APLICACIONES DE LAS CIENCIAS COMPUTACIONALES DURANTE LA PANDEMIA COVID19

María del Carmen Santiago Díaz

APLICACIONES DE LAS CIENCIAS COMPUTACIONALES DURANTE LA PANDEMIA COVID19

APLICACIONES DE LAS CIENCIAS COMPUTACIONALES DURANTE LA PANDEMIA COVID19

María del Carmen Santiago Díaz
Gustavo Trinidad Rubín Linares
Judith Pérez Marcial
Yeiny Romero Hernández
Ana Claudia Zenteno Vázquez
(Editores**)**

María del Carmen Santiago Díaz
(Coordinador)

María del Carmen Santiago Díaz, Gustavo Trinidad Rubín Linares, Judith Pérez Marcial, Yeiny Romero Hernández, Ana Claudia Zenteno Vázquez.
(editores BUAP)

María del Carmen Santiago Díaz
(coordinador BUAP)

María del Carmen Santiago Díaz, Gustavo Trinidad Rubín Linares, Ana Claudia Zenteno Vázquez, Yeiny Romero Hernández, Judith Pérez Marcial, María Blanca del Carmen Bermúdez Juárez , María de Lourdes Sandoval Solís, Luis Carlos Altamirano Robles, Irene Olaya Ayaquica Martínez, Abraham Sánchez López, María Josefa Somodevilla García, Maya Carrillo Ruíz, Hilda Castillo Zacatelco, Luis Enrique Colmenares Guillén, Josefina Guerrero García , Meliza Contreras González, Gabriel Juárez Díaz , José de Jesús Lavalle Martínez, Miguel Ángel León Chávez, Ivo Humberto Pineda Torres, Roberto Contreras Juárez, Graciano Cruz Almanza, Andrés Vázquez Flores, Darnes Vilariño Ayala, Héctor David Ramírez Hernández, Nelva Betzabé Espinoza Hernández , José Martín Estrada Analco, Rogelio González Velázquez, Pedro García Juárez, Beatriz Beltrán Martínez, Ana Luisa Ballinas Hernández, Nicolás Quiroz Hernández , Luz del Carmen Reyes Garcés, Alba Maribel Sánchez Gálvez, Alfredo Toriz Palacios, Rogelio Alfredo Campos Serapio.
(revisores locales)

Eden Belouadah, Jessica Nayeli López Espejel, Paola Eunice Rivera Salas, María Concepción Landa Arnaiz, Francisco Marroquín González, Oleg Starostenko Basarab, Raúl Antonio Aguilar Vera, Juan Pablo Ucán Pech, Carina Toxqui Quitl, Juan Alberto Guevara Jaramillo, Heidy Marisol Marín Castro, Julio Cesar Díaz Mendoza, Hermes Moreno Álvarez
(revisores externos)

Primera edición: 2021
ISBN: 978-607-8728-81-7

Montiel & Soriano Editores S.A. de C.V.
15 sur 1103-6 Col. Santiago Puebla, Pue.

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA
Rectora:
Dra. María Lilia Cedillo Ramírez

Secretario General:
Mtro. José Manuel Alonso Orozco

Vicerrector de Investigación y Estudios de Posgrado:
Dr. Ygnacio Martínez Laguna

Directora de la Facultad de Ciencias de la Computación:
M.I. María del Consuelo Molina García

Contenido

Prefacio

A más de dos años del inicio de la pandemia de COVID-19 el mundo no ha logrado contener sustancialmente la pandemia, muchos países han sufrido las variantes Delta, Omicron y se encuentran apenas saliendo de la cuarta ola de contagios y con la incertidumbre de cuántas olas más surgirán, cuántas variantes y mutaciones, y más aún cuántas vacunas o refuerzos de las existentes vendrán, incluso el tratamiento específico en los hospitales no es caso cerrado, hay una diversidad de estos. Aunque ha disminuido el número de defunciones no ha sido de forma tan radical y no hay garantía de que así sea, en conclusión estamos inmersos en un caos que nosotros mismos generamos, todos los estudios realizados no han podido ser concluyentes y no se ha parado de hacer esfuerzos tanto en generar herramientas de mitigación de la pandemia como para disminuirla, sin embargo, somos los herederos del método científico, la generación que debe resolver los problemas actuales y de generar más y mejor conocimiento así como su aplicación, por ello en este ejemplar se presentan algunos trabajos de Inteligencia Artificial, Ciencia de Datos, Internet de las Cosas, Ciberseguridad y otros, que buscan mostrar aplicaciones directas o indirectas en la actual contingencia del COVID-19.

Actualmente la premisa para generar y aplicar el conocimiento no es solo para brindar mejores condiciones de vida para los seres humanos, sino ha cambiado al punto de generar condiciones de supervivencia, nunca hubiéramos imaginado estar en éste punto, sin embargo, la pelea contra este enemigo invisible debe continuar, y por ello aunque se diga y suene bonito que “nunca hubo tanta información disponible, ni hubo tantos esfuerzos en torno a un mismo problema…”, si debemos continuar avanzando en el desarrollo tecnológico y científico, y sobre todo brindando a las siguientes generaciones las armas del método científico que aunque puede parecer de poca utilidad para algunos, sabemos que es el único capaz de permitirnos cambiar el destino que muchos creen que ya está escrito.

María del Carmen Santiago Díaz
Gustavo Trinidad Rubín Linares

Big Data Actual: Mayor Conocimiento y Mejores Decisiones

María de Lourdes Sandoval-Solís[2] ,Gladys Linares-Fleites[1]
Marcela Rivera-Martínez[2] ,Luis René Marcial-Castillo[2]
[1]Posgrado en Ciencias Ambientales, Benemérita Universidad Autónoma de Puebla.
[2]Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla.
[1]gladys.linares@correo.buap.mx [2]maria.sandoval@correo.buap.mx

Resumen. Inicialmente *Big Data* se refiere a conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño, complejidad y velocidad de crecimiento dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales**.** En este trabajo se introducen algunos conceptos que se manejan en la actualidad sobre *Big Data* y cómo esto se manifiesta en diversos problemas que se presentan en el mundo real, ya que pueden realizarse análisis que permiten obtener conocimientos que conduzcan a mejores decisiones.

Palabras Clave: Minería de Datos, Modelo Lineal, Estadística Bayesiana.

1 Introducción

En la primera década del actual siglo, en ciencias como la astronomía y la genética, se acuñó el término big data, "datos masivos". El concepto está trasladándose ahora hacia todas las áreas de la actividad humana. No existe ninguna definición rigurosa de los datos masivos. En un principio, la idea era que el volumen de información había aumentado tanto que ya no cabía en la memoria de los ordenadores para procesarla, por lo que se necesitaba modernizar las herramientas para poder analizarla [1].
Big Data es un término que describe un gran volumen de datos, tanto estructurados como no estructurados, pero no es la cantidad de datos lo que es importante. Actualmente, lo que verdaderamente importa es lo que puede hacerse con los datos, ya que con Big Data se pueden realizar análisis que permiten obtener conocimientos que conduzcan a mejores decisiones.

En las aplicaciones prácticas estamos interesados en fenómenos del mundo real, ya sean físicos, biológicos, ambientales o sociales, en los que existe una estructura sistemática y una variabilidad aleatoria, y deseamos representar, analizar e interpretar la forma de ambos [2]. Los modelos estadísticos brindan una manera de hacer esto y las probabilidades son entonces miradas como propiedades del sistema bajo estudio, por supuesto, idealizadas. Hay que reconocer que nuestro conocimiento es incierto y deseamos estudiar, entender y hacer cálculos sobre la incertidumbre presente en el fenómeno bajo estudio. Big data ayuda a dilucidar la complejidad de estos enfoques.

El objetivo de este trabajo es introducir algunos conceptos que se manejan en la actualidad sobre Big Data, y cómo se manifiesta en diversos problemas que se presentan en el mundo real. La estructura del trabajo es a siguiente: en el epígrafe 2 se exponen las ideas que se están desarrollando en la actualidad y en el epígrafe 3 se brindan dos temas

de aplicación de estas ideas, como son la simulación y los modelos lineales. Finalmente, se dan conclusiones y se brindan las referencias.

2 Big Data

La naturaleza compleja de Big Data se debe principalmente a la naturaleza no estructurada de gran parte de los datos generados por las tecnologías modernas, como los web logs, la identificación por radiofrecuencia, los sensores incorporados en dispositivos, la maquinaria, los vehículos, las búsquedas en Internet, las redes sociales como Facebook, computadoras portátiles, teléfonos inteligentes y otros teléfonos móviles, dispositivos GPS y registros de centros de llamadas.

A continuación, se hace un análisis sobre este tema.

2.1 ¿Qué es actual *Big Data*?

Antes, cuando hablamos de Big Data nos referíamos a conjuntos de datos o combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales, tales como bases de datos relacionales y estadísticas convencionales o paquetes de visualización, dentro del tiempo necesario para que sean útiles. [3]. Actualmente, al trabajar en tiempo real, las decisiones deben tomar en cuenta la dinámica de cambio del fenómeno estudiado.

Las herramientas cuantitativas más usadas en Big Data son: el "Clustering", la Regresión, la Simulación, la Detección de Anomalías, la Predicción Numérica y la Optimización.

En muchas ocasiones, los investigadores de las ciencias empíricas no logran sacar suficiente ventaja de los resultados estadísticos de sus investigaciones y pierden la oportunidad de discutir ciertas cantidades que enriquecerían las interpretaciones. Las técnicas de simulación estadística, posible en diversos programas de cómputo, permiten extraer mayor información de los datos numéricos. Para mejorar las interpretaciones de los estudios estadísticos, se pueden simular los parámetros del modelo, los valores predichos, los valores esperados y las primeras diferencias, esto es, una primera diferencia entre dos valores esperados. Esto puede realizarse por diversos métodos, como son la simulación de Monte Carlo y el método "bootstrap".

No cabe duda, con la alianza entre el tratamiento de los datos, la estadística, la minería de datos y la inteligencia artificial ofrecen una impresionante herramienta para la previsión y, en su caso, solución de muchos problemas.

3 Algunos Temas Relacionados con Big Data

Llegamos al siglo XXI en medio de dos revoluciones: la de la calidad y la innovación, así como, la de las telecomunicaciones y la informática. También el Cambio Climático ha estado planteando desafíos importantes a la Humanidad y, más recientemente, a principio de 2021, el desafío mayor ha sido la pandemia global, originada por el coronavirus, SARS-CoV-2, comúnmente conocido como COVID19.

A continuación, esbozamos las relaciones de estos dos últimos temas con Big Data y algunas de las experiencias que se han desarrollado.

3.1 Big Data y la lucha contra el cambio climático

No hace falta decir que, sin grandes bases de datos y análisis predictivos, las políticas o planes sobre el cambio climático reducirán mucho su eficacia. Por otra parte, gracias a la simulación, técnica del ***Big Data*** **la lucha del cambio climático** podrá ser mejor calculada y analizada [4]. Un aspecto se refiere a que el cálculo de la cantidad de ***emisiones de carbono*** que se deben controlar en todo el mundo será especificado de manera más exacta. Un corte de emisiones inadecuado significa un calentamiento global en aumento, enfermedades y otros problemas. A través de técnicas de Monte Carlo vía Cadenas de Markov (MCMC), tecnología del *Big Data* se podrá contar con una base de datos más potente y funcional.

Los mecanismos de cambio climático, los impactos, los riesgos, la mitigación, la adaptación y la gobernanza son ampliamente reconocidos como el problema más grande e interconectado que enfrenta la humanidad [5]. Los enfoques de minería de *Big Data* integrados e interdisciplinarios y transdisciplinarios resultantes están surgiendo, en parte con la ayuda del desafío climático de *Big Data* de las Naciones Unidas, algunos de los cuales se recomiendan ampliamente como nuevos enfoques para la investigación del cambio climático.

3.1.1 Predicción de emisiones de CO_2 en Coatzacoalcos, Veracruz, México.

En [6] se desarrolla un caso que persigue el propósito de estu*diar las asociaciones entre las* emisiones de dióxido de carbono (CO_2*) y diversas* propiedades de suelos perturbados en el Parque Ecológico Jaguaroundi en Coatzacoalcos, Veracruz, México. Para ello, se midieron las emisiones de CO_2 y las propiedades del suelo en determinados sitios de muestreo y se decidió utilizar la inferencia estadística bayesiana [7] para el caso de un único predictor: el nitrógeno. De acuerdo a esta metodología, se debe asignar una distribución inicial a la variable respuesta (CO_2) y a los parámetros del modelo de regresión, describiendo así, la incertidumbre sobre los verdaderos valores de estos.
La distribución final, solución Bayesiana a un problema de inferencia, se obtiene a través del Teorema de Bayes, el cual puede re-escribirse:

$$\pi(\theta|x) \;\; \alpha \;\; p(x|\theta)\pi(\theta) \tag{1}$$

donde $p(x|\theta)$ es el modelo probabilístico, para un conjunto de observaciones dado y, $\pi(\theta)$ es la distribución inicial.

La mayor parte del esfuerzo en los procedimientos Bayesianos, se concentra en el cálculo de ciertas características (media, mediana, moda) de la distribución posterior del parámetro de interés.

La implementación de las técnicas de Monte Carlo vía Cadenas de Markov (MCMC) permiten generar, de manera iterativa, observaciones de distribuciones multivariadas que difícilmente podrían simularse utilizando métodos directos. Los dos algoritmos más empleados para construir Cadenas de Markov con las propiedades deseadas son el de Metrópolis-Hastings y el muestreo de Gibbs, este último ha dado un impulso extraordinario a las aplicaciones de las técnicas bayesianas.
El Muestreador de Gibbs parte de los valores iniciales asignados a los parámetros y va generando valores de los mismos; en el primer paso obtiene el del primer parámetro, teniendo en cuenta los valores iniciales de los restantes; en el segundo paso obtiene la estimación del segundo parámetro teniendo en cuenta la estimación actual del primer parámetro y los valores iniciales de los restantes, y así sucesivamente hasta formar una

Cadena de Markov. La idea básica es sencilla: construir una cadena de Markov que sea fácil de simular y cuya distribución de equilibrio corresponda a la distribución final que es de interés.
En el módulo "normal.bayes: Bayesian Normal Linear Regression" del paquete Zelig en R [8], el modelo es implementado usando el Muestreador Gibbs, utilizando las semi-conjugadas priori siguientes:

$$\beta \sim Normal_k(b_0, B_0^{-1}) \quad (2)$$

$$\sigma^2 \sim Inverse\ Gamma\left(\frac{c_0}{2}, \frac{d_0}{2}\right) \quad (3)$$

Para estudiar la convergencia, esto es, para verificar que la Cadena de Markov converge a su distribución estacionaria, realiza varios test diagnósticos, entre los que se encuentra el de Geweke. El test diagnóstico de Geweke prueba la hipótesis nula que la Cadena de Markov está en la distribución estacionaria y produces estadísticos Z para cada parámetro estimado.
Los valores estimados a través de la técnica bayesiana con la priori mencionada antes son: β_0= 15.93, β_1= 401.25 siendo estos valores la media de la distribución a posteriori.
Zelig tiene la posibilidad adicional de simular otras cantidades de interés tal como E(Y|X).

En resumen, una cuestión clave en las aplicaciones es la incorporación del conocimiento de la materia objeto de investigación. La extensión en la que es posible y deseable hacerlo puede variar grandemente entre los diferentes campos de aplicación. Por otra parte, la simulación estadística es recomendada como un método fácil para calcular cantidades de interés y sus incertidumbres y, puede ayudar a los investigadores a mejorar sus interpretaciones.

Debe destacarse la importancia que tienen los estudios sobre el dióxido de carbono (uno de los gases causantes del efecto invernadero) debido a su relación con el fenómeno del calentamiento global. Estos gases son continuamente emitidos y removidos en la atmósfera por procesos naturales sobre la Tierra, pero las actividades antropogénicas causan cantidades adicionales de los mismos, incrementando sus concentraciones en la atmósfera, lo que tiende a sobrecalentarla. En la actualidad existe gran interés científico en modelar la relación que guardan las emisiones del dióxido de carbono con las propiedades del suelo. En el caso presentado, a través de la regresión lineal y utilizando la inferencia bayesiana, se obtiene un modelo para la predicción del carbono orgánico almacenado en suelo y, con el empleo de la simulación, se logra una representación de la ley de distribución los valores esperados E(Y|X), donde sólo permanece la incertidumbre de la estimación.

Como hemos destacado antes, Big Data no solo se refiere a conjuntos de datos o combinaciones de conjuntos de bases de datos y, se puede apreciar en este caso que la aplicación de técnicas de Monte Carlo vía Cadenas de Markov (MCMC), apoyan la predicción del dióxido de carbón y su relación con los suelos.

3.2 Big Data y la epidemia Covid 19

F. J. Sanz, 2020 [9], refiriéndose a Mayer-Schönberger y Cukier [1], plantea: “me impresionó mucho su primer capítulo, donde se contaba cómo en el 2009 se descubrió el virus de la gripe H1N1 y cómo, antes de que los Centros de Control y Prevención de Enfermedades (CDC) de los Estados Unidos lo hubieran detectado, unos ingenieros de Google publicaron un artículo en Nature en el que explicaba cómo este gigante de Internet podía predecir con el tratamiento del Big Data de su buscador la propagación del virus,

gracias a las consultas de sus usuarios sobre temas relativos a los indicios de esta enfermedad".

Comprender la propagación y el crecimiento futuros de COVID19 se complica aún más por los problemas de calidad de los datos [10]. Uno de los desafíos para los investigadores que buscan modelar la epidemia de COVID19 es la mala calidad de los datos sobre el número de infecciones por COVID19. Es probable que esto se deba a una combinación de factores, como la naturaleza nueva de la enfermedad, la falta de pruebas adecuadas y la prevalencia de casos asintomáticos.

Hay dos clases amplias de modelos que se utilizan para este propósito, modelos empíricos y modelos mecanicistas [10]. Los modelos empíricos ajustan una superficie de respuesta a una variable dependiente o función objetivo de respuesta múltiple utilizando múltiples variables de entrada. Si bien los modelos empíricos pueden proporcionar pronósticos precisos a corto plazo, este enfoque tiene varios inconvenientes. Los modelos empíricos pueden estar sobre ajustados, donde las variables de entrada están altamente correlacionadas o son falsas. Estos modelos también se basan en las condiciones actuales observadas y pueden no ser datos precisos observados en el pasado o si las condiciones subyacentes cambian. Los modelos mecanicistas son descripciones matemáticas de un fenómeno o proceso basadas en una comprensión o teoría de cómo se comporta un sistema, en las que la estructura del modelo restringe la forma de la superficie de respuesta potencial. También se parametrizan fácilmente en términos de comportamientos que afectan la tasa de reproducción de la enfermedad, por lo que se prestan a la modelización de escenarios de salud pública. Sin embargo, generalmente son muy sensibles a las variaciones en las condiciones iniciales,

Hay que señalar que existen muchos otros casos de Big Data que tienen problemas con la calidad de los datos y que son similares al problema de modelado de epidemias y, en particular, del modelado de COVID19.

3.3 Ajuste de datos del COVID 19 en el estado de Puebla, México.

La epidemia de COVID 19, es una enfermedad social. La amortiguación de los contagios depende del comportamiento de cada individuo de la Sociedad; se tiene que reducir la interacción entre los individuos y tomar medidas de higiene personal como lavarse las manos constantemente, usar cubre bocas, y llevar a cabo estornudos de etiqueta. También la disminución de los contagios depende de medidas colectivas, como son el mantener desinfectadas las superficies y la ventilación de las áreas cerradas.

Con el objetivo de mostrar el comportamiento social de esta epidemia, en [8] se realiza un estudio de contagios confirmados de COVID 19, a través de datos publicados por la Dirección General de Epidemiologia, de la Secretaria de Salud de México, para el Estado de Puebla, en ciertos intervalos de fechas. El comportamiento de la sociedad se va a reflejar en el cambio de la curva epidemiológica.

Para realizar el ajuste de datos se supone que las curvas epidémicas son gaussianas, esto es, tienen un valor máximo en la media. Por otra parte, se realiza el cálculo del "acmé", que es el nombre que se da en epidemiologia al día en el que se tiene el mayor número de contagios, en este caso es la media de la función gaussiana.

El ajuste de datos de contagios confirmados se realiza con el método de Mínimos Cuadrados Lineales a una exponencial cuadrática. El ajuste se obtiene al minimizar el error cuadrático medio del modelo cuadrático equivalente a ajustar [11], es decir:

$$\begin{aligned} y = M(x; a, b, c) &= \exp(ax^2 + bx + c) \\ \ln(y) = \ln\left(\exp(ax^2 + bx + c)\right) & \\ \mathrm{Y} = \ln(y) &= ax^2 + bx + c \end{aligned} \tag{4}$$

En [11] se realiza con detalles la obtención de las ecuaciones normales para el modelo cuadrático es un sistema de ecuaciones lineales en tres variables, a b y c, y su correspondencia con la función gaussiana.

La figura 1, muestra el ajuste para el estado de Puebla con los datos de los casos confirmados de contagio diario publicados por la Dirección General de Epidemiología del 23 de abril al 23 de mayo del 2020.

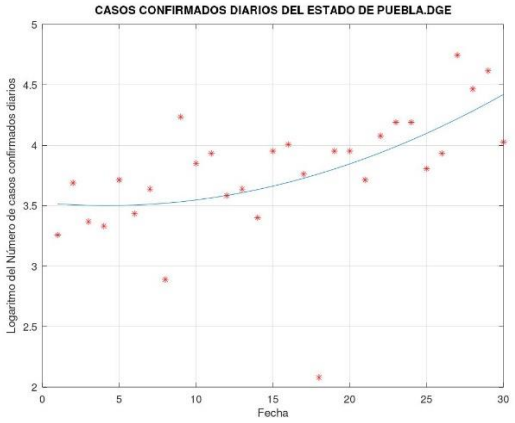


Fig. 1. Gráfica del logaritmo de los Casos Confirmados diarios de COVID 19 para el Estado de Puebla, del 23 de abril al 23 de mayo de 2020, publicados por la Dirección General de Epidemiología

Se muestra en la Fig. 2., la Gráfica del número de Casos Confirmados Diarios del COVID 19 para el Estado de Puebla, del 15 de abril al 31 de julio de 2020, publicados por la Dirección General de Epidemiología.

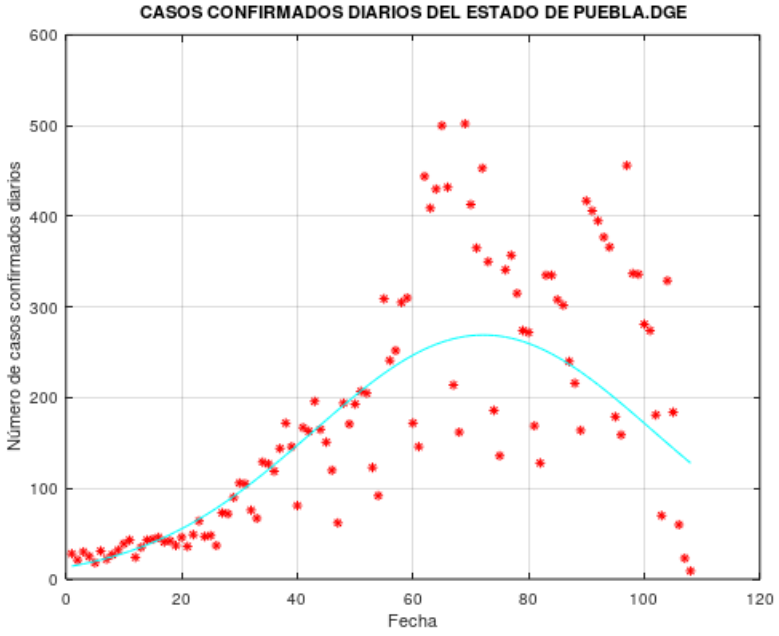


Fig. 2. Gráfica del número de Casos Confirmados Diarios del COVID 19 para el Estado de Puebla, del 15 de abril al 31 de julio de 2020, publicados por la Dirección General de Epidemiología**.**

Se puede concluir que para obtener la campana Gaussiana del comportamiento de los datos de contagio diario confirmados de COVID 19 y determinar el ACMÉ, es posible si se cuenta con suficientes datos de los contagios diarios confirmados

A través de la técnica de Mínimos Cuadrados Lineales se logra ajustar la campana gaussiana a una exponencial cuadrática y calcular el día en el que se tiene el mayor número de contagios (*acmé*).

4 Conclusiones y Trabajos Futuros

Se concluye que, en la actualidad, aprender de datos masivos (Big Data) requiere integrar métodos estadísticos, de investigación operativa y de ciencias de la computación.
La estadística, en su alianza con la computación y las matemáticas, enfrenta el reto de analizar grandes volúmenes de datos en tiempo real, poniendo a disposición de los tomadores de decisiones, los elementos de conocimiento significativo.
No existe aún una teoría sobre cómo hacer la analítica de los problemas donde se trata con Big Data. En particular, la problemática planteada a la Informática lleva al desarrollo de softwares que sean especialmente eficientes.

Referencias

1. Mayer-Schönberger, V., Cukier, K. *Big Data. La revolución de los datos masivos.* Turner Publicaciones S.L., Madrid (2013).
2. King, Gary, Michael Tomz and Jason Wittenberg. *Making the Most of Statistical Analyses: Improving Interpretation and Presentation.* American Journal of Political Science, Vol. 44, No. 2, pp 341-355, 2000.
3. Bouza, C., N. *TIC`s, Big Data, Data Mining : sus Tecnologías y la Estadística* (2019) https://www.researchgate.net/publication/331274879
4. Fernández-Esteban, C. *Cómo el Big Data puede ayudar a luchar contra el cambio climático* (2018). https://www.ticbeat.com/innovacion/como-el-big-data-puede-ayudar-a-luchar-contra-el-cambio-climatico
5. Zhang, Z., Li, J. *Big-Data-Mining-for-Climate-Change*. Elsevier Inc. (2020). .
6. López, A., Linares, G, y Saldaña, J.A. (2010). *Regresión Bayesiana para la predicción de emisiones de CO_2.* 3ra. SIEP. Puebla, Pue., México. Silva, L. C. *El enfoque bayesiano: otra manera de inferir*. Gac. Sanit., 15(4), 341-348, 2001.
7. Imai , Kosuke, Gary King, and Olivia Lau. (2009) ``*Zelig: Everyone's Statistical Software,"* http://gking.harvard.edu/zelig, 1995
8. Sanz, F..J (2020). COVID-19, «big data» y privacidad. https://www.lavozdegalicia.es/noticia/opinion/2020/
9. McCulloh, I., Kiernan, K y Kent, T. (2020). Inferir tasas reales de infección por COVID19 a partir de muertes. Frente. Big Data, 15 de octubre de 2020 https://doi.org/10.3389/fdata.2020.565589/
10. Sandoval, M. de L., Romero, E., Rivera, M. y Linares, G. (2020). *Ajuste de datos del covid 19*. Memorias de la XIII Semana Internacional de la Estadística y la Probabilidad, FCFM-BUAP.

Estimación de Costos de los Sistemas Computacionales usando Descomposición Funcional

Mario Rossainz-López, Carmen Cerón-Garnica, Etelvina Archundia-Sierra
Benemérita Universidad Autónoma de Puebla, Avenida San Claudio y 14 sur, San Manuel, Puebla, Puebla, 72000, México
{mario.rossainz, carmen.ceron, etelvina.archundia}@correo.buap.mx

Resumen. El presente trabajo muestra tres métodos para estimar costos de desarrollo software a través de la descomposición de la funcionalidad. El primer método estima el tamaño del software usando las Líneas de Código (LDC); el segundo método estima el tamaño del software usando los Puntos de Función (PF) y el tercer método muestra como estimar software haciendo uso del modelo constructivo de costos o COCOMO. Se estiman el esfuerzo, el tiempo de desarrollo, el personal requerido y el costo monetario para desarrollar un software. Se utiliza un estudio de caso como ejemplo. Se hace uso del estadístico del Valor Esperado (VE) o de los "Tres Puntos" como un estadístico normalizado para obtener estimaciones optimistas, probables y pesimistas y garantizar que la probabilidad de ocurrencia de las estimaciones realizadas sea muy cercana al 100% con un rango de error mínimo. Regularmente los datos que se utilizan para hacer los cálculos son tomados de bases de datos históricas, sin embargo, en nuestro caso y dado que las estimaciones realizadas en el presente trabajo se realizan a través de un estudio de caso, generaremos datos hipotéticos que simulan dichas bases de datos históricas. Finalmente se muestran las estimaciones realizadas con LDC, PF y COCOMO y se calcula la probabilidad de ocurrencia de la estimación de las LDC del sistema del ejemplo utilizando la distribución normal de probabilidades.

Palabras Clave: LDC, PF, COCOMO, Estimación, Descomposición Funcional.

1 Introducción

En nuestra vida diaria todos los días llevamos a cabo mediciones de algo para así realizar tomas de decisiones importantes respecto a la situación que en ese momento se nos presenta, de forma que podamos seleccionar las mejores alternativas posibles [1]. En la Ingeniería de Software la medición es un proceso fundamental para garantizar la calidad del producto software que se esté desarrollando. Necesitamos cuantificar de alguna manera los objetivos que nos proponemos en la realización del software para que las tomas de decisión que hagamos se basen en hechos cuantificables que garanticen certeza. Podemos categorizar a las mediciones de un sistema software en dos grupos: las mediciones directas y las mediciones indirectas.

Algunas mediciones directas de los sistemas software son: Líneas de código producidas, Velocidad de ejecución, Tamaño en memoria, Defectos encontrados durante el uso del software en un periodo de uso determinado. Algunas mediciones indirectas de los sistemas software son: Funcionalidad, Complejidad, Eficiencia, Fiabilidad, Facilidad de

mantenimiento. Para poder medir un software se requiere tener claro 3 conceptos básicos según definiciones de la IEEE [2]: "Medida" que es un indicador cuantitativo de extensión, cantidad, dimensiones, capacidad y tamaño de algunos atributos de un proceso o un producto; "Métrica" que es cualquier medida o conjunto de medidas destinadas a conocer o estimar ciertas características de un software para realizar comparativas o para la planificación de proyectos de desarrollo e "Indicador" que es una métrica que proporciona una visión a detalle del proceso, del producto o del proyecto software en sí. El dominio de la medición del software se divide en métricas del Producto, del Proceso y del Proyecto. El presente escrito centra su atención en los métodos de estimación de costos del desarrollo de software que utilizan las métricas de Producto, particularmente aquellas que se utilizan para estimar tamaño, esfuerzo y tiempo de desarrollo de un sistema software, para llevar a cabo una toma de decisión importante que consiste en saber en cuanto venderlo, pues estas variables servirán para hacer una estimación del costo del sistema que se pretende desarrollar una vez que se han levantado los requerimientos iniciales del cliente. Se utilizan ejemplos clásicos de los textos más representativos de la Ingeniería del Software y se particulariza en un ejemplo como estudio de caso para aplicar los métodos que se deben seguir al hacer una estimación de costos de desarrollo de sistemas software usando LDC, PF y COCOMO.

2 Modelo Empírico y Modelo Numérico para la Estimación de Software

El uso de modelos en la estimación del software es necesario para generar abstracciones de la realidad que permitan identificar entidades particulares representadas mediante modelos numéricos para la interpretación correcta de las mediciones que se lleven a cabo para dicha estimación [3]. El modelo empírico representa el contexto del mundo real del sistema software a implementar, mientras que el modelo numérico representa el modelo formal de las medidas obtenidas del contexto empírico al aplicar la estimación. El modelo empírico puede ser representado de muchas formas. Quizá la forma más simple sea a través de bases de datos históricas. Esas medidas son utilizadas en el modelo numérico para que, mediante la aplicación de estadísticos, se obtenga una estimación numérica de la aproximación que se quiere encontrar y una vez obtenida esa estimación poder hacer la interpretación del resultado de dichos cálculos para la toma de decisiones. Las bases de datos históricas pueden contener información diversa y varía de empresa a empresa o de persona a persona, dependiendo de lo que se quiera registrar como medidas que sirvan posteriormente para estimar cosas.

3 Métricas Orientadas al Tamaño

Una métrica de software es una medida o conjunto de medidas que se utilizan para estimar alguna característica del nuevo software a desarrollar y poder llevar a cabo una toma de decisión basada en una determinada probabilidad de éxito. Las métricas consideradas en el presente trabajo para estimar la complejidad de un software en cuanto al tamaño son dos: Líneas de código (LDC) y Puntos de Función (PF). Para la estimación es necesario hacer una descomposición funcional del software que se está estimando.

3.1 El valor esperado o estimación de los tres puntos

Sin importar la variable de estimación deberemos aplicar un estadístico al rango de valores utilizados para cada funcionalidad en la descomposición del software que se realice dentro del domino de la información. Este estadístico llamado de los tres puntos o valor esperado (VE) es una combinación estadística de 3 valores que se obtienen de la base de datos histórica de nuestro modelo Empírico, la estimación empírica "optimista" (Sopt), "más probable" (Sm) y "pesimista" (Spess); de tal manera que el VE para una variable de estimación cualquiera se calcula como la media ponderada de las estimaciones optimista, más probable y pesimista a través del estadístico de los tres puntos que tiene una distribución normal de probabilidades [4]: VE = (Sopt + (4*Sm) + Spess) /6.

3.2 Líneas de código (LDC)

Existe mucha controversia en el uso de las LDC como métrica para la estimación de medidas en el desarrollo de software como puede ser el tamaño de este y que a partir de ahí se pueda estimar el costo de su realización. Algunos opinan que es una buena métrica pues en la mayoría de las bases de datos empíricas de las empresas incluyen en sus datos empíricos las LDC del software que han desarrollado y que han sido normalizados con el uso del VE. Sin embargo, la principal desventaja del uso de las LDC como métrica de estimación es la dependencia que ésta tiene con el lenguaje que se utiliza para desarrollar el software [5].

3.3 Puntos de función (PF)

Los Puntos de Función (PF) son una métrica orientada a la estimación del tamaño de un software que es independiente de la tecnología que se utilice para su implementación. Se derivan de una relación empírica según las medidas registradas en las bases de datos históricas. Los PF miden el tamaño de un sistema software en términos de la funcionalidad que tenga el mismo y el uso de esta métrica da como resultado un valor normalizado. Los puntos de función se calculan obteniendo las medidas que se muestran en la Fig.1, de 5 componentes básicos del dominio de la funcionalidad que todo sistema software posee y obtener una Cuenta Total que representa el cálculo de los PF sin ajuste.

		Factor de Ponderación					
Parámetro de Medición	**Cuenta**		**Simple**	**Medio**	**Complejo**		
Núm. de entradas de usuario		x	3	4	6	=	
Núm. de salidas de usuario		x	4	5	7	=	
Núm. de consultas de usuario		x	3	4	6	=	
Núm. de archivos		x	7	10	15	=	
Num. de Interfaces externas		x	5	7	10	=	
Cuenta= Total →							

Fig. 1. Componentes para el cálculo de los PF sin ajuste [Creación Propia]

La fórmula que se utiliza para el cálculo de los PF es: Cuenta_Total*[0.65+ (0.01*∑Fi)]. Donde Cuenta_Total es la suma de todas las entradas de los 5 componentes de la tabla mostrada en la Fig 1, y ∑Fi (con i=1..14) son los valores de ajuste de la complejidad

(desempeño) del software según las respuestas de 14 preguntas destacadas. Cada pregunta se responde con alguno de los siguientes valores: 0 = No influye / No se requiere, 1 = Incidental, 2 = Moderado, 3 = Medio, 4 = Significativo, 5 = Esencial (ver [5] para detalles).

4 Ejemplo como Estudio de Caso

Para ilustrar el método que se utiliza para estimar el tamaño, esfuerzo, tiempo y costo de un sistema software tanto con LDC como con PF se utilizará el ejemplo recuperado de [2] como estudio de caso y consiste en suponer la implementación de un sistema software de Diseño Asistido por Computadora (CAD) de componentes mecánicos. Después de levantar los requisitos iniciales del sistema, la revisión indica que: El sistema CAD deberá ejecutarse en una workstation y debe interactuar con varios dispositivos periféricos como: ratón, digitalizador, pantalla a color de alta resolución y una impresora láser; aceptará datos geométricos en 2D y 3D por parte del usuario quien se conectará y controlará el sistema por medio de una GUI, y el sistema CAD manejará una base de datos que contendrá todos los datos geométricos y la información de soporte del sistema. Además, se deberán desarrollar módulos de análisis de diseño de componentes mecánicos para producir la salida requerida que podrá visualizarse en los dispositivos gráficos contemplados y el sistema deberá estar diseñado para controlar e interconectarse con los dispositivos periféricos considerados.

4.1 Estimación usando LDC

PASO 1: Se lleva a cabo una descomposición del sistema CAD y se les asigna una nomenclatura abreviada para su fácil manejo: Interfaz de usuario gráfica y facilidades de control (IUFC), Análisis Geométrico 2D (AG2D), Análisis Geométrico 3D (AG3D), Gestión de BD (GBD), Facilidades de Presentación gráfica de computadora (FPGC), Control de Periféricos (CP), Módulos de Análisis de Diseño (MAD).
PASO 2: Supondremos a través de datos propios simulados, una base de datos histórica de 5 proyectos CAD implementados con anterioridad y similares al que se está estimando de forma que se tienen las medidas de las funcionalidades descritas en dicha base de datos para poder hacer los cálculos de la estimación del número de LDC aproximadas para codificar cada funcionalidad como se indica en Tabla 1.

Tabla 1. Base de datos Histórica simulada de datos propios de las funcionalidades del sistema CAD.

Columna1	IUFC	AG2D	AG3D	GBD	FPGC	CP	MAD	TOTAL LDC
Proy Cad 1	2100	4900	4600	3300	5210	1900	8000	30010
Proy Cad 2	1998	5300	5100	4500	4100	2450	7995	31443
Proy Cad 3	2500	5100	8600	4450	5600	2510	9000	37760
Proy Cad 4	1850	5150	7550	3890	3900	1800	8900	33040
Proy Cad 5	2345	4980	500	3910	5500	2000	9900	33635
Proy Cad 6	2164	5091	6313	3973	4825	2140	8822	33328
Promedio	2159	5086	6170	4010	4862	2132	8759	

PASO 3: Se estiman las LDC aproximadas de cada funcionalidad del "PROY CAD 6" en base a las medidas registradas de los anteriores sistemas CAD utilizando el

estadístico del VE. Por simplicidad sólo mostramos el cálculo realizado para la funcionalidad del AG3D (ver Fig. 2): *Sopt*= 4600 LDC, *Spess*= 8600 LDC, *Sm* = 6170 LDC, *VE = (4600 + (4*6170) + 8600) /6 = 6313 LDC estimadas*

PASO 4: Se suman las LDC estimadas que se obtuvieron para cada funcionalidad y obtenemos las LDC totales estimadas del sistema PROY CAD 6 que en el ejemplo es de 33328 LDC estimadas (ver Tabla 1).

PASO 5: Suponiendo una productividad media de LDC por parte del equipo de trabajo (según base de datos histórica) de 620 LDC/PM, donde PM significa Persona-Mes y una tarifa laboral de $900.00 US/mes podemos estimar el costo de una LDC aproximada en $900.00 US /620 LDC = $1.45 US. Por lo tanto, el costo total estimado del sistema "PROY CAD 6" es de: 33328 LDC * $1.45 US = $48325.60 US. Suponiendo un equipo de trabajo de 5 personas, el esfuerzo de trabajo estimado que realizarían cada una de ellas es de: 33328 LDC / 620 LDC/PM = 54 PM. Con el esfuerzo calculado podemos estimar el tiempo de desarrollo del sistema en: 54 PM / 5 Personas = 10.8 meses.

4.2 Estimación usando PF

PASO 1: Se lleva a cabo una descomposición funcional del software centrada en los valores o medidas del dominio de la información en base a los componentes de la Fig. 1, para el cálculo de los PF sin ajuste. La Base de Datos histórica hipotética y simulada que se utilizó se muestra en la Fig. 2.

	Entradas de usuario					
	Optimista	Probable	Pesimista	Valor Esperado	Peso	Cuenta PF
PROY CAD 1	22	29	30	28	4	112
PROY CAD 2	18	20	30	21	6	128
PROY CAD 3	20	25	28	25	4	99
PROY CAD 4	18	23	30	23	4	93
PROY CAD 5	19	25	31	25	3	75
PROY CAD 6	20	24	30	24	4	96

	Salidas de Usuario					
	Optimista	Probable	Pesimista	Valor Esperado	Peso	Cuenta PF
PROY CAD 1	10	13	21	14	4	55
PROY CAD 2	12	14	23	15	5	76
PROY CAD 3	15	17	24	18	7	125
PROY CAD 4	13	15	21	16	4	63
PROY CAD 5	12	16	19	16	5	79
PROY CAD 6	12	15	22	16	5	80

	Consultas / Peticiones / Interacciones					
	Optimista	Probable	Pesimista	Valor Esperado	Peso	Cuenta PF
PROY CAD 1	12	13	28	15	6	92
PROY CAD 2	15	33	35	30	4	121
PROY CAD 3	15	16	18	16	3	49
PROY CAD 4	18	19	21	19	4	77
PROY CAD 5	24	29	32	29	3	86
PROY CAD 6	16	22	28	22	4	88

	Archivos					
	Optimista	Probable	Pesimista	Valor Esperado	Peso	Cuenta PF
PROY CAD 1	5	6	6	6	15	88
PROY CAD 2	3	4	6	4	7	29
PROY CAD 3	4	5	6	5	7	35
PROY CAD 4	3	3	4	3	10	32
PROY CAD 5	3	3	4	3	7	22
PROY CAD 6	4	4	5	4	10	40

	Interfaces Externas					
	Optimista	Probable	Pesimista	Valor Esperado	Peso	Cuenta PF
PROY CAD 1	2	3	6	3	5	17
PROY CAD 2	1	1	2	1	7	8
PROY CAD 3	1	2	3	2	7	14
PROY CAD 4	4	5	1	4	10	42
PROY CAD 5	1	2	1	2	7	12
PROY CAD 6	2	2	3	2	7	14

Fig. 2. Datos históricos simulados para sistemas CAD de componentes mecánicos.

PASO 2: Para el "PROY CAD 6" se realiza la estimación de sus 5 componentes para la obtención de los PF sin ajuste. Tomemos como ejemplo la tabla de "Entradas de Usuario" de la Fig. 2, y estimemos las medidas optimista, probable y pesimista del "PROY CAD 6", aplicando el estadístico del VE o de los Tres Puntos: *Optimista* = [18+(4*Prom(22,18,20,18,19)+22]/6=20, *Probable* = [20+(4*Prom(29,20,25,23,25)+29]/6 =24, *Pesimista* = [28+(4*Prom(30,30,28,30,31)+31]/6=30. De la misma manera estimemos el Peso o ponderación: *Sopt*=3, *Sm*=Prome(4,6,4,4,3), *Spess*=6, VE= [3+(4*Prom(4,6,4,4,3) +6]/6=4. Con los resultados obtenidos obtengamos el valor esperado del número de entradas de usuario que seguramente se necesitarán implementar como parte del desarrollo del PROY CAD 6 y usemos este valor multiplicado por el peso o ponderación estimada para calcular la CUENTA TOTAL de PF sin ajuste de esta funcionalidad: Sopt =Optimista=20, Sm=Probabl =24, Spess=Pesimista=30, Valor Esperado=[20+(4*24)+30]/6=24. CUENTA TOTAL PF=Valor

Esperado*Peso=24*4=96. De la misma forma obtenemos las cuentas totales de los PF sin ajuste del resto de los componentes en las tablas de la Fig. 2, para obtener la cuenta total de todo el sistema CAD, la cual se muestra en la tabla 2.

Tabla 2. Tabla que muestra la CUENTA-TOTAL de los PF sin ajuste de todo el sistema CAD

	Optimista	Probable	Pesimista	VE	Ponderación	Cuenta PF
Entradas de Usuario	20	24	30	24	X 4	96
Salidas de Usuario	12	15	22	16	X 5	80
Consultas	16	22	28	22	X 4	88
Archivos	4	4	5	4	X 10	40
Interfaces Externas	2	2	3	2	X 7	14
Cuenta-Total						318

PASO 3: Supongamos que obtenemos la sumatoria de las respuestas de las 14 preguntas destacadas con un valor de 52.
PASO 4: Estimamos los PF del sistema CAD usando la fórmula 5: PF=*318*[0.65+(0.01* 52)]=372*.
PASO 5: Suponiendo una productividad media de PF por parte del equipo de trabajo según base de datos histórica de 6.5 PF/PM, donde PM significa Persona-Mes y una tarifa laboral de $900.00 US/mes podemos estimar el costo de un PF aproximado en $900.00 US /6.5 PF/PM = $138.46 US. Por lo tanto, el costo total estimado del sistema "PROY CAD 6" es de: 372 PF * $138.46 US = $51507.12 US. Suponiendo un equipo de trabajo de 5 personas, el esfuerzo de trabajo estimado que realizarían cada una de ellas es de: 372 PF / 6.5 PF/PM = 57.23 PM. Con el esfuerzo calculado podemos estimar el tiempo de desarrollo del sistema en: 57.23 PM / 5 Personas = 11.5 meses.

5 Estimación de Costos usando COCOMO

COCOMO (COnstructive COst MOdel) es un modelo empírico de estimación de costos creado por Bohem en 1981 que proporciona estimaciones directas del esfuerzo y tiempo de desarrollo de un software [6].

5.1 Submodelos de COCOMO

Modelo 1: El modelo COCOMO básico calcula el esfuerzo y el costo del desarrollo de software en función del tamaño del programa, expresado en LDC. Modelo 2: El modelo COCOMO intermedio calcula el esfuerzo del desarrollo de software en función del tamaño del programa y de un conjunto de "conductores de costo".

Modelo 3: El modelo COCOMO avanzado incorpora todas las características de la versión intermedia y lleva a cabo una evaluación del impacto de los conductores de costo en cada fase del proceso de ingeniería de sofware. El modelo COCOMO II de Bohem [7] incorpora en su actualización 81 elementos adicionales que permiten hacer mejores estimaciones en función de las técnicas y tecnologías de desarrollo de software que existen en la actualidad. Para el presente trabajo adoptamos el modelo COCOMO básico para aplicarlo al ejemplo del sistema CAD. Éste se clasifica en: ORGÁNICO: Desarrollo del software en un entorno estable, con poca innovación técnica, con pocas presiones de tiempo y tamaño relativamente pequeño (<50 KLDC). EMPOTRADO: Desarrollo de Software con requisitos muy restrictivos, con gran volatilidad de requerimientos,

complejo, en un entorno con gran innovación técnica. SEMI-LIBRE (SEMIACOPLADO): Situaciones entre el modo orgánico y el empotrado. En la Tabla 3, se muestran los estadísticos del COCOMO básico que se utilizan para estimar Esfuerzo y Tiempo de desarrollo de software en sus tres tipos (orgánico, semi-libre y empotrado) [6, 7]. La estimación del esfuerzo se basa en KLDC y un multiplicador de atributos (M) que por simplicidad vale 1 (consultar [8] para los detalles).

Tabla 3. Estadísticos del COCOMO básico para estimar esfuerzo y tiempo de desarrollo en software orgánico, semi-libre y empotrado

Modo de Desarrollo	Personas/mes (Nominal)	Tiempo de Desarrollo (nominal)
Orgánico	$E= 3.2 * KLDC^{1.05} *M$	$D = 2.5 * E^{0.38}$
Semi-Libre	$E= 3.0 * KLDC^{1.12} *M$	$D = 2.5 * E^{0.35}$
Empotrado	$E= 2.8 * KLDC^{1.20} *M$	$D = 2.5 * E^{0.32}$

5.2 Estimación del sistema CAD usando COCOMO

Tomemos las LDC estimadas del sistema CAD de la sección 4.1 y estimemos el esfuerzo, tiempo de desarrollo y número de personas necesarias para implementar el sistema [9], utilizando la tabla 3. Obtenemos entonces los resultados de la Tabla 4.

Tabla 4. Estimación del Esfuerzo, Tiempo de Desarrollo y Personal-Requerido del Sistema CAD utilizando COCOMO básico con LDC

Sistema CAD	33328 LDC estimadas		
COCOMO BÁSICO	**Esfuerzo (E) Personas-mes (nominal)**	**Tiempo de Desarrollo (D) – meses (nominal)**	**Personal Necesario N=E/D**
Orgánico	127	16	8
Semi-libre	152	15	10
Empotrado	188	13	14

6 Resultados de la Estimación de Costos del Sistema CAD

La tabla 5, muestra las estimaciones obtenidas con LDC y PF del Esfuerzo, Tiempo de Desarrollo, Personal Requerido y Costo, incluyendo los resultados obtenidos con COCOMO básico.

Tabla 5. Estimaciones obtenidas del sistema CAD utilizando LDC, PF y COCOMO básico

Cálculo estimado de costos del sistema CAD					
	LDC/PF	**Esfuerzo (PM)**	**Tiempo Desarrollo (meses)**	**Equipo Humano**	**Costo $US**
LDC	33328	54	10.8	5 personas	48325.6
PF	372	57	11.5	5 personas	51507.12
LDC COCOMO	33328	127	16	8 personas	48325.6
PF COCOMO	39804	153	17	9 personas	51507.12

Finalmente, si calculamos la distribución normal de probabilidades de las LDC totales de cada uno de los proyectos CAD registrados en la base de datos histórica del ejemplo incluyendo el sistema CAD estimado obtenemos la gráfica que se muestra en la Fig. 3.

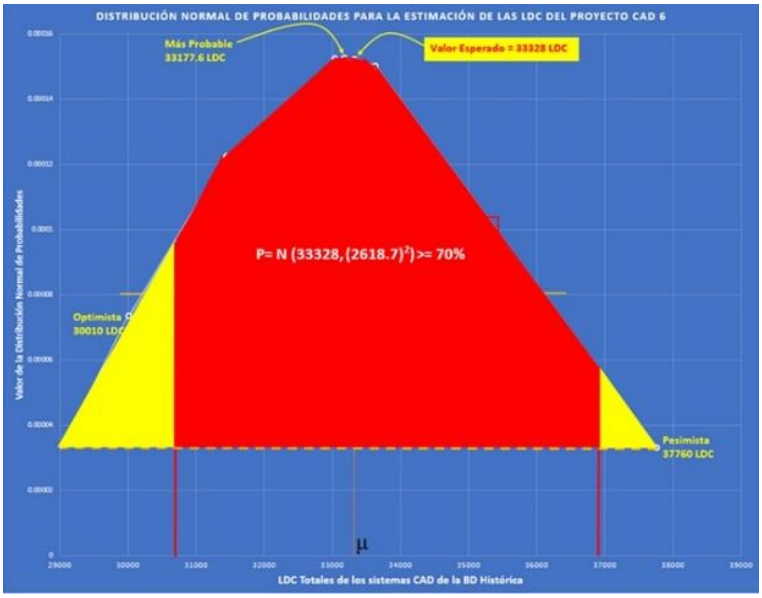


Fig. 3. Probabilidad de ocurrencia del VE=33328 LDC del sistema CAD con una distribución normal de probabilidades [creación propia]

7 Conclusiones

Hemos mostrado tres métodos de estimación de costos en el desarrollo de sistemas computacionales basados en la descomposición funcional: el método que utiliza LDC para la estimación, el método que utiliza los PF y el método que hace uso de los estadísticos del modelo COCOMO. En todos ellos los pasos de cada metodología hacen uso del mismo ejemplo como estudio de caso: El desarrollo de un Sistema Asistido por Computadora de componentes mecánicos para ingenieros. A través de dicho ejemplo, se realizaron además estimaciones de Esfuerzo (personas/mes), Tiempo de desarrollo (en meses), Personal Necesario para la implementación del sistema y costo económico del desarrollo de dicho sistema. Para todos los cálculos se utilizan estadísticos normalizados y de forma específica el uso del VE o de los tres puntos como un estadístico importante que ajusta y normaliza con una distribución normal de probabilidades los resultados, producto del uso de las bases de datos históricas hipotéticas que representan la parte empírica o la experiencia registrada de la empresa, equipo de trabajo o persona que realiza la estimación. Finalmente calculamos la probabilidad de ocurrencia de las LDC estimadas del ejemplo del sistema CAD para demostrar que el rango de error de esa estimación es pequeño comparado con el resultado probabilístico lo que da certeza y seguridad de que los cálculos están bien hechos y muy cercanos a lo que realmente debiera pasar con el desarrollo del sistema.

Referencias

1. Piattini V., García Rubio F. (2019). *Medición de Software*. Editorial Ra-Ma. México.
2. Pressman R.S. (2010*). Ingeniería del Software. Un enfoque Práctico*. Séptima Edición. Editorial

McGraw-Hill. México.
3. Fenton, N.E. y Pfleeger, S.L., (2015). *Software metrics. A rigorous & practical approach* , Third Edition. CRC-Press USA.
4. Gascón Busio (2019). *Todo PMP & Agile*. Recuperado de: https://todopmp.com/estimar-la-duracion-las-actividades/
5. Villareal D. (2007). *Estimación de Esfuerzo y Costo en la Producción de Software hecho en Venezuela: Estudio de casos. Proyecto de Pregrado.* Universidad de los Andes. Venezuela. Recuperado de http://bdigital.ula.ve/storage/pdftesis/pregrado/tde_arquivos/8/TDE-2007-06-29T10:19:15Z-327/Publico/Dixon%20Villareal.pdf
6. Boehm, B. (1981). *Software Engineering Economics*, Prentice Hall. USA.
7. Boehm, B. (2000), et al. (2000). *Software Cost Estimation in COCOMO II*, Prentice Hall. USA.
8. Garita González G., Lizano Madriz F. (2018). *Estimación de Costo de Software. Una propuesta de aplicación pedagógica de COCOMO*. Universidad Nacional de Costa Rica. Revista UNICIENCIA. Volumen 32, Número 1.
9. Fiallos Ordoñez A. (2015). *Improving Productivity Software through the adaptation of an agile development framework.* Enfoque UTE, V.6-N.2, Jun.2015, pp.117 – 134. e-ISSN: 1390-6542 / p-ISSN: 1390-9663. Recuperado de http://scielo.senescyt.gob.ec/pdf/enfoqueute/v6n2/1390-6542-enfoqueute-6-02-00117.pdf

Ataques de Web Scraping

Angel T. De la Cruz, Ana C. Zenteno, Ma. del Carmen Santiago, Yeiny Romero, Judith Pérez, Gustavo T. Rubín

Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Ciudad Universitaria, 14 sur y Avenida San Claudio, Fraccionamiento Jardines de San Manuel, CP. 72570, Puebla, Pue; México.
angel.delacruzt@alumno.buap.mx,{ana.zenteno, marycarmen.santiago, yeiny.romero, judith.perez, gustavo.rubin}@correo.buap.mx

Resumen. En este trabajo se realiza análisis sobre los conceptos de Spidering y Web Scraping y cómo estos se pueden utilizar para realizar ataques a sitios web los cuales se discuten detalladamente, en particular un caso de prueba realizado mediante una herramienta diseñada para realizar spidering. Aunque estos ataques no parezcan tan peligrosos a simple vista es verdad que este tipo de ataques se pueden utilizar para obtener diferente información de un sitio web para diferentes propósitos como por ejemplo el obtener ventaja competitiva en el mercado.

Palabras Clave: Araña**,** Ataque**,** Crawler**,** Robot, Web Scraping.

1 Introducción

La navegación a través de sitios web se ha convertido en una actividad tan sencilla y cotidiana de realizar a través de los años, donde con solo un click realizamos búsquedas exitosas de lo que se quiere encontrar. Para ello los motores de búsqueda utilizan herramientas que les permiten de forma automática explorar en muy poco tiempo una gran cantidad de sitios web para desplegar un conjunto de resultados relevantes para la búsqueda realizada.

Los motores de búsqueda como Google utilizan Web Crawlers (Rastreadores Web) para indexar las búsquedas que el usuario realiza, estos web crawlers también conocidos como Arañas se dedican a buscar información a través de diversos sitios web para al final mostrar los resultados de una búsqueda ingresada por el usuario. A grandes rasgos es el principal propósito que tienen que cumplir estas arañas cada vez que se realiza una búsqueda en el navegador, su funcionamiento detallado se explicará más a profundidad en la sección 2 y 2.1.

Hasta ahora se explicó el caso de una búsqueda promedio realizada por un usuario, pero este concepto puede llegar más lejos para realizar una serie de búsquedas automatizadas, como ya se había mencionado previamente el usuario desea obtener la información que sea relevante para él pero cuando el usuario hace una investigación más profunda dentro de un sitio web este es un proceso muy lineal y tardado, por esta razón los analistas y computólogos utilizan técnicas de Scraping con el fin de obtener grandes cantidades de información en el menor tiempo posible, ya que como sabemos una búsqueda para una persona puede tener muchas variantes que ralentizan el tiempo invertido en ello, lo que para una búsqueda de manera automatizada le puede tomar segundos, los fines más comunes de este tipo de prácticas suelen ser el data mining (Minería de datos) y una forma de espiar a la competencia en ámbitos orientados al mercado electrónico.

2 Web Scraping

Cuando hablamos de navegación web se hace referencia al proceso en el cual un usuario visita un sitio web y realiza una exploración a través del mismo, observa la información en el sitio, sigue diferentes hipervínculos dentro del sitio, copia información o guarda la información que le sea de interés.
Este es un concepto con el que hemos estado familiarizados ya que el Web Scraping[1,2,3] o Raspado Web tiene sus fundamentos en la navegación a través de un sitio web de manera sistemática.

El Web Scraping es una técnica que busca y copia todo el contenido de un sitio web, básicamente copia toda la estructura del sitio web y genera un mapa del mismo, por lo que conoce muchas cosas sobre el sitio, como son el contenido HTML, scripts, estilos, imágenes, etc.

Este proceso se realiza principalmente por las arañas o Crawlers en los motores de búsqueda, por lo que no hay que confundirlo con el proceso de scraping[4] pero también puede ser utilizado por otro tipo de herramientas con diferentes fines, entre los que destaca el robo de información con derechos de autor, Marketing, Comercio electrónico, e inclusive Phishing, ya que al tener toda la información del sitio es posible hacer un análisis de la competencia para posicionar mejor a la empresa en el mercado, pero el uso scraping puede ser aplicado a la creación de sitios web falsos para la obtención de datos personales o sensibles, como tarjetas de crédito, correos, direcciones, entre otros datos ya que el usuario no podría diferenciar entre el sitio legítimo y el sitio falso.

3 Arañas (Crawler)

Las arañas o también conocidas como Robots o Crawlers[5, 6, 7], son las encargadas de visitar los diversos sitios web, su funcionamiento consiste principalmente en comenzar a explorar un conjunto de URLs y cada una es llamada semilla, ya que comienza con una semilla, descarga el contenido del sitio web y lo analiza, obteniendo así más hipervínculos con nuevas URL que visitar, con cada hipervínculo añadido, la araña va realizando una lista de URLs visitadas la cual va a ser la lista de resultados desplegada al usuario.
Este proceso se puede resumir como se muestra en la fig 1.

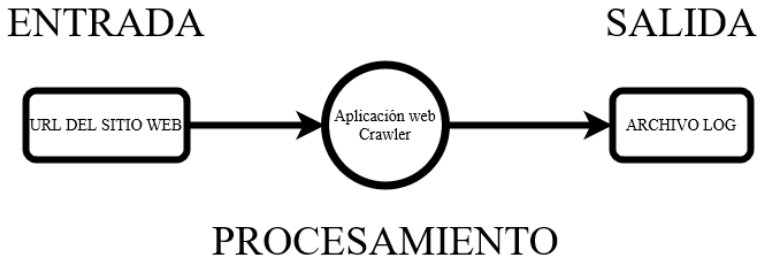


Fig 1. Procedimiento de búsqueda

Dado que existe una cantidad de Sitios Web que la araña podría visitar, se suele guiar por una serie de políticas para discernir entre qué páginas son o no relevantes, como lo son el número de referencias que se le hacen al sitio, el número de visitas e inclusive que la estructura del sitio web sea correcta, esto se puede ver en la Figura 2 y 3. Las aplicaciones crawlers también deciden qué Sitios Web deben explorar con base en el protocolo de

archivo robots.txt (también conocido como protocolo de exclusión de robots) la araña revisará este archivo el cual debe estar almacenado en la raíz del Sitio Web donde se encuentra la página principal del sitio.

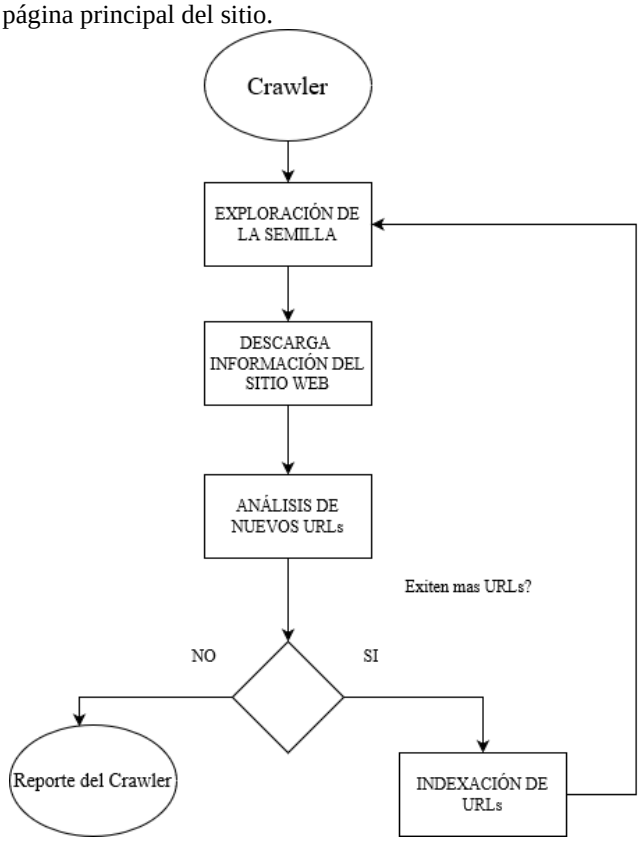


Fig. 2. Diagrama de funcionamiento de un crawler

Un archivo robots.txt [8] es un archivo de texto que especifica las reglas de acceso para las arañas y les indica qué páginas del sitio web pueden o no indexar y qué links pueden utilizar, lo cual principalmente está pensado para las arañas de motores de búsqueda comunes, los bots pensados para atacar sitios web están programados para ignoraran este archivo.

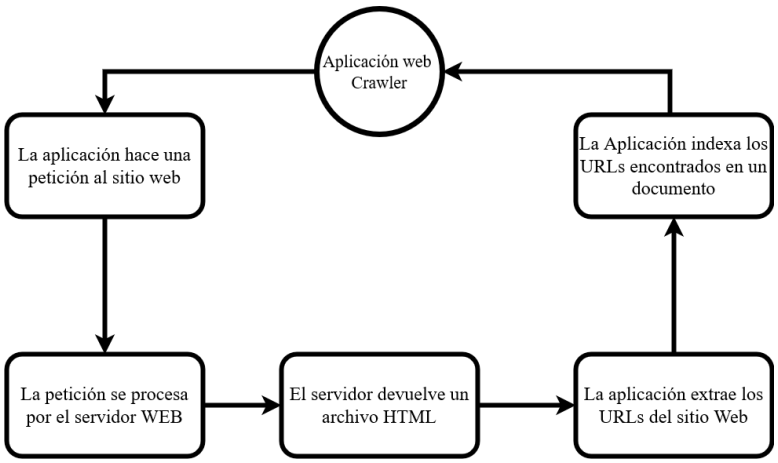


Fig. 3.. Proceso de indexado

4 Pruebas

Para la realización de las pruebas, se utilizó la herramienta ZAP[9] la cual al ser una herramienta de auditoría nos permite realizar un ataque de Crawling así como encontrar vulnerabilidades de un sitio web de nuestra elección, Para las pruebas realizadas se decidió atacar un sitio web almacenado en un hosting gratuito y se atacó mediante ZAP para buscar y descargar toda la información del sitio web, como se puede observar en la figura 4.
Para comenzar con el ataque, se debe introducir la URL del sitio que se desea atacar, esta URL será nuestra primera semilla para que la araña comience su búsqueda a través del sitio web, conforme vaya explorando el sitio irá añadiendo más semillas que explorar.

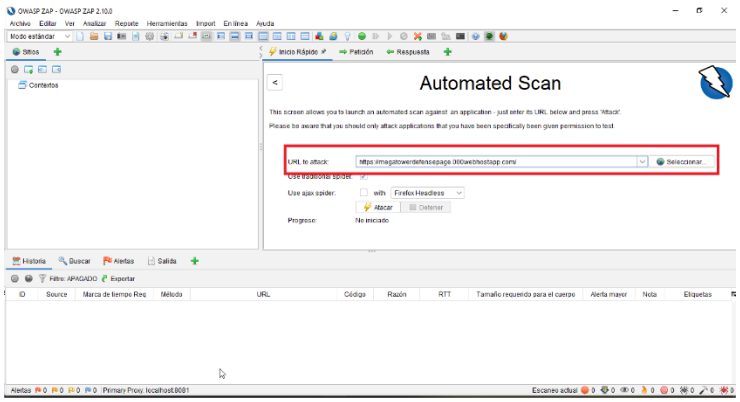


Fig. 4. Ataque Crawling al sitio web con ZAP

Ya que se iban explorando las URL, se fueron agregando más semillas, esto se puede observar en la figura 5.

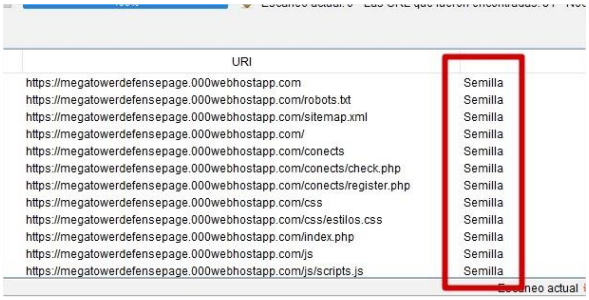


Fig. 5.. Semillas encontradas

Este ataque al sitio web nos obtuvo todos los archivos del sitio, desde los archivos de conexión en PHP hasta los estilos y scripts, así como imágenes asociadas.

Con el ataque Scraping efectivamente se obtuvo la información del sitio atacado consiguiendo las carpetas desde las carpetas css, js,icons, y desplegó una petición get del archivo robots.txt como se muestra en la figura 6.

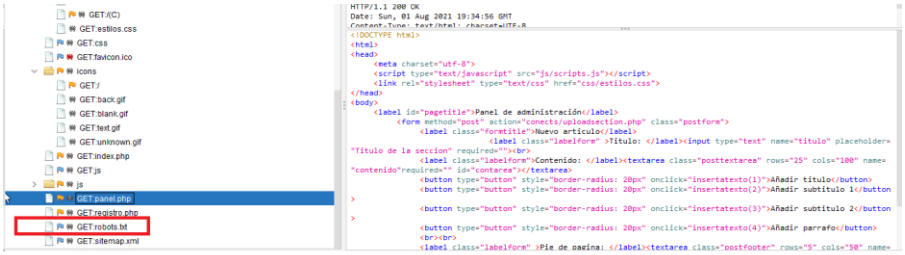


Fig. 6. Scraping del sitio

A la hora de realizar el ataque la herramienta también nos muestra la versión de HTTP, la fecha, el tipo de contenido, el tamaño del contenido, el tipo de conexión, última modificación, el servidor, así como el tipo de seguridad del hosting de cada uno de los elementos que desplegó como se muestra en la figura 7.

Fig. 7. Información de seguridad y conexión el sitio web

Una parte importante además de la herramienta utilizada es que nos muestra una serie de alertas a manera de reporte de lo sucedido durante el ataque como se puede observar en la figura 8 estas alertas se van clasificando por tipo de vulnerabilidad y nos dan información detallada del ataque.

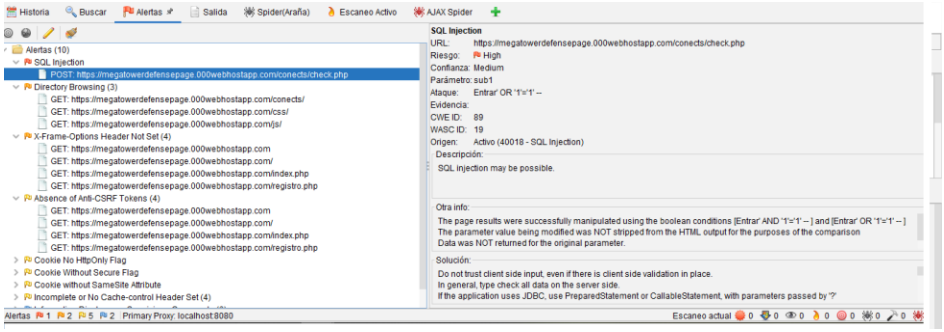


Fig. 8. Alertas y su Descripción

5 Resultados

Con base en las pruebas realizadas en ZAP, observamos que es muy fácil obtener todo un sitio web y copiarlo, ya que estas herramientas consultan todo el contenido del sitio y obtienen el contenido de las mismas. Existen además herramientas de Parsing para realizar el análisis sintáctico de los resultados encontrados por medio del scraping y de

esta forma posteriormente realizar el análisis apropiado y obtención de la información que más interese.
Es importante notar que principalmente las páginas que contienen etiquetas HTML son las que se pudieron obtener por medio de la herramienta como se puede observar en la figura 9.

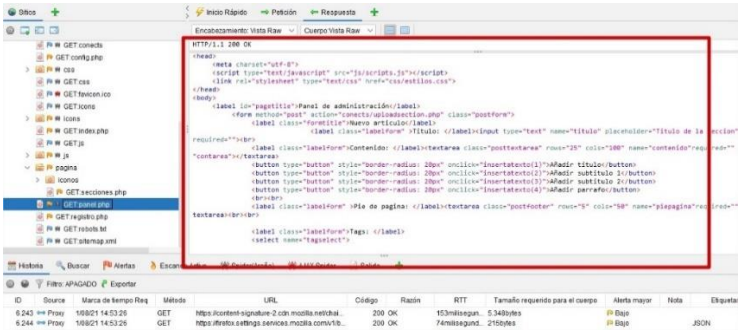
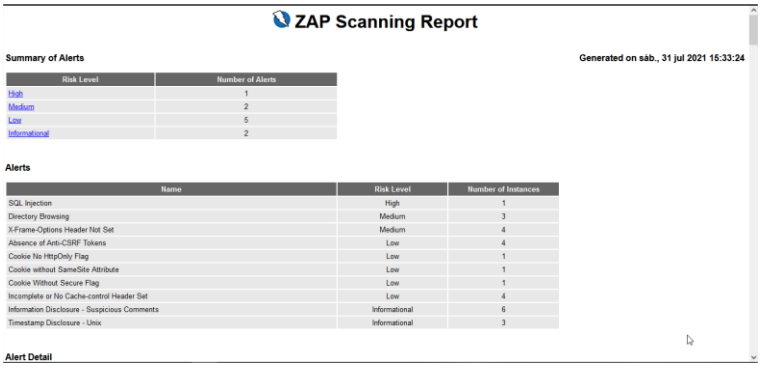


Fig. 9. Resultados de búsqueda.

Estas páginas principalmente se pueden utilizar para crear sitios falsos como se explicó al comienzo del documento, ya que es fácil copiar todo este contenido y hacerlo pasar por una página legítima.

Por lo que montar un servidor con un sitio web falso, por lo menos mostrando las primeras páginas del sitio, podría ser sencillo.
Estos resultados, orientados a una herramienta de Parsing se pueden utilizar para analizar el contenido del sitio y así realizar de forma correcta el análisis correspondiente de la información que busca el interesado.

Como resultado de las pruebas realizadas la herramienta es capaz además de generar reportes de las vulnerabilidades que se encontraron cuantificándolas y clasificándolas en una tabla como se puede observar en la figura 10 y desglosándolas por prioridad dando una descripción profunda de cada vulnerabilidad y cómo aprovecharla para un ataque o que hacer en caso de contrario, para prevenir el ataque.



ZAP Scanning Report

Summary of Alerts

Generated on sáb., 31 jul 2021 15:33:24

Risk Level	Number of Alerts
High	1
Medium	2
Low	5
Informational	2

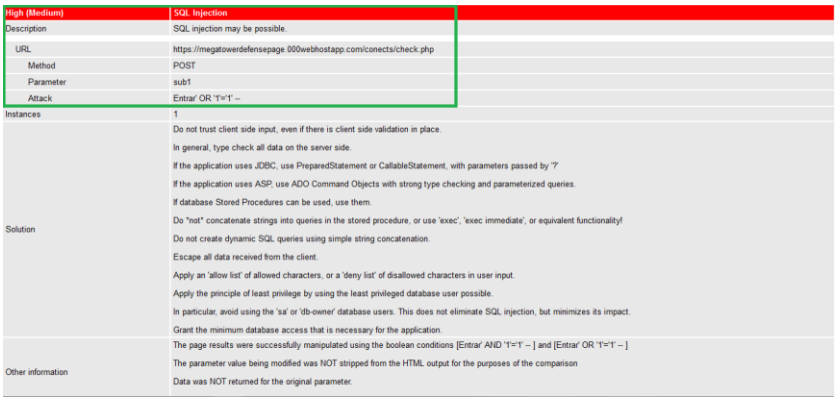
Alerts

Name	Risk Level	Number of Instances
SQL Injection	High	1
Directory Browsing	Medium	3
X-Frame-Options Header Not Set	Medium	4
Absence of Anti-CSRF Tokens	Low	4
Cookie No HttpOnly Flag	Low	1
Cookie without SameSite Attribute	Low	1
Cookie Without Secure Flag	Low	1
Incomplete or No Cache-control Header Set	Low	4
Information Disclosure - Suspicious Comments	Informational	6
Timestamp Disclosure - Unix	Informational	3

Alert Detail

Fig. 10. Resultados del ataque

Como podemos observar en la figura 11 se nos indica más información sobre el ataque realizado, la forma en que se realizó y posibles soluciones.



High (Medium)	SQL injection
Description	SQL injection may be possible.
URL	https://megatowerdefensepage.000webhostapp.com/conects/check.php
Method	POST
Parameter	sub1
Attack	Entrar' OR '1'='1' --
Instances	1
Solution	Do not trust client side input, even if there is client side validation in place. In general, type check all data on the server side. If the application uses JDBC, use PreparedStatement or CallableStatement, with parameters passed by '?' If the application uses ASP, use ADO Command Objects with strong type checking and parameterized queries. If database Stored Procedures can be used, use them. Do *not* concatenate strings into queries in the stored procedure, or use 'exec', 'exec immediate', or equivalent functionality! Do not create dynamic SQL queries using simple string concatenation. Escape all data received from the client. Apply an 'allow list' of allowed characters, or a 'deny list' of disallowed characters in user input. Apply the principle of least privilege by using the least privileged database user possible. In particular, avoid using the 'sa' or 'db-owner' database users. This does not eliminate SQL injection, but minimizes its impact. Grant the minimum database access that is necessary for the application.
Other information	The page results were successfully manipulated using the boolean conditions [Entrar' AND '1'='1' --] and [Entrar' OR '1'='1' --] The parameter value being modified was NOT stripped from the HTML output for the purposes of the comparison Data was NOT returned for the original parameter.

Fig. 11. Reporte de Vulnerabilidades

6 Conclusiones

A pesar de que el Web Scraping no es considerado ilegal ya que es el principal proceso que utilizan los motores de búsqueda, puede derivar en actos que sí lo sean [10], cada vez el mercado se vuelve más competitivo por lo que las empresas suelen utilizar nuevas técnicas para estar al tanto de los movimientos de la competencia y se ha vuelto una herramienta ideal para creación de sitios orientados al phishing. Por lo que no es de extrañar que será una práctica que se seguirá ejerciendo y para la cual no existe una herramienta como tal que proteja tu sitio web de este tipo de ataques. Por esto es importante que la seguridad del sitio web sea lo más completa posible por parte del servidor, ya que es fácil copiar el sitio web, pero uno puede prevenir peticiones maliciosas al servidor y uno como usuario debe estar atento al navegar en internet y por sitios web que soliciten datos personales.

Referencias

1. Zhao, Bo. (2017). Web Scraping. 10.1007/978-3-319-32001-4_483-1.
2. Zyte (formerly Scrapinghub) #1 Web Scraping Service. 2021. *What Is Web Scraping And How Does It Work? | Zyte.com*. [online] Disponible en: https://www.zyte.com/learn/what-is-web-scraping/ [Consultado el 30 de Julio 2021].
3. Thomas, D. & Mathur, S. (2019). Data Analysis by Web Scraping using Python. Amity Institute of Information Technology. pp 1-6
4. OpenWebinars.net. 2021. *Diferencias entre Scraping, Crawling y Parsing*. [online] Disponible en: https://openwebinars.net/blog/diferencias-entre-scraping-crawling-y-parsing/ [Consultado el 30 Julio 2021].
5. Bhatt D. & Vyas A. D. & Pandya S. (2015). Focused Web Crawler. Krishi Sanskriti Publications, Volumen 2 (11), pp. 1-6

6. Gupta, Anish & Singh, K & Singh, R. (2021). WEB CRAWLING TECHNIQUES AND ITS IMPLICATIONS. Globus An International Journal of Management & IT. 9. 7.
7. Cloudflare.com 2021 What is a Web Crawler | How web spiders work [online] Disponible en: https://www.cloudflare.com/learning/bots/what-is-a-web-crawler [Consultado el 30 Julio 2021].
8. Semrush Blog. 2021. *A Beginners Guide to Robots.txt: Everything You Need To Know*. [online] Disponible en: https://www.semrush.com/blog/beginners-guide-robots-txt/?kw=&cmp=LM_SRCH_DSA_Blog_Core_BU_EN&label=dsa_pagefeed&Network=g&Device=c&utm_content=484830561024&kwid=dsa-1057183199915&cmpid=11799892963&agpid=112575465577&BU=Core&extid=167384957085&adpos=&gclid=EAIaIQobChMIza-Jx--N8gIVcWxvBB2qUgnCEAAYASAAEgL_2fD_BwE [Consultado el 31 de Julio 2021].
9. Owasp.org. 2021. *OWASP ZAP Zed Attack Proxy | OWASP*. [online] Disponible en: https://owasp.org/www-project-zap/ [Consultado el 30 de Julio 2021].
10. Krotov, Vlad & Johnson, Leigh & Silva, Leiser. (2020). Legality and Ethics of Web Scraping. Communications of the Association for Information Systems. 47. 539-563. 10.17705/1CAIS.04724.

Prueba de T-Student Aplicada a Personas Contagiadas de COVID-19 por Género

M. González[2], Carlos Palomino[2], A. Hernández[1], R. Lima[2]

[1]Facultad de Ciencias Químicas, [2]Facultad de Ciencias de la Computación,
Benemérita Universidad Autónoma de Puebla.
marcos.gonzalez@correo.buap.mx, carlos_cpj@hotmail.com, ximikad09@gmail.com,
rox_li3@hotmail.com,

Resumen. En este trabajo se hace un análisis de las personas contagiadas por COVID-19 por género y se consideró también la edad para agrupar las frecuencias observadas de las personas contagiadas, con el objetivo de identificar si hombres y/o mujeres clasificados por edades, muestran diferencias en cuanto al número de contagiados, con el propósito de encontrar posibles diferencias significativas, para lo cual se tomó como muestra los datos proporcionados por el gobierno federal de la página http://coronavirus.gob.mx y la de https://datos.covid-19.conacyt.mx/ del día 12 de julio de 2021, por ser el día de inicio de este trabajo, esta investigación se realizó en 3 etapas, la primera fue verificar que la muestra tenía distribución normal, situación que se demostró aplicando la prueba de Shapiro-Wilk,, la segunda parte consistió en verificar la igualdad de las varianzas, lo cual se demostró aplicando la prueba F de Snedecor y la tercera y última etapa, consistió en la aplicación de la metodología de la prueba de t-Student, la cual permitió plantear 2 hipótesis, la nula y la alternativa, en dónde se verificó la aceptación de la hipótesis nula, la cual consistió en preguntar si no había diferencias en cuanto a los valores medios de las muestras por generó, aceptando dicha hipótesis; para el análisis estadístico se utilizó la hoja de cálculo de EXCEL de Microsoft 365, utilizando los complementos Xrealstats y Análisis de Datos**.**

Palabras clave. COVID-19, *t*-Student, Shapiro-Wilk, Normalidad, F de Snedecor, Coronavirus, Xrealstats, Frecuencias Observadas, Prueba de Hipótesis

1 Introducción

La distribución de t-Student se utiliza cuando nos encontramos con la dificultad de no conocer la desviación estándar poblacional y la muestra es menor de 30. Es similar a la curva normal, pero la distribución t tiene mayor área a los extremos y menos en el centro. Sus funciones se basan en establecer un intervalo de confianza, utilizando un nivel de confianza y los grados de libertad, obteniendo valores de una tabla dada con respecto a estas variables y aplicarla en la fórmula, se utiliza para probar hipótesis y también para saber si dos muestras tienen diferencias significativas.[8]

Cuando comparamos dos medias nos suele interesar responder a estas tres preguntas: **1º ¿Podemos afirmar que hay una diferencia?** A esta pregunta se responde mediante la t de Student. Es lo mismo que preguntar si la diferencia es estadísticamente significativa, o si es mayor de lo puramente aleatorio (diferencia distinta de cero en la población). Razón por la cuál se determinó aplicar la prueba de t-Student para la investigación.

2º ¿Es grande la diferencia? A un mayor valor de t no corresponde de manera sistemática una mayor diferencia; un valor grande de t sólo nos dice que tenemos mucha seguridad para poder afirmar que la diferencia entre las medias de las poblaciones no es cero, que hay una diferencia, pero un valor grande de t no nos permite afirmar que la diferencia es grande.
3º ¿Es importante la diferencia? La respuesta a esta pregunta supone un juicio cualitativo, pero depende en buena medida de las respuestas a las dos preguntas anteriores. Lo más frecuente es responder a la primera pregunta, y exponer, como datos necesarios y suficientes, los valores de t y de p (la probabilidad asociada al valor de t). A veces esto puede ser suficiente, pero no lo es habitualmente. El limitarse a exponer y comentar los valores de t y p puede llevar a interpretaciones equívocas, insuficientes o a exagerar la importancia de la diferencia. En buena parte las interpretaciones limitadas, e incluso equívocas, de las diferencias estadísticamente significativas radican en las mimas limitaciones del paradigma que utilizamos en el contraste de medias.[6]

Los coronavirus son una familia de virus que causan enfermedades desde el resfriado común hasta enfermedades respiratorias más graves y circulan entre humanos y animales. En este caso, se trata del SARS-COV2. Apareció en China en diciembre del 2019 y provoca una enfermedad llamada COVID-19, que se extendió por el mundo y fue declarada pandemia global por la Organización Mundial de la Salud[1,13].

La mayoría de las personas infectadas por el virus de la COVID-19 presentan cuadros respiratorios de leves a moderados y se recuperan sin tratamiento especial. Las personas mayores y las que padecen afecciones médicas subyacentes, como enfermedades cardiovasculares, diabetes, enfermedades respiratorias crónicas o cáncer, tienen más probabilidades de presentar un cuadro grave [2].

A la fecha del día 12 de julio de 2021, la información del gobierno federal índica que la población que se ha vacunado ha consistido de cuatro etapas, las cuales se describen a continuación: en la etapa 1 el 100% de la población de personal de salud se vacunó, el número fue de 1,229,391, para la etapa 2 se ha vacunado al 73% de la población de adultos mayores, cuyo número es de 11,049,352, la etapa 3 ha vacunado al 67% de la población de adulto entre 50 y 59 años, cuyo número es 18,585,732 y por último la etapa 4 ha vacunado al 33% de la población entre 40 y 49 años, cuyo número es de 3,848,500 [1,13,14]. Mucho se ha hablado de quienes son las personas que se contagian más y en base a los datos se ha creado un mito de que son los hombres los que se están contagiando más que las mujeres de COVID-19 y es la razón de investigación del presente trabajo, determinar si hay diferencias significativas en cuanto al promedio de las muestras, referentes a contagiados por COVID-19 de hombres y mujeres agrupados por edades, a través del rechazo o la aceptación de pruebas de hipótesis planteadas para la prueba de t-Student y para la igualdad de varianzas.

2 Fundamentos Teóricos

Con mucha frecuencia el propósito de la investigación va más allá de describir el comportamiento de la variable en la muestra y se debe generalizar o inferir los resultados obtenidos en la muestra a la población o universo, los datos casi siempre son recolectados de una muestra o población y sus medidas importantes tales como la media y varianza muestral reciben el nombre de estadísticas. Por otro lado, las medidas representativas de la población, media y varianza poblacional, casi siempre desconocidos, reciben el nombre de parámetros.

Una hipótesis científica es el resultado de un pensamiento creativo y tal vez inspirado, mientras que la hipótesis estadística es la expresión de una fase de la comprobación empírica de la hipótesis científica [12].

Hipótesis estadística. Una hipótesis estadística es un enunciado o proposición respecto a uno o más parámetros de la población. Una hipótesis estadística puede ser simple o compuesta. Es simple, cuando la proposición caracteriza completamente a la distribución de la variable aleatoria y en caso contrario se denomina compuesta. A fin de probar una proposición, es preciso formular una hipótesis denominada nula juntamente con otra denominada hipótesis alternativa, en este trabajo se utilizan las pruebas de hipótesis para determinar igualdad de varianzas y diferencias de medias [12].

El supuesto de homogeneidad de varianzas, también conocido como supuesto de homocedasticidad, considera que la varianza es constante (no varía) en los diferentes niveles de un factor, es decir, entre diferentes grupos.

A la hora de realizar contrastes de hipótesis o intervalos de confianza, cuando los tamaños de cada grupo son muy distintos ocurre que:

- Si los grupos con tamaños muestrales pequeños son los que tienen mayor varianza, la probabilidad real de cometer un error de tipo I en los contrastes de hipótesis será menor de lo que se obtiene al hacer la prueba. En los intervalos, los límites superior e inferior reales son menores que los que se obtienen. La inferencia será por lo general más conservadora.
- Si, por el contrario, son los grupos con tamaños muestrales grandes los que tienen mayor varianza, entonces se tendrá el efecto contrario y las pruebas serán más liberales. Es decir, la probabilidad real de cometer un error de tipo I es mayor que la devuelta por la prueba y los intervalos de confianza verdaderos serán más amplios que los calculados [15].

  Existen diferentes pruebas que permiten evaluar la distribución de la varianza. Todos ellos consideran como hipótesis nula que la varianza es igual entre los grupos y como hipótesis alternativa que no lo es. La diferencia entre ellos es el estadístico de centralidad que utilizan:
- Las pruebas que trabajan con la media de la varianza son los más potentes cuando las poblaciones que se comparan se distribuyen de forma normal.

  El F-test, también conocido como contraste de la razón de varianzas, contrasta la hipótesis nula de que dos poblaciones normales tienen la misma varianza. Es muy potente, detecta diferencias muy sutiles, pero es muy sensible a violaciones de la normalidad de las poblaciones. Por esta razón, no es una prueba recomendable si no se tiene mucha certeza de que las poblaciones se distribuyen de forma normal, pero como en nuestro caso de investigación los datos muestrales tienen comportamiento normal, es la prueba que se aplicó [15].

La prueba t-Student se utiliza para contrastar hipótesis sobre medias en poblaciones con distribución normal. También proporciona resultados aproximados para los contrastes de medias en muestras suficientemente grandes cuando estas poblaciones no se distribuyen normalmente (aunque en este último caso es preferible realizar una prueba no paramétrica). Para conocer si se puede suponer que los datos siguen una distribución normal, se pueden realizar diversos contrastes llamados de bondad de ajuste, de los cuales el más usado es la prueba de Kolmogorov. A menudo, la prueba de Kolmogorov es referida erróneamente como prueba de Kolmogorov-Smirnov, ya que en realidad esta última, sirve para contrastar si dos poblaciones tienen la misma distribución. Otros tests empleados para la prueba de normalidad son debidos a Shaphiro y Wilks. Existen dos versiones de la prueba t-Student: una que supone que las varianzas poblacionales son iguales que es el caso de estudio del presente trabajo y otra versión que no asume esto último. Para decidir si se puede suponer o no la igualdad de varianza en las dos poblaciones, se debe realizar previamente la prueba F-Snedecor de comparación de dos varianzas. La prueba t-Student fue desarrollada en 1899 por el químico inglés William Sealey Gosset (1876-1937) [3].

La distribución F es conocida con este nombre gracias al matemático americano George W. Snedecor (1882-1974) quien la bautizó de este modo en honor de R. A. Fisher (1890-1962) que ya la había estudiado anteriormente en 1924. Las pruebas de bondad de

ajuste mencionadas son debidas a Nikolai Vasil'yevich Smirnov (1890-1966), Andrei Nikolaevich Kolmogorov (1903-1987) gran teórico probabilista que fundó las bases de la teoría de la medida en 1929 y finalmente Samuel S. Shapiro (actualmente profesor de matemáticas en los EE. UU) y Martin.B. Wilk (matemático canadiense) que publicaron sus hallazgos en la revista "Biometrika" en 1965 [3].

Pruebas con muestras pequeñas para comparar dos medias poblacionales supuestos: Muestras independientes de poblaciones normales con $\sigma_1^2 = \sigma_2^2$

$$H_0 = \mu_1 - \mu_2 = D_0 \tag{1}$$

$$H_1 = \begin{cases} \mu_1 - \mu_2 > D_0 \ (alternativa\ de\ cola\ superior) \\ \mu_1 - \mu_2 < D_0 \ (alternativa\ de\ cola\ inferior) \\ \mu_1 - \mu_2 \neq D_0 \ (alternativa\ de\ dos\ colas) \end{cases}$$

Estadístico de prueba:

$$T = \frac{\bar{X}_1 - \bar{X}_2 - D_0}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{2}$$

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \tag{3}$$

Región de rechazo (RR) : $\begin{cases} t > t_\alpha \ (RR\ de\ cola\ superior) \\ t < -t_\alpha \ (RR\ de\ cola\ inferior) \\ |t| > t_{\frac{\propto}{2}} (RR\ de\ \ dos\ colas) \end{cases}$

Aquí $P(T > t_\alpha) = \alpha$. [4,16]

3 Metodología

1.- Probar que cada una de las muestras tiene una distribución normal
2.-Obtener para cada una de las muestras:
  a) el tamaño de las muestras (n1 y n2)
  b) sus respectivas medias (m1 y m2)
  c) sus varianzas (v1 y v2);
3. Probar que las varianzas sean homogéneas;
4. En caso de homogeneidad en esas varianzas:
  a) establecer la diferencia entre las medias: m1-m2
  b) calcular la varianza agrupada de las dos muestras.
    Es decir, la varianza agrupada (vc) es igual a un promedio pesado de las varianzas de las dos muestras en donde los pesos para ese promedio son iguales al tamaño, menos uno (n-1) para cada una de las muestras.
  c) con esa varianza agrupada, se calcula el error estándar de la diferencia de las medias
    ESM= √ ((vc) (n1 + n2) /(n1n2))
    Donde vc es la varianza agrupada.
5. Finalmente, la t-Student es igual al cociente de la diferencia de medias entre el ESM anterior

6. De acuerdo con nuestra hipótesis nula y alterna se debe demostrar que existe diferencia entre las medias de las muestras, se consulta una tabla de t-Student con grado de libertad igual a n1 + n2-2 y se calcula el valor de P. [8,9,10,11].

4 Resultados

En este trabajo se consideraron los datos proporcionados por el gobierno federal en la página http://coronavirus.gob.mx y la de https://datos.covid-19.conacyt.mx/ , se tomó como muestra poblacional, la del día 12 de julio de 2021, porque fue el día en que se inició la investigación del presente y se buscó la información referente al número de contagiados COVID-19 de hombres y mujeres, estas muestras se utilizaron para realizar el análisis de este trabajo y demostrar si hay diferencias significativas entre los contagiados por COVID-19 por género y clasificados en grupos de edades, para lo cual se comparó la media muestral de cada muestra, con el objetivo de determinar que la enfermedad no tiene género y derribar el mito de si son los hombres los que se contagian más.

La información se agrupó por intervalos de edad y se consideró una amplitud de 10 para cada intervalo, tal como se ha estado utilizando para aplicar las vacunas contra el COVID -19, y son los siguientes: 0-9,10-19,20-29,30-39,40-49,50-59,60-69,70,79,80-89. Tabla 1. [1,17,20,21]

Tabla 1. Datos de contagiados por COVID-19 del día 12 de julio de 2021 de hombres y mujeres.

Edades-día 12-julio	Hombres_Covid-19	Mujeres_Covid_19
[0-9]	16021.00	16021.00
[10-19]	55948.00	58440.00
[20-29]	228623.00	240527.00
[30-39]	278455.00	281969.00
[40-49]	259613.00	265881.00
[50-59]	214719.00	215385.00
[60-69]	135994.00	124354.00
[70-79]	74746.00	62970.00
[80-89]	28538.00	24650.00
[90-94]	3390.00	3443.00

A continuación aplicaremos la metodología mencionada en el punto 3, para lo cual empezaremos por determinar si hay normalidad en las muestras de la tabla 1, utilizaremos el proceso de Shapiro-Wilk[17,7], para determinar la normalidad, paro lo cual se utilizó el complemento de Excel-Microsoft 365 denominado Real Statistics[5] (Figura 1), una vez seleccionada, se elige la opción Descriptive Statistics and Normality y se pulsa OK, posteriormente se muestra la figura 2 y en esta opción se permite elegir la prueba de Shapiro-Wilk, la prueba de Grubbs', Estadística Descriptiva y Gráficos, una vez seleccionado el campo Shapiro-Wilk para nuestro caso de estudio, se presiona OK y los resultados obtenidos se muestran en la tabla 2. [5,17]

Fig. 1. Menú del complemento Real Statistics de Excel-Microsoft 365

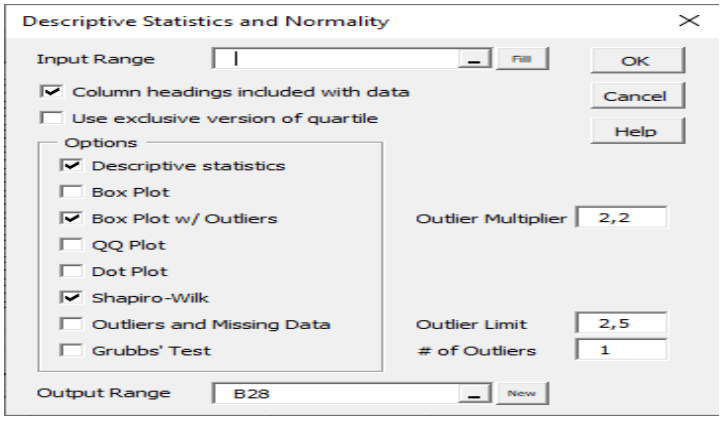


Fig. 2. Opción que permite elegir la Prueba de Shapiro-Wilk.

Tabla 2. Resultados de la prueba de Shapiro-Wilk.

Shapiro-Wilk Test

	Hombres_Covid-19	***Mujeres_Covid_19***
W-stat	0.884665378	0.867203228
p-value	0.147551757	0.092723591
alpha	0.05	0.05
normal	**yes**	**yes**

d'Agostino-Pearson

DA-stat	4.125960202	4.440948603
p-value	0.12707471	0.108557608
alpha	0.05	0.05
normal	**yes**	**yes**

Los resultados de la tabla 2 evidencian que las muestras tienen distribución normal y que por lo tanto podemos seguir con el siguiente punto de la opción 3, es decir demostrar la igualdad de varianzas, cabe hacer mención que junto con la prueba de Shapiro-Wilk también se realiza la prueba de d'Agostino-Pearson, en donde también se evidencia que las muestras tienen una distribución normal, aunado a los resultados anteriores también se puede corroborar viendo el gráfico 1 de la figura 3 que es un histograma, en donde se puede evidenciar que prácticamente los datos tienen una Distribución Normal , tienen la forma de la campana de Gauss, con esto se concluye el primer paso de la metodología, las muestras tienen distribución normal.

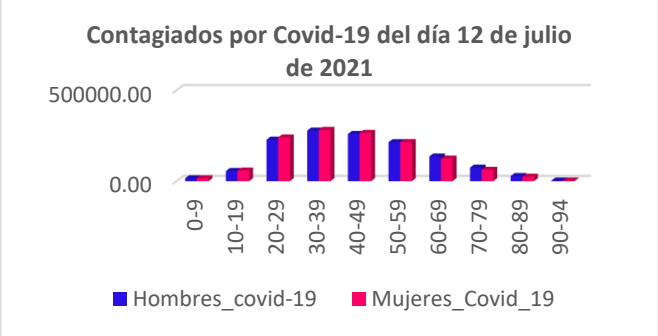


Fig. 3. Histograma de contagiados por Covid-19 por género.

El siguiente paso ejemplificará la igualdad de varianzas, para lo cual utilizaremos también el complemento de Excel-Microsoft 365, para lo cual se seleccionó Datos →Análisis de Datos→ Prueba F para varianzas de dos muestras, tal como se muestra en la figura 4.

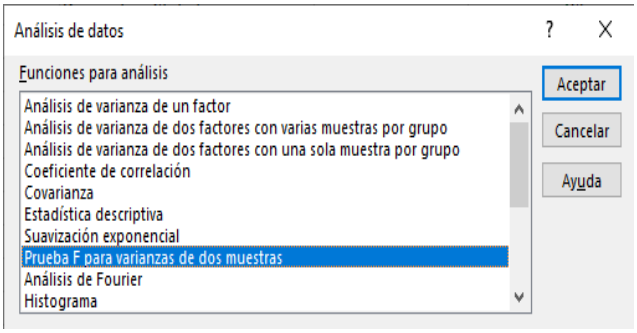


Fig. 4. Opción Análisis de datos y Prueba F para varianzas de dos muestras.

Una vez seleccionada la opción Prueba F, entonces hay que proporcionar el rango para la muestra 1 y muestra 2, incluir rótulos, con un nivel de significancia del 5% y colocar el rango de salida, tal como se muestra en la figura 5.

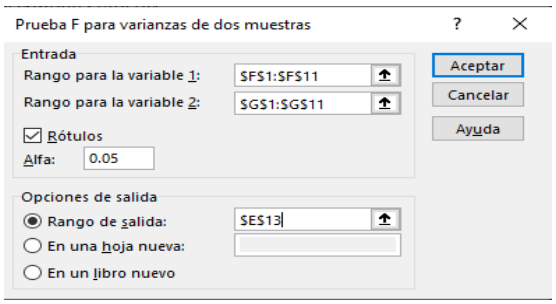


Fig. 5. Prueba F para varianzas de dos muestras.

Tabla 3. Resultados de la prueba de F para dos muestras.

Prueba F para varianzas de dos muestras

	Hombres_Covid-19	***Mujeres_Covid_19***
Media	129604.7	129364
Varianza	11514348056	12312581680
Observaciones	10	10
Grados de libertad	9	9
F	**0.935169273**	
P(F<=f) una cola	**0.461051976**	
Valor crítico para F (una cola)	0.314574906	

De acuerdo con los datos de la tabla 3, se plantearon las hipótesis para determinar si había igualdad de varianzas

Pruebas de hipótesis para determinar igualdad de varianzas

$$H_0: Las\ varianzas\ del\ número\ de\ contagiados\ de\ COVID - 19\ de\ hombres\ y\ mujeres\ son\ iguales.$$
$$\sigma_1^2 = \sigma_2^2$$
$$H_1: Las\ varianzas\ del\ número\ de\ contagiados\ de\ COVID19\ de\ hombres\ y\ mujeres\ son\ diferentes.$$
$$\sigma_1^2 \neq \sigma_2^2$$

El valor de la F de Snedecor de acuerdo a la tabla 3 es F=0.935169273, la prueba de hipótesis alternativa se acepta si el p-valor es menor que el nivel de significancia del 5% y si el p-valor es mayor que el nivel de significancia del 5% se acepta la hipótesis nula, por lo que viendo la tabla 3 se observa que el valor -p que se puede identificar como P(F<=f) de una cola, con un valor de 0.461051976 que convirtiéndolo a porcentaje queda 46.10519759% que obviamente es mayor que el 5%, por lo que se acepta la hipótesis nula, es decir la varianzas son iguales.[12,15,18,19]

Por último, realizaremos la prueba de t-Student en virtud de que se cumplió la condición de normalidad y la igualdad de las varianzas, para lo cual se seleccionó Datos →Análisis de Datos →Prueba t para dos muestras suponiendo varianzas iguales, tal como se ve en la figura 6.

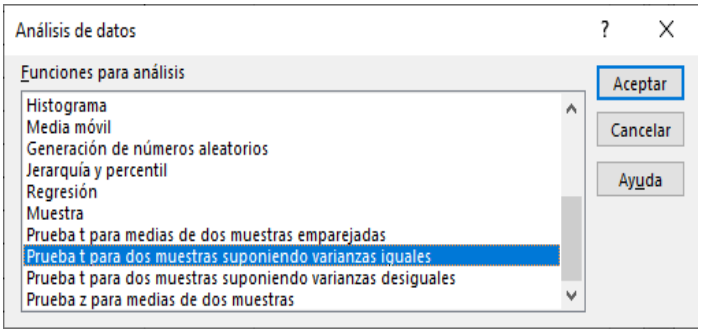


Fig. 6. Opción Prueba t para dos muestras suponiendo varianzas iguales.

Una vez seleccionada la prueba t, se procede a elegir el rango de las muestras, se incluyen rótulos, se considera un nivel de significancia del 5% y se elige el rango de salida tal como se observa en la figura 7.

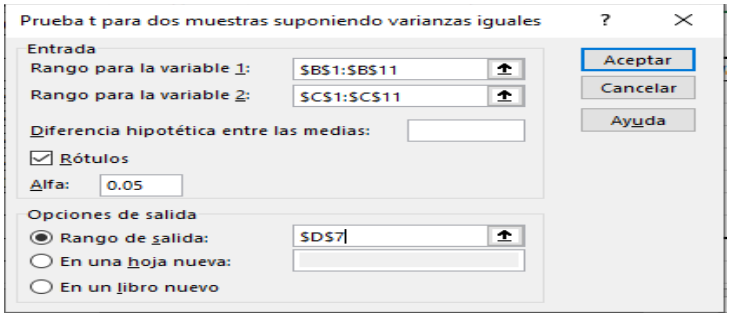


Fig. 7. Opciones para aplicar la prueba de t-Student.

Una vez seleccionadas las opciones para la prueba t, se da clic en aceptar y se obtienen los resultados mostrados en la tabla 4.

Tabla 4. Aplicación de la prueba de t-Student

Prueba t para dos muestras suponiendo varianzas iguales

	Hombres_Covid-19	*Mujeres_Covid_19*
Media	129604.7	129364
Varianza	11514348056	12312581680
Observaciones	10	10
Varianza agrupada	11913464868	
Diferencia hipotética de las medias	0	
Grados de libertad	18	
Estadístico t	**0.00493108**	
P(T<=t) una cola	**0.498059912**	
Valor crítico de t (una cola)	1.734063607	
P(T<=t) dos colas	0.996119824	
Valor crítico de t (dos colas)	2.10092204	

Sabemos que si el estadístico $t > 0$ (es positivo) $\rightarrow H_1$ será mayor y si t < 0 (es negativo) $\rightarrow H_1$ será menor; ahora bien, como el estadístico t=0.00493108 entonces es positivo, para lo cual se plantean la hipótesis nula y alternativa como sigue:

Prueba de Hipótesis para la t-Student

H_0: *El número promedio de contagiados hombres de covid* 19 *es igual al número promedio de contagiadas mujeres covid* − 19

$$\mu_1 = \mu_2$$

H_1: *El número promedio de contagiados hombres de covid* 19 *es mayor al número promediode contagiadas mujeres de covid* − 19

$$\mu_1 - \mu_2 > 0 \rightarrow \mu_{Hombres} > \mu_{Mujeres} \quad [12]$$

Se identifica el p-valor en la tabla 4 como P(T<=t) una cola, cuyo valor es 0.498059912, convirtiéndolo a porcentaje se tiene 49.80599118%.

Si el valor del p-valor es menor del 5% se acepta H_1 y si es mayor del 5% entonces se acepta H_0, entonces como el p-valor es del 48.80599118% que es mayor que el 5%, entonces se acepta la Hipótesis nula.

Podemos entonces concluir que no hay diferencias significativas en el número de contagiados de covid-19 por género.

5 Conclusión

Este trabajo se realizó en 3 etapas, la primera fue verificar que la muestra tenía distribución normal, situación que se demostró aplicando la prueba de Shapiro-Wilk,, la segunda parte consistió en verificar la igualdad de las varianzas, lo cual se demostró aplicando la prueba F de Snedecor y la tercera y última etapa, consistió en la aplicación de la prueba de t-Student, la cual permitió plantear 2 hipótesis, la nula y la alternativa, en dónde se verificó la aceptación de la hipótesis nula, la cual consistió en preguntar si no había diferencias en cuanto a los valores medios de las muestras por generó, aceptando dicha hipótesis. Podemos concluir que este trabajo permitió destacar que aunque los datos numéricos indican que hay más contagiados hombres que mujeres, estadísticamente se demostró que no hay diferencias significativas con respecto a su promedio, esto nos permite difundir, que el COVID-19 no tiene género y podrá contagiarse tanto en hombres como en mujeres, por lo que todavía es conveniente mencionar que hay que seguir aplicando los criterios de protección para no contagiarse y por su puesto en los momento de vacunarse acudir a hacerlo, podemos hacer notar que de acuerdo a la información obtenida es importante resaltar que el número mayor de contagiados con respecto a cualquier bloque de edades se da en el intervalo de 30-39 años, que es precisamente dónde el número de vacunados apenas empieza, recordando que nos referimos al mes de Julio de 2021. para trabajos futuros, aumentaremos las variables, tal como si ya estaban vacunados y aun así se contagiaron, la edad considerándola ya como parte de la muestra, esto permitirá aplicar la distribución t-Student multivariada.

Referencias

1. Gobierno de México, Coronavirus (2019), https://www.coronavirus.org.mx, accedido el 12 de julio de 2021.
2. Organización Mundial de la Salud. https://www.who.int/es/health-topics/coronavirus#tab=tab_3, accedido el 12 julio de 2021.
3. Universidad Pedagógica y Tecnológica de Colombia https://virtual.uptc.edu.co/ova/estadistica/docs/libros/tstudent.pdf, accedido el 19 de julio de 2021.
4. Sánchez Turcios A., Revista mexicana de cardiología Vol. 26 no. 1 México ene/mar, versión impresa ISSN 0188-2198 http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-21982015000100009, accedido el 19 de julio de 2021.
5. Real Statistics Resource Pack, https://www.real-statistics.com/free-download/real-statistics-resource-pack/ accedido el 5 de julio de 2021.
6. Morales Vallejo P., Universidad Pontificia Comillas, Madrid (3 de octubre de 2012) https://web.upcomillas.es/personal/peter/investigacion/Tama%f1oDelEfecto.pdf, accedido el 26 de julio de 2021.
7. Flores Tapia C., Flores Ceballos K., Societas. Revista de ciencias sociales y humanidades, Universidad de Panamá, http://portal.amelica.org/ameli/jatsRepo/341/3412237018/index.html, accedido el 27 de julio de 2021.
8. Scientific European Federation of Osteopaths (2014 SEFO) https://www.scientific-european-federation-osteopaths.org/wp-content/uploads/2019/01/Distribucion_tStudent.pdf, accedido 28 de julio de 2021.
9. Morten Wang F., Oslo University Hospital, ResearchGate (2008) https://www.researchgate.net/publication/225376415_T-tests_non-parametric_tests_and_large_studiesa_paradox_of_statistical_practice, accedido el 28 de junio de 2021.

10. SlidesShare from Scribd,https://es.slideshare.net/torimatcordova/distribucion-t-de-student-28545004, accedido el 26 de julio de 2021.
11. Revista mexicana de cardiología, (2015) http://www.scielo.org.mx/pdf/rmc/v26n1/v26n1a9.pdf, accedido el 26 de julio de 2021.
12. Universidad Nacional Mayor de San Marcos, (Perú) https://sisbib.unmsm.edu.pe/bibvirtualdata/libros/Matematicas/inferencia/pdf/a03.pdf, accedido el 29 de julio de 2021.
13. Geo-Hub COVID-19-Information System for the Region of the Americas, https://paho-covid19-response-who.hub.arcgis.com/pages/paho-mexico-covid-19-response, accedido el 12 julio de 2021.
14. COVID-19 Tablero México-CONACYT (2020),https://datos.covid-19.conacyt.mx/, accedido el 12 julio de 2021.
15. Amat Rodrigo J., Ciencia de Datos, Estadística, Machine Learning y Programación. https://www.cienciadedatos.net/documentos/9_homogeneidad_de_varianza_homocedasticidad.html, accedido el 30 de julio de 2021.
16. Dennis D. Wackerly., William Mendenhall., Richard L. Scheaffer., "Estadística Matemática con Aplicaciones". Séptima Edición, Ed. Cengage Learning, 2008, EU.
17. Lawrence S. Meyers, Glenn Gamst, A.J. Guarino, "Applied Multivariate Research", first Edition. Edit. Sage Publication, Inc. 2006, Shapiro Wilk 68,86,100,264,390,420.
18. Douglas C. Montgomery "Diseño y análisis de experimentos", segunda edición, Universidad Estatal de Arizona, Ed. Limusa Wyley, 2004.
19. Mario F. Triola," Estadística", Editorial Pearson Education, décima edición, Dutchess Community College

2009

20. George C. Canavos, "Probabilidad y Estadística, Aplicaciones y Métodos", Editorial McGrawHill, Virginia Commonwealth University, 1988.
21. Roberto H. Sampieri, Carlos F. Collado, Pilar B. Lucio," Metodología de la Investigación", Tercera Edición, Ed. McGrawHill, IPN,2003.

Medidas y Análisis de Prevención Ante Amenazas de Ransomware

Oscar M. González, Ana C. Zenteno, Ma. del Carmen Santiago, Yeiny Romero, Judith Pérez, Gustavo T. Rubín
Facultad de Ciencias de la Computación en la Benemérita Universidad Autónoma de Puebla, 14 sur esquina con Av. San Claudio, CP. 72570, Puebla, Pue. México
oscar.gonzalezra@alumno.buap.mx, {ana.zenteno, marycarmen.santiago, yeiny.romero, judith.perez, gustavo.rubin}@correo.buap.mx

Resumen. El ransomware es un tipo de software malicioso que sigue siendo una de las mayores amenazas en internet [1]. De acuerdo con el crecimiento del mismo se estima que será el principal factor en generar ganancias ilícitas multimillonarias. Algunos mercados son particularmente propensos al ransomware y a pagar el rescate, un claro ejemplo son los hospitales debido a que se tiene la salud de los pacientes y no pueden arriesgar o tomar decisiones a largo plazo y es por eso por lo que muchas veces se opta por pagar el rescate [2]. Esto puede evitarse con buenas medidas de seguridad y una buena implementación en la estructura computacional donde se trabaja. Conociendo herramientas de nueva generación y respaldando toda información junto con lo anterior mencionado se minimizará los daños, así como la vulneración de este tipo de software malicioso.

Palabras Clave: Ransomware, Antivirus, Ciberataque, Phishing, Ingeniería Social.

1 Introducción

Debido al incremento de herramientas, así como de servicios que le facilitan al usuario su día a día, existe también la competitividad y la saturación de estos, la rapidez con la que buscan posicionarse en el mercado hace que ignoren su implementación de seguridad, así como su respaldo de información, de tal forma que, suelen ser canales o vías de riesgo para el consumidor, afectándolo indirecta o directamente.

La historia de este tipo de software malicioso ha cambiado y evolucionado mucho en estos últimos años, puesto que ha sido el tipo de código más dañino y más destructivo en la última década [3].

A lo largo del escrito profundizaremos en análisis e implantaciones de seguridad para evitar ser víctimas de este tipo de software malicioso, así como la utilización de servicios que gratuitamente nos proporcionan los antivirus puesto que son los principales proveedores de seguridad.

Se tiene que aclarar que las medidas son tanto individuales como empresariales puesto que este tipo de ciberataque se propaga por phishing o ingeniería social y es vital capacitar al personal para que reconozcan este tipo de amenazas y se pueda reducir el riesgo de infección [4].

1.1 Tipo de Víctima

De acuerdo a análisis forenses que se han realizado a víctimas por parte de este tipo de software malicioso, se determinó que el principal canal o vía de infección radica mayormente en la poca capacitación que se les dirige a los usuarios. Esto tanto a nivel empresarial como a nivel individual puesto que se asegura que en ningún momento seremos víctima de este tipo de software malicioso, claramente hasta que esto sucede.

Se tiene que entender el factor que representa uno mismo ante la sociedad o a nivel personal, ya que depende de eso la integridad de su sistema y sus datos, específicamente hablamos de ser una empresa de renombre o privada o ser un usuario común con archivos personales, laborales, etc.

Tomando en consideración lo anterior podremos implementar medidas cautelares para evitar incidencias con este tipo de software malicioso, evitando así daños colaterales o críticos puesto que muchas veces uno mismo funciona como propagador de este mal.
Dicho lo anterior trabajaremos con los perfiles mencionados, cabe aclarar que serán medidas preventivas y que, si bien este escrito está enfocado a un sistema operativo comercial, sirve de referencia para aplicarlo o implementarlo en diversificación de sistemas operativos.

1.2 Personal

Actualmente este tipo de software malicioso está dirigido al ramo empresarial, puesto que la inversión que se hace es relativamente grande por lo tanto se busca tener una retribución monetaria mayor. La pérdida de datos a nivel personal no genera algún quebrando económico la mayoría de las veces, puesto que en muchas de las ocasiones la información es irrelevante y no puede generar una gran afluencia monetariamente al delincuente.

Eso no impide que el mismo usuario sirva como puente para afectar a una empresa o a algún otro objetivo, se tiene que recordar que este tipo de programa malicioso se propaga sobre la red local o interna, tratando de ubicar a la mayor parte de dispositivos.

Las intenciones u objetivos de los criminales es en muchas ocasiones lograr un tipo de chantaje emocional, tratando de que la víctima no tenga otra opción más que pagar el rescate, con la intención de recobrar su información. En otras situaciones el objetivo es que el usuario realice acciones que perjudiquen a terceros por lo tanto analizaremos en el siguiente punto algunas recomendaciones, para después tomar medidas de implementación más robustas y tener un entorno de trabajo más seguro.

1.3 Empresarial

Como se había mencionado anteriormente el principal objetivo de este tipo de amenazas es el ramo empresarial que, si bien pueden presumir de tener una infraestructura sólida y bien implementada, el factor humano siempre va a prevalecer, puesto que un gran porcentaje de evidencias demuestran que debido a phishing o ingeniería social han sido los principales motivos de vulneración. Normalmente las empresas líderes en su rubro tienen sus propios departamentos de seguridad, muchas veces denominados Blue Team o Red Team, dedicadas a salvaguardar la integridad de la empresa.

2 Mecánica del Uso de Ransomware

Debido a que este estudio es bastante extenso y riguroso depende de esta línea la integridad de la empresa, así como de su prestigio, generalizaremos y tomaremos en cuenta la funcionalidad y el modus operandi de cómo se implementa.

2.1 Modus operandi

Existe una amplia diversificación al momento de implementar este tipo de ransomware, si bien normalmente los objetivos y los procedimientos son personificables, existe un protocolo base que se toma en consideración véase Figura.1

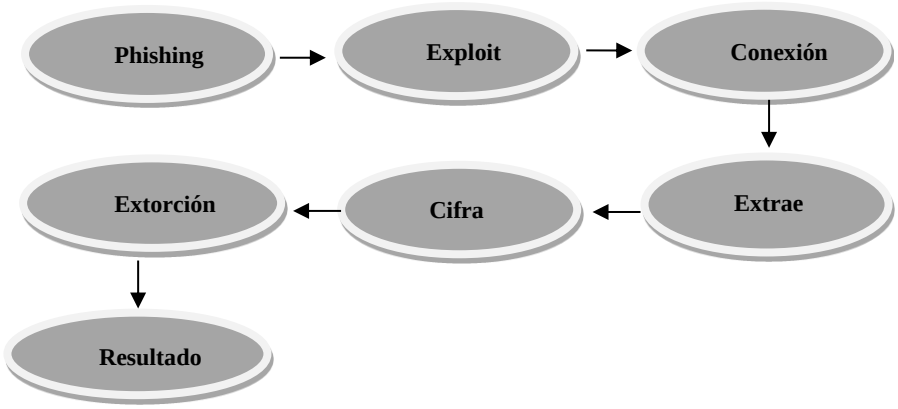


Fig. 1. Mecánica que utilizan los ciberdelincuentes para tomar posesión de una víctima a través de una vulnerabilidad, provocada por phishing o ingeniería social.

3 Medidas Preventivas Nivel Usuario

Algunas recomendaciones esenciales son las que veremos a continuación, sirven para tomarlas de base y a si mejorarlas o adicionarlas con algunas otras

3.1 Medidas de uso personal

Se tiene que considerar que las medidas que a continuación vamos a nombrar resultan en su mayoría efectivas, si se practican de manera correcta, cabe destacar que aun que se tenga una buena práctica, implementación o herramientas de nueva generación, el factor humano siempre prevalecerá, tanto para bien como para mal

El factor radica en detectar alguna vulnerabilidad, sufrir de algún ataque de ingeniería social, caer en la trampa de falsos positivos, nuevas herramientas de implementación, etc. Es así como consideramos lo anterior como punto de partida para analizar los siguientes puntos:

- *Reconocimiento de estación de trabajo:* Tanto en el ámbito empresarial como personal, se tiene conocimiento del funcionamiento de nuestros equipos de cómputo, por tal motivo al detectar discrepancias o comportamientos diferentes es factible aislar el equipo y analizarlo con un antivirus.
- *Evitar descargas ilícitas*: Se tiene que reconocer que este hábito es un acto delictivo y que muchas veces sirve para propagar algún tipo de ransomware por tal motivo evitar este tipo de prácticas.
- *Generar respaldos de información*: Una de las mejores soluciones a futuro es generar respaldo de información externas, es decir, no solo basta con generar la copia de seguridad y dejarla en el dispositivo si no exportarla a la nube o a un dispositivo de almacenamiento externo, a si los daños serán de un rango definido.
- *Ignorar páginas de dudoso contenido*: En el momento en que el corpus de un enlace muestre ciertas ambigüedades de lo que tiene que aparentar, es mejor no ingresar,

puesto que en la mayoría de las ocasiones se generan cargas de scripts y descarga de pequeños fragmentos que pueden ser perjudiciales es mejor evitarlos.

- *Autenticación doble*: Entre más seguro y robusto sea nuestro mecanismo de seguridad mejor estaremos protegidos puesto que se nos avisara en el momento cualquier intento de intrusión, así como la hora y el lugar, esto debe ser implementado en todas las plataformas en las cuales tengamos acceso y permitan la opción.
- *Autenticación triple*: Si bien hoy en día es bastante debatible el por qué es que toman en cuenta los datos biométricos, son fundamentales y cruciales a la hora de la validación, el sensor de huella, así como la retina y la cara juegan un papel muy importante a la hora de la seguridad
- *Intercalar caracteres en nuestras contraseñas*: A pesar de que existen métodos muy sofisticados de fuerza bruta, es irrelevante para el atacante invertir tiempo y dinero en descifrar una contraseña que congenia con números, caracteres especiales, mayúsculas, espacios etc. Por lo cual es vital robustecer este aspecto, cambiar contraseñas cada determinado tiempo, utilizando la variedad de caracteres especiales, evitar utilizar fechas especiales y aunar en la ambigüedad.

Otro aspecto muy crucial y delicado no mencionado anteriormente es el filtrado de correos electrónicos, se tiene que estudiar muy bien tanto el remitente como la calidad del correo, puesto que normalmente los delincuentes no invierten demasiado en estos aspectos, por lo tanto, si es sorpresivo un correo electrónico no esperado es mejor analizarlo y cotejarlo por otro medio si es verídico o no.

3.2 Implementación de medidas cautelares

Para el usuario inexperto con las medidas preventivas anteriormente mencionadas, serán más que suficiente para tener una sana navegación en internet, aunque para los usuarios especializados y avanzados en el ámbito computacional es recomendable, retroalimentarse periódicamente con cultura de este ámbito.

Se implementarán medidas avanzadas para el usuario común, estas servirán de base para tomar en cuenta en el futuro. Por lo cual analizaremos la Figura 2.

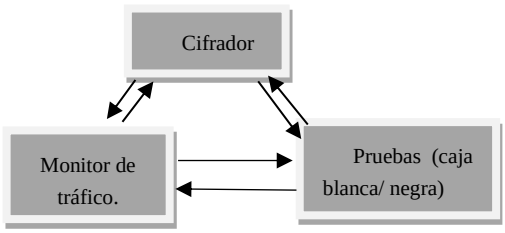


Fig. 2. Bases para confrontar tanto en sospechas como en afirmaciones de ransomware.

- *Monitor de tráfico:* Además de tener un antivirus, siempre es recomendable tener un monitor de tráfico, esto es para detectar las entradas y salidas de datos que se generen durante tu navegación un gran ejemplo es Wireshark que analiza los protocolos TCP y UDP verifica su validez y en dado caso de que se tenga una incidencia de mostrará la procedencia y datos relevantes.
- *Cifrador:* Usando una referencia ambigua pero que sirve de ejemplo es tener un programa o script, que en dado caso de que surja alguna alerta o discrepancia en la red o en el dispositivo se ejecute, de tal forma que tus archivos o información se cifre, esto quiere decir que en caso de que el ciberdelincuente absorbe tu información o planee hacerlo tomara documentos inválidos, imposibilitando el objetivo principal del

ransomware, que es el de extorsionara la víctima, una herramienta a tomar en cuenta es Folder Lock, fácil de usar y con versión libre y de pago.

- *Pruebas (Caja negra/blanca):* Aquí nos apoyaremos de las características que nos proporcionan los antivirus y que pocas veces se toman en cuenta, las pruebas de baúl o caja negra/blanca, consiste en simular la ejecución del archivo en un entorno controlado por los antivirus, analizan su comportamiento y de acuerdo a eso te informan si tiene alguna alteración o no, combinándolo con el monitor de red, consolidamos si el archivo es sano o genero alguna conexión.

Realizando los pasos anteriores minuciosamente, lograremos que sea menos probable sufrir de este tipo de software malicioso, puesto que abarcamos el área de red, el área administrativa por parte del antivirus y un acto defensivo como lo es el cifrado.

Si bien puede mejorarse y restructurar la implementación de medidas cautelares, quedara en el lector tomar en cuenta las medidas mencionadas para reforzar esto.

4 Medidas Preventivas Nivel Empresarial.

Las recomendaciones que veremos a continuación, en su mayoría son implementadas en pequeñas y medianas empresas, cuando se tiene una cercanía con el personal y el control de seguridad cercano.

4.1 Medidas de uso empresarial.

Listaremos las medidas más básicas y que son mas representativas en el ámbito empresarial, si bien podemos encontrar infinidad de las mencionadas estas podrían ser bases para implementar nuevas.

No entraremos en detalles, pero si dejaremos en énfasis su aplicación.

- *Exámenes de confianza:* Analizar las intenciones de los trabajadores.
- *Delimitar dispositivos:* Evitar celulares memorias aparatos que se malinterpreten.
- *Restringir redes:* Condicionar la navegación del personal.
- *Inicios de sesión:* Verificar hora, fecha y lugar del inicio de un usuario.
- *Historial de uso:* Analizar uso y manejo de los dispositivos.
- *Implementar protocolos:* Analizar medidas de seguridad en caso de alguna situación.
- *Doble autenticación:* Utilizar diversificación de contraseñas.
- *Instrucciones de uso:* Validar que se use el material de manera correcta.
- *Respaldos de información:* Generar recurrentemente respaldos de datos.
- *Generar limpieza de datos:* Realizar limpieza de navegación o datos o innecesarios.

Estas son medidas básicas que hoy en día son pilares en cualquier empresa, a pesar de esto muchas veces se presume que se practican, lamentablemente las estadísticas de ataque de ransomware dicen lo contrario.

En análisis forenses muestran que el usuario es el principal responsable, ya sea que por equivocación abrió o ejecuto algún vínculo contaminado, o es parte del grupo de cibercriminales.

Existen innumerables medidas de prevención, si tenemos como base las anteriores y les adicionamos las necesarias para la empresa robusteceremos este ámbito.

4.2 Recomendación de medidas cautelares.

Si bien en este apartado vamos a generalizar medidas e implementaciones, quedara constatado que algunas herramientas son vitales debido a su factor de uso.

En la Figura.1 que visualizamos anteriormente, constatamos que la mecánica de implementación del ransomware es la que normalmente se aplica, por lo tanto, tomaremos en referencia la Figura 3 esto para destacar algunas herramientas e implementaciones cruciales para tomar de base.

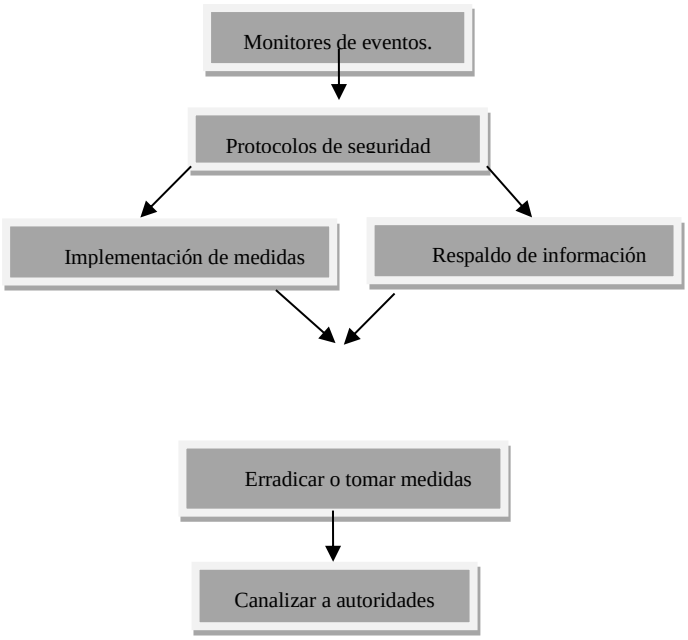


Fig. 3. Ejemplo de una estructura defensiva y preventiva orientada a ransomware, esta puede diferenciar de las comerciales.

- *Monitores de eventos:* Si bien no se cuenta con un Blue Team o Read Team se tiene que contar con diversos analizadores o escáneres de dispositivos y redes, sin profundizar en el tema listaremos algunas herramientas vitales que cumplen con el objetivo: Manage Engine, IBM QRadar, Splunk, Rapid7, InsightIDR, LogRhyth.
- *Protocolos de seguridad:* Dependerá del requerimiento de la empresa los protocolos a seguir en caso de evidenciar alguna intrusión, ya sea desde clausurar redes, bloquear computadoras o restringir accesos.
- *Implementación de medidas:* Aquí es crucial el tipo de implementación puesto que es vital la seguridad e integridad de la empresa, recomendablemente es consultar con un analista en ciberseguridad la opción de implementar un algoritmo de cifrado de datos, puesto que en caso de existir alguna incidencia blindar la información de tal forma que no pueda ser extraída o en dado caso dejarla inservible. Si bien existen infinidad de cifrados, recomendablemente es combinarlos, preferentemente utilizar estructuras binarias y lógicas para que la integridad de la información quede nula.
- *Respaldo de información*: En paralelo a el punto anterior es recomendable que se establezca un efecto en cadena que permita cifrar los archivos y aparte respaldarlos, el punto es estudiar el entorno donde se suscitaron las incidencias, de acuerdo con eso se toman medidas y sirven de que se establece una investigación.

- *Erradicar o tomar medidas:* Una vez reestablecido el orden procederemos a tomar medidas, ya sea de restructuración, limpieza, implementación forense, todo dependiendo de la situación que emergió.
- *Canalizar a autoridades:* Es factible canalizar toda información acerca de la intrusión o intento de intrusión a las autoridades, seguirán al pendiente de estas situaciones y se tendrá un respaldo extra, entre mas conocidos se hagan los intentos de ransomware estaremos contribuyendo a una sana cultura de la ciberseguridad.

Reiteraremos que estas medidas de implementación son opcionales y bajo el criterio de quien este al mando del área de redes o seguridad. Cada empresa personaliza su forma de tal manera de que sea sana y cómoda la seguridad implementada.

Algunas herramientas mencionadas son de costo, dependerán de ciertas circunstancias y entornos, así como las características que brindan. Por lo tanto, se considera, tomarlas de referencia puesto que con el avance tecnológico podrían emerger nuevas herramientas de acuerdo a nuestras necesidades.

5 Conclusión

Bajo la investigación que se realizó se considera que tomando de referencia una estructura sólida bajo el entendimiento de cómo funciona un ransomware se pueden implementar medidas, así como implementaciones de protocolos de seguridad, abordar la situación que se presente y a si tener una estructura sólida, funcional y segura.

Referencias

1. Steve Ranger. (2019). *Ransomware: 11 steps you should take to protect against disaster.* 13 de mayo del 2021, de zdnet Sitio web: https://www.zdnet.com/article/ransomware-11-steps-you-should-take-to-protect-against-disaster/
2. Josh Fruhlinger. (2020*). Ransomware explained: How it works and how to remove it*. 13 de mayo del 2021, de CSO Sitio web: https://www.csoonline.com/article/3236183/what-is-ransomware-how-it-works-and-how-to-remove-it.html
3. Centro Criptológico Nacional. (2018). *Medidas de seguridad contra ransomware. España: Gobierno de España*.
4. Jason Firch. (2020). How To Prevent Ransomware Attacks: An Expert Guide. 13 de mayo del 2021, de Purplesec Sitio web: https://purplesec.us/identify-prevent-remove-ransomware-attacks/
5. Amos Kingatua . (2021). Las 9 mejores herramientas de respuesta a incidentes de seguridad para pequeñas y empresas. 12 de mayo del 2021, de geekflare Sitio web: https://geekflare.com/es/security-incident-response-tools/
6. Daniel Kundro. (2020). Análisis del código fuente de un ransomware escrito en Python. 10 de mayo del 2021, de Eset Sitio web: https://www.welivesecurity.com/la-es/2020/07/29/analisis-codigo-fuente-ransomware-escrito-python/
7. Redacción CIO México. (2020). ¿Cuáles son las señales de una empresa que está a punto de ser atacada con ransomware?. 1 de mayo del 2021, de cio Sitio web: https://cio.com.mx/cuales-son-las-senales-de-una-empresa-que-esta-a-punto-de-ser-atacada-con-ransomware/
8. Amrit Singh. (2021). Ransomware: How to Prevent or Recover From an Attack. 1 de mayo del 2021, de backblaze Sitio web: https://www.backblaze.com/blog/complete-guide-ransomware/
9. Brenda Facundo. (2020). PRÁCTICAS RECOMENDADAS CON FIREWALLS PARA BLOQUEAR EL RANSOMWARE. 28 de abril del 2021, de idric Sitio web: https://www.idric.com.mx/blog-post/practicas-recomendadas-con-firewalls-para-bloquear-el-ransomware
10. JEFF PETTERS. (2020). How To Prevent Ransomware: The Basics. 25 de abril del 2021, de varonis Sitio web: https://www.varonis.com/blog/how-to-prevent-ransomware/

11. Pallavi Dutta. (2021). Top 5 Ransomware Attacks to Watch Out for in 2021. 14 de mayo del 2021, de securityboulevard Sitio web: https://securityboulevard.com/2021/05/top-5-ransomware-attacks-to-watch-out-for-in-2021/
12. Vanson Bourne. (2020). EL ESTADO DEL RANSOMWARE 2020. 3 de mayo del 2021, de sophos Sitio web: https://www.sophos.com/es-es/medialibrary/Gated-Assets/white-papers/sophos-the-state-of-ransomware-2020-wp.pdf
13. Josh Fruhlinger. (19 de junio del 2020). Ransomware explained: How it works and how to remove it. 22 de enero del 2021, de csoonline Sitio web: https://www.csoonline.com/article/3236183/what-is-ransomware-how-it-works-and-how-to-remove-it.html
14. Ivan Belcic. (31 de marzon del 2021). How to Remove Ransomware from Android Devices. 21 de enero del 2021, de avast Sitio web: https://www.avast.com/c-how-to-remove-ransomware-android
15. Cassius Puodzius. (13 de septiembre 2016). Cómo y por qué el cifrado moldeó al ransomware criptográfico. 20 de enero del 2021, de welivesecurity Sitio web: https://www.welivesecurity.com/la-es/2016/09/13/cifrado-ransomware-criptografico/

Ubicación de Instalaciones por Medio del Modelo P- Centro Usando Código Lingo

Rogelio González-Velázquez[1], Erika Granillo-Martínez[2], M. Beatriz Bernábe-Loranca[1], Jairo E. Powell-González[1]
[1]Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla
[2] Facultad de Administración, Benemérita Universidad Autónoma de Puebla, Prol. 14 sur Esq. Av. Sn. Claudio, C.P 72590, Puebla, México
[1]{rogelio.gzzvzz,beatriz.bernabe}@gmail.com, [1]jairoe.powell@viep.com.mx
[2]erika.granillo76@gmail.com

Resumen. El problema P- centro, es un problema clásico de optimización combinatoria que pertenece a la clase de los NP- hard del tipo de localización – asignación, además, puede ser planteado como un problema entero binario (PEB). En el presente trabajo se diseñó un código en el lenguaje del software Lingo para el PEB del P-centro. Se implementó un caso hipotético tomando una región geográfica como instancia de prueba para encontrar la ubicación óptima de los centros de servicios de salud (primer nivel servicios básicos) y su respectiva asignación de usuarios. Los principales resultados se obtuvieron al ejecutar el código para 34 ciudades con 5 y 10 centros, obteniendo dos particiones en un mapa para la atención eficiente de los usuarios.

Palabras Clave: P- Centro, Ubicación Asignación, NP-hard, Lingo, Optimización.

1 Introducción

En el presente trabajo se aplica el modelo de programación entera binaria del problema del P-centro con el propósito de producir una solución a una instancia de dimensión adecuada que pueda ser resuelta por un software como Lingo [1] que utiliza un método de ramificación y acotamiento. Si bien existen diferentes modelos que se pueden aplicar para este tipo de problemas, por ejemplo: el problema de la P- mediana, el problema de ubicación de planta y el problema de cobertura de conjuntos. Se seleccionó el modelo de P-centro que ha sido probado para optimizar redes logísticas [6] obteniendo resultados satisfactorios.

Se abordó un caso de localización de instalaciones y asignación de usuarios para una instancia de prueba real como es el ejemplo del Estado de Puebla. El objetivo consiste en minimizar la distancia recorrida en el sistema. El problema de P-centro es un problema de optimización combinatoria perteneciente a los NP- hard [6,7], es decir, no existe un algoritmo exacto que lo puede resolver en un tiempo polinomial.

En estos problemas se persigue el objetivo de encontrar las ubicaciones del menor número de instalaciones que son necesarias para cubrir toda la demanda de los nodos [1,2]. Se ha demostrado para casos prácticos que el número de instalaciones utilizadas para cubrir toda la demanda de los nodos dentro de una distancia específica son bastante extensas. Aunado a esto, los modelos de ubicación caen en la cuenta que las demandas de los diversos nodos difieren unos de otros [1,2]. Para contrarrestar el problema se han diseñado modelos aplicativos de máxima cobertura para los problemas de ubicación.

Para el modelo de máxima cobertura [17] para los problemas de ubicación, es necesario asociar el nivel de demanda con la demanda de cada nodo para encontrar la

ubicación del número fijo de instalaciones que maximice el número de demandas cubiertas. De modo que, el modelo de máxima cobertura de ubicaciones, suaviza el requerimiento de la demanda de los nodos que serán cubiertos posteriormente.

Con el fin de abordar las deficiencias del modelo se han implementado estrategias en general donde se requiere que todas las demandas de los usuarios sean cubiertas. Sin embargo, para tales efectos, en lugar de utilizar distancias específicas de cobertura y pedirle al modelo minimizar el número de instalaciones necesarias para cubrir todas las demandas de los nodos, le pedimos al modelo minimizar la cobertura de las distancias de cada demanda del nodo de tal manera, que cada demanda de los nodos sea cubierta dentro de una determinada distancia por una de las instalaciones. A este modelo se le conoce como el problema de P- centro o problema minimax [1,2, 17], debido a que se minimiza la máxima distancia entre una demanda y la instalación más cercana a la demanda.

Cabe mencionar, que existe una distinción entre los problemas en los cuales las instalaciones son ubicadas en cualquier punto de la red y los problemas en los cuales las instalaciones se ubican solo en los nodos de la red. Por lo tanto, la primera categoría de problemas conocida como centro absoluto, son aquellos donde las instalaciones son ubicadas en cualquier punto de la red. Por otra parte, los problemas donde las instalaciones solo se encuentran en los nodos de la red se les conoce como centro del vértice. En este trabajo se ubicará de acuerdo a la primera categoría perteneciente al modelo P-centro.

La mayor contribución de este trabajo es la aplicación del modelo P- centro planteado como un problema de programación entera binaria para la ubicación de instalaciones en un caso hipotético, considerando una región geográfica como el Estado de Puebla siendo este el grafo y los municipios como los nodos, la demanda de los nodos se representan por la población de cada municipio.

El presente trabajo se encuentra organizado y presentado de la siguiente forma: sección 1, introducción, seguido de la revisión de la literatura en sección 2, posteriormente, en sección 3, se encuentra el modelo matemático de P-centro, seguido del código de Lingo en sección 5. Los resultados se muestran en la sección 6. Finalmente, las conclusiones y el trabajo futuro aparecen en la sección 7.

2 Revisión de la Literatura

Las actividades logísticas hoy en día son de carácter esencial para la operatividad y la solución de múltiples problemas de transporte, inventarios, distribución, ubicación entre otros. Los problemas de ubicación surgen de la necesidad de definir el sitio más conveniente para ubicar instalaciones como centros de distribución, plantas productivas, almacenes, centros comerciales, centros de servicios, basureros, estaciones de policía o bomberos, hospitales o supermercados [3].

Pero, ¿Cómo se asegura que una instalación se encuentre en la mejor ubicación? Para responder esta cuestión, es necesario definir la mejor ubicación, lo que permite a todos los actores involucrados en las operaciones logísticas de ubicación contar con las mejores ventajas en distancias, costos, tránsitos y tiempo. En otras palabras, se comprueba que los sujetos involucrados trabajan en el óptimo de sus actividades minimizando los elementos antes mencionados [4,5].

El problema de P- centro [6,7] es un problema de localización de los modelos de máxima distancia que pertenece al grupo de los NP-hard, también es conocido como problema de mínima ¨x¨ o como un problema de punto objetivo ya que minimiza la distancia de cada ubicación de demanda al sitio de la instalación como un objetivo separado, de modo que, existe un objetivo por cada ubicación de demanda. El principal objetivo es minimizar la máxima distancia de un nodo de demanda a su instalación más cercana, dado el número predeterminado de instalaciones a ubicar [8].

Desde que el problema de P-centro fue utilizado como modelo en la solución de ubicación [8,7] se han encontrado aplicaciones potenciales como: distribución para almacenes de producto terminado, materia prima, cross dock, donde se resuelven las tareas de traslado de mercancías agrupadas a diferentes sitios [9].

Por otro lado, las aplicaciones de servicios de redes informáticas tienen el propósito de solucionar la ubicación de archivos con bases de datos, otras actividades básicas cotidianas como: servicios médicos (emergencias, hospitales, centros de salud), estaciones policiacas, servicio de bomberos entre otros. En cuanto a la parte de ubicación y designación de instalaciones se encuentran: centrales de autobuses, áreas, férreas y marítimas [10]. Además, para fines gubernamentales se encuentra las siguientes: oficinas del gobierno, centros recreativos, hoteles, parques, centros comerciales y centros educativos. Finalmente, para objetivos militares y bélicos [11].

Otros trabajos que modelan la ubicación de instalaciones se relacionan con modelos de accesibilidad espacial de geografía asociados a la salud pública que mejoran la distribución socio territorial. Para la modelación de estos problemas se utiliza la regresión geográfica ponderara (GWR) [12].

Por otra parte, se han encontrado métodos de dos fases para resolver el problema de P-centro para determinar la ubicación de centros de distribución [13].

Adicionalmente, existen trabajos que usan métodos de aproximación robusta basados en procesos estocásticos para el análisis de la información [14]. Otra variante del P-centro se enfoca en grafos mejor conocidos como Vertex P-centre que consiste en seleccionar p-centros entre un conjunto finito de candidatos y asignarle un conjunto de clientes con el objetivo de minimizar la máxima disimilaridad entre un cliente y su centro asociado [15].

3 Modelo Matemático de P-centro

Para la formulación del modelo matemático de localización se siguió la propuesta por [12], que se presenta a continuación: donde

I= conjuntos de nodos de demanda *i*

J= conjuntos ubicaciones de las instalaciones *j*

d_{ij} = distancia entre la demanda del nodo *i* y sus instalaciones ubicadas en el sitio *j*

h_i = demanda de nodo *i*

P= número de instalaciones para ubicar

Definiendo las siguientes variables de decisión:

X_i= 1 si se ubica en el sitio *j*; 0 en otros casos

Y_{ij}= Si la demanda del nodo *i* es asignada a la instalación ubicada en el sitio *j*; 0 en otros casos

W = máxima distancia entre un nodo de demanda y su instalación a la cual el nodo es asignado.

El P- centro se formula con la siguiente expresión matemática:

$$\text{Min } W \tag{1}$$

Sujeto a:

$$\sum_{i \in J} X_j = p \tag{2}$$

$$\sum_{i \in J} Y_{ij} = 1 \qquad \forall i \in I \tag{3}$$

$$Y_{ij} - X_j \leq 0 \qquad \forall i \in I \; \forall j \in J \tag{4}$$

$$w - \sum_{j \in J} h_i d_{ij} Y_{ij} \geq 0 \qquad \forall i \in I \tag{5}$$

$$X_j \in \{0,1\} \qquad \forall j \in J \tag{6}$$

$$Y_{ij} \in \{0,1\} \qquad \forall j \in J \; \forall i \in I \tag{7}$$

La función objetivo en (1) minimiza la máxima demanda – distancia entre cada nodo de demanda a la instalación más cercana, la restricción en (2) establece que existe un P número de instalaciones para ser ubicadas, la restricción en (3) requiere que cada nodo de demanda sea asignado exactamente a una instalación. Por su parte, la restricción en (4) solo permite que la demanda de un nodo sea asignada a una instalación abierta, la restricción en (5) estipula la máxima distancia entre el nodo *i* y la instalación en el sitio *j*, denotado por W que es más grande que la distancia entre cualquier nodo *i* y la instalación ubicada en el sitio *j*.

Finalmente, en cuanto al conjunto de restricciones en (6) y (7) se establece la naturaleza binaria de las variables de decisión.

4 Metodología de Implementación P-centro

En esta sección se muestra el código en lenguaje de modelación Lingo [1,13], con la implementación del modelo matemático del problema P-centro.

Para ejecutar el algoritmo se introduce una matriz simétrica de distancias entre los nodos de un grafo en la línea 11 de la sección “DATA”, donde algunos nodos son centros potenciales de ubicación de instalaciones. Cada nodo de la matriz tiene una demanda asociada que se coloca como vector de entrada en la línea 10.

En la línea 5 se encuentran las variables de asignación y en la línea 12 se introduce el número de centros *k* que son requeridos a priori (ver algoritmo 1).

La recolección de datos se efectuó seleccionado 34 ciudades del Estado de Puebla investigando su población [18] y la distancia entre estas ciudades marcada en kilómetros , el mapa de estas ciudades es análogo a un grafo, donde las cabeceras municipales forman los nodos y las líneas de distancias simulan los arcos, de esta manera se genera la instancia de prueba con una matriz de distancia y un vector de demanda cuyo tamaña no es adecuado para la versión limitada de Lingo que se ejecutó en este trabajo. Con los datos e información antes mencionada se cubre el diseño requerido para el modelo P-centro. A continuación, se presenta el algoritmo:

Algoritmo 1 . Código Lingo para solución de P-centro

```
1 MODEL:
2 SETS:
3   CD: APERTURA.
4   CLIENTES: DEMANDA.
5  ARCOS (CD, CLIENTES): ASIGNACION, DISTANCIA.
6 ENDSETS
7  DATA:
8   CD;
9   CLIENTES.
10   DEMANDA.
11   DISTANCIA=@file('DISTANCIA.txt');
12   P= k
13 ENDDATA
14  MIN= R;
15  @FOR(CLIENTES(J): R>=@SUM(CD(I): DEMANDA(J)*DISTANCIA (I,
   J) *ASIGNACION (I, J)));
```

```
16  @FOR(CLIENTES(J): @SUM(CD(I): ASIGNACION (I, J)) =1);
17  @SUM (CD(I): APERTURA(I)) = P;
18  @FOR (ARCOS (I, J): ASIGNACION (I, J) <=APERTURA(I));
19  @FOR (CD: @BIN(APERTURA));
20   @FOR (ARCOS: @BIN(ASIGNACION));
21 END
```

5 Caso Específico

Para representar la aplicación del modelo P- centro como una problemática real, fue necesario obtener la información correspondiente a los puntos de referencia [18], para este caso en específico se utilizaron 34 cabeceras municipales del estado de Puebla [18], con el propósito de que cada una representará a los nodos de demanda así como las ubicaciones potenciales de las instalaciones (servicios médicos primer nivel). En la siguiente tabla 1 se presenta la matriz de distancia con algunas cabeceras de municipios.

Tabla 1. Muestra de la matriz de distancia de municipios en kilómetros

Mpio \ Mpio	Puebla	Tehuacán	San M. Texmelucan	Atlixco	San Pedro Cholula	Amozoc
Puebla	0	132	46.6	31.2	13.5	24.9
Tehuacán	132	0	167	166	152	112
San M Texmelucan	46.6	167	0	64.4	9.1	54.7
Atlixco	31.2	166	64.4	0	24.5	54.6
San Pedro Cholula	13.5	152	9.1	24.5	0	40.7
Amozoc	24.9	112	54.7	54.6	40.7	0

Continuando con la aplicación del modelo, se observa en la tabla 2, que para cada nodo fue necesario agregar el valor de demanda. Cabe mencionar, que el problema de estudio se enfocó en la ubicación de nuevos centros de salud de primer nivel con la finalidad de brindar los servicios médicos esenciales a la población de cada región debido a la premura por la reciente pandemia del Covid-19 [19]. Por consiguiente, el valor de la demanda de cada nodo se representa por el número de habitantes de cada municipio.

Tabla 2. Vector de demanda

Puebla	Tehuacán	San M. Texmelucan	Atlixco	San Pedro Cholula	Amozoc
1539819	274906	141112	127062	120459	100964

Una vez definidos los datos anteriores, se implementó el modelo P-centro en código de Lingo para obtener el resultado de la ubicación de las instalaciones, así como de la asignación de los municipios a cada instalación seleccionada.

Se utilizó Lingo como lenguaje de modelación para problemas de optimización diseñado por la empresa Lindo Systems [1]. El lenguaje de Lingo es muy similar al modelo matemático de los problemas de optimización, de modo que es fácil y conveniente de usar, además, dentro de los diferentes softwares de optimización es considerado uno de los más robustos.

6 Resultados

El código Lingo fue ejecutado para el problema de P-centro en una computadora laptop i7 Toshiba. Las ejecuciones se realizaron para dos valores de P=5, P=10 para comparar resultados, es decir, si se requiere observar la distribución de la población en 5 instalaciones o en 10 instalaciones. Los valores de P son valores de prueba que ayudan a verificar la ejecución del código. Respecto al valor de P, se espera que sea propuesto y designado por un planeador estratégico gubernamental con conocimientos de designación de presupuestos que para este caso se aplica al sector salud.

Los resultados para P= 5 (centros), indican que los municipios seleccionados para la ubicación de instalaciones son: Puebla, Tehuacán, San Pedro Cholula, Huachinango y Zacapoaxtla. Para el caso de Puebla se le asignaron los siguientes municipios: San Pedro Cholula, Amozoc, San Andrés Cholula, Cuautlancingo, Tepeaca, Izúcar de Matamoros, Huejotzingo, Acajete, Acatzingo, Quecholac, Tlahuapan, Coronango, Chietla, Chignautla.

Para Tehuacán se asignaron: Tecamachalco, Ajalpan, Tlacotepec de Benito Juárez, Chalchicomula de Sesma, Palmar de Bravo, Acatlán, Tlachichuca.

Seguido de San Pedro Cholula con la siguiente asignación: San Martín Texmelucan, Atlixco. Posteriormente, Huachinango con el resultado de: Zacatlán, Xicotepec, Chignahuapan. Finalmente, Zacapoaxtla con las asignación siguiente: Tlatlauquitepec, Cuetzalan del Progreso, Xiutetelco, Libres.

Los resultados para P= 10 (centros) se encontró la solución de esta forma: Puebla, Tehuacán, San Martín Texmelucan, Atlixco, Huachinango, Teziutlán, Acatlán, Tecamachalco, Zacapoaxtla, Zacatlán.

El desglose de las asignaciones se presenta a continuación: *Puebla* con una asignación de 12 municipios: San Pedro Cholula, Amozoc, San Andrés Cholula, Cuautlancingo, Huejotzingo, Acajete, Acatzingo, Tlahuapan, Coronango, Chietla, Libres. *Tehuacán* con 5 municipios: Ajalpan, Tlacotepec de Benito Juárez, Palmar de Bravo, Tlachichuca. *San Martín Texmelucan* se quedó con su misma ubicación, *Atlixco* con 2 municipios: Atlixco e Izúcar de Matamoros. A *Huachinango* se le asignaron 4 municipios: Huauchinango, Zacatlán, Xicotepec, *Teziutlán* quedó con 4 municipio: Teziutlán, Zacapoaxtla, Tlatlauquitepec, Xiutetelco. A *Zacatlán* se le asignaron 2 municipios: Zacatlán y Chignahuapan. Por otra parte, *Tecamachalco* se quedó con 4 sitios: Tepeaca, Tecamachalco, Quecholac, Chalchicomula de Sesma, el siguiente fue *Zacatlán* con 2 puntos: Cuetzalan del Progreso y Chignautla, finalmente, *Acatlán* que se quedó en su mismo sitio Acatlán.

El método usado en Lingo [1] para resolver problemas enteros se basa en ramificación y acotamiento (Branch & Bound) [1], método usado por excelencia por otros softwares para resolver el problema de la P –centro.

Debido a que el tamaño de la matriz es de tamaño menor, no se esperan cambios significativos en cuanto a la solución.

7 Conclusiones y Trabajos Futuros

En el presente trabajo se aplicó el modelo de P-centro que tiene la característica desde su creación de minimizar la máxima distancia [17] como se describe en diversas publicaciones [2,3,7] se busca la ubicación de los centros potenciales de abastecimiento de servicios y la asignación de usuarios de tal manera que se minimice la distancia recorrida por los usuarios en el sistema por medio de minimizar la máxima distancia de un punto de demanda hasta su centro de servicio más próximo.

Los resultados de la investigación arrojaron que el municipio de Puebla tiene asignado el mayor número de nodos de demanda en comparación de las demás instalaciones. Es notorio observar que no existe un balance equitativo entre las cargas (demandas).

Cabe mencionar que los resultados obtenidos en el trabajo coadyuban en la toma de decisiones estratégicas a largo plazo. Sin embargo, la ubicación es solo un componente de un proyecto integral, donde se requiere de otros análisis para verificar elementos, por ejemplo: el uso de suelo, la plusvalía y vías de comunicación. Adicionalmente, si se selecciona una ciudad como punto de ubicación, todavía se requiere hacer un análisis de escala de ubicación del sitio para la instalación.

Como trabajo futuro, se buscará implementar el modelo P-centro con una metaheurística, ya que es un problema NP- hard, por ejemplo: GRASP o Recocido Simulado a fin de generar una herramienta de tecnologías de información que sea utilizada de manera abierta para buscar soluciones aproximadas a la óptima para instancias de gran escala.

Referencias

1. Lindo Systems Inc.: Lingo: The Modeling Language and Optimizer. Lindo Systems Inc. Chicago, Illinois (2020). Lindo Homepage, https://www.lindo.com/downloads/PDF/LINGO.pdf, last accessed 2021/10/06
2. Daskin, M.S.: Chapter 5 Center Problems. Network and Discrete Location: Models, Algorithms and Applications. John Wiley& Son, Inc. (1995).
3. Daskin, M. S.: Network and Discrete Location: Models, Algorithms, and Applications. 2nd ed. John Wiley & Sons. Michigan, (2013).
4. Revelle, C.: A perspective of location Science. Location Science, 5, 3–13 (1997).
5. Granillo, M.E., González, V. R., Bernabé, L. MB., & Martinez, F. JL.: FOOT-TRAFFIC ANALYSIS: EVIDENCE FROM A MALL IN MEXICO. Revista Internacional de Administración y fianzas. 10(4), 71-79 (2017).
6. Romero, M.M., González, V. R., Martinez, F., Bernabé, L. MB., & Estrada,A.M.: Solution Search for the Capacited P- Median Problem using Tabu Search. International Journal of Combinatorial Optimization Problems and informatics. 10(2), 17-25 (2019).
7. Hakimi, S.: Optimum location of switching centers and the absolute centers and medians of a graph. Operations Research, 12, 450-459 (1964).
8. Hakimi, S.: Optimum location of switching centers in a communications network and some related graph theoretic problems. Operations Research, 13, 462-475 (1965).
9. Hale, T., & Moberg, Ch.: Location science research: A review. Annals of Operations Research, 123(1), 21-35(2003).
10. Calik, H.: "Exact solution methodologies for the p-center problem under single and multiple allocation strategies" Ph.D. dissertation, Operation Research, Bilkent University Ankara, Turkey, (2013).
11. Araneda, M. R. H., & Moraga, R. J.: La Decisión de Localización en la Cadena de Suministro. Revista de ingeníeria Industrial. 3(1), 57-67 (2004).
12. Martínez, B.M and Rojas, Q. C.: Regresión Geográficamente Ponderada para la modelación de la Accesibilidad a la Red Hospitalaria en el área Metropolitana de Concepción. Rev. geogr. Valpso 52, 28-39 (2015).
13. Azzindani, M.R., Kusuma Ningrum, N.F., Sudiar, M.R., & Redi, A.A.: Two-Phase Optimization Method for Determining Distribution Center Locations and Distribution Routes. IPTEK Journal of Proceedings Series. 5, 44-48 (2020).
14. Du, B., & Zhou, H.: A Robust Optimization Approach to the Multiple Allocation p-Center Facility Location Problem. Symmetry. 10 (588), 1-15 (2018).
15. Contardo, C., Iori, M. and Kramer, R.: A scalable exact algorithm for the vertex p-center problem. Computers & Operations Research. 103, 211-220 (2019).
16. Biazaran, M., & SeyediNezhad, B.: Facility location: Concepts, Models, Algorithms and Cases of Study. R.Z. Farahani & M. Hekmatfar (Eds). Springer, Verlag Berlin Heidelberg (2020).

17. Current, J., Daskin, M., & Schilling, D.: Discrete Network Location Models. In: Facility Location Theory: Applications and Methods, 85-112. Z. Drezner and H. Hamacher (Eds). Berlin: (2001).
18. INEGI Homepage, https://www.inegi.org.mx/, last accessed 2021/10/08.
19. Bernábe, L.M. B., González, V.R., Granillo, M.E., Ruíz, V.J., and Carrillo, C.A.: Towards an Approach of the Contagion Curve for COVID-19 in México. Intelligent Systems Design and Applications. 553-566 (2021)

Pruebas de Ataques Basados en Inyección SQL

Marisol Soriano, Ana C. Zenteno, Ma. del Carmen Santiago, Yeiny Romero, Judith Pérez, Gustavo T. Rubín
Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Ciudad Universitaria, 14 sur y Avenida San Claudio, Fraccionamiento Jardines de San Manuel, CP. 72570, Puebla, Pue; México.
marisol.soriano@alumno.buap.mx,{ana.zenteno, marycarmen.santiago, yeiny.romero, judith.perez, gustavo.rubin}@correo.buap.mx

Resumen. En el presente documento se abordará un tema de seguridad en redes, una de las vulnerabilidades más comunes de los sitios web son las inyecciones SQL. Hablaremos de lo que es una inyección SQL, los tipos de inyecciones que existen actualmente, abordaremos y utilizaremos algunas herramientas automatizadas de ataques SQL de código libre para evaluar su efectividad dentro de ambientes simulados y reales , por último realizaremos el análisis de los ataques realizados a los sitios web.

Palabras Clave: Ataque, Inyección SQL, IPS, Seguridad en Redes

1 Introducción

El uso de internet a lo largo de los años se ha vuelto cada vez más cotidiano, tanto para realizar actividades del día a día como visitar un sitio web, actividades más complejas como pagos en línea, uso de servicios de streaming, por citar algunos. Conforme este tipo de servicios avanzan, se ha vuelto cada vez más indispensable tener almacenada la información personal de los usuarios en línea, que es igual de importante que la información almacenada de forma física, esta información válida nuestra identidad dentro de algún sitio o servicio en línea, por lo que al igual que un documento físico con nuestra información, este debe ser resguardado de una forma segura en línea.

Pero al igual que en la vida real, siempre existirá alguien que intente sacar provecho de información del usuario, de un sitio o aplicación web, a pesar de que la información se encuentre en un servidor en línea este puede llegar a ser susceptible a ataques informáticos con el único afán de obtener información y beneficiarse de la misma. Por lo que conforme avanza la tecnología, se ha vuelto cada vez más importante el uso de técnicas que nos brinden la mayor seguridad posible contra estos ataques.

La mejor forma de saber cómo proteger un sitio web es conocer las posibles amenazas a las que este podría ser susceptible.

Existen según la OWASP [1] las 10 vulnerabilidades más relevantes de las que un sitio web podría ser víctima, en este documento se abordará la vulnerabilidad listada como la número uno, inyecciones SQL.

2 Ataques de Inyección SQL

Estos ataques de inyección SQL [2] buscan vulnerar al motor del gestor de bases de datos a través del mismo motor en sí, mediante el uso de Querys (consultas) que parezcan válidas, o que el mismo motor pueda interpretar como válidas cuando en realidad son

consultas a las que se les añadió instrucciones extra para que el motor le muestre al usuario información de la base de datos a la que usualmente no debería tener acceso.

Por mencionar algunos ejemplos sencillos, consideremos un sitio web en el que se solicite al usuario su ID para ingresar, el flujo de datos normal sería que el usuario solo ingrese su ID y la base de datos consulte este ID.

De existir retornará todos los datos del usuario con el ID ingresado, pero ¿qué pasaría si no se valida este dato?

Es aquí donde el usuario puede agregar más texto que el motor del gestor de bases de datos interpreta como válido, si el usuario añadiera *OR 1=1* en el campo ID, la consulta se realizaría.

El *OR* añadido convierte toda la consulta en verdadera, ya que 1=1 siempre va a ser verdadero, lo que retorna todos los usuarios de la tabla UserId.

Mencionando otro ejemplo un poco más complejo, imaginemos que ahora se debe ingresar el ID y el password para validar a un usuario, en dos campos distintos, uno para el usuario y otro para ingresar el password, suponiendo que se desee realizar un ataque de inyección SQL, el atacante podría ingresar en ambos campos *" or ""=" lo que generaría una consulta exitosa*.

Esto siempre será verdadero ya que la operación *" or ""="* siempre es verdadera, porque esta consulta retorna todo el contenido de la tabla Users.

Este tipo de inyección suele ser el más básico y por tanto no muy utilizado actualmente debido a que los sitios web suelen estar preparados para este tipo de ataques.

Entre los ataques de inyección SQL existentes más complejos podemos encontrar tres tipos de ataques. Los *Union Based Attacks*, *Error Based Attacks* y los *Blind Based Attacks* este último puede ser de dos tipos, ya sea *Boolean Based* y *Time Based* cómo se puede apreciar en la Figura 1.

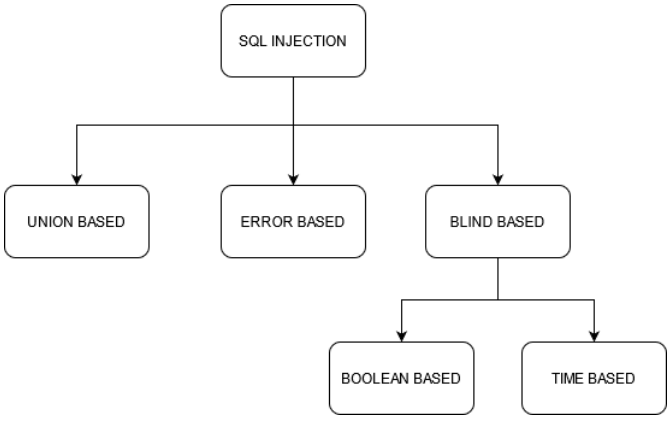


Fig.1. Tipos de inyecciones SQL

2.1 Union based

Es una técnica de inyección SQL que aprovecha el operador UNION SQL[3] para combinar los resultados de dos o más declaraciones SELECT en un solo resultado que luego se devuelve como parte de la respuesta HTTP.

Los ataques basados en UNION permiten al evaluador extraer fácilmente información de la base de datos. Debido a que el operador UNION solo se puede usar si ambas consultas tienen exactamente la misma estructura, el atacante debe crear una declaración SELECT similar a la consulta original.

Esta consulta SQL devolverá un único conjunto de resultados con dos columnas, que contienen los valores de las columnas a y b en table1 y las columnas c y d en table2.

2.2 Boolean based

Este tipo de ataques así como los Error Based pertenecen a la categoría de los ataques Blind Based[4] se basa en la observación de los resultados de diferentes consultas, esto ocurre cuando el sitio web no maneja los errores que se pueden mostrar, y al ocurrir un error con la base de datos este los muestra en el navegador. Por medio de esto, el atacante puede enviar Querys para observar en qué situaciones el servidor no regresa ningún error (Query Verdadero) o en cual el sitio muestra un error (Query Falso). Con esto puede conocer cómo está estructurada la base de datos.

2.3 Error based

Este tipo de ataques son parecidos a los *Boolean Based* con la diferencia es que estos buscan cualquier error que pueda mandarse al sitio web por medio de la base de datos, estos mismos errores harán que el atacante poco a poco vaya observando la estructura de la base de datos.

3 Herramientas Utilizadas

SQLMap[5] es una herramienta de código abierto para la inyección SQL. Esta herramienta hace fácil la explotación de vulnerabilidades de una aplicación web y toma la base de datos del servidor. Además de tener incluido un poderoso motor de detección de vulnerabilidades SQL.

Entre las bases de datos que soporta se encuentran MySQL, Oracle, PostrgreSQL, Microsoft SQL server, Microssoft Acces, IBM DB2, SQLite, Firebird, Sybase, SAP MaxDB y HSQLDB.

Una buena característica de la herramienta es su sistema de reconocimiento de password hash. Ayuda a identificar la contraseña hash y entonces descifra la contraseña realizando un ataque de diccionario.[6]

Esta herramienta te permite descargar o cargar cualquier archivo desde la base de datos del servidor si el servidor de base de datos es MySQL, PostgreSQL o Microsoft SQL Server. Y solo para esos tres servidores de bases de datos, también te permite ejecutar arbitrariamente comandos y recuperar su salida estándar en el servidor de la base de datos.

Después de conectarse al servidor de la base de datos, esta herramienta también te permite realizar búsquedas de una base de datos específica, tablas específicas o columnas específicas en toda la base datos del servidor. Esto es una característica muy útil cuando se quiere buscar una columna en específico pero la base de datos del servidor es muy grande y contiene muchas bases de datos y tablas.

SQLIV[7] es una herramienta que permite la búsqueda automatizada de sitios web que tengan algún tipo de vulnerabilidad propensos a una inyección SQL está basado en python2, y se usa más comúnmente con Kali Linux.

Esta herramienta se analizará posteriormente en la sección de pruebas realizadas y se detalla de mejor forma el uso de la misma.

4 Pruebas Realizadas

Se realizaron diversas pruebas tanto en un servidor en localhost como en un servidor en línea, a continuación, se especifican los detalles para cada caso y como se implementaron.

4.1 Pruebas en localhost

Para realizar las pruebas de vulnerabilidad se utilizó la herramienta Sqlmap en un sitio creado por nosotros para fines de prueba. Este sitio está corriendo en localhost bajo el servicio que ofrece Wamp[8] para desplegar nuestro sitio web.

Cabe mencionar que nuestro sitio web está funcionando sin configuraciones extra tanto en la base de datos como en el sitio web.

4.2 Estructura de la página

La estructura del sitio web es sencilla, se tiene la página para el login de usuarios y otra para el login de administradores.

Una vez ingresados el usuario o el administrador, se muestran las páginas correspondientes y pertinentes para cada tipo de usuario, en este caso para el usuario común se le despliega el sitio web de uso común, donde se muestran secciones de noticias, sus configuraciones de usuario, etc.

Mientras que para el administrador se le muestra el panel de administración, donde puede gestionar a los usuarios y el contenido de la página.

Nuestro ataque por medio de Sqlmap será al administrador, por medio de una de las páginas desplegadas al administrador.

Al ingresar en esta página con los datos que se nos presentan, se nos redirecciona a otra página en donde se hace la conexión a la base de datos para confirmar que el administrador que ingresa con los datos dados en el formulario de login son correctos y se le de acceso al panel de administración, esta página de conexión corresponde a *check_admin.php* en la subcarpeta del sitio llamada *conects.*

En caso de encontrar al administrador en la base de datos, se redirige al administrador al panel de administración, de lo contrario se regresa a la página de login.

4.3 Vulnerabilidades en localhost

La página de conexión a la base de datos, al realizar la petición para consultar los datos lo que hace es hacer la consulta por medio del método *GET* lo que como sabemos, es inseguro, debido a que se envían los parámetros de consulta en la misma URL, por lo que al realizar la consulta la URL llevaría los parámetros: adm=admin&password=123456789&sub1=Entrar

Como se puede observar, se están enviando los parámetros “adm” y “password” los cuales se verifican en la base de datos, al ser visibles ambos campos en la url se convierten en las vulnerabilidades que explotaremos por medio de Sqlmap para obtener los datos de las bases de datos, no solo de ese sitio, sino de otros sitios más que se tienen en nuestra base de datos local.

4.4 Ataque realizado a localhost

Para realizar las pruebas, trabajamos con el link vulnerable que se mostró anteriormente, en sqlmap debemos ingresar la URL mediante el comando -u, que en este caso fue el link previamente mostrado, y para obtener las bases de datos alojadas en el servidor se utiliza el comando -dbs

Y solo se dejó el parámetro “adm” para trabajar.

Y como se puede observar, nos mostró los nombres de las bases de datos que se tienen en el servidor.

Ahora para obtener las columnas de una de las bases de datos, esto lo obtuvimos por medio del comando -D para ingresar la base de datos de la que queríamos analizar, en este caso página, y -tables para obtener los nombres de las tablas además del comando -dump para que se muestran en la consola las tablas como se muestra en la figura 2.

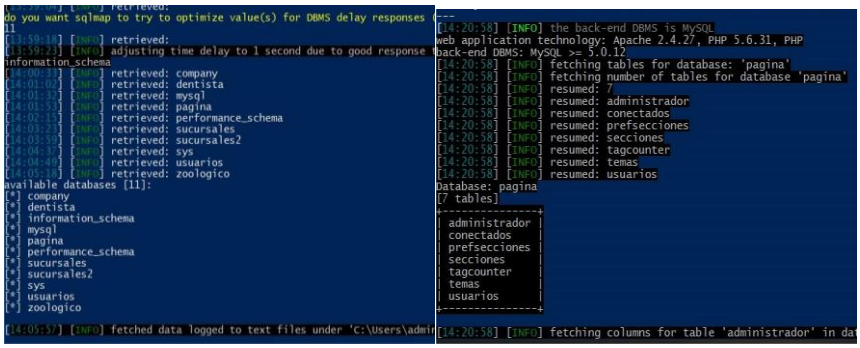


Fig. 2. Muestra de las bases de datos encontradas y tablas obtenidas

4.5 Pruebas de vulnerabilidad en hosting gratuito

Se subió nuestra página a un hosting gratuito donde se cargaron los archivos y se cambiaron las credenciales de las bases de datos con las de la base de datos del hosting.

Una vez realizados los cambios, se probó que el sitio estuviera correctamente en línea para proceder a realizar los ataques como se realizaron en localhost.

4.6 Ataque en hosting

Se realizó una primera prueba de conexión para saber si la herramienta podría enlazarse a la página y obtener algún tipo de información como con las pruebas de localhost intentamos obtener los nombres de las bases de datos con el mismo comando que se utilizó anteriormente.

Pero los resultados que nos arrojó la herramienta nos mostraron que el parámetro “adm” no era inyectable y se nos mostró un error el cual indicaba que no se podía establecer conexión con la página debido a una falta de contenido.

Por lo que no se pudo obtener ningún dato de la base de datos del servidor ya que el sitio cuenta con una seguridad X-xss[9] en modo bloque, el cual previene la visualización de la página si llegara a detectar algún ataque por medio del navegador.

Y tiene además un content-type configurado como nosniff.

4.7 Pruebas con SQLIV

Para estas pruebas de vulnerabilidad, trabajamos con SQLIV la cual nos permite buscar sitios que tengan alguna vulnerabilidad de seguridad que pueda ser explotada. La herramienta buscará sitios que contengan una secuencia de caracteres en la url, como se puede ver en la figura 3. Se realizó una búsqueda de sitios web que tuvieran la secuencia “index.php?id=” en su URL

Como hemos visto antes, esto es muy peligroso ya que es fácil manipular estas URL para realizar consultas a la base de datos del sitio web.

Por medio de SQLIV se realizó una búsqueda de sitios web vulnerables pertenecientes al dominio Bing, se seleccionó un sitio web de la lista para realizar una consulta sin fines de dañar el sitio web ni obtener información sensible, por lo que se consultó el gestor de bases de datos y nombre de tablas.

Cabe recalcar que no se extrajo ninguna información de las tablas, solo se realizó una consulta para conocer el nombre de las mismas.

Fig. 3. Búsqueda de páginas vulnerables con sqliv.py

Una vez elegido el sitio para realizar el ataque, se procedió a hacer la inyección con sqlmap como se puede ver en la figura 4.

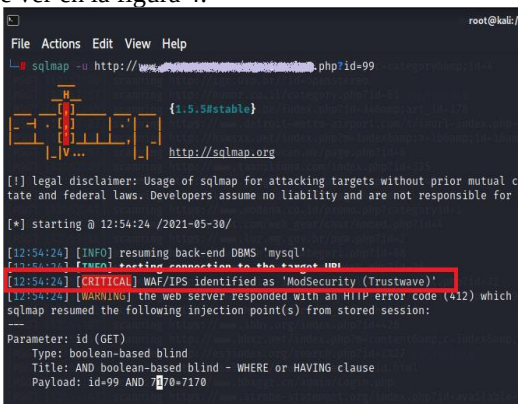


Fig. 4. Intento fallido de conexión por seguridad WAF/IPS

En el primer ataque no se pudo obtener información, la herramienta nos muestra sugerencias de comando cuando un ataque es fallido en este caso nos sugirió --random-agent en caso de que el sitio contuviera algún tipo de seguridad como WAF/IPS

Por lo que se utilizó esta opción para el segundo ataque realizado, esto se puede observar en la figura 5. Que muestra cómo se pudo vulnerar ese tipo de seguridad y pudimos consultar la información de la base de datos actualmente utilizada en el sitio, los resultados de la siguiente consulta se explican más a detalle en la sección de resultados.

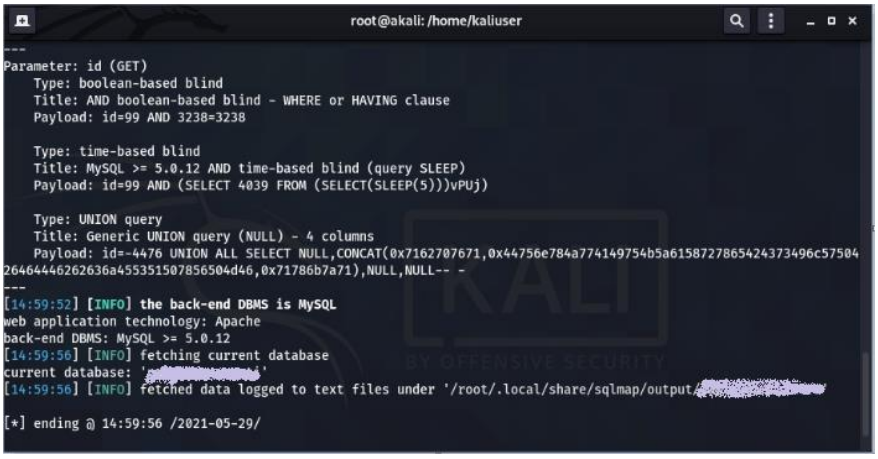


Fig. 5. Ataque con sqlmap

5 Resultados

Pudimos observar que el sitio web almacenado en el hosting gratuito por diversas pruebas que se realizaron a pesar de que se utilizaban los ataques de fuerza bruta hasta el límite de la capacidad de la herramienta no fue posible conseguir la conexión al sitio web, como se explicó este sitio web cuenta con una configuración de seguridad muy robusta, la cual en localhost era prácticamente nula.

Esto indica que el sitio web está debidamente configurado para este tipo de ataques como se puede observar en la figura 6.

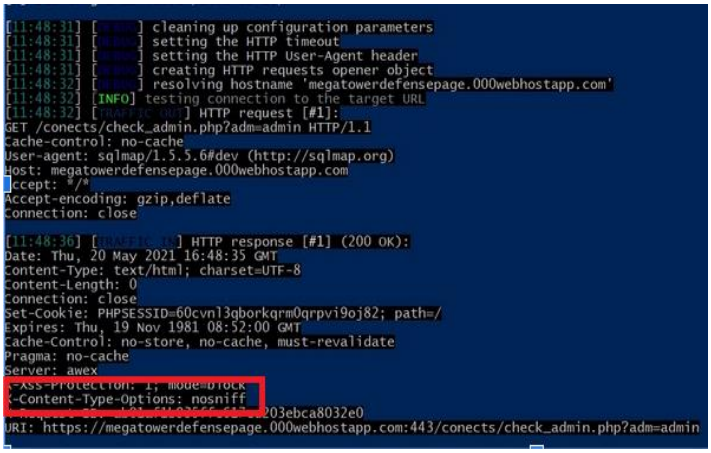


Fig. 6. Configuración de seguridad

Por lo que obtener más información de este sitio fue imposible.

Posteriormente en las pruebas con SQLIV fue donde se tuvo más éxito, ya que existe una gran cantidad de sitios web potencialmente vulnerables en su seguridad, por lo que trabajar con SQLIV para buscar un sitio web y SQLMAP para atacar el mismo, supone una gran facilidad para encontrar y atacar sitios vulnerables de manera automática.

Al atacar el sitio en la primera prueba este rechazó correctamente el intento de inyección SQL pero una vez agregado otro algoritmo de ataque la protección WAF/IPS se pudo vulnerar y se hizo conexión con el sitio web y la base de datos.

Una vez que se obtuvo la base de datos asociada al sitio web, se realizó una nueva consulta para obtener el nombre de las tablas que maneja el sitio y se utilizó la opción random-agent para asegurarnos que se conectara correctamente al sitio en cada consulta. Esto lo podemos observar en la figura 7.

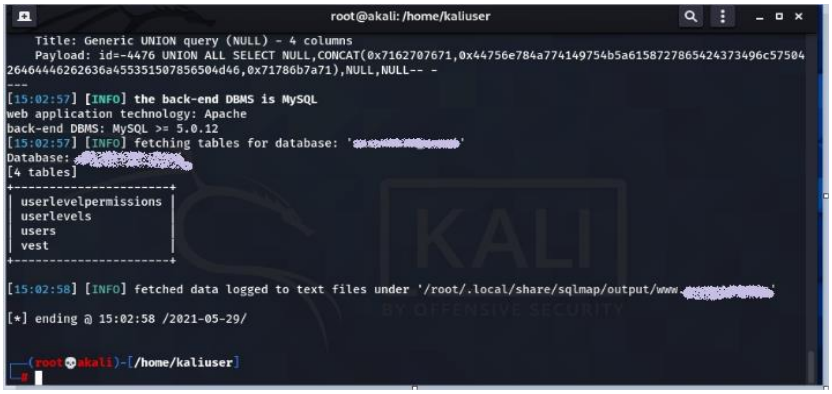


Fig. 7. Tablas del sitio vulnerable

Como se pudo observar, en este nuevo intento ya no se tuvo problema alguno para obtener los nombres de las tablas.

Esto nos indica que a pesar de contar con diferentes mecanismos de seguridad en un sitio web, ya sea el proxy, un WAF/IPS, etc. si no están debidamente configurados y actualizados, deja vulnerable nuestro sitio web a cualquier tipo de ataque.

6 Conclusiones

De acuerdo a la investigación que se hizo mediante los ataques realizados al sitio web tanto en el servidor de localhost como puesto en línea en un hosting, así como los ataques realizados a sitios web vulnerables que se encontraron con SQLIV hacen notar que cada vez es más importante el uso de herramientas y técnicas más sofisticadas para la detección de estos ataques, el uso de inteligencia artificial como parte de la estrategia de clasificación de tráfico dentro de nuestra red sería una buena implementación a futuro para complementar la seguridad de un servicio en línea.

Podemos concluir que existen muchas herramientas de seguridad, como los IDS/IPS, firewall, proxy, etc. y es importante priorizar la información más importante ya que existen muchos tipos de ataques y se puede implementar una topología de seguridad como nosotros lo prefiramos, pero eso no garantiza que sea infalible a todos los ataques que se le puedan realizar a un sitio web.

Referencias

1. Owasp.org. 2021. *SQL Injection | OWASP*. [online] Disponibles en: https://owasp.org/www-community/attacks/SQL_Injection [Consultado el 13 May 2021].
2. Binnie, Chris. (2016). SQL Injection Attacks. 10.1002/9781119283096.ch10.
3. Medium. 2021. *SQL injection UNION attack*. [online] Disponible en: https://medium.com/@nyomanpradipta120/sql-injection-union-attack-9c10de1a5635 [Consultado el 14 May 2021].
4. Owasp.org. 2021. *Blind SQL Injection | OWASP*. [online] Disponible en: https://owasp.org/www-community/attacks/Blind_SQL_Injection[Consultado el 13 May 2021].
5. Sqlmap.org. 2021. *sqlmap: automatic SQL injection and database takeover tool*. [online] Disponible en: https://sqlmap.org/ [Consultado el 16 Mayo 2021].
6. Cert.org.mx. 2021. *Ataque de diccionario | UNAM-CERT*. [online] Disponible en: https://www.cert.org.mx/glosario/ataque-de-diccionario [Consultado el 13 Mayo 2021].
7. GitHub. 2021. *the-robot/sqliv*. [online] Disponible en: https://github.com/the-robot/sqliv [Consultado el 20 Mayo 2021].
8. WampServer. 2021. *WampServer*. [online] Disponible en: https://www.wampserver.com/en/ [Consultado el 20 Mayo 2021].
9. Developer.mozilla.org. 2021. X-XSS-Protection - HTTP | MDN. [online] Disponible en: https://developer.mozilla.org/es/docs/Web/HTTP/Headers/X-XSS-Protection [Consultado el 1 Julio 2021].
10. 2021. [online] Disponible en: https://www.cloudflare.com/es-la/learning/ddos/glossary/web-application-firewall-waf/ [Consultado el 1 Junio 2021].

Plan de Actualización Profesional para Cumplir el Sexto Atributo de Egreso para la Acreditación por CACEI para las Carreras de ISC e IADyEV del TESCHI

José Hernández Santiago[1,2], José Sergio Ruiz Castilla[2], Beatriz Hernández Santiago[2]
[1] Tecnológico de Estudios Superiores de Chimalhuacán, Primavera s/n, col. Santa María Nativitas, C. P. 56330, Chimalhuacán Edo. de México.
[2] Posgrado e Investigación UAEMEX (Universidad Autónoma del Estado de México), Av. Jardín Zumpango s/n, fracc., el Tejocote, C. P. 56259, Texcoco, Edo. de México.
jhernandezs@uaemex.mx,betty_hsbt@hotmail.com, jsergioruizc@gmail.com

Resumen. El Consejo de Acreditación de la Enseñanza de la Ingeniería (CACEI) es un organismo encargado de validar que los programas de ingeniería cumplan con los estándares internacionales de calidad educativa. De los 4859 Programas de Estudio de Ingeniería en México, el 24% han obtenido el reconocimiento de calidad [1]. En esta investigación se presenta la implementación de un plan de actualización profesional durante el periodo 2019-2 al 2021-1 para las carreras de Ingeniería en Sistemas Computacionales (ISC) e Ingeniería en Animación Digital y Efectos Visuales (IADyEV) del Tecnológico de Estudios Superiores de Chimalhuacán (TESCHI). Cumpliendo con el sexto atributo de egreso requerido por CACEI y contribuyendo en la acreditación del programa de ISC con vigencia de mayo del 2020 a mayo del 2025, mientras que la carrera de IADyEV continua en proceso de evaluación y se espera que las evidencias de la actualización profesional ayuden también a la acreditación de este programa.

Palabras clave: Acreditación, Plan de Actualización, Proyecto Integrador, CACEI, ISC, IADyEV, Ingeniería.

1 Introducción

La acreditación es un proceso creado para garantizar la calidad y pertinencia de los programas educativos, buscando que éstos cumplan los estándares mínimos internacionales reconocidos para los programas de buena calidad en ingeniería y se promueva en las instituciones la cultura de la mejora continua de los programas educativos, incorporando las tendencias internacionales para la formación de ingenieros [2].

CACEI es uno de los organismos reconocidos por el Consejo para la Acreditación de la Educación Superior (COPAES) y se encarga de validar que los programas de ingeniería cumplan con los estándares internacionales de calidad educativa.

El proceso de acreditación que se realiza en México requiere una gran inversión, de $141 752.0 con IVA incluido de acuerdo a la página de CACEI [3] y es de carácter voluntario para las instituciones educativas, inconvenientes que de acuerdo con estudios realizados en [1] por el CACEI, del total de Programas de Estudio de Ingeniería y Técnico Superior Universitario Acreditados (4 859), solo el 24% (1 156) han obtenido el

reconocimiento de calidad de CACEI. En el área de computación 144 programas de estudio de 927 cuentan con el reconocimiento, representando un 16 % a nivel nacional. En el caso de la carrera de Ingeniería en Sistemas Computacionales a nivel nacional 77 programas públicos cuentan con acreditación CACEI (Información a Mayo 2021) [4].

De las ventajas que se obtienen al contar con un programa educativo acreditado se destacan las siguientes: Los estudiantes tendrán la certeza de que lo aprendido es pertinente y actualizado, la institución incrementará el reconocimiento social y prestigio académico. Se contribuye a la formación de egresados satisfechos en su nivel académico y con mayor seguridad para afrontar los retos de ejercer una profesión, al obtener competencias con capacidad de adaptación en un medio que requiere actualización constante, repercutiendo en mejores condiciones laborales y la posibilidad de auto superación, además de permitirles cursar un posgrado en cualquier otra IES acreditada por alguno de los demás signatarios de los 18 países miembros del Washington Accord al cual pertenece CACEI [2].

Con la finalidad de acreditar las carreras de ISC y ADyEV en el TESCHI, en el año 2019 se inició un proceso de alineación entre los objetivos educacionales de los planes de estudio y los objetivos educacionales de CACEI.

En este trabajo se presenta el desarrollo de un plan de actualización profesional mediante un proyecto integrador para cumplir con el sexto de los siete atributos de egreso definidos por CACEI, el cual se enfatiza en la necesidad permanente que tienen los profesionales sobre adquirir y aplicar conocimiento actual.

2 Trabajos Relacionados

La importancia de que las universidades cuenten con programas acreditados se ha abordado por diversos investigadores. En [5] se resalta que las carreras profesionales son impartidas por diversas instituciones del sector público y privado, requiriéndose una forma de evidenciar la heterogeneidad de la calidad de la formación educativa respecto a una misma profesión mediante un proceso de acreditación.

Los organismos acreditadores en México y que además tienen alcance internacional, se describen en [6]. Un beneficio de la acreditación radica en obtener financiamiento mediante la vinculación como se menciona en [7], donde la Universidad Autónoma de Tamaulipas definió un plan estratégico 2003-2007 para mejorar su sistema educativo. En [8], se indica que el Centro de Estudios Superiores del Estado de Sonora logro que 66% de sus profesores de tiempo completo contara con nivel de posgrado, incrementando su productividad en investigación a nivel nacional e internacional.

En 2003 el Instituto Tecnológico de Aguascalientes obtuvo la acreditación ante CACEI, requiriendo 6500 horas hombre en un periodo de 31 meses, contando con un equipo de 19 maestros y personal de los diferentes departamentos [9]. Otro programa acreditado por CACEI es el de Ingeniero Mecánico Administrador de la Universidad Autónoma de San Luis Potosí, que en el 2012 realizó su tercera re acreditación preparándose con un instrumento de autoevaluación mediante un Diagnóstico Operacional a partir de la experiencia de sus anteriores acreditaciones [10]. En el 2020 el programa de Ingeniero en Mecatrónica de la Universidad Autónoma de Nuevo León también fue acreditado por CACEI [11].

En [12], el Instituto Tecnológico Superior de Huachinango indicó que el problema de realizar la acreditación como requerimiento puede no conducir a la calidad. En [13] se argumenta que la calidad es una cualidad imprescindible e inherente en la educación que se pretende alcanzar con las acreditaciones; sin embargo, se requiere capacitar evaluadores ya que en ocasiones solo se limitan a realizar una lista de cotejo en base a los criterios establecidos.

Finalmente, en [14] remarca que la equidad educativa significa ofrecer igualdad de oportunidades educativas de buena calidad para todos, sin importar que provengan de diferentes estratos sociales. Es aquí donde la acreditación de la educación superior

permite detectar desigualdades de calidad en los programas educativos, propiciar una mejora continua y asegurar la calidad en los mismos.

• Preliminares

CACEI

El Consejo de Acreditación de la Enseñanza de la Ingeniería, Asociación Civil (CACEI), se constituye formalmente el 6 de julio de 1994 como la primera instancia acreditadora de México con el objetivo de apoyar a la sociedad mexicana en la promoción de un desarrollo social, basado en la formación de ingenieros que egresen de programas educativos pertinentes y de calidad reconocida [2].

Para cumplir con el objetivo anteriormente mencionado, CACEI toma en cuenta los criterios y estándares internacionales aceptados por el Consejo para la Acreditación de la Educación Superior (COPAES) y Washington Accord, desarrollando un marco de referencia 2018 y un proceso metodológico, evaluados por dos organismos de acreditación de ingeniería reconocidos internacionalmente: el Accreditation Board for Engineering and Technology (ABET) de Estados Unidos y el Canadian Engineers Accreditation Board (CEAB) de Canadá.

La acreditación realizada por CACEI reconoce la calidad de los programas educativos considerando estándares definidos para un programa educativo de ingeniería de buena calidad que cuenten con al menos una generación de egresados y por egresar la segunda. La evaluación es a través de estándares y criterios de calidad establecidos y difundidos previamente por el organismo acreditador. El procedimiento incluye una autoevaluación del programa, así como una evaluación por un equipo de expertos externos o pares académicos. En todos los casos es una validación temporal, por tres o cinco años [2].

CACEI cuenta con siete atributos de egreso; sin embargo, en este trabajo solo se enfocará en el sexto, definido como: "Reconocer la necesidad permanente de conocimiento adicional y tener la habilidad para localizar, evaluar, integrar y aplicar este conocimiento adecuadamente" [2].

3 Proyecto Integrador

Un proyecto integrador es una estrategia didáctica que consiste en realizar un conjunto de actividades articuladas entre sí, con un inicio, un desarrollo y un final con el propósito de identificar, interpretar, argumentar y resolver un problema del contexto, y así contribuir a formar una o varias competencias del perfil de egreso, teniendo en cuenta el abordaje de un problema significativo del contexto disciplinar–investigativo, social, laboral–profesional, etc. [15].

4 Metodología

a. Desarrollo del Proyecto Integrador

El objetivo general de la carrera de Ingeniería en Sistemas Computacionales (ISC) del Tecnológico de Estudios Superiores de Chimalhuacán es: "Formar profesionistas líderes, analíticos, críticos y creativos, con visión estratégica y amplio sentido ético, capaces de diseñar, implementar y administrar infraestructura computacional para aportar soluciones innovadoras en beneficio de la sociedad, en un contexto global, multidisciplinario y sustentable" [16].

Para cumplir con la acreditación, los objetivos educacionales y atributos de egreso de la carrera de ISC se alinearon con los atributos de egreso de CACEI, en este caso el quinto

objetivo educacional y quinto atributo de egreso de la carrera de ISC son los que se relacionan con el sexto atributo de egreso de CACEI definidos a continuación.

El quinto objetivo educacional de la institución define: "Los egresados serán autónomos en su formación profesional con habilidades y actitudes que les permita actualizarse y adquirir conocimientos para entender y adaptarse a las nuevas tecnologías y entornos, buscando mantenerse vigentes a través del aprendizaje constante" [16].

El quinto atributo de egreso del programa educativo define: "Desarrolla habilidades y actitudes para su actualización permanente en su área profesional, que le permita adquirir conocimientos para adaptarse a las nuevas tecnologías y entornos, buscando mantenerse vigente a través del aprendizaje constante" [16].

CACEI cuenta con siete atributos de egreso; sin embargo, en este trabajo solo se enfocará en el sexto, definido como: "Reconocer la necesidad permanente de conocimiento adicional y tener la habilidad para localizar, evaluar, integrar y aplicar este conocimiento adecuadamente" [2].

Una vez alineados el quinto objetivo educacional y atributo de egreso con el sexto atributo de egreso de CACEI, se desarrolló un proyecto integrador, definido en [17] por CACEI como una estrategia curricular que relaciona las competencias profesionales de los planes de estudio del Tecnológico Nacional de México (TecNM), es una estrategia metodológica y evaluativa de investigación, direccionada al planteamiento y solución de problemas relacionados con la práctica profesional y calidad de vida; requiriendo de la articulación de asignaturas del nivel y disciplina o carrera.

El proyecto integrador se desarrolló en julio del 2019, se planteó como un plan de actualización profesional, agrupando las materias de los últimos tres semestres de la carrera, que corresponden a la especialidad y que requieren una actualización constante debido al desfase con el plan de la carrera ISIC-2010-224. Las áreas que se definieron para ser actualizadas son: Bases de datos, sistemas programables, redes (Fig. 1) y programación (Fig. 2).

Fig. 1. Certificados de ejemplo para el área de redes. URL de verificació https://capacitateparaelempleo.org/verifica/3qlf3aw9x/

https://capacitateparaelempleo.org/verifica/9zj9ovjmc/

https://capacitateparaelempleo.org/verifica/zkb50y52e/

Fig. 2. Certificados de ejemplo para el área de programación e Inteligencia Artificial

El documento del proyecto integrador tiene extensión de siete páginas, en las que se incluyen el atributo de egreso del programa educativo, el atributo de CACEI relacionado, la descripción del proyecto, la distribución de los cursos por semestre que han sido verificados por la academia, con sus respectivos enlaces a las plataformas formadoras y que el alumno deberá realizar, aspectos técnicos, la forma de evidenciar los entregables, los aspectos para una evaluación sobresaliente, el tiempo destinado para realizar el proyecto y realizar su revisión, mínimos requeridos para acreditar la evaluación, el formato de seguimiento de los cursos y la rúbrica de evaluación.

En la Figura 3 se muestra el formato elaborado por la academia de profesores de la carrera para llevar el seguimiento de los cuatro cursos validados que deberán realizar los alumnos, uno por semestre, desde séptimo a noveno, hasta terminar la carrera y cubrir todas las áreas de actualización. El alumno al egresar, habrá adquirido el hábito y conciencia de cumplir con una actualización constante.

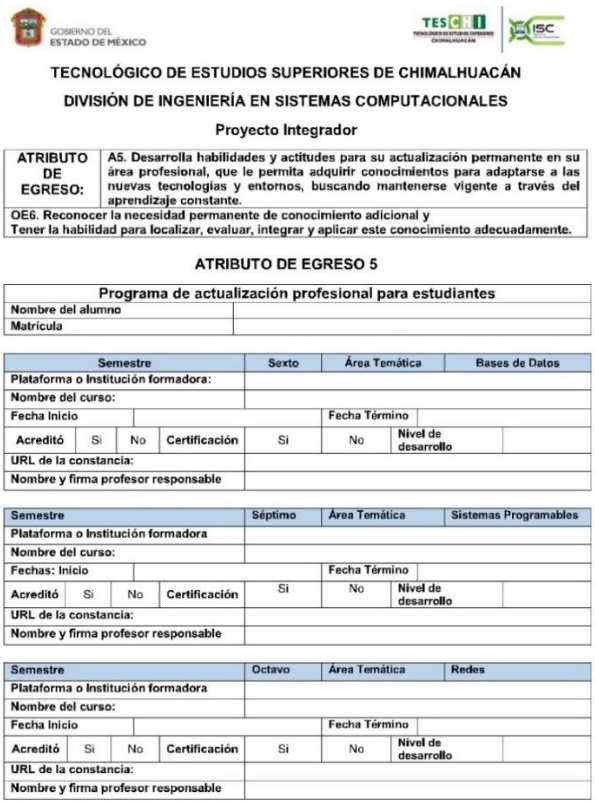
En el formato de seguimiento se solicitan datos del alumno y datos del curso como el nombre, plataforma o institución formadora, fecha de inicio y termino, URL pública de la constancia, nombre y firma del profesor que validó el curso. Los cursos de actualización definidos por la academia de la carrera, son de acceso gratuito y afines a las áreas de especialidad de los últimos tres semestres. Entre las plataformas que ofertan cursos gratuitos de actualización profesional, se encuentran: Capacítate para el empleo de la Fundación Carlos Slim, Google Activate, SOLO LEARN, entre otras.

En el caso del plan IAEV-2012-238 que también esta desactualizado, se desarrolló un plan de actualización similar, definiendo las siguientes áreas para actualizarse: Imágenes creadas por computadora (CGI); Videojuegos, RA y RV; Pre, pro y post producción y para el último semestre el área de Motion Graphics Avanzado.

b. Rúbrica de Evaluación

Una rúbrica o matriz de evaluación es una forma de registro que selecciona aquellos aspectos que se quieran evaluar ordenados por niveles de ejecución o calidad mediante descriptores precisos. Tiene el objetivo de apoyar a los evaluadores del CACEI en la revisión del documento de autoevaluación y las evidencias, así como la visita al Plan de Estudios, buscando contar con información que permitan que el proceso sea válido y confiable [18].

El siguiente paso es definir una rúbrica para evaluar el proyecto integrador. De acuerdo con CACEI en su Marco de Referencia 2018, la escala para evaluar los indicadores se basa en el nivel de cumplimiento, dividiéndose en: no se alcanza, se alcanza parcialmente, se alcanza y se supera [18].



GOBIERNO DEL ESTADO DE MÉXICO

TECNOLÓGICO DE ESTUDIOS SUPERIORES DE CHIMALHUACÁN

DIVISIÓN DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

Proyecto Integrador

ATRIBUTO DE EGRESO:	A5. Desarrolla habilidades y actitudes para su actualización permanente en su área profesional, que le permita adquirir conocimientos para adaptarse a las nuevas tecnologías y entornos, buscando mantenerse vigente a través del aprendizaje constante.
OE6. Reconocer la necesidad permanente de conocimiento adicional y Tener la habilidad para localizar, evaluar, integrar y aplicar este conocimiento adecuadamente.	

ATRIBUTO DE EGRESO 5

Programa de actualización profesional para estudiantes	
Nombre del alumno	
Matrícula	

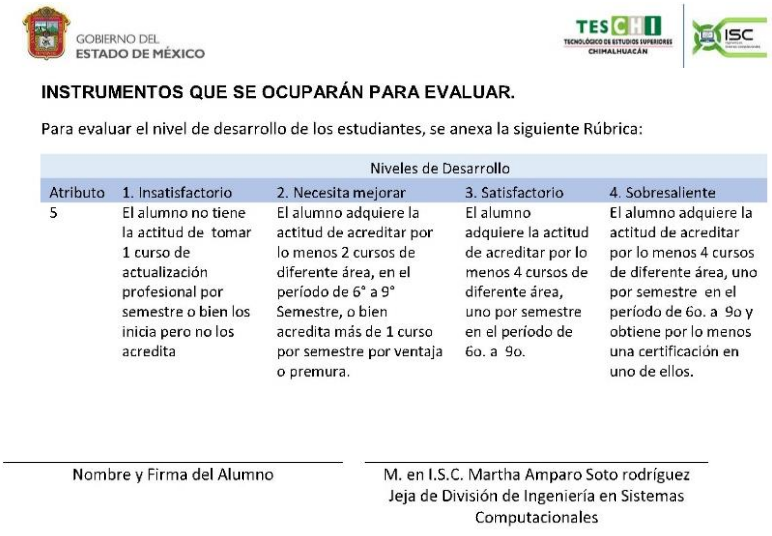
Semestre				Sexto	Área Temática		Bases de Datos
Plataforma o Institución formadora:							
Nombre del curso:							
Fecha Inicio					Fecha Término		
Acreditó	Si	No	Certificación	Si	No	Nivel de desarrollo	
URL de la constancia:							
Nombre y firma profesor responsable							

Semestre				Séptimo	Área Temática		Sistemas Programables
Plataforma o Institución formadora							
Nombre del curso:							
Fechas: Inicio					Fecha Término		
Acreditó	Si	No	Certificación	Si	No	Nivel de desarrollo	
URL de la constancia:							
Nombre y firma profesor responsable							

Semestre				Octavo	Área Temática		Redes
Plataforma o Institución formadora							
Nombre del curso:							
Fecha Inicio					Fecha Término		
Acreditó	Si	No	Certificación	Si	No	Nivel de desarrollo	
URL de la constancia:							
Nombre y firma profesor responsable							

Figura3. Formato para el seguimiento del plan de capacitación (página 4 del pı integrador)

En la Figura 4 se muestra la página 5 del proyecto integrador, donde se incluye la rúbrica de evaluación. El alumno alcanzará el nivel de sobresaliente solo cuando adquiera la actitud de acreditar por lo menos cuatro cursos de diferente área, realizando uno por semestre en el periodo de séptimo a noveno y obteniendo por lo menos una certificación en uno de ellos.



GOBIERNO DEL ESTADO DE MÉXICO

INSTRUMENTOS QUE SE OCUPARÁN PARA EVALUAR.

Para evaluar el nivel de desarrollo de los estudiantes, se anexa la siguiente Rúbrica:

	Niveles de Desarrollo			
Atributo	1. Insatisfactorio	2. Necesita mejorar	3. Satisfactorio	4. Sobresaliente
5	El alumno no tiene la actitud de tomar 1 curso de actualización profesional por semestre o bien los inicia pero no los acredita	El alumno adquiere la actitud de acreditar por lo menos 2 cursos de diferente área, en el período de 6° a 9° Semestre, o bien acredita más de 1 curso por semestre por ventaja o premura.	El alumno adquiere la actitud de acreditar por lo menos 4 cursos de diferente área, uno por semestre en el período de 6o. a 9o.	El alumno adquiere la actitud de acreditar por lo menos 4 cursos de diferente área, uno por semestre en el período de 6o. a 9o y obtiene por lo menos una certificación en uno de ellos.

Nombre y Firma del Alumno

M. en I.S.C. Martha Amparo Soto rodríguez
Jeja de División de Ingeniería en Sistemas Computacionales

Fig. 4. Formato para la rúbrica de evaluación (página 5 del proyecto integrador)

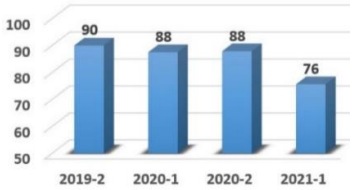
5 Resultados

El seguimiento del plan de actualización profesional y aplicación del proyecto integrador se realizó durante el periodo del 2019-2 al 2021-1. Durante los semestres de 7° hasta 9°, los alumnos seleccionaron un curso de actualización profesional validado por la academia de la carrera, de tal forma que al terminar la carrera hayan obtenido mínimo tres constancias de los cursos realizados, adquiriendo el hábito por mantenerse actualizado y obteniendo las competencias profesionales que podrán agregarse como evidencia en su currículum vitae (Fig. 1 y Fig. 2).

En la tabla 1 se pueden notar los grupos de la carrera de Ingeniería en Sistemas Computacionales, para la muestra se emplearon dos grupos. El contraste puede notarse en la Figura 5.a, donde el porcentaje de cumplimiento de las certificaciones se encuentra en el rango entre 88% a 90% para los primeros tres semestres de evaluación; sin embargo, el porcentaje bajó a 76% en el semestre 2021-1 debido a que después de un año de confinamiento por la pandemia, los alumnos han desertado o han reducido su tiempo de estudio al integrarse en algún empleo, reduciendo el número de alumnos a 41 por generación, de los cuales solo 31 cumplieron con una capacitación aprobatoria. Ver figura 5.b.

Tabla 1. Seguimiento de la actualización profesional para la carrera de ISC.

Año	Grupos	Total de alumnos	Certificados	Porcentaje
2019-2	8ISC21, 7ISC22	60	54	90%
2020-1	8ISC21, 8ISC22	72	63	88%
2020-2	9ISC21, 9ISC22	66	58	88%
2021-1	9ISC21, 8ISC22	41	31	76%



a) Porcentaje de alumnos certificados

b) Total de alumnos certificados

Fig. 5. Resultados del plan de actualización profesional para la carrera de ISC durante el periodo 2019-2 al 2021-1

La evaluación se realizó al inicio del año 2020, donde los primeros resultados del plan de actualización profesional evidenciaron el cumplimiento del sexto atributo de egreso de CACEI, contribuyendo para obtener la acreditación del programa de ISC (Figura 6) con vigencia del 7 de mayo del 2020 al 6 de mayo del 2025.

En el caso de la carrera de Animación Digital y Efectos Visuales, al inicio del 2020 cuando fue evaluado, se realizaron varias recomendaciones para poder obtener la

acreditación; sin embargo, el cierre de la institución por las medidas sanitarias frente a la pandemia, evitaron una segunda evaluación.


El Consejo de Acreditación de la Enseñanza de la Ingeniería, A.C.

Organismo reconocido por el Consejo para la Acreditación de la Educación Superior A.C.

ACREDITA

al programa de

Ingeniería en Sistemas Computacionales

del

Tecnológico de Estudios Superiores de Chimalhuacán

del día 07 de mayo del 2020 hasta el 06 de mayo del 2025, por cumplir los estándares internacionales de calidad educativa establecidos por el CACEI.

Ing. María Elena Barrera Bustillos
Directora General del CACEI, A.C.

http://www.cacei.org.mx/constancias/190-30-22A_1B263981E0.jpg

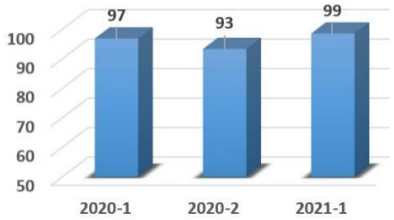
Fig. 6. Certificado de acreditación de la carrea de ISC

Tabla 2. Seguimiento de la actualización profesional para la carrera de ADyEV.

Año	Grupos	Total de alumnos	Certificados	Porcentaje
2020-1	8ADyEV22, 7ADyEV11	64	62	97%
2020-2	8ADyEV22, 7ADyEV11	75	70	93%
2021-1	8ADyEV22, 7ADyEV11	70	69	99%

El seguimiento del plan de actualización profesional para la carrera de IADyEV, puede notarse en la tabla 2. El contraste puede notarse en la Figura 7.a, donde el total de certificaciones realizadas se encuentra en el rango entre 62 y 70 por semestre. El porcentaje de cumplimiento de las certificaciones es de 96% en promedio para un total de 70 alumnos por semestre en promedio (ver la figura 7.b), en esta carrera el indicador se ha mantenido en niveles altos, sin embargo, la acreditación en el 2020 para esta carrera se vio interrumpida por la pandemia.

a) Total de alumnos certificados



b) Porcentaje de alumnos certificados

Fig. 7. Resultados del plan de actualización profesional para la carrera de ADyEV durante el periodo 2020-1 al 2021-1

6 Conclusión

En esta investigación se evidencio el seguimiento de un plan de actualización profesional enfocado en los alumnos, para cubrir el sexto atributo de egreso evaluado por CACEI.

Los resultados mostraron que del periodo 2019-2 al 2021-1 en promedio por semestre, se ha contado con 60 alumnos, de los cuales 85% se han actualizado constantemente durante cada semestre.

En la carrera de IADyEV, el plan de actualización profesional inició la aplicación hasta el semestre 2020-1, durante ese periodo se contó en promedio con 70 alumnos y un 96% de cumplimiento en la actualización constante por medio de certificaciones.

Finalmente, el sexto atributo de egreso de CACEI puede cumplirse mediante un proyecto integrador que se enfoque en un plan de actualización profesional como el que se ha planteado, fomentando el interés de los alumnos por estar actualizados en su área de especialidad hasta que se formen el hábito y la motivación por aprender. Se espera que el seguimiento de este plan pueda apoyar en culminar la acreditación de la carrera de IADyEV como sucedió con la carrera de ISC.

Un especial agradecimiento a los alumnos, profesores y jefes de división de las carreras de ISC e IADyEV del Tecnológico de Estudios Superiores de Chimalhuacán, cuyo trabajo permitió realizar el proceso de acreditación y darle seguimiento a esta investigación.

Referencias

1. Barrera, M. E. A., Jiménez, M.: Resumen estadístico de los programas de calidad de ingeniería evaluados por CACEI**.** Consejo de Acreditación de la Enseñanza de la Ingeniería, A. C. http://www.cacei.org.mx/nvpp/nvppdocs/resestpc.pdf (2021)
2. Barrera, M. E. A.: Marco de Referencia 2018 del CACEI en el Contexto Internacional (Ingenierías). Consejo de Acreditación de la Enseñanza de la Ingeniería, A.C. Versión 2 (2018).
3. Consejo de Acreditación de la Enseñanza de la Ingeniería, A. C.: Costo de Acreditación, http://www.cacei.org.mx/nvfs/nvfs02/nvfs0205.php (2021)
4. Consejo de Acreditación de la Enseñanza de la Ingeniería,A.C.:Catálogo de Programas Acreditados. http://www.cacei.org.mx/nvfs/nvfs04/nvfs0403.php (2021)
5. Marín, D. E.: La acreditación de carreras universitarias. Una tendencia actual en la formación de profesionales universitarios. Perfiles Educativos, (71) (1996)
6. Meza, A., Cárdenas, M. T. Y., Sánchez, M. L.: Organismos acreditadores internacionales, componentes innovadores en torno a los procesos de evaluación curricular. In: Debates en Evaluación y Currículum/Congreso Internacional de Educación 2017, 3(3). https://posgradoeducacionuatx.org/pdf2017/D048.pdf (Septiembre de 2017 a Agosto de 2018)
7. Boville, B., Argüello, N., Reyes, N. G.: La acreditación como proceso dinamizador hacia la calidad. Revista Electrónica "Actualidades Investigativas en Educación", 6(1) (2006)
8. Casas, E. V., Olivas, E.: El proceso de acreditación en programas de Educación Superior: un estudio de caso. Omnia, 17(2), pp. 53-70 (2011)
9. Villordo, J. A.: La Acreditación del programa de Ingeniería Química por el Consejo de Acreditación de la Enseñanza de la Ingeniería, A.C. Conciencia Tecnológica, (26) (2004)
10. Castillo, A., Izar, J. M., Hernández, V.: Modelo de autoevaluación para la acreditación del programa de Ingeniero Mecánico Administrador de la Facultad de Ingeniería de la Universidad Autónoma de San Luis Potosí. CPU-e, Revista de Investigación Educativa, (14), pp. 106-125 (2012)
11. Treviño, A., García, C., Martínez, A. R.: Estrategia para la Acreditación del programa Ingeniero en Mecatrónica, en la Universidad Autónoma de Nuevo León. Revista Cubana de Educación Superior, 38(1) (2019)
12. Soto-Leyva, Y., Bones-Martinez, R., Santos-Osorio, A.: La acreditación elemento clave en el fortalecimiento académico del Instituto Tecnológico Superior de Huauchinango (ITSH). In: Marroquín, A., Olivares, J., Cruz, L., Bautista, A. (Coord) Educación. Handbooks-©ECORFAN-México, Capítulo 3. (2020)
13. Barragan, J. N., Contreras, B. P.: La acreditación educativa en México: orígenes, evolución y contribución a la mejora de la educación. Aproximación conceptual. Daena: International Journal of Good Conscience. 15(1), pp. 142-158 (2020)
14. Rubio Oca, J.: La evaluación y acreditación de la educación superior en México: un largo camino aún por recorrer. Reencuentro, 50, pp. 35-44 (2007)
15. López Rodríguez, N. M.: El proyecto Integrador: Estrategia didáctica para la formación de competencias desde la perspectiva del enfoque socioformativo. Gafra Editores, México. (2012)

16. Tecnológico de estudios superiores de Chimalhuacán: Ing. en Sistemas Computacionales,https://www.teschi.edu.mx/alumnos/oferta_educativa/sistemas/ (2021)
17. Tecnológico Nacional de México: Proyectos integradores para la formación y desarrollo de competencias profesionales del tecnológico nacional de México. Segunda edición (2014). http://www.itsjuanrodriguezclara.edu.mx/Document/Transparencia/875/FRACCION_I/Proyectos_Integradores_2da_edicion.pdf
18. Consejo de Acreditación de la Enseñanza de la Ingeniería, A. C.: Rúbrica para evaluadores. Versión 6. http://cacei.org.mx/docs/rubrica_ing_2018.pdf (2018)

Creación de un Ecosistema Tecnológico con Inteligencia Artificial y Robótica para Garantizar Zonas Libres de COVID-19 en CU-BUAP

Ma. del Carmen Santiago, Ana C. Zenteno, Judith Pérez, Yeiny Romero, Gustavo T. Rubín, Antonio E. Álvarez
Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación, Ciudad Universitaria, 14 sur y Avenida San Claudio, Fraccionamiento Jardines de San Manuel, CP. 72570, Puebla, Pue; México.
{marycarmen.santiago, ana.zenteno, judith.perez, yeiny.romero, gustavo.rubin}@correo.buap.mx, antonio.alvarez@alumno.buap.mx

Resumen. A dos años del inicio de la pandemia de COVID-19 la sociedad conoce las indicaciones básicas para disminuir los contagios, sin embargo, hay bastante rebeldía o relajación en el cumplimiento de éstas, por ello en este trabajo se describe una propuesta tecnológica que vigile el cumplimiento de los protocolos de seguridad en esta contingencia sanitaria. Éste entorno está constituido por robots autónomos, aplicaciones móviles y un sistema de procesamiento de información basado en Inteligencia Artificial y Ciencia de Datos, mediante los cuales se realiza el análisis de los elementos anteriores y se generan los entornos que garanticen la seguridad e higiene, además de generar un seguimiento de una posible cadena de contagios y el análisis de los diferentes entornos con la suficiente antelación para tomar medidas que mitiguen la pandemia en la zona que se quiere asegurar.

Palabras Clave: Covid19, Ecosistema Tecnológico, Robótica, Ciberseguridad.

1 Introducción

La aparición en China de una nueva enfermedad a finales de 2019 se veía como una enfermedad que remotamente llegaría a América y sobre todo a México. Pero a principios del 2020 el coronavirus (SARS-CoV-2 o COVID-19) impuso en cuestión de semanas una realidad dura jamás imaginada. Los gobiernos decretaron el estado de alarma y el confinamiento domiciliario por la expansión de la enfermedad convirtiéndose ésta en una Pandemia [1].

Siendo más precisos los coronavirus son una familia de virus que causan enfermedades como resfriado común e incluso enfermedades respiratorias más graves dicho virus puede transmitirse entre humanos y también entre animales. Pero, la aparición del COVID-19 fue declarada grave dado que es transmitida rápidamente llegando a ser mortal si o se tata a tiempo, la Organización Mundial de la Salud el 30 de enero de 2020 la declara Pandemia, lo que significa que la epidemia se ha extendido por varios países, continentes o todo el mundo afectando a millones de personas. [2]

La COVID 19 es una enfermedad infecciosa grave con 2,660,507 casos estimados de contagios, de los cuales se tienen 1969,137 personas recuperadas y 231,151 defunciones confirmadas [10], algunas personas infectadas por el virus pueden tener cuadros respiratorios de leves a moderados y se recuperan sin tratamiento, en cambio personas

mayores al igual que personas con enfermedades cardiovasculares, diabetes, enfermedades respiratorias crónicas o cáncer, tienen una alta probabilidad de presentar un cuadro grave y morir.[3]

Una persona que tiene COVID-19 presenta síntomas entre los días 2 a 14 después de la exposición al virus. como:[2], [5]

- Fiebre o Escalofríos
- Dolor de cabeza
- Tos
- Estornudos
- Dolor de garganta
- Escurrimiento nasal
- Ojos rojos
- Dolor de articulaciones o músculos
- En casos más graves, dificultad para respirar
- Fatiga
- Pérdida reciente del olfato o el gusto
- Náuseas o vómitos
- Diarrea

A diferencia de la influenza que es una enfermedad con tratamiento y existen vacuna, esta enfermedad no tiene una inmunidad previa conocida, no hay tratamiento específico, todas las personas son susceptibles al virus. Este virus daña las células de los pulmones y las del sistema respiratorio. Es transmitida a otras personas vía ojos, nariz y boca [2], [4]. Una persona infectada transmite el virus fácilmente a través de:

- Expulsión de gotas de saliva al estornudar
- Toser sin cubrirse la boca y la nariz
- Al hablar o cantar
- Al saludar de mano o de beso a una persona
- Al tener contacto con una superficie contaminada
- Llevarse las manos sucias a la boca, nariz u ojos

Hasta este momento no existe un medicamento que cure la enfermedad, ya se cuenta con vacunas para prevenir la infección por el virus SARS-CoV2 (Pfizer BioNTech efectividad del 95%, Astra Zeneca efectividad del 82.4%, SinoVac efectividad del 52%, Spunik V efectividad del 91.6%, Cansino efectividad del 65.7%, Jansen efectividad del 66%), pero, es importante seguir las normas de sanidad para evitar contagios las cuales son [5][6][11][12][13]:

- Lavarse las manos
- Usar un gel hidroalcohólico con frecuencia
- No tocándose la cara
- Taparse la boca y nariz con el codo flexionado al estornudar
- Distanciamiento físico con cualquier persona que presente síntomas
- Consejos de limpieza y desinfección para el aula en caso de escuelas

Ante la propagación a nivel mundial del COVID-19, las Secretarías de Salud y de Educación Pública del Gobierno de México presentaron, las medidas de prevención y atención prioritarias donde destacan que dadas las circunstancias sanitarias que se vienen a nivel mundial se realiza el confinamiento de los estudiantes en todas sus etapas para salvaguardar su integridad. Indicando que se reanudarán las labores siempre y cuando, se cuente con todas las condiciones determinadas por la autoridad sanitaria federal en cada plantel escolar. [7] Sin embargo. una de las principales lecciones que ha dejado la pandemia es la importancia de los maestros para que los niños continúen su proceso de

aprendizaje siendo el aprendizaje a distancia la mejor opción de continuar con los programas educativos.

A más de un año del confinamiento escolar para reabrirse las escuelas, es importante tomar medidas de precaución tanto dentro de las aulas como fuera de ellas para evitar la propagación de la COVID-19. Se busca que los alumnos regresen en un entorno seguro y saludable para continuar su formación académica. [8]

Para prevenir la propagación del COVID-19, hay varias medidas de seguridad que las escuelas pueden tomar y que pueden ayudar a reducir el contagio. A continuación, algunas de estas medidas:

- Limpiar y desinfectar superficies de uso diario
- Lavarse las manos
- Distanciamiento entre mesas o pupitres
- Reducir el número de alumnos por salón de clases
- Cambios de salón de docentes para evitar movilidad en estudiantes
- Uso de espacios al aire libre
- Uso obligatorio de mascarillas en todo el personal escolar
- Comidas o almuerzos en pupitres para evitar movilidad.

1.1 La propuesta tecnológica

En este trabajo se presenta la propuesta para la creación de un ecosistema tecnológico con inteligencia artificial y robótica para garantizar zonas libres de covid-19 en CU-BUAP. La gran ventaja del proyecto radica en el aprovechamiento de diversos módulos, a saber: robótica integrada por un robot sanitizador, aplicación móvil en plataformas IOS y/o android, integración de métodos de seguridad para las sesiones en la aplicación y el manejo de los datos; tanto en hardware como en software lo que ofrecería grandes beneficios en la salud de la comunidad universitaria y un monitoreo completo por áreas de la institución; contrario a lo que sucede actualmente en donde existen empresas que brindan el servicio a un alto costo. Permitiendo un diseño a modo; que cubra las necesidades particulares y específicas de la infraestructura actual y coadyuvando al desarrollo de nueva tecnología y de recursos humanos del área de las ciencias computacionales.

Además, este proyecto se diseña para expandirse a sistemas más complejos y de mayores dimensiones que puede replicarse a nivel nacional e internacional con una infraestructura relativamente de bajo costo.

2 Metodología

Asegurar una zona libre de COVID-19 se relaciona directamente con dos aspectos:

a) Descontaminar o sanitizar la zona
b) Evitar que se contamine.

Para este fin se debe implementar una metodología de seguridad que brinde un rango de certeza confiable en los dos aspectos, para lo cual se proponen las siguientes etapas y elementos de infraestructura física y de software para generar el sistema de vigilancia sanitizante y de implementación de los protocolos de seguridad preventiva y descontaminante.

Es importante considerar que el sistema propuesto es 100% autónomo e inteligente, capaz de tomar las decisiones que han sido programadas para los diferentes escenarios, con un bajo índice de fallos y donde se evite el contacto de personas a fin de contagiarse.

2.1 Etapa de desinfección o descontaminación robótica

Para evitar contagios en esta etapa se requieren robots que lleven a cabo esta tarea. Los robots serán de dos tipos fijos y móviles, todos con sistemas de expulsión de rocío sanitizante y lámparas UV (ultravioleta), además de sensores para detección de obstáculos a fin de realizar la descontaminación cuando no se encuentren personas cerca de los robots.
Los robots tienen las siguientes características:

- Sensores para evasión de obstáculos
- Sensores de temperatura
- Sistema motriz giratorio en su eje y de desplazamiento (para robots móviles)
- Sistema de visión IR para detección de usuarios con fiebre o temperatura alta.
- Sistema de visión para reconocimiento de imágenes e identificación del uso incorrecto del cubrebocas.
- Sistema de rocío de solución sanitizante
- Sistema de emisión de radiación UV
- Sistema de comunicación bluetooth para identificación de los usuarios
- Lector de tarjetas para identificar usuarios.
- Sistema de comunicación Wi-Fi y RF, con otros robots y con el centro de operaciones.
- Sistema GPS
- Sistema de control y planeación de recorridos (para robots móviles)

Adicional a estos robots se requiere utilizar cámaras fijas para videovigilancia y ocasionalmente drones para cubrir de forma más eficiente las diferentes zonas cuando las cámaras no brindan la información necesaria.

2.2 Procesamiento de información y aplicación móvil

Una vez que se tienen los diversos elementos con la conectividad requerida se necesita desarrollar una plataforma para recopilar toda la información de los diferentes elementos y las variables que se procesarán con algoritmos de Ciencia de Datos e Inteligencia Artificial, sin olvidar que se requiere una aplicación móvil donde todos los usuarios que ingresan a esta "zona segura" deben agendar sus actividades previamente (aunque se pueden modificar durante la evolución de dichas actividades) a fin de que la información de todos los usuarios la conozca el sistema central de procesamiento y pueda generar los protocolos de seguridad para brindar bajas concentraciones de personas.

La aplicación móvil que se requiere debe contener las siguientes características:

- Identificación del usuario
- Identificación de su entorno social cercano (familiares y amigos) que han tenido COVID-19
- Mecanismo de contagio de su entorno social
- Actividades que realizara en su estancia en CU con hora y localización aproximada.
- Datos que genera el robot en CU al recibirlo (temperatura, uso de cubrebocas, etc.)

3 Resultados

En este momento no se cuenta con resultados obtenidos a nivel experimental dado que esta propuesta está en desarrollo en todas sus etapas, pero se espera obtener los siguientes resultados conforme se concluyan las diferentes etapas:

A corto plazo: Se contará con un primer prototipo de robot sanitizante que cuente con los elementos necesarios para su Navegación en entornos abiertos como sensores de proximidad y ubicación GPS, y además contar con los elementos de sanitización como sistema de rocío y lámparas UV.

También se tendrá el diseño de una y una primera versión de la aplicación móvil que se utilizará para que los usuarios realicen la carga de información a la central de procesamiento.

A mediano plazo: Se generarán los algoritmos de planeación de trayectorias y navegación del robot y se iniciará el procesamiento de información que este genere, así como también los algoritmos que permitan una sanitización con el rocío y las lámparas UV de forma automática y segura para los usuarios.

Se analizarán las variables que generan las cámaras termográficas en cuanto al uso adecuado del cubrebocas, medición de temperatura y/o identificación de zonas con exceso de personas sin conservar la sana distancia.

Se implementarán y procesaran los sensores de temperatura en el robot y de comunicación con los usuarios vía bluetooth y esta información se empezará a procesar para incrementar la base de datos que se analizara.

A largo plazo: Diseño e implementación de la red virtual para llevar a cabo el procesamiento tanto de los usuarios de la aplicación móvil, los robots y las cámaras fijas y ubicadas en los robots, permitiendo aplicar las técnicas de Ciencia de datos e Inteligencia Artificial para procesar la información y generar los escenarios preventivos que permitan asegurar una zona libre de COVID-19

Cada uno de los elementos principales descritos anteriormente debe cumplir los siguientes requerimientos:

Robot móvil sanitizante:

- Identificación y técnicas de medición de las variables del entorno involucradas (temperatura, humedad, proximidad y cámaras de reconocimiento)
- Diseño y construcción de la estructura del robot sanitizante.
- Desarrollo del módulo de comunicaciones
- Diseño del sistema de potencia.
- Diseño del módulo de control (planeación de movimientos, seguridad, etc)

Aplicación MOVIL_COVID BUAP

- Delimitación de requerimientos. En esta etapa se identifican los requerimientos de la aplicación móvil COVID de la población universitaria de la BUAP para estudiantes, docentes, administrativos y personas con acceso a las instalaciones, la cual permite prevenir los contagios por SARS-CoV-2 entre la comunidad.
- Modelado de la App. Utilizando el Lenguaje de Modelado Unificado se podrá crear los diagramas siguientes:
- Diagrama de Clases
- Diagrama de casos de uso
- Diagrama de actividades
- Diagramas de iteración

- Diagrama de componentes.
- Se utilizará la herramienta Lucidchart para UML y tiene un costo estimado $1680 anual.
- Análisis de tecnologías de desarrollo de la App. Se llevará acabo un análisis de requerimientos de desarrollo de aplicaciones móviles Se adquirirá Visual Studio Profesional para Desarrollo de aplicaciones móviles Android con Visual C++ .
- Implementación de la App. El desarrollo de la App involucra la creación de los Mockups en Adobe. La implementación misma de la App consiste en generar un prototipo interactivo para modelar la solución de la aplicación móvil y se desarrollará, de acuerdo con el punto 3, en Visual Studio para desarrollo de móviles IOS/Android. Equipo de cómputo para desarrollo con 16GB de RAM, 1TB SSD, Core i7, 1 Becario -
- Pruebas. Las pruebas que aplicaremos a la App desarrollada consisten en generar casos de prueba dentro de ciudad universitaria. Equipos móviles IOS/Android
- Liberación de la App. La distribución de la aplicación móvil será a través de Google Play y Apple Store, y se pondrá a disposición de los estudiantes, profesores, administrativos y personas que tienen acceso.

Desarrollar una Metodología de Análisis de Información mediante IA y Ciencias de Datos, así como su implementación.

- Implementación de los algoritmos de ciencia de datos e IA
- Implementación de Algoritmos de reconocimiento de imágenes y patrones
- Diseño y normalización de la base de datos e implementación de algoritmos de tratamiento de información

Análisis e implementación de algoritmos de IA y aprendizaje automático

Implementación de una red virtual para diseñar estrategias de ciberseguridad con protocolos de comunicación seguros entre la aplicación móvil, el robot y los usuarios.

- Diseño e implementación de red para el robot, aplicación en dispositivo móvil y servidor.
- Análisis de protocolos de comunicación seguros para redes inalámbricasEl Análisis de Algoritmos de cifrado/encriptación y seguridad - SHA, uso de
- certificados e intercambio de llaves públicas y/o privadas permitirá
- Diseño e implementación de red virtual de integración de robot, servidor y dispositivos móviles.

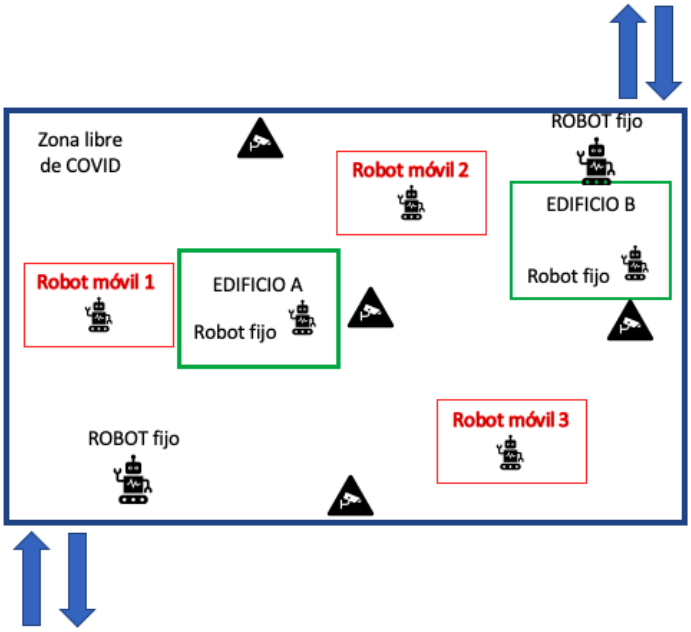


Fig. 1. Diagrama esquemático que describe el escenario prototipo de implementación del sistema para garantizar una zona libre de Covid y sus elementos constitutivos.

4 Conclusiones

Garantizar una zona libre de Covid podría parecer una tarea sencilla, pero no es así, está comprobado después de más de 2 años de pandemia que la sociedad no es capaz de seguir protocolos de forma ordenada y estricta, por lo tanto se requieren medidas en las cuales el criterio de los usuarios no esté en juego, es decir que sea un sistema automático el que verifique que se cumplan estos protocolos de forma estricta, esto no significa darle el control a la tecnología, sino utilizarla para que sea más eficiente el cumplimiento de un protocolo de seguridad.

Al concluir la redacción de este trabajo ya se cuenta con un avance del 70% en todos los módulos para poder implementarse y generar los resultados esperados a corto plazo.

Referencias

1. Tu, Z., Zheng, S. & Yuille, A. (2008). Shape matching and registration by data-driven EM. Computer Vision and Image Understanding, 109(3), 290–304.
2. Yin, P.Y. (2000). A tabu search approach to polygonal approximation of digital curves. International Journal of Pattern Recognition and Artificial Intelligence, 14(2), 243–255.
3. OMS: Neumonía de Causa desconocida- China [Online]. Available: https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/es/, Accessed on: May. 31, 2020
4. Novel Coronavirus(2019-nCoV) Situation Report-10 [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200130-sitrep-10-ncov.pdf?sfvrsn=d0b2e480_2, Accessed on: May.31, 2020

5. CIDE y Stanford desarrollan modelo matemático de proyecciones sobre COVID-19.[Online] Available:https://www.cide.edu/saladeprensa/cide-y-stanford-desarrollan-modelo-matematico-de-proyecciones-sobre-covid-19/, Accessed on: May.31, 2020
6. Así evoluciona la curva del coronavirus en México, Colombia, Chile, Argentina y el resto de Latinoamérica [Online]. Available: https://elpais.com/sociedad/2020/04/07/actualidad/1586251212_090043.html, Accessed on: May.31, 2020
7. Burgos, Pablo. (2020). COVID-19 Modelo numérico de casos de infección y estimaciones epidémicas Modelo Asimétrico -GOMPERTZ. 10.13140/RG.2.2.19440.40969.
8. Jiang S, Zhou Q, Zhan X, Li Q. BayesSMILES: Bayesian Segmentation Modeling for Longitudinal Epidemiological Studies. Preprint. medRxiv. 2021;2020.10.06.20208132. Published 2021 Jan 18. doi:10.1101/2020.10.06.20208132
9. Prediction of COVID-19 transmission dynamics using a mathematical model considering behavior changes in Korea. [Online]. Available: https://pesquisa.bvsalud.org/portal/resource/es/mdl-32375455 Accessed on: May.31, 2020
10. Chaos theory applied to the outbreak of COVID-19: an ancillary approach to decision making in pandemic context. [Online]. Available:https://pesquisa.bvsalud.org/portal/resource/es/mdl-32381148 Accessed on: May.31, 2020
11. Propagation analysis and prediction of the COVID-19 [Online]. Available:https://www.sciencedirect.com/science/article/pii/S2468042720300087 Accessed on: May.31, 2020
12. Mathematical Modeling of Epidemic Diseases; ACase Study of the COVID-19 Coronavirus [Online]. Available: https://arxiv.org/pdf/2003.11371.pdf Accessed on: May.31, 2020
13. Model the transmission dynamics of COVID-19 propagation with public health intervention [Online]. Available:https://www.medrxiv.org/content/10.1101/2020.04.22.20075184v1

Aplicaciones de las Ciencias Computacionales Durante la Pandemia COVID19
se terminó de editar en Diciembre de 2021 en la
Facultad de Ciencias de la Computación
Av. San Claudio y 14 Sur Jardines de San Manuel
Ciudad Universitaria
C.P. 72570

Aplicaciones de las Ciencias Computacionales Durante la Pandemia COVID19
Coordinado por María del Carmen Santiago Díaz