la recuperación no aplica puesto que se accede desde el índice invertido solo a las notas que contienen las palabras disparadoras; es decir, se tiene un 100% de precisión.

5. Conclusiones

El objetivo de este trabajo es anonimizar de manera automática las notas de trabajo social del Instituto Mexicano del Seguro Social. La anonimización

Tabla 1. Comparación de resultados entre el algoritmo y una identificación manual.

Tipo de datos	ADPNM	Identificados manualmente	Recuperación
Teléfono	266	300	88.66 %
Domicilio	56	68	82.35 %
Nombres	133	164	81.09 %
Ciudad	67	79	84.81 %
General			85.43 %

está enfocada a eliminar los datos personales de los pacientes y de sus familiares directos de dichas narrativas.

En este trabajo se presenta el avance que se tiene actualmente para conseguir esta anonimización. Los resultados obtenidos muestran que la recuperación general ronda el 85 %, sin embargo, aún podría mejorar.

De manera general es un problema interesante. Para los domicilios resulta un poco más complejo, puesto que es difícil identificar cuando se deja de hablar de él; o cuándo se sigue hablando de él, saber si incluirá la colonia, o hasta la ciudad, o si sólo dirá el nombre de la calle, etc.

El índice invertido creado también será empleado para analizar las notas médicas desde el punto de vista del agrupamiento en NTS semejantes. Lo cual es parte del trabajo futuro de este proyecto. Esto podrá llevarse a cabo con la vectorización de los documentos haciendo uso de conceptos como *frecuencia del documento*, *frecuencia del término*, *idf*, *tf-idf*. El agrupamiento en las NTS podría revelar información sobre el tipo de seguimiento a cada paciente y probablemente el impacto de éste.

Referencias

1. Badue, Claudine Santos, Ricardo A. Baeza-Yates, Berthier Ribeiro-Neto, and Nivio Ziviani. 2001. Distributed query processing using partitioned inverted files. In Proc. SPIRE, pp. 10–20.
2. Blanco, Roi, and Alvaro Barreiro. 2007. Boosting static pruning of inverted files. In Proc. SIGIR. ACM Press.
3. C. D. Manning, P. Raghavan, H. Schütze.: “Introduction to information retrieval”. 2008. Ed. Cambridge University Press
4. M. Kayaalp, A. C. Browne, Z. A. Dodd, C. J. McDonald. “De-identification of Address, Date, and Alphanumeric Identifiers in Narrative Clinical Reports”, Kayaalp M, Browne AC, Dodd ZA, Sagan P, McDonald CJ. De-identification of Address, Date, and Alphanumeric Identifiers in Narrative Clinical Reports. AMIA Annu Symp Proc. 2014 Nov 14;2014:767-76. PMID: 25954383; PMCID: PMC4419982.
5. B. Malin. “Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule”. 2012. Recuperado de: <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html>

CAPÍTULO 12

*Modelo difuso para la evaluación de cortes
carreteros*

JUAN CARLOS GALLEGOS CORNEJO, LUIS ALBERTO MORALES ROSALES, JAIME SAAVEDRA ROSALES, DANIEL RAYA GAMIÑO Y VIRGILIO LÓPEZ MORALES

Modelo Difuso para la Evaluación de Cortes Carreteros*

Juan Carlos Gallegos Cornejo¹, Luis Alberto Morales Rosales², Jaime Saavedra Rosales¹, Daniel Raya Gamiño¹, Virgilio López Morales³

¹ Universidad Michoacana de San Nicolás de Hidalgo

² Conacyt-Universidad Michoacana de San Nicolás de Hidalgo

³ Universidad Autónoma del Estado del Hidalgo

Abstract. La gestión de una red vial necesita modelos de evaluación del desempeño que manejen la incertidumbre de los datos de sus diferentes tipos de activos viales, con el fin de identificar problemas y/o daños. Este trabajo se centra en los activos geotécnicos, cuyos principales elementos son cortes y terraplenes. La mayor parte de los sistemas de evaluación existentes están limitados debido a la necesidad de un experto que evalúe los datos obtenidos, así como la falta de clasificación de las propiedades fijas y dinámicas de los cortes. En este artículo se propone el diseño de un modelo difuso para la evaluación de estado de diferentes tipos de cortes de carreteros en función de su material. Para el desarrollo del modelo se propone clasificar la información de evaluación en dos conjuntos: Inventario, que incluye las características fijas y valores iniciales de las características dinámicas. Performance, compuesto por parámetros dinámicos de desempeño. Lo que hace posible evaluar la evolución de daños o deterioros durante la fase de operación carretera. Con estos elementos, se representa de mejor manera la realidad, haciendo posible realizar una mejor evaluación de su estado. Al momento que permite estandarizar una evaluación que actualmente es influenciada por la expertiz de quien la realiza.

Keywords: CLQ· management · cortes carreteros · evaluación de estado

1 Introducción

Un activo geotécnico es una obra de suelo o roca, natural o hecha por el hombre adyacente a la carretera [5]. Entre ellos se encuentran: terraplenes, cortes, taludes reforzados y/o estabilizados, subrasantes y cimentaciones estructurales [6]. Los activos geotécnicos refuerzan el sistema de transporte, disminuyendo tiempos, acortando caminos y manteniendo pendientes apropiadas. Su mal funcionamiento puede causar retrasos significativos. La falla de estos puede generar daños significativos a la infraestructura y costos de operación muy elevados [7].

El valor del factor de seguridad en cortes carreteros varía dentro de un intervalo de valores que depende de la imprecisión, dispersión y errores sistemáticos

* CONACYT

cometidos en la obtención de los datos de entrada [1]. Además de la incertidumbre en los parámetros del suelo, la evaluación de estabilidad en los cortes debe considerar las incertidumbres y las ambigüedades climáticas [2].

Reedy (2016) [3] propone el uso de la lógica difusa para el manejo de la incertidumbre de los datos al determinar la estabilidad y factor de seguridad en cortes carreteros. La lógica difusa permite hacer un manejo apropiado de la imprecisión de los datos con base en la interpretación de expertos. [4] La estructuración y el grado de daño de las medidas de tratamiento en taludes puede influir directamente en la estabilidad de taludes de suelos expansivos [8]. Salcedo (2018) [9] propone un análisis de factores causales a través de la función: $f(x)=0.379 x_1+0.246 x_2+0.159 x_3+0.102 x_4+ 0.0545+ 0.036 x_6+ 0.024x_7$, donde las variables son pendiente, litología, precipitación, distancia de fallas geológicas, distancia a ríos, distancia a carreteras y cobertura vegetal.

En este artículo se presenta el desarrollo de un modelo difuso de evaluación de cortes carreteros. El modelo determina el estado del corte considerando sus condiciones de funcionamiento a partir de sus propiedades geológicas, geométricas y de estabilidad. La información es clasificada y analizada como información de inventario, y de performance. El modelo estandariza el proceso de evaluación, limitando las diferencias que pueden existir por la influencia de la expertiz del evaluador.

2 Objetivo

Desarrollar un modelo difuso de evaluación de condiciones de cortes carreteros en rocas sedimentarias, considerando las fases de construcción o de apertura, así como, operación carretera.

3 Estado del Arte

Actualmente en México la gran mayoría de administradores carreteros no realizan una gestión de los activos geotécnicos.

En 2016 el Dr. Paul Garnica [10] analizó los factores de desempeño de cortes y carreteros con la metodología de factores exponenciales propuesta y aplicada por Highways England [5]. Entre 2018 y 2019 autores como Jian Zhang [8], Daniela Salcedo [9] y Xiaorong Zhou [2] investigaron la implementación de modelos difusos para la evaluación de estabilidad de taludes.

El modelo presentado en este artículo realiza la evaluación de estado de inventario y performance con lógica difusa, considerando la estabilidad y el desempeño de los cortes carreteros.

4 Metodología

La metodología propuesta, esquematizada en la figura 1, establece las bases del proceso de evaluación de estado de activos viales.

4.1 Colección y clasificación de datos.

Consiste en obtener la información relativa a la evaluación del activo. A partir de su origen esta se clasifica en: 1) Inventario, obtenida en la primera implementación del modelo de evaluación, establece las propiedades fijas e iniciales, y 2) Performance, obtenida a partir de mediciones y evaluaciones visuales en la etapa de operación, permite identificar y dimensionar la aparición y evolución de daños y/o deterioros en el corte.

4.2 Agrupamiento de los datos.

Los datos obtenidos son agrupados en submódulos de evaluación a partir de las relaciones que existen entre estos. Con esto, se simplifica el análisis de influencia entre los diferentes parámetros evaluados al disminuir el número de variables. Además se genera la base de la evaluación difusa. La determinación de las relaciones entre los parámetros es establecida a partir de un sistema de evaluación que considera los datos necesarios para determinar el estado del corte.

4.3 Procesamiento de los datos.

El procesamiento de datos consiste en la interpretación de los valores obtenidos y de la influencia que existe entre ellos. De ahí que sea susceptible a la expertiz del evaluador. Con la finalidad de limitar esta susceptibilidad, el sistema de evaluación es fuzzificado, desarrollando un modelo difuso por cada submódulo de evaluación.

Desarrollo del modelo difuso. Para el desarrollo de la evaluación difusa, se propone el uso del mecanismo de inferencia Mamdani. Este mecanismo es útil cuando se tiene un número reducido de variables. Además de ser adecuado cuando los procesos de evaluación o determinación dependen de la participación humana (experto) ya que sus resultados se basan en las reglas definidas. Las etapas del modelado difuso contemplan: Fuzzificación de entradas y salidas, definición y evaluación de reglas, defuzzificación e interpretación de resultados

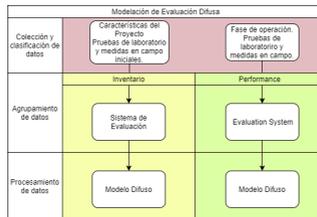


Fig. 1. Metodología de evaluación de activos viales.

5 Avances

Con base en las propuestas de los autores mencionados en el estado del arte. Se definieron dos sistemas de evaluación, uno de inventario y uno de performance. En estos, las propiedades que definen el estado del corte fueron agrupadas en submódulos de evaluación en función de las relaciones existentes entre estas, partiendo de los parámetros evaluados y formando finalmente los módulos de evaluación para el inventario y performance como se muestran en la figura 2.

Sedimentary Rock-Cutting	Inventory	Structure	Buzamiento	Interstratification	Thickness						
			Geology	Grooves	# Soft stratum	Deeper					
					Width						
					Spacing						
			Geometry	Height							
			Inclination angle								
			Safety Factor								
		Improvement	Stabilization	Deeper							
			Shallow								
			Protection	Captation Area							
			Screen								
	Performance	Structure	Geometry	Height							
				Inclination angle							
				Body	Height						
				% area							
			Vegetal Coverage	Crown	Height						
			% area								
			Body Surface	Scour	Deeper						
				Width							
				Spacing							
			Grooves	Deeper							
			Width								
			Spacing								
	Improvement	Stabilization	Global or deeper	Bolts	Bolt	Integrity					
					Body cut						
					Width						
		Reinforced concrete	Cracks	Spacing							
			Spalls	Depth							
		Shallow	Bolts	Bolt	Integrity						
					Body cut						
			Shot Concrete	Cracks	Spacing						
					Body cut						
					Width						
					Spacing						
			Mesh	Integrity							
					Dampness						
			Road Protection	Captation Area	Cleaning						
					Body						

Fig. 2. Clasificación de la información

A partir de los submódulos definidos se desarrollaron los modelos difusos. Los conjuntos difusos de cada modelo se diseñaron a partir de funciones triangulares y trapezoidales según cual represente mejor el comportamiento de cada parámetro. Se emplearon 4 etiquetas de evaluación “muy mal, mal, bien y muy bien”. La evaluación de los factores comunes en inventario y performance se realiza a partir de la comparación entre el valor obtenido en cada fase con la finalidad de observar la evolución de daños.

Los límites de los parámetros cuantitativos se definieron a partir de las clases y límites propuestos por Garnica [10]. Para los factores cualitativos y las respuestas se consideró una escala de evaluación del 0 (peor condición) al 10 (mejor condición). Las reglas de membresía definidas son del tipo Si-Entonces y se definieron a partir de la combinación entre las etiquetas lingüísticas de los factores miembros de un mismo submódulo. Observando tres tipos de interacción. El impacto de los factores: se suma, se contrapone o es independiente.

Como salida el modelo asigna una calificación a partir de los valores particulares de cada uno de los factores evaluados. Para la interpretación de la calificación se realizó un análisis lógico para generar la descripción de las respuestas obtenidas.

Como parte del trabajo futuro, los modelos generados se validarán a partir de su aplicación en la evaluación de un tramo carretero, verificando que las respuestas ofrezcan el cumplimiento de los estándares definidos en las normas correspondientes. Finalmente, se analizarán las ventajas de la implementación del modelo propuesto respecto a la actual forma de evaluar las carreteras en México.

6 Conclusión

El uso de la lógica difusa es una herramienta valiosa en la evaluación de cortes carreteros. Permite determinar su estado a partir de la interpretación de medidas y observaciones en campo de sus características, así como de las relaciones que existen entre estas. Limitando la ambigüedad de resultados inherente a la expertiz del evaluador. La aplicación del modelo propuesto en este artículo permitirá a evaluadores, administradores carreteros o cualquier interesado describir sistemáticamente las condiciones en que se encuentran los cortes carreteros que se desee evaluar.

References

1. Habibagahi, Ghassem, and M. Meidani. "Reliability of slope stability analysis evaluated using fuzzy set approach." Proc. of the 5th Int. Conf. on Civil Eng. Vol. 36. sl; Ferdowsi University, 2000.
2. Zhou, Xiaorong, et al. "Research on Highway Slope Stability Based on Hierarchical Fuzzy Comprehensive Evaluation Method." IOP Conference Series: Materials Science and Engineering. Vol. 569. No. 3. IOP Publishing, 2019.
3. Anuma Reddy, Sai Samanth. Fuzzy Theory Application for Analysis and Assessment of Slope Stability. Agosto 2016.
4. Fuzzy Logic Toolbox MATLAB. MathWorks, 2021.
5. England, Highways. "HD41. Maintenance of highway geotechnical assets." (2015).
6. Auriol, J. C., et al. "Having indicators representative of the condition of geotechnical structures for road assets management." PIARC Activity Report. PIARC World Road Association, 2007. 2004-2007.
7. Daly, Tony, et al. "Use of Geotechnical Asset Data Within Highways England: The Journey so Far and the Future." International Conference on Information technology in Geo-Engineering. Springer, Cham, 2019.
8. Zhang, Jian, et al. "Risk assessment model of expansive soil slope stability based on Fuzzy-AHP method and its engineering application." Geomatics, Natural Hazards and Risk 9.1 (2018): 389-402.
9. Salcedo, Daniela, et al. "Landslide susceptibility mapping using fuzzy logic and multi-criteria evaluation techniques in the city of Quito, Ecuador." Natural Hazards and Earth System Sciences Discussions (2018): 1-33.
10. Anguas, Paúl Garnica, José Antonio Ramírez Culebro, and Carlos Pérez García. "Propuesta de indicadores para la gestión de taludes de corte y de terraplén." (2016).

Índice de Autores

A

Alarcón Ibarra Jorge, 43
Aldeco-Pérez Rocío, 14
Álvarez Cleto, 61
Antolino Hernández Anastacio,
37
Arias Lizarraga Dulce María, 25

B

Barradas Daniela, 19
Bravo Méndez Jesus Alfredo, 37

C

Cantero Ramírez Rosario de La
Luz, 55

D

Daniel Raya Gamiño, 67

F

Fernández Gregorio Sofia Isabel,
31
Ferreira Medina Heberto, 37
Figueroa Karina, 61

G

Gómez Anel, 61
Gallegos Cornejo Juan Carlos, 43
González Huitrón Víctor Alejandro,
25
Gonzalez Bello Victor Manuel,
7

H

Hernández Rodríguez María de
Lourdes, 7

I

Imperial Rojo Rafael, 31

J

Jaime Saavedra Rosales, 67
Jiménez Miramontes José Antonio,
14
Juan Carlos Gallegos Cornejo, 67

L

López Arturo, 61
Luis Alberto Morales Rosales, 67

M

Marin-Castro Heidy M., 49
Mercado Jesús, 61
Meza Aguilar Marco Antonio, 55
Miranda Torres Luis Alberto, 25
Molina De la Fuente José Antonio,
1
Morales Rosales Luis Alberto, 31,
43
Morales Sandoval Miguel, 1
Morales Turrubiates Elvia, 19
Moreno Arely, 49

R

Radilla Flor, 19

Raya Gamiño Eduardo Daniel, 43
Rodríguez González Ansel Yoan,
55
Rodríguez Mata Abraham Efraim,
25, 31
Rodríguez Rangel Héctor, 25, 31

S

Saavedra Rosales Jaime, 43

V

Vazquez-Santacruz Eduardo, 19
Virgilio López Morales, 67

Z

Zavala Emily, 19