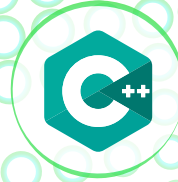


REVECOM

REVISTA VENEZOLANA DE COMPUTACIÓN

ML • Java • Python • C# • C++ • Scala
Perl • Haskell • Lisp • JavaScript



Sociedad Venezolana de Computación

Vol. 6, No. 1
Junio 2019



ISSN: 2244-7040



REVECOM

Revista Venezolana de Computación

**Sociedad Venezolana
de Computación**

**Editores:
Eric Gamess, Wilmer Pereira, Yudith Cardinale**

ISSN: 2244-7040

Vol. 6, No. 1
Junio 2019

Editorial

La Gestión de Proyectos de Software como Factor de Éxito

En la actualidad, se observa una competencia intensa entre las distintas organizaciones, en un escenario más interconectado e interdependiente, acompañado de una renovación acelerada de aspectos informáticos, normalmente formalizados en proyectos de software, que son utilizados en la promoción de los cambios más radicales en la gestión pública y privada. También se ve el aumento en la difusión y acceso a la información en todas las organizaciones, con una gran cantidad de clientes y usuarios quienes exigen y presionan para obtener oportunamente un producto con características de calidad, o recibir un servicio de calidad, conjuntamente con las respectivas mejoras en sus procesos.

Sin embargo, aunque ya hemos recorrido alrededor de 50 años desde la llamada “crisis del software”, los grandes proyectos tecnológicos enfrentan en todo el mundo enormes dificultades, altas tasas de fracaso parcial o total, resistencias al cambio y carencia de dirección estratégica por parte de las autoridades superiores que suelen creer que esta situación es sólo responsabilidad del jefe del departamento de informática, la carencia o indebida utilización de metodologías de trabajo, entre otros. En algunos casos, aunque se han institucionalizado diversas metodologías para gestionar los proyectos, se utilizan generalmente más por formalismo o protocolo y no como lo que verdaderamente constituyen: el eje o piedra angular del proyecto. En cuanto a los criterios de éxito para la gestión del proyecto, varían de un proyecto a otro, por ejemplo, entregar el software al cliente en el tiempo acordado; mantener costos dentro del presupuesto general, cubrir las expectativas del cliente al momento de la entrega del software y mantener un equipo de desarrollo cohesionado y en buen funcionamiento.

Vale la pena señalar que los proyectos han formado parte de la historia de la humanidad desde sus inicios. Sin embargo, fue hasta principios del siglo XX cuando se empezó a hablar de éstos como la base para una ciencia administrativa: la gerencia de proyectos. Uno de los precursores de esta disciplina fue Henry Gantt, autor de la gráfica de calendarización que lleva su nombre, y que es empleada en la actualidad, como una herramienta de planificación y control de proyectos. Otro de los precursores fue Henri Fayol, quien declaró que las empresas están caracterizadas por seis tipos de áreas: producción, comercio, finanzas, seguridad, contabilidad y gestión.

En este sentido, aunque en la Ingeniería de Software los aspectos gerenciales se observaron por primera vez desde los años 80, en la especificación de Modelos de Desarrollo de software como el Espiral de Barry Bohem, y en los años 90 en la definición de un marco metodológico como RUP (*Rational Unified Process*) donde la gestión de proyectos se considera una disciplina, es en los últimos años cuando se ha visto más interés en esta actividad. Esto debido a que se ha determinado que es un factor crítico en el éxito y fracaso de los mismos, porque aunque una buena gestión no puede garantizar el éxito; la mala gestión, por lo general, da como resultado un proyecto no exitoso.

Adicionalmente, la gestión de proyectos es un área de conocimiento del Cuerpo de Conocimiento de la Ingeniería del Software, la cual cada vez más recibe mayor atención dada la complejidad actual de los proyectos de software, lo que hace necesario la planificación, calendarización, gestión de riesgos y estimación de costos, entre otros. De hecho, para maximizar la eficiencia operativa y aumentar las probabilidades del éxito de los proyectos, se ha consolidado un conjunto de estándares internacionales para orientar el área de la Gestión de Proyectos, tales como PMI (Project Management Institute,

Editorial

PMBOK), IPMA (International Project Management Association), PRINCE2 (Projects in Controlled Environments 2), y P2M (Project and Program Management for Enterprise Innovation). Éstos proponen una serie de guías, protocolos, fundamentos, teorías, procesos, herramientas y características, como un referente de buenas prácticas para el ciclo de vida de los proyectos.

Otro aspecto que vale la pena destacar es la existencia de un rol fundamental dentro de un proyecto de software, el gerente del proyecto, quien tiene una tarea particularmente desafiante debido a que el software es un producto, considerado como intangible. Éste se encarga de la planificación, ejecución y cierre de proyectos, incluyendo su definición, y el establecimiento del plan de trabajo, duración y presupuesto mediante análisis de riesgos y análisis de costo beneficio, usando las metodologías y herramientas adecuadas que permitan alcanzar una meta establecida de forma exitosa, con los recursos necesarios y de manera puntual y creativa. También, debe encargarse de la administración de los recursos técnicos y recursos humanos para dichos proyectos, eliminando los obstáculos que se puedan presentar, motivando a los miembros de los equipos y haciéndolos más productivos. Por último, es responsable del manejo diplomático de las expectativas de los clientes mediante la comunicación efectiva con todas las partes involucradas en el proyecto.

Asimismo, entre las habilidades blandas o *soft skills* más interesantes que un Gerente de Proyecto están: liderazgo, comunicación, capacidad de análisis, toma de decisiones, trabajo en equipo, confianza, planificación de proyectos, motivación, influencia, conocimientos políticos y culturales, y negociación. Estas habilidades también pueden ser un factor de ÉXITO de un proyecto de software.

Finalmente, en el contexto internacional, se observan los cambios empresariales que buscan implementar estándares más exigentes de competitividad, calidad, agilidad de gestión y rigor. Están creando una tendencia hacia la gestión holística o integral en la gestión de proyectos, la cual incluye áreas como la innovación y el direccionamiento estratégico, entre otros factores. Es decir, el desafío es abordar los proyectos de software en forma integral considerando la interacción de las personas, procesos, y tecnologías de crecimiento acelerado, el contexto normativo y estratégico, la implantación, la cultura organizacional y el liderazgo.

Dra. Dinarle Ortega
Profesora del Departamento de Computación-FaCyT
Universidad de Carabobo

Revista Venezolana de Computación

ReVeCom (Revista Venezolana de Computación) es la primera revista venezolana arbitrada, periódica, digital, orienta a la publicación de resultados de investigación en el campo de la computación. ReVeCom fue creada por la SVC (Sociedad Venezolana de Computación) y tiene entre sus objetivos hacer conocer los trabajos de alta calidad investigativa que se realizan a nivel nacional, latinoamericano e internacional. La revista permite la divulgación de artículos con aporte original en castellano o inglés.

ReVeCom es una revista abierta para una mayor difusión de los resultados de investigación. Cuenta con una página web (<http://www.svc.net.ve/revecom>), donde se encuentran los trabajos publicados e información sobre la revista. La revista promueve la pluralidad de intereses, dando cabida a la divulgación de trabajos de todos los campos del conocimiento inherentes a la computación.

Este volumen de ReVeCom (Vol. 6, No. 1) corresponde a artículos de investigación en el campo de la computación que fueron seleccionados a través de un riguroso arbitraje por expertos del área. En esta oportunidad, el comité evaluador fue compuesto por:

Nombre	Afiliación
Ana Aguilera	Universidad de Carabobo – Venezuela
Andrés Arcia-Moret	Xilinx Cambridge – Reino Unido
Yudith Cardinale	Universidad Simón Bolívar – Venezuela
Ernesto Coto	University of Oxford – Reino Unido
Ruby De Valencia	Archivo Nacional de Arte Rupestre – Venezuela
Irvin Dongo	Université de Bordeaux – Francia
Eric Gamess	Universidad Central de Venezuela – Venezuela
Francisca Losavio	Universidad Central de Venezuela – Venezuela
Jonas Montilva	Universidad de Los Andes – Venezuela
Masun Nabhan	Universidad Simón Bolívar – Venezuela
David Padua	University of Illinois – USA
Wilmer Pereira	Instituto Tecnológico Autónomo de México – México
Eugenio Scalise	Universidad Central de Venezuela – Venezuela

Directorio de la Sociedad Venezolana de Computación

Presidente:

Dr. Leonid Tineo (Universidad Simón Bolívar – Venezuela)

Vicepresidente:

Dr. Eric Gamess (Universidad Central de Venezuela – Venezuela)

Secretario:

Dr. Wilmer Pereira (Instituto Tecnológico Autónomo de México – México)

Tesorero:

Dr. David Coronado (Universidad Simón Bolívar – Venezuela)

Coordinadora de Educación e Investigación:

MSc. Mildred Luces (Universidad Nacional Experimental de la Gran Caracas – Venezuela)

Coordinadora de Publicaciones:

Dra. Yudith Cardinale (Universidad Simón Bolívar – Venezuela)

Coordinadora de Eventos:

MSc. Soraya Carrasquel (Universidad Simón Bolívar – Venezuela)

Edición

Comité Editorial

Director:

Dr. Eric Gamess - Universidad Central de Venezuela, Venezuela
Redes de computadores, computación de alto desempeño, simulación.

Coordinador del Comité Editorial:

Dr. Wilmer Pereira - Instituto Tecnológico Autónomo de México, México
Inteligencia artificial, robótica autónoma, aprendizaje automatizado.

Jefe de Redacción:

Dra. Yudith Cardinale - Universidad Simón Bolívar, Venezuela
Computación paralela, computación de alto desempeño, sistemas distribuidos, computación en la nube, arquitecturas paralelas, servicios web, web semántica.

Miembros del Comité Editorial

Dr. Carlos Acosta - Universidad Central de Venezuela, Venezuela
Computación paralela, computación de alto desempeño, computación reconfigurable y FPGAs, simulación paralela y distribuida, BigData.

Dr. Andrés Arcia-Moret - Xilinx Cambridge, Reino Unido
Simulación de redes, protocolos de transporte, redes inalámbricas.

Dr. Ernesto Coto - University of Oxford, Reino Unido
Computación gráfica, visualización científica, procesamiento digital de imágenes.

Dra. Francisca Losavio - Universidad Central de Venezuela, Venezuela
Ingeniería del software, arquitecturas y calidad del software, producción industrial de software.

Dr. Francisco Luengo - Universidad del Zulia, Venezuela
Computación social, minería de texto.

Dr. Jonas Montilva - Universidad de Los Andes, Venezuela
Ingeniería del software, sistemas de información.

Dra. Masun Nabhan - Universidad Simón Bolívar, Venezuela
Inteligencia artificial, minería en datos, aplicaciones de inteligencia artificial para educación y discapacitados.

Dra. Dinarle Ortega - Universidad de Carabobo, Venezuela
Ingeniería del software, arquitectura del software, arquitecturas empresariales, modelado de procesos de negocio.

Dr. David Padua - University of Illinois, USA
Compiladores, computación de alto desempeño.

Dr. Leonid Tineo - Universidad Simón Bolívar, Venezuela
Bases de datos, lógica difusa, lenguajes artificiales, minería de datos.

Tabla de Contenido

Editorial	ii
Revista Venezolana de Computación	iv
Directorio de la Sociedad Venezolana de Computación	v
Comité Editorial	vi
Tabla de Contenido	vii
1. Predicción del Sentimiento de Mercado para los Crudos Pivotes Venezolanos con base en el Análisis de Noticias	1-9
Arlán Briceño, Aneriz Rodríguez, Haydemar Núñez	
2. Einalu'u: Videojuego de Simulación de Construcción y Gestión de Culturas Indígenas Prehispánicas	10-18
Adriana Urdaneta, Esmitt Ramírez	
3. Manual-design of Blocks: Una Herramienta para Gestionar Segmentaciones Manuales de Páginas Web	19-27
Jean Garcia, Andres Sanoja	
4. Sistema de Apoyo a la Terapia Física basado en el Internet de las Cosas	28-38
Francisco Moreno, Walter Hernández	
Índice de Autores	39

Predicción del Sentimiento de Mercado para los Crudos Pivotes Venezolanos con base en el Análisis de Noticias

Arlán Briceño¹, Aneriz Rodríguez¹, Haydemar Núñez²

arlan.briceno@gmail.com, aneriz.rodriguez@gmail.com, haydemar.nunez@ciens.ucv.ve

¹ Escuela de Computación, Universidad Central de Venezuela, Caracas, Venezuela

² Centro de Ingeniería de Software y Sistemas, Universidad Central de Venezuela, Caracas, Venezuela

Resumen: En este estudio se utilizan diversas técnicas de la minería de datos para emitir una recomendación basada en la clasificación del sentimiento de un grupo de noticias. La idea de esta recomendación es apoyar la toma de decisión sobre la dirección del precio futuro del petróleo crudo venezolano de exportación. Específicamente, se trabajó con minería de texto para clasificar noticias recientes del mercado petrolero a partir del aprendizaje de noticias previamente clasificadas. En este artículo se plantea y resuelve el problema de clasificación de noticias desde el punto de vista de un proyecto de minería de datos. El proceso de minería contempla la recolección de los datos, la limpieza y transformación de los mismos, aplicación de diferentes algoritmos de clasificación, y la fase de difusión y uso. Los resultados demuestran que el clasificador *Máquina de Vectores de Soporte (SVM)* ofrece el mejor desempeño para la correcta clasificación de las noticias.

Palabras Clave: Minería de Texto; Análisis de Términos más Frecuentes; Balanceo de Clases; Selección de Variables; Máquina de Vectores de Soporte.

Abstract: In this study, various data mining techniques are used to make a recommendation based on the sentiment classification of a newsgroup. The idea of this recommendation is to support decision making about the direction of the future price of Venezuelan export crude oil. Specifically, text mining was used to classify recent oil market news based on learning from previously classified news. This article discusses and solves the news classification problem from the point of view of a data mining project. The mining process contemplates the collection of data, the cleaning and transformation of it, the application of different classification algorithms, and the dissemination and use phase. The results show that the *Support Vector Machine (SVM)* classifier offers the best performance for the correct classification of news.

Keywords: Text Mining; Frequently Asked Terms Analysis; Class Balancing; Variables Selection; Support Vector Machine.

I. INTRODUCCIÓN

El desarrollo de fórmulas para valoración de crudos es de gran importancia en la industria internacional del petróleo ya que estas permiten optimizar los ingresos por exportación de materia prima no renovable [1]. Un aspecto a destacar de este proceso es su complejidad, debido a la necesidad de análisis y ponderación (generalmente expuesto a críticas y revisiones de expertos) de una gran cantidad de aspectos técnicos, económicos y políticos en un contexto internacional, y a la insuficiencia de información pertinente.

En el caso de Venezuela, las fórmulas de precio usadas para los 4 crudos pivotes (*Santa Bárbara, Mesa 30, Mery 16 y Boscán*) consideran un factor de ajuste K , de estimación mensual, el cual refleja el comportamiento coyuntural del mercado y la

estrategia de colocación de los crudos de exportación [2].

La estimación del factor K consiste en realizar un ajuste a las fórmulas de precios para los crudos pivote en los mercados de interés, a fin de evitar posibles distorsiones que pueden generarse debido a la disparidad de condiciones políticas, financieras, económicas y técnicas entre los países consumidores, productores y exportadores de crudos [2]. La respectiva estimación involucra un análisis global del mercado petrolero y la ponderación de factores cualitativos relevantes.

En este sentido, se desea determinar mensualmente el sentimiento del mercado para ajustar las fórmulas de precio de los crudos pivotes venezolanos *Santa Bárbara, Mesa 30, Mery 16 y Boscán*, en los mercados de la Costa Estadounidense del Golfo de México, Noroeste de Europa y Asia. La propuesta

para realizar esta tarea es utilizar la minería de datos [3][4] para el análisis de noticias económicas que, según los expertos, pueden influir en el mercado petrolero.

Este documento se encuentra estructurado de la siguiente forma: en la Sección II se describen algunos trabajos relacionados con el tema de interés. En la Sección III se detallan el contexto del problema y los aspectos que se consideran relevantes a este. Luego, en la Sección IV, se describe cómo fue aplicado el proceso de minería de datos para la construcción de modelos de predicción del sentimiento del mercado. Por último, se presentan las conclusiones derivadas de este trabajo.

II. TRABAJOS PREVIOS

Las noticias de la web pueden ser usadas para rastrear con precisión no sólo varios fenómenos sociales sino también las tendencias financieras [5][6]. Los sistemas de los mercados financieros son complejos y, por lo tanto, las decisiones comerciales suelen basarse en información sobre una gran variedad de temas socioeconómicos y acontecimientos sociales.

Tradicionalmente los inversionistas, economistas y periodistas revisan publicaciones mensuales de datos macroeconómicos relacionados con las condiciones económicas. El gran problema con estos datos es que la información está disponible con un desfase que aumenta muy rápido. De hecho, los datos de un mes dado se publican generalmente a mediados del mes siguiente y suelen revisarse unos meses después. Ésta es la razón por la que en los últimos años se ha estudiado ampliamente un nuevo enfoque, basado en el aprendizaje automático, por su potencial para predecir la dirección de los mercados financieros.

Aplicando la teoría y las herramientas de la tecnología de la información y las comunicaciones (*TIC*), la estadística y la econometría, se han desarrollado sistemas capaces de recoger y analizar grandes cantidades de datos que están disponibles gratuitamente en la red. Por ejemplo, el análisis de sentimientos de la opinión expresada por los usuarios de las redes sociales es un campo de aplicación muy estudiado en los mercados financieros.

En este sentido, en [7] propusieron un método de predicción de la tendencia de los precios del petróleo con base en sentimientos. En el método propuesto se incluyen tres pasos principales, a saber, el análisis de sentimientos, el análisis de relaciones y la predicción de tendencias. En el análisis de sentimientos, este se extrae siguiendo un enfoque basado en diccionarios para capturar la información relevante en línea sobre los mercados petroleros y los factores conductores. En el análisis de relaciones, se aplica el test de causalidad de *Granger* para explorar cómo el sentimiento impacta el precio del petróleo. En la predicción de tendencias, el sentimiento se utiliza como una variable independiente importante, y se usan algunos modelos de predicción populares, como por ejemplo, regresión logística, máquina de vectores de soporte, árbol de decisión y redes neuronales, en particular el Perceptron Multicapa.

En [8] analizaron, en un periodo de dos años, la relación

entre los precios diarios del crudo *WTI* (*West Texas Intermediate*) y las múltiples variables extraídas de Twitter, Google Trends, Wikipedia y la base de datos *Global Data on Events, Language, and Tone (GDELT)*. Se aplicó un análisis basado en semántica para estudiar el sentimiento, la emotividad y la complejidad del lenguaje utilizado. Los modelos *ARIMAX* se utilizaron para hacer predicciones y confirmar valores de variables.

En [9] propusieron un método de predicción del precio del crudo basado en la minería de texto en línea, con el objetivo de captar los antecedentes más recientes de las fluctuaciones de precios en el mercado. Específicamente, se trata de un intento inicial de aplicar técnicas de aprendizaje profundo para la predicción del precio del petróleo crudo y extraer patrones ocultos dentro de los medios de comunicación en línea utilizando una red neuronal convolucional (*CNN*). Si bien las características del sentimiento de las noticias y las características extraídas por el modelo de *CNN* revelan relaciones significativas con el cambio de precios, es necesario agruparlas según su tópico, usando el modelo de tópicos *Latent Dirichlet Allocation (LDA)*, para distinguir los efectos sobre el precio de varios tópicos de noticias en línea y obtener una mayor precisión en la predicción.

Recientemente, en [10] propusieron un modelo de predicción del consumo de petróleo basado en datos en línea que utiliza las tendencias de Google, el cual refleja varios factores relacionados basados en una enorme cantidad de resultados de búsqueda. Este modelo implica dos pasos principales, el análisis de relaciones y la mejora de la predicción. En primer lugar, se realizan pruebas de cointegración y el test de causalidad de *Granger* con el fin de probar estadísticamente el poder predictivo de las tendencias de Google, en términos de tener una relación significativa con el consumo de petróleo. En segundo lugar, las tendencias efectivas de Google se introducen en los métodos populares de predicción para pronosticar tanto las tendencias como los valores del consumo de petróleo.

III. PLANTEAMIENTO DEL PROBLEMA

Partiendo de los datos de precios de crudos y productos refinados, flujos de crudos y productos a nivel mundial, y las características del sistema de refinación de cada región estudiada; se procede a identificar y analizar el comportamiento de los factores cuantitativos y cualitativos que influyen en la formación de precios del crudo [11].

Los aspectos principales que regularmente se consideran para la estimación del factor *K* varían mensualmente dependiendo de las condiciones y acontecimientos específicos del mercado [1], y pueden obtenerse mediante suscripciones a servicios internacionales de información petrolera. Estos aspectos son agrupados en 4 categorías las cuales son: *refinación*, *crudo*, *estacionalidad* y *asfaltos* [2].

En la categoría *refinación* se considera el volumen de crudo a consumir (porcentaje de procesamiento en refinación); temporada de mantenimiento (capacidad incorporada o desincorporada para el mes en estudio); cierre de refinerías; adecuaciones

en sistemas de refinación; paradas de refinación no programadas; situaciones de fuerza mayor en refinación; oferta y demanda de productos refinados principales (naftas, gasolinas, destilados medios y fueloil); márgenes de refinación; craqueo, inventarios y arbitraje de productos refinados entre cuencas.

Con respecto a la categoría *crudo*, se considera la incorporación o desincorporación de volúmenes de crudo de competidores, programas de exportación de los crudos competidores, inventarios de crudos, situaciones de fuerza mayor para la exportación de crudos, almacenamiento flotante de crudos, expectativas de disponibilidad de oferta de crudos, expectativas de demanda de crudos, situación de oleoductos y arbitrajes de crudos entre cuencas.

En cuanto a la categoría *estacionalidad*, se consideran temporadas de gasolina o diésel para calefacción, generación termoeléctrica durante el verano en Arabia Saudita, monzones en Asia y huracanes en costa este de los Estados Unidos de América.

Para los *asfaltos* se consideran precios, temporada de asfaltado por región, expectativas de demanda, expectativas de oferta o disponibilidad, inventarios, mantenimiento o cierres no programados de refinerías asfalteras y arbitraje entre cuencas.

Como se puede notar, son muchos los aspectos que se deben considerar para estimar el factor de ajuste K . El proceso de estimación del factor K , como se aborda en este trabajo, consiste en realizar un análisis de las noticias agrupadas según las categorías mencionadas anteriormente y según los mercados de interés, para extraer el sentimiento de mercado petrolero.

IV. PROCESO DE MINERÍA DE DATOS

A continuación se describen los pasos de la minería de datos [12] que se siguieron para la resolución del problema:

A. Recopilación e Integración

En la etapa de recopilación e integración de datos se tiene el problema planteado, los objetivos que se desean lograr, y la fuente de datos que se utilizará. Esta fase inicial se centra en la comprensión de los objetivos y requisitos del proyecto desde el punto de vista del negocio, para luego usar este conocimiento y definir el problema en términos de minería de datos, así como un plan preliminar que permita alcanzar los objetivos propuestos.

1) *Objetivo del Negocio*: Determinar mensualmente el sentimiento del mercado para ajustar el factor K de las fórmulas de precio de los crudos pivotes venezolanos *Boscán*, *Mesa 30*, *Merey 16* y *Santa Bárbara*, en los mercados de la Costa Estadounidense del Golfo de México, Noroeste de Europa y Asia.

2) *Fuentes de Datos*: Los datos son obtenidos por suscripción a los siguientes servicios globales de información y precios sobre energía: *Argus* (<https://www.argusmedia.com>) y *S&P Global Platts* (<https://www.spglobal.com>) (fuentes externa de datos). Los expertos en analizar el mercado petrolero

recopilan manualmente las noticias y luego las clasifican según el mercado destino, categoría de noticia y sentimiento. En la sección *Conociendo los Datos* se describe con mayor detalle el conjunto de datos.

3) *Criterios de Éxito del Negocio*: Se desea tener una precisión igual o superior a la que están proporcionando los expertos, 75% para al menos un crudo pivote en un mercado y, a su vez, reducir el tiempo que tardan los expertos en obtener el análisis completo del mercado.

4) Evaluación de la Situación:

- El negocio estima con esfuerzo propio el factor K obteniendo un margen de error aceptable.
- Cuenta con personal propio para realizar las actividades de los analistas del mercado petrolero.
- Dispone del acceso necesario a los principales proveedores globales de información sobre energía.
- Dispone de acceso a un software propietario de programación lineal para estimar precios de crudos vía valor bruto en productos.
- Cuenta con el asesoramiento de expertos en el negocio petrolero.
- Ausencia de un método para validar los juicios de experto en el análisis y la ponderación de factores cualitativos.

Esta situación es la que lleva a la búsqueda de una solución por medio de la minería de datos, a fin de obtener una herramienta de predicción que permita apoyar la toma de decisiones en cuanto al factor de ajuste K de la fórmula de precio de los crudos, así como para reducir el tiempo de análisis de los expertos.

5) *Objetivos de la Minería de Datos*: Los objetivos desde el punto de vista de la minería de datos son:

- Predecir, para cada noticia, la dirección en la que influye sobre el factor de ajuste K .
- Tarea de Minería de Datos: Clasificación.

6) *Criterios de Éxito de la Minería de Datos*: Desde la visión de la minería de datos, el criterio de éxito es obtener un rendimiento de predicción superior al 75%.

B. Preparación de los Datos

En este contexto, la preparación de los datos consiste en conocer el significado de todo el conjunto, seleccionar los datos adecuados para el estudio y realizar las transformaciones necesarias para convertir los documentos a una representación estructurada para poder aplicar los algoritmos de aprendizaje. Por último, aplicar técnicas de selección de variables para la reducción de la dimensionalidad.

1) Conociendo los Datos:

- El cuerpo (corpus) de entrada está constituido por 779 documentos de noticias que pueden influir en el sentimiento del mercado petrolero venezolano.
- Cada noticia se acompaña de los atributos *Fecha*, *Mercado*, *Categoría* y el *Sentimiento* (clase) asociados a cada

crudo pivote venezolano, cuya descripción se muestra a continuación:

- *Fecha*: denota el mes y año de la noticia.
- *Mercado*: los distintos mercados en los que influye la noticia. Estos pueden ser *CEGM* (Costa Estadounidense del Golfo de México); *NOE* (NorOeste de Europa); *Asia*.
- *Categoría*: las distintas categorías en las que se agrupan las noticias. Estas son *Refinación*; *Crudo*; *Estacionalidad*; *Asfaltos*.
- *Sentimiento Crudo x*: valor nominal en el conjunto $\{S : \text{Sube}, N : \text{Neutro}, B : \text{Baja}\}$ que denota si la noticia influye en el alza (*S*) del precio del crudo *x*, o es una noticia neutral (*N*) que no afecta su precio, o la noticia influye a la baja (*B*) del precio. Este valor fue proporcionado por los expertos.

Una muestra de estos datos se puede ver en la Tabla I, y un resumen se muestra en la Tabla II.

2) *Selección de los Datos*: Se tomó todo el conjunto de datos disponible, Dic2015 - Mar2017, para un total de 16 meses de información.

3) *Transformación de los Datos*: Esta etapa consiste en aplicar todas las transformaciones necesarias para lograr una representación estructurada a partir de los textos, siguiendo el modelo vectorial basado en bolsa de palabras. Los pasos que se aplicaron se mencionan a continuación:

- Traducción por parte de los expertos de algunas noticias desde el idioma inglés al idioma español.
- Corrección ortográfica de palabras.
- Corrección semántica de oraciones.
- Separación de noticias en *tokens* (tokenización).
- Eliminación de caracteres no imprimibles, acentos, números, símbolos, signos de puntuación y palabras vacías (*stopwords*).
- Adicionalmente, los expertos aportaron una lista de 500 palabras vacías aproximadamente, lo que permitió reducir la dimensionalidad de la matriz documento-término de 37.133 a 25.876 tokens, lo que representa una reducción del 30,32%. Estos tokens son la unión de palabras, bigramas y trigramas. Luego, se consideró solo analizar aquellos tokens que aparecieran en al menos el 1% del conjunto de noticias y en a lo sumo el 90% de las mismas. Esto se hizo con la idea de evitar aquellos tokens que sean específicos de una noticia (< 1%) o comunes a todas las noticias (> 90%). Esta restricción permitió reducir la dimensionalidad de la matriz de 25.876 a 615 tokens, lo que representa una reducción del 97,62%.
- Conversión a minúsculas.
- Lematización (*stemming*): aplicación del algoritmo *Porter Stemmer* para el español.
- Ejecución de 89 *expresiones regulares* para homologar expresiones escritas de diferentes maneras, como por ejemplo:
 - betun → bitumen
 - grado liviano → crudo liviano

– usa → EEUU

Finalmente, para construir una representación estructurada de las noticias se construye una matriz documento-término donde cada elemento $TfIdf_{ij}$ se calcula de la siguiente manera:

$$TfIdf_{ij} = TfN_{ij} * Idf_i \quad (1)$$

$$TfN_{ij} = \frac{F(t_i, d_j)}{T(d_j)} \quad (2)$$

$$Idf_i = \log_2\left(\frac{m}{D(t_i)}\right) \quad (3)$$

donde:

- *m*: cantidad de documentos
- *n*: cantidad de términos
- *t_i*: término *i* para $i = 1, 2, \dots, n$
- *d_j*: documento *j* para $j = 1, 2, \dots, m$
- $TfIdf_{ij}$: métrica $TfIdf$ del *t_i* en el *d_j*
- TfN_{ij} : frecuencia relativa o normalizada del *t_i* en el *d_j*
- $F(t_i, d_j)$: frecuencia absoluta o cantidad de veces que aparece el *t_i* en el *d_j*
- $T(d_j)$: cantidad de términos en el *d_j*
- Idf_i : componente Idf del *t_i*
- $D(t_i)$: cantidad de documentos que contienen el *t_i*

4) *Selección de Variables*: Para la selección de variables o tokens de la matriz de documento-término, se utilizaron tres métodos de selección de atributos por ranqueo (*Chi Cuadrado*, *Gain Ratio* e *Information Gain*) [13], los cuales seleccionaron exactamente las mismas variables para cada crudo, quedando distribuida la cantidad de atributos seleccionados de la siguiente manera:

- 404 para el crudo *Merey 16*.
- 401 atributos para el crudo *Boscán*.
- 370 atributos para el crudo *Mesa 30*.
- 351 atributos para el crudo *Santa Bárbara*.

Otro de los métodos utilizados para la selección de variables fue el *análisis de términos más frecuentes*. Para aplicar éste método, primero se totalizó cada término según su frecuencia de ocurrencia en la colección de documentos de noticias $TfIdf$, luego se ordenan los términos descendientemente según la totalización del paso anterior, y por último, para cada crudo se seleccionan los *n* primeros términos más frecuentes, donde *n* denota la cantidad de atributos elegidos mediante los métodos por ranqueo. Con éste método, los atributos seleccionados difieren de los atributos elegidos por los métodos de ranqueo. Los autores consideran que esta diferencia se debe a la propia naturaleza de ambas familias de métodos, es decir, mientras que los métodos por ranqueo utilizados se basan en la entropía de los atributos para determinar su relevancia, en el análisis de términos más frecuente se usa la frecuencia relativa de cada término con respecto a la colección de documentos.

5) *Construcción de Nuevos Atributos*: Se realizó una transformación en el atributo clase, cambiando su tipo de dato real en el rango $[-1, 1]$ a nominal. Así, se cambiaron los valores negativos al valor *B*, los valores de 0 a *N*, y los valores positivos al valor *S*.

Tabla I: Muestra de los Datos

FECHA	MERCADO	CATEGORÍA	NOTICIA	SANTA_BARBARA	MESA	MEREY	BOSCAN
01/12/2015	CEGM	Refinación	Se espera que el mantenimiento de crudo REFQ4 sea relativamente ligero en el futuro y se mantendrá así durante los primeros meses del primer trimestre de 2016. Dada la reciente recuperación de margen esperamos que CEGM continúe aumentando las carreras hasta muy por encima del 90%	B	B	B	B
01/12/2015	CEGM	Crudo	Se espera que el crudo de WTS Midland regrese a una prima contra WTI Cushing una vez que las operaciones de reanudación del oleoducto y mantenimiento local disminuyan. Esto aumentará los precios de la fórmula venezolana	N	N	N	N
01/12/2015	CEGM	Estacionalidad	Una abundancia de fondos de torre de vacío baratos (VTBs) en el USG agregó presión a la baja a los mercados de asfalto en masa. La gran cantidad de compradores que buscan recibir entregas finales para terminar la temporada de pavimentación estaban retrasando las compras con la esperanza de pagar precios más bajos la próxima semana. Barcazas del Golfo de Estados Unidos se escucharon vendiendo tan bajo como \$210-220/st en la parte posterior de la demanda débil una disminución de \$20/st de la semana pasada. Una importante refinería de la costa del Golfo anunció que se estaría manteniendo en unidades no reveladas lo que podría aumentar el exceso de VTBs en la región y establecer el escenario para más debilitamiento a corto plazo en la mayoría de los mercados	N	N	N	B
01/01/2016	NOE	Asfalto	En Alemania Francia y Benelux se acordó mantener los precios domésticos a los niveles de enero. La actividad permanece débil como es usual en la época	N	N	N	B

Tabla II: Resumen de los Datos

VARIABLE	TIPO	AUSENCIAS	MODA	VALORES
Fecha	Fecha	0	01/03/17	01/12/2015..01/03/2017
Mercado	Nominal	0	Asia	CEGM, NOE, ASIA
Categoría	Nominal	0	Crudo	REFINACIÓN CRUDO ESTACIONALIDAD ASFALTO
Noticia	Nominal	0	NA	-
Santa Bárbara	Nominal	0	N	S, N, B
Mesa	Nominal	0	N	S, N, B
Merrey	Nominal	0	N	S, N, B
Boscán	Nominal	0	N	S, N, B

6) *Análisis Exploratorio*: Con el análisis exploratorio se busca un entendimiento básico de los datos y de las relaciones existentes entre las variables. Para lograrlo, primero se deben organizar y preparar los datos de manera que puedan aplicarse métodos estadísticos para la detección de fallas en el diseño y recolección de los mismos, tratamiento y evaluación de datos ausentes, identificación de casos atípicos (*outliers*) y comprobación de los supuestos subyacentes en la mayor parte de las técnicas multivariantes (normalidad, linealidad, homocedasticidad). Es importante señalar que el análisis exploratorio se llevó a cabo con todas las palabras y haciendo uso de los procedimientos gráficos descritos en [14] para manipular, resumir y visualizar características del texto.

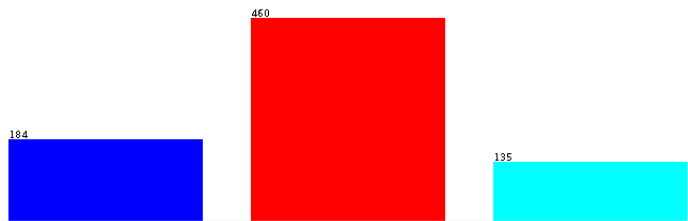


Figura 1: Histograma de Clase para Santa Bárbara

La Figura 1 muestra la distribución de las clases para el crudo Santa Bárbara, con cada clase descrita por un color: Azul = B (Baja), Rojo = N (Neutro), Cyan = S (Sube). En la Figura 2 se aplicó el análisis de términos más frecuentes para conocer cuáles eran los términos más frecuentes en el conjunto de noticias. Específicamente, se utilizaron las medidas Tf y $TfIdf$ para graficar los 10 primeros términos más frecuentes.

La Figura 3 muestra los términos con una correlación mayor o igual al 95% que están asociados al mercado de Asia. Mientras mayor sea la correlación, más gruesa es la línea entre los términos.

También se aplicó el análisis de términos más frecuentes para pares de palabras (*bigramas*). Los resultados se muestran en

Términos Globales Más Frecuentes

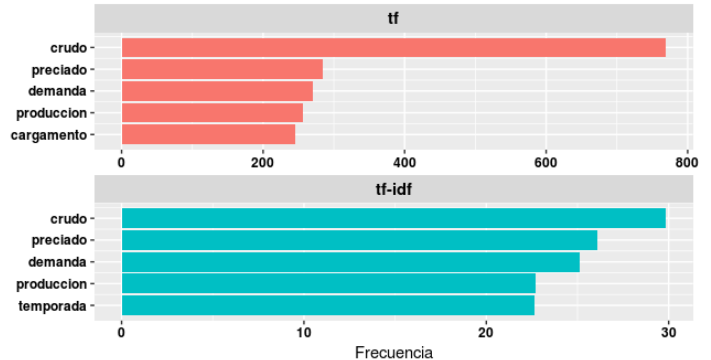


Figura 2: Términos más Frecuentes según Tf y $TfIdf$

Red de Correlación de términos para Asia



Figura 3: Red de Correlación, Mercado de Asia

Bigramas más frecuentes

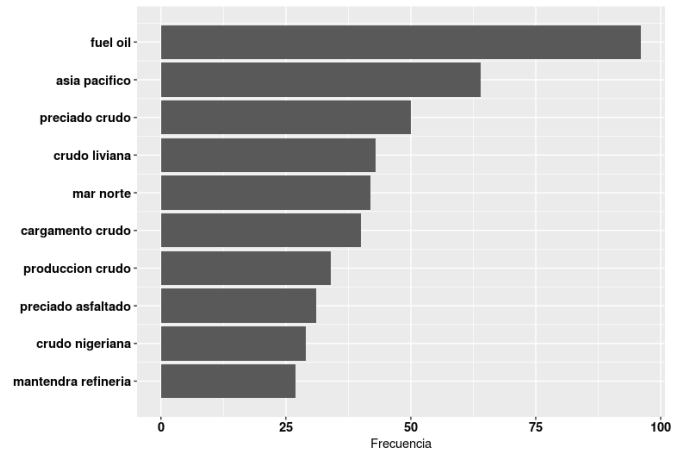


Figura 4: Bigramas más Frecuentes

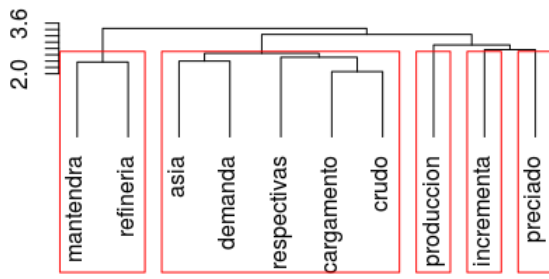


Figura 5: Dendrograma

la Figura 4. La Figura 5 muestra las relaciones más fuertes entre las palabras más frecuentes entre todas las noticias.

Similitud de Términos

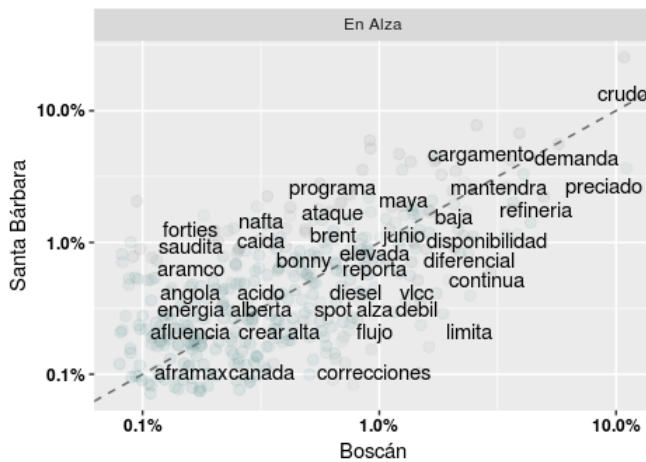


Figura 6: Similitud de Términos entre Santa Bárbara y Boscán

Para visualizar cuáles son los términos que están mayormente asociados a un crudo dado en comparación con otro crudo, se emplea el gráfico de similitud por proporción de palabras. La Figura 6 muestra este gráfico para los crudos Santa Bárbara y Boscán, agrupado por sentimiento al alza.

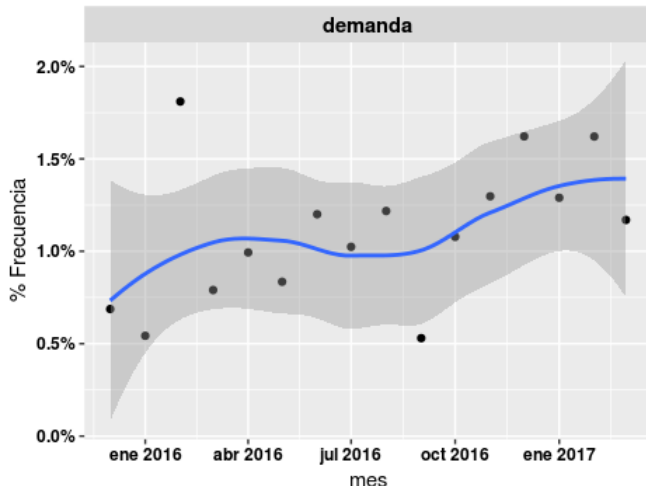


Figura 7: Variación de Frecuencia en el Tiempo

Sentimiento Mensual

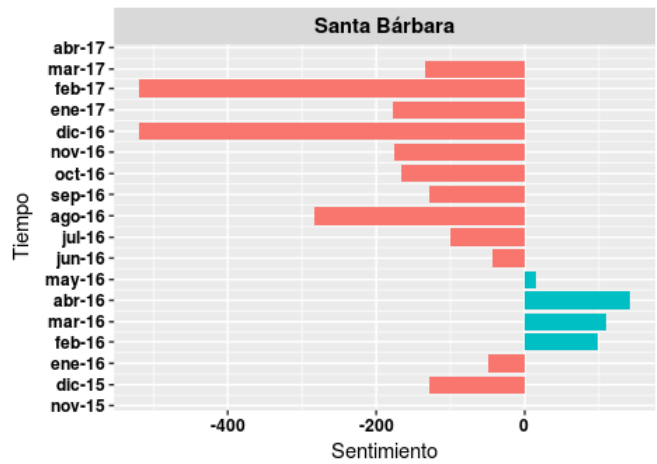


Figura 8: Sentimiento Mensual - Santa Bárbara

La Figura 7 ilustra la variación de la frecuencia en el tiempo para el término *demanda*, y un gráfico sobre el sentimiento mensual para el crudo *Santa Bárbara* se muestra en la Figura 8.

7) *Vista Minable*: El último paso en la preparación de los datos es la construcción de la vista minable. En este estudio, se construyeron cuatro vistas minables, una por cada uno de los crudos. Debido a que las clases estaban desbalanceadas, se utilizó el método *SMOTE* [15] para generar muestras sintéticas de las clases minoritarias.

A continuación se resumen los pasos generales que se siguieron para construir una vista minable con métrica *TfIdf*.

- Calcular la matriz de documento-término $DT_{m \times n}$ con métrica $TfIdf_{ij}$.
- Eliminar atributos redundantes. Resultado $DT_{m \times n'}$. Los atributos redundantes son aquellos que están altamente correlacionados, aplicando la correlación de Pearson mayor a 0.75.
- Balancear las clases minoritarias. Resultado $DT_{m' \times n'}$. El balanceo se realizó aplicando el algoritmo *SMOTE*. Las clases minoritarias para algunos crudos pivotes estaban por debajo del 15% y para otros crudos estaban por debajo del 19%.
- Seleccionar atributos por importancia o frecuencia. Resultado $DT_{m' \times n''}$.

C. Minería de Datos

En este estudio se buscó predecir, para cada noticia, la dirección (*S, N, B*) en la que influye sobre el factor de ajuste *K*. Para lograrlo, es necesario contar con un conjunto de noticias previamente clasificadas y aprender de ellas. A continuación se presentan las distintas máquinas de aprendizaje utilizadas para la tarea de clasificación de noticias, las cuales fueron entrenadas haciendo uso de la herramienta *Weka*:

- Bayes Ingenuo con Estimador Núcleo
- K-Vecinos para $K = 3, 5, 7$

- RIPPER (JRip) con $F = 3$, $N = 2$, $O = 2$, $S = 1$ y *Pruning* activado
- C4.5 (J48) con $C = 0.25$ y $M = 2$
- Random Forest con $P = 100$, $l = 100$, $slots = 1$, $K = 0$, $M = 1$, $V = 0.001$ y $S = 1$
- Máquina de Vectores de Soporte (SVM) con $K=NormalizedPolyKernel$ y *BuildCalibrationModels* activado

La aplicación de estos algoritmos se hizo con validación cruzada de K-particiones (*K-fold cross validation*) con $k = 10$. Considerando que para el negocio es más importante reconocer cuando una noticia influye en el alza de los precios, entonces la clase positiva es S . Para seleccionar la máquina de aprendizaje con los mejores resultados de clasificación, se toma aquella que presente las siguientes métricas más altas en función de la clase positiva:

- Instancias Correctamente Clasificadas (ICC).
- Tasa de Verdaderos Positivos (TVP).
- Precisión.
- Sensibilidad.
- Medida-F.
- G-Media.

Es importante señalar que para cada crudo, las máquinas de aprendizaje se aplicaron utilizando el conjunto de atributos seleccionados por los métodos de ranqueo (*Chi Cuadrado*, *Information Gain* y *Gain Ratio*) así como el conjunto de atributos seleccionados por el análisis de términos más frecuentes con *TfIdf*.

D. Evaluación

Las Tablas III y IV muestran los resultados para cada una de las máquinas de aprendizaje. Se observa que la Máquina de Vectores de Soporte (SVM) es la que exhibe, en general, los mejores valores de las métricas de rendimiento. Tomando en cuenta que se utilizó para cada crudo dos conjuntos de atributos como consecuencia de los métodos de selección de variables, el modelo de predicción se construyó de la siguiente manera:

- Para los crudos *Boscán*, *Merey 16* y *Santa Bárbara*, la Máquina de Vectores de Soporte (SVM) se entrenó con el conjunto de atributos seleccionados por los métodos de ranqueo (*Chi Cuadrado*, *Gain Ratio*, *Information Gain*).
- Para el crudo *Mesa 30*, la Máquina de Vectores de Soporte (SVM) se entrenó con el conjunto de atributos seleccionados mediante el análisis de términos más frecuente con *TfIdf*.

La Tabla III muestra las métricas de las máquinas de aprendizaje usando los atributos seleccionados por los métodos de ranqueo para el crudo *Boscán*, mientras que la Tabla IV muestra las métricas de las máquinas de aprendizaje usando los atributos seleccionados mediante el análisis de términos más frecuente con *TfIdf* para el crudo *Boscán*.

Aunque en la literatura puede parecer que *Random Forest* obtiene, en general, un mejor desempeño en tareas de clasi-

ficación, nuestros resultados demuestran que esto no es así cuando los datos son textos.

E. Difusión y Uso

El modelo de predicción recibe como entrada un grupo de noticias a clasificar organizado como sigue: *MERCADO*, *CATEGORIA* y *NOTICIA*. A este grupo se le aplica la Máquina de Vectores de Soporte (SVM) para asignarle una clase (S, N, B) o sentimiento de mercado a cada una de las noticias. Luego que todas las noticias hayan sido clasificadas, el modelo emite una recomendación sobre si se debe subir, mantener o bajar el valor del factor k , según el siguiente criterio:

Para Todo $c_i \in C$ **Hacer**
Si $T(c_i) > 1,5 * mg$ **Entonces**
 $R(c_i)$
Fin Si
Fin Para

donde:

- $C = \{c_i / c_1 = S, c_2 = N, c_3 = B\}$.
- $T(c_i)$: cantidad de noticias clasificadas como c_i
- $mg = \sqrt[3]{T(c_1) * T(c_2) * T(c_3)}$ (media geométrica)
- $R(c_i)$: emitir recomendación según c_i

Adicionalmente, el modelo calcula la probabilidad de que cada noticia pertenezca a una clase y genera una gráfica que muestra la tendencia del sentimiento del mercado como lo ilustra la Figura 9. En la actualidad, el sistema se encuentra en período de prueba.

V. CONCLUSIÓN

Los resultados demuestran que tanto los objetivos como los criterios de éxito desde el punto de vista del negocio y de la minería de datos fueron logrados. En este sentido, se superó ampliamente el objetivo del negocio ya que se logró predecir con una precisión cercana al 90% el sentimiento del mercado para los 4 crudos pivote en los 3 mercados de interés. Cuando los expertos no lograron un acuerdo sobre el sentimiento de alguna noticia, estos utilizaron la clasificación hecha por el modelo para decidir sobre el sentimiento de la noticia siempre y cuando la probabilidad de que una noticia perteneciera a una clase superara el 90%. Así, se logró reducir en un 50% aproximadamente el tiempo de análisis del mercado mediante el uso del modelo anteriormente descrito para la predicción y recomendación. Adicionalmente, se logró el objetivo de la minería de datos mediante la clasificación de las noticias con una precisión cercana al 90%.

Los autores consideran que los aportes de este trabajo están relacionados con la aplicación de técnicas de aprendizaje automático en el contexto económico petrolero venezolano, y cómo pueden ser utilizadas para apoyar la toma de decisiones en un sector tan importante para el país.

Como el modelo propuesto fue entrenado a partir de juicios de expertos y este juicio cambia en función de las condiciones del mercado petrolero, es necesario re-entrenar el modelo cada

Tabla III: Métricas basadas en Métodos de Ranqueo

Crudo Pivote		Boscán	Métodos de Selección de Atributos		Chi Cuadrado, Information Gain y Gain Ratio *				
# Atributos		401	# Instancias		1217				
Métrica		Bayes Ingénúo (Estimador Núcleo)	SMV (SMO) (NormalizedPolyKernel) (BuildCalibrationModels)	KNN (K=3)	KNN (K=5)	KNN (K=7)	RIPPER (Jrip)	C4.5 (J48) (BinarySplit)	Random Forest
ICC		87,35%	92,19%	70,99%	67,38%	62,12%	74,94%	81,35%	89,65%
IIC		12,65%	7,81%	29,01%	32,62%	37,88%	25,06%	18,65%	10,35%
TVP	B	92,70%	90,40%	98,50%	95,90%	92,70%	65,20%	77,80%	84,20%
	N	80,00%	92,10%	39,80%	35,70%	28,20%	87,70%	84,80%	98,00%
	S	94,60%	94,30%	96,50%	92,70%	89,20%	62,90%	79,00%	80,60%
TFP	B	6,90%	2,90%	26,50%	31,30%	35,30%	5,50%	5,50%	0,80%
	N	4,10%	7,00%	1,10%	1,80%	1,80%	30,40%	18,40%	17,20%
	S	7,40%	2,70%	12,60%	12,30%	15,50%	6,30%	6,40%	0,70%
Precisión	B	84,10%	92,50%	59,20%	54,50%	50,60%	82,30%	84,70%	97,60%
	N	94,30%	91,80%	97,00%	94,30%	92,90%	71,10%	79,70%	82,90%
	S	81,60%	92,50%	72,70%	72,50%	66,70%	77,60%	81,10%	97,70%
Sensibilidad	B	92,70%	90,40%	98,50%	95,90%	92,70%	65,20%	77,80%	84,20%
	N	80,00%	92,10%	39,80%	35,70%	28,20%	87,70%	84,80%	98,00%
	S	94,60%	94,30%	96,50%	92,70%	89,20%	62,90%	79,00%	80,60%
Medida-F	B	88,20%	91,40%	74,00%	69,50%	65,50%	72,80%	81,10%	90,40%
	N	86,60%	92,00%	56,50%	51,80%	43,30%	78,50%	82,20%	89,90%
	S	87,60%	93,40%	82,90%	81,30%	76,40%	69,50%	80,10%	88,30%
G-Media		88,86%	92,25%	72,32%	68,21%	61,55%	71,12%	80,48%	87,29%

* Los 3 métodos seleccionaron los mismos atributos

Tabla IV: Métricas basadas en TfIdf

Crudo Pivote		Boscán	Métodos de Selección de Atributos		Término más Frecuente				
# Atributos		401	# Instancias		1217				
Métrica		Bayes Ingénúo (Estimador Núcleo)	SMV (SMO) (NormalizedPolyKernel) (BuildCalibrationModels)	KNN (K=3)	KNN (K=5)	KNN (K=7)	RIPPER (Jrip)	C4.5 (J48) (BinarySplit)	Random Forest
ICC		82,09%	91,13%	76,58%	70,50%	68,69%	74,28%	79,21%	88,58%
IIC		17,91%	8,87%	23,42%	29,50%	31,31%	25,72%	20,79%	11,42%
TVP	B	90,60%	90,10%	98,00%	94,40%	92,40%	64,30%	77,20%	82,70%
	N	72,10%	91,60%	53,40%	47,30%	45,70%	86,60%	81,40%	97,70%
	S	90,50%	91,40%	94,60%	85,70%	83,80%	63,20%	77,50%	78,70%
TFP	B	11,00%	3,40%	24,50%	30,60%	32,10%	6,70%	5,60%	0,90%
	N	6,80%	8,70%	1,80%	4,70%	4,60%	30,40%	19,00%	18,90%
	S	8,50%	2,30%	6,50%	6,70%	7,80%	6,00%	8,80%	0,80%
Precisión	B	76,40%	91,10%	61,00%	54,70%	52,90%	78,90%	84,30%	97,30%
	N	90,00%	90,00%	96,10%	89,50%	89,50%	70,80%	78,50%	81,50%
	S	78,70%	93,20%	83,50%	81,80%	79,00%	78,70%	75,50%	97,30%
Sensibilidad	B	90,60%	90,10%	98,00%	94,40%	92,40%	64,30%	77,20%	82,70%
	N	72,10%	91,60%	53,40%	47,30%	45,70%	86,60%	81,40%	97,70%
	S	90,50%	91,40%	94,60%	85,70%	83,80%	63,20%	77,50%	78,70%
Medida-F	B	82,90%	90,60%	75,20%	69,20%	67,30%	70,90%	80,60%	89,40%
	N	80,10%	90,80%	68,70%	61,90%	60,50%	77,90%	79,90%	88,90%
	S	84,20%	92,30%	88,70%	83,70%	81,40%	70,10%	76,50%	87,00%
G-Media		83,93%	91,03%	79,11%	72,60%	70,73%	70,60%	78,68%	85,99%

vez que las condiciones varían. En otras palabras, el modelo debe re-entrenarse en épocas de alta volatilidad de los precios porque existe un cambio en las condiciones que desestabiliza el mercado.

REFERENCIAS

[1] T. Rifai, *The Pricing of Crude Oil*. Economic and Strategic Guidelines for an International Energy Policy, Praeger Publishers, 1975.
 [2] A. Briceño y J. Suárez, *Informe de Posicionamiento sobre las Actividades de la Oficina para la Determinación de los Precios de Crudo de Exportación (ODPCE) del Ministerio del Poder Popular de Petróleo*. Dirección Ejecutiva de Automatización, Informática y Telecomunicaciones, Petróleos de Venezuela, S.A. 2016, Sin publicar.
 [3] C. Aggarwal, *Data Mining. The Text Book*, Springer, 2015.
 [4] S. Weiss, N. Indurkha, and T. Zhang, *Fundamentals of Predictive Text Mining*, Second Edition. Springer. 2015.
 [5] H. Choi and H. Varian, *Predicting the Present with Google Trends*, Technical report, 2009.
 [6] T. Preis, H. Moat, and H. Stanley, *Quantifying Trading Behavior in Financial Markets using Google Trends*, Scientific Report, 3, 1684, 2013.
 [7] L. Jian, X. Zhenjing, Y. Lean, and T. Ling, *Forecasting Oil Price Trends with Sentiment of Online News Articles*, Procedia Computer Science, vol. 91, pp. 1081-1087, <https://doi.org/10.1016/j.procs.2016.07.157>, 2016.

[8] E. Mohammed, F. Andrea, B. Elisa, and A. Peter, *Using Four Different Online Media Sources to Forecast Crude Oil Price*, 2017.
 [9] L. Xuerong, S. Wei, and W. Shouyang, *Text-based Crude Oil Price Forecasting: A Deep Learning Approach*, International Journal of Forecasting, vol. 35, no. 4, pp. 1548-1560, <https://doi.org/10.1016/j.ijforecast.2018.07.006>, 2019.
 [10] Y. Lean, Z. Yaqing, T. Ling, and Y. Zebin, *Online Big Data-driven Oil Consumption Forecasting with Google Trends*, International Journal of Forecasting, vol. 35, no. 1, pp. 213-223, <https://doi.org/10.1016/j.ijforecast.2017.11.005>, 2019.
 [11] J. R. Zanoni, *El Precio del Petróleo sus Determinantes y su Fijación por la OPEP*, Ediciones FaCES/UCV, 1981.
 [12] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, *CRISP-DM 1.0, Step-by-Step Data Mining Guide*, 2000.
 [13] P. Sethi and M. Jain, *A Comparative Feature Selection Approach for the Prediction of Healthcare Coverage*, ICISTM 2010, CCIS 54, pp. 392-403. Springer-Verlag Berlin Heidelberg, 2010.
 [14] J. Silge and D. Robinson, *Text Mining with R. A Tidy Approach*, Publisher: O'Reilly Media, 2017.
 [15] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, *SMOTE: Synthetic Minority Over-sampling Technique*, Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, <https://doi.org/10.1613/jair.953>, 2002.

Clasificador de Noticias del Mercado Petrolero

Se recomienda mantener el valor del factor k para el crudo Boscán

NOTICIAS CLASIFICADAS

Show 5 entries

Search:

MERCADO	CATEGORÍA	NOTICIA	Sentimiento	Prob_Baja	Prob_Neutro	Prob_Alza
CEGM	Refinación	Se espera que el mantenimiento de crudo REFO4 sea relativamente ligero en el futuro y se mantendrá así durante los primeros meses del primer trimestre de 2016. Dada la reciente recuperación de margen esperamos que CEGM continúe aumentando las carreras hasta muy por encima del 90%	Baja	99.71	0.29	0.00
CEGM	Refinación	Los grados de Permian se mantuvieron con descuentos contra el WTI de Midland debido principalmente al intensivo mantenimiento de las refinerías en la zona central de EEUU. A pesar de las caídas en los números de taladros en los EEUU la producción sigue creciendo de acuerdo con los datos del gobierno de dicho país debido a que las áreas más productivas se han convertido en el foco de las operaciones. El número de taladros en Permian ha caído en cada una de las seis últimas semanas.	Neutro	11.14	88.85	0.01
CEGM	Crudo	Se espera que el crudo de WTS Midland regrese a una prima contra WTI Cushing una vez que las operaciones de reanudación del oleoducto y mantenimiento local disminuyan. Esto aumentará los precios de la fórmula venezolana.	Neutro	0.00	99.69	0.31
CEGM	Crudo	La comercialización de crudo LLS para diciembre comenzó la sesión con una prima de 85 centavos por barril sobre el WTI. El Mars para diciembre se comercializó con descuentos de \$4/barril a \$3.75/barril fortaleciéndose junto al LLS después de haber sido comercializado un día antes a -4.20/-4.10 \$/b. Ambos crudos el LLS y el Mars estuvieron lo más fuertes que han estado desde el 27 de octubre durante la segunda sesión de los cargamentos de las primeras ventanas de diciembre.	Neutro	0.22	99.78	0.00
CEGM	Crudo	La prima del crudo de Alaska Slope Norte (ANS) con respecto al WTI CMA Nymex de diciembre se fortaleció en unos 13 centavos por barril después del fin de semana para reflejar el descuento de \$2.03/barril contra el CMA ICE Brent de Diciembre establecido la última vez que fue comercializado este crudo para entrega en diciembre en la Costa Oeste de los EEUU el 27 de Octubre.	Neutro	0.21	99.79	0.00

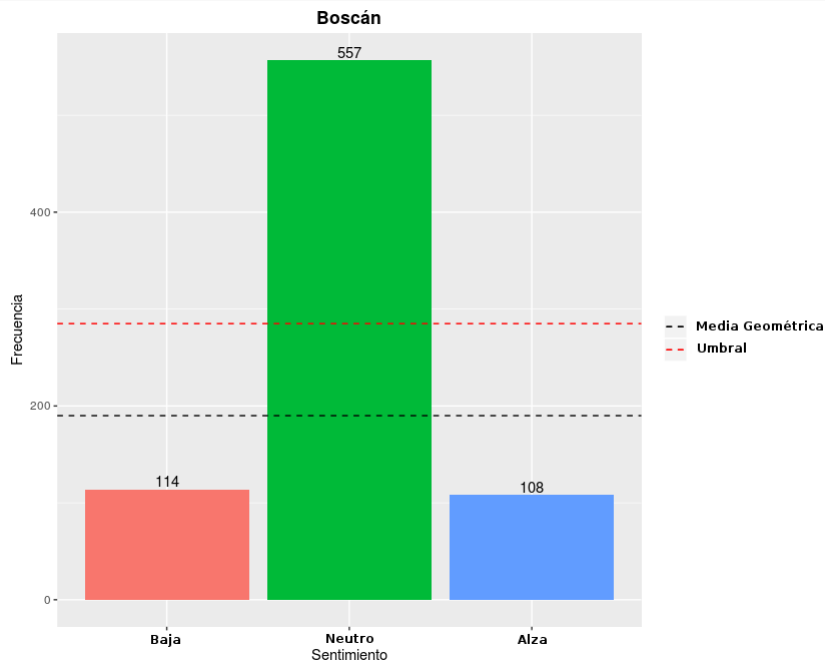


Figura 9: Sistema de Recomendación sobre el Sentimiento del Mercado Petrolero

Einalu'u: Videojuego de Simulación de Construcción y Gestión de Culturas Indígenas Prehispánicas

Adriana Urdaneta¹, Esmitt Ramírez^{1,2}
addeurme@gmail.com, esmitt.ramirez@ciens.ucv.ve

¹ Escuela de Computación, Universidad Central de Venezuela, Caracas, Venezuela

² Centre de Visió per Computador, Universitat Autònoma de Barcelona, Barcelona, España

Resumen: La historia es parte de la identidad de una nación. El conocimiento público de su historia y sus orígenes es importante, por ello existen iniciativas que contribuyen a su difusión. En Latinoamérica, las culturas indígenas precolombinas aportan un gran valor a las identidades de sus naciones. Particularmente, la cultura venezolana se forjó como producto del intercambio cultural entre europeos, africanos y los nativos pobladores de América. Actualmente, es posible identificar elementos culturales de los pueblos indo-originarios de Venezuela, aun cuando muchas etnias hayan desaparecido y exista un desconocimiento generalizado de las mismas. Así, con el fin de preservar las culturas indígenas prehispánicas, se justifican las propuestas tecnológicas que contribuyen a su difusión. En este trabajo se plantea una propuesta basada en un videojuego de simulación de construcción y gestión, de un solo jugador, empleando arte 2D dimétrico basado en *tiles*. En Einalu'u, nombre del juego, el jugador elige una etnia y una ubicación para jugar, construir y gestionar los elementos hasta alcanzar un hito poblacional. La metodología usada fue el diseño iterativo, cuyo enfoque de desarrollo ágil permite mejorar el producto de manera incremental en cada iteración. Se utilizó un sistema de encuestas para evaluar el impacto del videojuego sobre los jugadores enfocados en la difusión de la cultura indígena. Los resultados muestran que en Einalu'u, un jugador encuentra una amplia cantidad de elementos culturales, naturales y geográficos que, en muchos casos, son desconocidos y de su interés; sin embargo, la densidad etnográfica y la carencia de algunos elementos identificativos en el juego obstaculizó su experiencia lúdica.

Palabras Clave: Indígenas; Videojuegos; Simulación; Biomas; Ecorregiones.

Abstract: History is part of the identity of a nation. Public knowledge of its history and its origins is crucial, therefore, they are initiatives that contribute to its dissemination. In Latin America, pre-Colombian indigenous cultures bring great value to nations identities. Particularly, Venezuelan culture was forged as a product of the cultural exchange between Europeans, Africans and the native people of America. Currently, it is possible to identify cultural elements of the indigenous people of Venezuela, even when many ethnic groups have disappeared and there is a widespread ignorance of indigenous cultures. Thus, in order to preserve pre-Hispanic indigenous cultures, technological proposals that contribute to their dissemination are justified. In this paper, we propose an application for a single player for construction and management of a simulation videogame, using 2D tiles-based dimetric art. In Einalu'u, name of the game, the player chooses an ethnic group and a location to play, build and manage their elements until success a population milestone. The methodology used was the iterative design, whose agile development approach allows to improve the product incrementally on each iteration. A survey system was used to assess the impact of the videogame on players focused on the dissemination of indigenous culture. The results show that in Einalu'u, a player finds a large number of cultural, natural and geographical elements that, in many cases, are unknown and of interest; however, ethnographic density and the lack of some identifying elements in the game hindered his playful experience.

Keywords: Natives; Videogame; Simulation; Biomes; Ecoregions.

I. INTRODUCCIÓN

Dentro de los aspectos socio-culturales de los habitantes de una nación se encuentra la historia de dicha nación. Esta constituye un hecho fundamental en el proceso de formación de la identidad ciudadana. Particularmente, en Venezuela, parte de su cultura fue gestada a raíz del intercambio cultural entre

los indígenas, habitantes originarios de las tierras americanas, los conquistadores europeos y los africanos traídos como esclavos a América. Los indígenas, nativos lugartenientes, fueron diezmados y aculturizados viéndose en la necesidad de movilizarse de sus hogares luego de la llegada de los europeos a América.



Figura 1: Capturas de Pantalla durante la Ejecución de Einalu'u: (a) Instante justo antes de Iniciar a Jugar, (b) Información de la Materia Prima que el Jugador Puede Consultar y, (c) Vista de la Evolución del Juego con los Elementos sobre el Mapa

De acuerdo con el registro oficial más reciente [1], para el año 2011, en Venezuela había un poco más de 700.000 habitantes indígenas que equivale a un 2,7% de su población. Así, la cultura venezolana se formó con un matiz bien marcado de la cultura occidental, dificultando el conocimiento y la valoración de su cultura indígena autóctona, que es un aspecto fundamental para entender e identificar la herencia cultural y el legado de los pueblos indo-originarios que la habitaron siglos atrás.

Las iniciativas que contribuyen a la difusión de las culturas indígenas son de gran importancia en el proceso de inclusión y valoración de dicho legado a la población venezolana. Principalmente, este proceso va dirigido a aquellos habitantes que actualmente no forman parte de dichos pueblos y desconocen, en gran medida, las características culturales de las diferentes etnias que existen en Venezuela, y aún más, las que dejaron de existir.

Considerando la importancia de la difusión de los valores autóctonos, la existencia de una propuesta tecnológica que facilite el acercamiento a las culturas indígenas se justifica en el impacto que podría representar en las personas interesadas en documentarse acerca de las etnias indo-originarias. Los juegos de video o videojuegos son una solución tecnológica que permite hacer una experiencia exploratoria, inmersiva y divertida para sus usuarios. El impacto que podría significar el desarrollo de videojuegos de temática indígena reposa en el hecho de que éstos constituyen no sólo una herramienta de entretenimiento, sino una forma de expresión artística que permite al jugador conectar con los contenidos.

En Venezuela, existen algunas iniciativas basadas en videojuegos que reivindican y promocionan la diversidad cultural. Entre dichas propuestas cabe destacar el videojuego El Canto del Autana [2], que aborda aspectos culturales de los pueblos autóctonos. En dicho trabajo, se trata la temática indígena sin hacerla su enfoque principal, apostando por un carácter nacionalista. Igualmente, se resalta la cultura indígena y muestra algunas de sus creencias, partiendo de mecánicas con objetivos como la exploración, recolección de elementos y mini-juegos cuyo fin lúdico es superar un reto aislado representativo. Al mismo tiempo, mezcla tradiciones, leyendas y deidades de diferentes culturas, con elementos venezolanos e incluso occidentales. Estos aspectos motivan el interés y el acercamiento hacia algunos elementos de la cultura indígena

venezolana enmarcados con un fin turístico.

Asimismo, el videojuego El Sueño de Bolívar desarrollado en el INCES (Instituto Nacional de Capacitación y Educación Socialista) dentro del Proyecto Integral Socialista Desarrollo Educativo en 3D del Estado Zulia, tiene como objetivo rescatar los símbolos patrios de Venezuela [3]. Otro ejemplo, es el juego Aventura Venezuela, un videojuego serio de concienciación enfocado en la fauna silvestre de Venezuela en peligro de extinción [4]. Ambas propuestas, presentan aspectos lúdicos de dar conocimiento a sus jugadores sobre situaciones específicas en Venezuela. Sin embargo, ninguna de estas propuestas favorece el conocimiento de las culturas indígenas desde un punto de vista histórico.

En este trabajo se presenta Einalu'u, un videojuego de simulación y gestión centrado en la cultura de los diferentes pueblos indígenas que habitaron en el territorio de Venezuela durante el período neo-indio. Este género de juego permite emular una realidad histórico-social, invitando al jugador a controlar la gestión y el crecimiento de una tribu, dentro de la cultura de una etnia específica. Con este enfoque, se busca que el jugador se sienta identificado con los valores y el modo de vida de dicho pueblo.

La organización de este artículo se presenta a continuación. En la Sección II se describen aspectos del videojuego como la dinámica y mecánica del juego, así como las características del mapa. La Sección III muestra de forma concisa el simulador, ente que gestiona la interacción de los elementos dentro de Einalu'u. Algunos aspectos artísticos relevantes son explicados en detalle en la sección IV. La Sección V contiene la experimentación realizada para demostrar nuestra propuesta. Finalmente, la Sección VI, muestra las ideas finales de la investigación realizada, analizando el proceso de desarrollo.

II. EINALU'U

En lengua Wayuunaiki [5] Einalu'u significa *principio*, con la idea de capturar la esencia exótica del indigenismo, por la suavidad y variedad de sus fonemas. El diseño del videojuego fue creado para que las metáforas y simbolismos plasmados se representen dentro de todo el contexto lúdico. Al emplear conceptos como el origen o el principio, se hace referencia a la población que inicialmente habitó en Venezuela, a la cultura indo-originaria, a los nativos americanos.

Einalu'u consiste en un videojuego de simulación de construcción y gestión para la plataforma PC, centrado en el período histórico neo-indio, justo antes de la llegada de los conquistadores europeos a Venezuela, en la Figura 1 se muestran unas escenas del videojuego. El simulador abarca una gran cantidad de aspectos culturales para que el jugador explore diversas etnias y conozca detalles de sus formaciones socioeconómicas, modos de producción y algunos datos etnográficos como la alimentación, la producción de herramientas, enseres, armas y artesanías, ligadas o no al sector primario. El videojuego está inspirado en juegos de simulación de construcción y gestión, de temática histórica, como la saga Anno [6] de Ubisoft y la serie City Building [7], de Sierra Entertainment.

Así, el videojuego propuesto presenta una simulación donde cada elemento requiere de un tiempo para producir un nuevo elemento. Por ejemplo, para producir una hamaca se requiere de x unidades de fibra de moriche, y personas dedicadas a su fabricación y z minutos en tiempo de producción. Así, existe una función que se ejecuta cada z minutos que verifica que las condiciones estén dadas para la construcción de una hamaca: personas y unidades de moriche. En este caso, es el jugador que decide qué producir basados en los recursos disponibles en un instante de tiempo durante el juego.

A. Dinámica

La dinámica del juego se centra en la gestión de una tribu cuyas opciones viables corresponden a las características culturales de la etnia elegida por el jugador. La tribu seleccionada comenzará con una pequeña cantidad de población, y el jugador deberá alcanzar un hito poblacional mediante una buena administración de los recursos.

Para ello, el juego muestra indicadores de progreso que señalan los avances en la gestión de los recursos así como los logros en el crecimiento de la comunidad. A diferencia de otros juegos del género, no se maneja una moneda de juego ni se hace énfasis en la administración de dinero o en la distribución de las riquezas en diferentes clases, puesto que la vida indígena se mueve más dentro del equilibrio sistémico de las comunidades en relación con el ecosistema. Entonces, se plantea al jugador una serie de dificultades debido a la adquisición de materia prima y alimentos, los medios de producción, la organización social y la distribución del recurso humano como motor principal de la vida tribal.

La gestión de las acciones dentro del juego sucede en tiempo real, bajo una escala que busca representar el transcurrir de los procesos históricos y que considera al tiempo como un costo dentro del balance del juego. El jugador explora diferentes ámbitos culturales de una sociedad indo-originaria venezolana, intentando superar retos de supervivencia y crecimiento poblacional de la comunidad que gestiona. De esta manera, se mantiene al jugador en una constante necesidad de generar mejores estrategias a medida que va descubriendo y conociendo los elementos que forman parte de la vida indígena.

La dualidad existente entre la curiosidad ante lo diferente y el

dominio de este conocimiento, crea en el jugador una conexión y un vínculo que le permitirá sensibilizarse e identificarse con la riqueza cultural de los pueblos indígenas venezolanos.

B. Mecánica

Einalu'u es un videojuego del tipo un solo jugador, enfocado en las personas interesadas en videojuegos de estrategia, de simulación y gestión y juegos con temática histórica, donde los jugadores se sientan motivados ante la idea de planificar estrategias, organizar elementos y administrar recursos. El juego también está dirigido a personas interesadas en la cultura indígena o en la cultura venezolana en general (i.e. motivación intrínseca).

No existe una limitante en cuanto a edad o género para sentir interés en el juego. Sin embargo, dada la dinámica de estrategia y gestión de recursos, la tendencia se inclina a jugadores mayores de 13 años (clasificación PG+13 según ESRB [8]). Es probable que los venezolanos sientan más afinidad por el juego que personas de otras latitudes, dado que el tema central del videojuego es la diversidad cultural de los indígenas de Venezuela.

Al comienzo del juego, el jugador podrá elegir una ubicación geográfica inicial que determina el bioma del lugar, la etnia a la que representará y los retos y posibilidades que ha de tener para alcanzar una meta poblacional. Una vez iniciado el juego, éste podrá explorar el lugar navegando en la escena, construir infraestructuras y asignar personas a diferentes tareas.

La mecánica global del juego consta de una serie de mecánicas atómicas a partir de las cuales el jugador llevará a cabo sus estrategias de juego. Como parte de los comandos disponibles están explorar el área de juego, construir y mejorar infraestructuras, visualizar el estado de cada elemento de simulación, y gestionar las tareas en la población.

C. Mapa de Juego

El videojuego está compuesto por niveles que constituyen un mapa distinto dentro del juego. Un mapa de juego representa un bioma existente en Venezuela. Un bioma se define como un área del planeta que comparte una serie de condiciones ecológicas como clima, fauna y flora. El Fondo Mundial para la Naturaleza (World Wildlife Fund for Nature, WWF) [9] propone tres grandes grupos biomáticos: biomas terrestres, biomas de agua dulce y biomas marinos. Dentro de los biomas terrestres, se distinguen 14 grupos distintos de acuerdo con la biodiversidad, las condiciones climáticas y el relieve topográfico continental.

Por otro lado, el WWF define el concepto de ecorregión como una porción extensa de agua o tierra que contiene un conjunto de especies, comunidades naturales y condiciones ambientales únicas. Las distintas ecorregiones del mundo se agrupan en biomas que son áreas bióticas en una determinada parte del planeta que comparte el clima, flora y fauna. En Venezuela existen 22 ecorregiones, con una amplia representación de biodiversidad dado que existen 7 biomas distintos a lo largo de

su territorio. La importancia de las ecorregiones radica en el interés de los científicos en los ecosistemas y su funcionamiento, para el estudio de la biodiversidad y la conservación de las especies. En esta investigación, se utilizó el criterio de subdivisión en biomas para la selección de especies animales y vegetales que se incluyen dentro de la simulación, de acuerdo con las características biomáticas de cada ecorregión.

Existen siete mapas, cada uno compuesto de una cuadrícula de 70×70 *tiles* donde cada mapa de juego representa a cada uno de los siete tipos de terrenos diferentes, distribuidos tal que puedan representar al bioma. Dentro de cada terreno, se identifican territorios mediante un color como se muestra en la Figura 2.

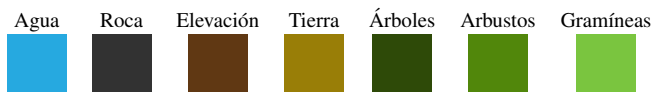


Figura 2: Colores de los Territorios dentro de un Mapa

Nótese que las especies vegetales son agrupadas como árboles, arbustos o gramíneas (herbáceas). A todas las posiciones del esquema lógico correspondiente a alguno de estos tres valores, se le asigna de forma pseudoaleatoria una especie vegetal válida para esa región. Esto determina la cantidad de especies explotables para el jugador dentro de cada escenario, siendo un factor importante en las posibilidades del jugador dentro de su relación ecosistémica con el entorno y la explotación de la materia prima.

Según el área geográfica escogida por el jugador se configura el mapa de juego en función del diseño para el bioma al que pertenece dicha localidad. Los posibles mapas en Einalu'u son:

- **NT01:** Bosques latifoliados húmedos tropicales y subtropicales (ver Figura 3a)
- **NT02:** Bosques latifoliados secos tropicales y subtropicales (ver Figura 3b)
- **NT07:** Pastizales, sabanas, y matorrales tropicales y subtropicales (ver Figura 3c)
- **NT09:** Praderas y sabanas inundadas (ver Figura 3d)
- **NT10:** Pastizales y matorrales montanos (ver Figura 3e)
- **NT13:** Desiertos y matorrales xéricos (ver Figura 3f)
- **NT14:** Manglares (ver Figura 3g)

Una representación visual de cada mapa se puede observar en la Figura 3, donde se organiza como una rejilla o malla de tamaño $n \times n$.

D. Detalles de Implementación

Einalu'u fue desarrollado en Unity (www.unity.com), empleando scripts en el lenguaje de programación C#. Los mapas y el arte fue construidos empleando Photoshop (www.adobe.com/photoshop).

La estructura de un mapa viene dada por una malla de *tiles*, donde la posición es relativa a la fila f y columna c que se encuentre. El mapa 2D es desplegado en vista isométrica, transformando cada coordenada cartesiana (f, c) a su correspondiente en coordenada isométrica [10]. Esta

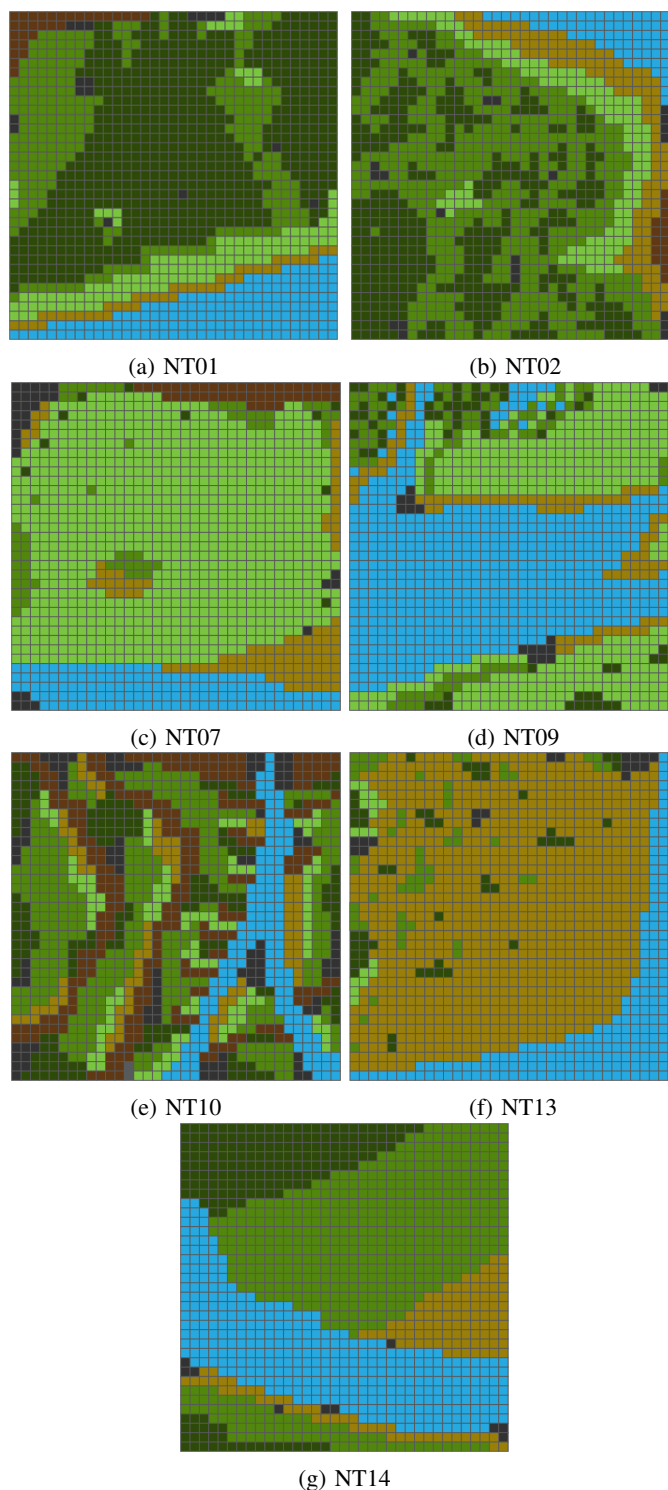


Figura 3: Mapas de los Siete Biomas Diseñados

correspondencia se va realizando continuamente dado que se actualizan las posiciones visualizadas de acuerdo con el desplazamiento del jugador (i.e. en tiempo real).

En el contexto de esta investigación, un *tile* se implementa como un *isometric tile*, parte de un *tilemap* que es una estructura provista por Unity [11]. De esta manera, el diseño de los mapas se realizó en 2D (como se muestra en la Figura 3) y la implementación como *isometric tiles*.

La malla de dimensión 70×70 de cada nivel requiere 4.900 instancias a ser desplegadas. Sin embargo, esto afecta el rendimiento del juego debido a los recursos computacionales requeridos para su despliegue. Asumiendo la posición de una cámara 3D dentro de un nivel, solo se despliegan las *tiles* visibles al usuario. Para todas las posiciones posibles de la cámara, la cantidad máxima de *tiles* visibles es de 1.296 (solo el 26% del total).

En la reducción de la cantidad visibles de *tiles* se emplearon las técnicas de *backface culling* y niveles de detalle de acuerdo con la distancia de la cámara con los objetos 3D presentes en la escena. Por otro lado, la composición de los mapas que se muestran en la Figura 3 indican qué objetos deben ser renderizados. Así, existen mapas de distintos niveles de complejidad basados en el número de polígonos (e.g. mapa NT10, mapa NT02). Esto hace que exista una técnica que pueda manipular a todos por igual.

La técnica empleada para equilibrar los cuadros por segundo en el despliegue consiste en una estructura de datos basada en listas, donde se agregan los índices a los modelos empleados en cada mapa. Así, el despliegue de mapas es eficiente dado que solo se encuentran alojados en memoria los modelos que son utilizados por cada mapa junto con una matrix que almacena las posiciones del jugador dentro del mapa.

III. SIMULADOR

El Simulador es un ente que gestiona cuatro tipos de elementos: 1) la materia prima, 2) los alimentos, 3) los productos manufacturados, y 4) las infraestructuras. Éste controla el tiempo de regeneración de las materias primas a medida que un jugador vaya explotando estos recursos en un mapa y ecorregión en particular. De este modo, un jugador cuenta con una serie de menús y ventanas que le permiten la asignación de personas a la extracción, producción o construcción de cada uno de estos elementos.

A. Elementos

La materia prima puede ser de tipo vegetal, animal o mineral y la gestión de la existencia de cada uno de los elementos de este tipo se calcula a partir de las configuraciones de escenarios. La materia prima es base para el sustento de la población y para la extracción, producción y construcción de otros elementos de simulación.

Los alimentos pueden ser de recolección, caza, pesca, cultivo, ganadería o alimentos preparados o cocinados. La obtención de alimentos puede depender no sólo de otros elementos de materia prima sino también de herramientas, utensilios u otros elementos para su preparación. Cada unidad de alimento puede sustentar una cantidad de personas por un tiempo determinado, los excedentes ayudarán al crecimiento poblacional y la carestía a la merma de los habitantes.

Los elementos de producción pueden ser de tipo alfarería, tejidos, ornamentos, herramientas o armas. Estos elementos suelen ser demandados en la extracción de materia prima, producción de alimentos o construcción de infraestructuras.

Los elementos de producción tienen un tiempo de vida útil y pueden ser confeccionados con diferentes materiales y tecnologías, determinando su efectividad, su durabilidad y su tiempo de producción.

Por último, las infraestructuras pueden ser viviendas, depósitos, centros sociales o infraestructuras relacionadas a la tecnología agrícola. Estos elementos requieren siempre de otros para su construcción y tienen un tiempo de construcción y de mantenimiento (expansión o mejora de la infraestructura), y tienen asociado una capacidad limitada de almacenamiento.

B. Etnias y Crecimiento Poblacional

En Einalu'u se consideraron los siguientes grupos étnicos: Timoto-Cuicas, Caribes Occidentales y de Oriente, Wayúu, Arawacos, Waraos, Llanos, Guayana y los del Norte-Central. Cada una de las etnias tiene una configuración que determina los elementos que puede producir: el tipo de materia prima que extraen, los alimentos que consumen, los productos que manufacturan y las infraestructuras que construyen. Desde la perspectiva del jugador, al iniciar se le indica la población de la etnia seleccionada y la población que debe alcanzar para ganar en esa instancia del juego.

No todos los elementos de simulación que forman parte de la cultura de la etnia están disponibles desde el inicio del juego, algunos son incorporados durante el transcurso de este. Esto se hace con la intención de representar la diversidad de los procesos históricos de cada etnia, dado que el enriquecimiento cultural ocurría de manera diacrónica y no siempre contenía las mismas características.

Con el fin de controlar este enriquecimiento cultural e incorporar nuevas tecnologías en las etnias, se propone un sistema de niveles sustentado en el crecimiento poblacional. La Tabla I presenta cuatro curvas de crecimiento poblacional, con diferentes tasas de crecimiento y diferentes valores de población inicial. Estas curvas están equilibradas para que las variaciones de población en un mismo nivel no sean disímiles, empleando factores de crecimiento de 0.706, 0.681, 0.66 y 0.643 para poblaciones de 20, 25, 30 y 35 respectivamente.

El equilibrio en las curvas de crecimiento poblacional no es equivalente para representar la diacronía cultural de las etnias; sin embargo no difieren demasiado entre sí para hacer de todos los escenarios opciones viables para el jugador. Un jugador puede encontrar diferentes recursos naturales para la extracción de materia prima y la recolección, equilibrados de acuerdo con las características biomáticas de cada ecorregión y a los requerimientos de cada etnia para el aprovechamiento indígena.

La relación costo-beneficio que define el equilibrio en cuanto a las tecnologías que el jugador puede usar, están determinados por la cantidad de elementos requeridos para la producción de dicha tecnología en función de la cantidad de elementos que puede producir. En ese sentido, la elección de una tecnología con respecto a otra por parte del jugador se introduce como un cambio en las reglas del juego para aumentar la efectividad del tiempo de producción (producción más rápida) o la cantidad

Tabla I: Crecimiento Poblacional para Cuatro Valores de Población Inicial en distintos Niveles de Einalu'u

Nivel	Población Inicial			
	20	25	30	35
1	34	42	50	58
2	58	71	83	94
3	99	119	137	155
4	169	200	228	255
5	289	336	378	419
6	493	564	628	688
7	841	948	1042	1131
8	1435	1594	1730	1859
9	2448	2679	2871	3054
10	4177	4504	4767	5017
11	7125	7572	7912	8243
12	12156	12728	13135	13543
13	20738	21395	21804	22251
14	35378	35966	36194	36559
15	60355	60458	60082	60066

de productos generados (mayor cantidad con respecto a lo habitual). Así, la incorporación de nuevas tecnologías es habilitada a medida que el jugador adquiere nuevos niveles culturales en función del crecimiento poblacional.

IV. ASPECTOS ARTÍSTICOS

Uno de los elementos de la cultura indígena que inspiró el diseño del menú de Einalu'u fue los petroglifos. Para ello, se realizó una investigación y documentación en petroglifos y pintura lítica de Venezuela [12]. De acuerdo con Rojas [13], el arte rupestre se refiere al arte hecho sobre piedra. La etimología de la palabra petroglifo, viene de *petro* que significa piedra y *glifo* significa grabado. Además, los petroglifos no son necesariamente una manifestación de arte: pueden ser también registros, notas, símbolos comunicacionales o señalizaciones que los indígenas plasmaron sobre piedras, en caminos transitados o lugares importantes.

En los menús del juego, se emplearon formas que hacen referencia a los petroglifos de los indígenas venezolanos basado en la dualidad cultural de los mismos, como una manifestación artística y como un medio informativo. En la Figura 4 se muestra el menú principal y el logo del juego. El logo fue concebido cumpliendo el doble rol de comunicar una idea a través de una expresión artística, empleando un color plano característico de las pinturas corporales indígenas. Igualmente, el trazo es accidentado como suele ocurrir en los grabados empleando la unión de elementos, los círculos y espirales.

El menú de configuración se observa en Figura 5, donde el jugador puede elegir la ubicación sobre un mapa ilustrado de Venezuela. El panel informativo tendrá al inicio un texto indicando al jugador que debe elegir una localidad. Tanto la imagen de fondo del panel como el marco del menú tienen motivos de petroglifos indígenas venezolanos reales, algunos de ellos presentes en la compilación de Padilla [14]. El fondo



Figura 4: Menú Principal de Einalu'u

del panel utiliza matices claros de azul, que, basado en la teoría del color, transmite serenidad, invitando al jugador a elegir su localidad con calma. El color azul mantiene la sintonía en el menú y no se percibe como un elemento ajeno; y los tonos marrones en los botones y el marco hacen referencia a los colores de la tierra y la naturaleza.



Figura 5: Menú de Configuración para Establecer la Ubicación de una Etnia

Finalmente, cuando un jugador elige un escenario de juego, aparece en el panel la información relativa a dicho escenario: el nombre y la descripción de la etnia, una imagen de fondo de un paisaje natural, y una imagen de un indígena a la derecha.

A. Escenarios

Cada escenario es el resultado de la combinación de una ecorregión con una etnia indígena. En la Figura 6 se destacan dos elementos gráficos que forman parte del menú de configuración, y que comunican visualmente el escenario elegido: un paisaje natural que representa la ecorregión y un indígena cuyas características fenotípicas e indumentaria caracterizan a la etnia seleccionada.

Además, cada paisaje emplea una paleta de color distinta acorde al bioma que representa. Cada paleta corresponde a cada uno de los siete biomas presentes en Venezuela. La paleta se construyó eligiendo un color principal para cada uno de los elementos predominantes en los paisajes (agua y cielo, vegetación, tierra y rocas), y un color para las luces y sombras de cada elemento.



Figura 6: Menú Posterior una vez Seleccionada la Etnia

B. Iconografía

El sistema de ventanas y menús del simulador utiliza una serie de íconos para identificar gráficamente a cada uno sus elementos. Estos íconos son basados en vectores con detalles de sombreado dibujados posteriormente sobre el vector. El arte conceptual de los íconos de botones y menú se basa en los petroglifos como ideogramas los cuales representan ideas más abstractas. La Figura 7 muestra los cuatro botones principales del menú, que corresponden a cada grupo de los elementos de simulación.

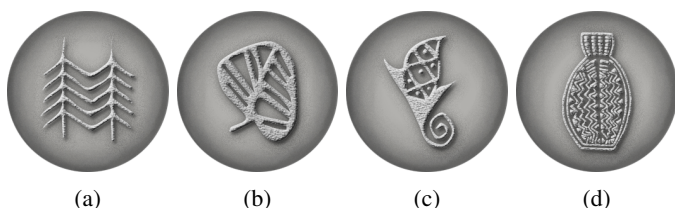


Figura 7: Iconos de (de izq. a der.) Infraestructura, Materia Prima, Alimentos y Producción

Los íconos que corresponden a cada elemento de simulación buscan ser más precisos y menos metafóricos. Esto se hace con la intención de que el jugador pueda reconocer algunos elementos ya conocidos y concebir una idea de cómo lucen aquellos elementos desconocidos.

C. Música

La música incluida son obras indigenistas de compositores venezolanos. La selección de estas obras tiene un doble propósito, agregar un elemento artístico conceptualmente coherente a la temática del juego, y contribuir a la difusión de obras de compositores venezolanos, especialmente las de temática indigenista. La obra principal cuyo nombre es homónimo al juego fue concebida por el compositor Abraham Urdaneta especialmente para el juego, creada digitalmente y sintetizada. La obra combina melodías indígenas interpretadas por distintos tipos de flautas e instrumentos de percusión, acompañadas de varias secuencias rítmicas y sonidos de la naturaleza. Tiene una duración total de once minutos y está compuesta por cuatro *tracks*, los dos últimos fueron transcripciones de cantos waraos.

De acuerdo con el compositor, se utilizan varios sistemas de afinación para las flautas, caracterizados por temperamentos distintos al estandarizado en Occidente (llamado Temperamento Igual) y parámetros de afinación más graves a 440Hz. Cada flauta utiliza un sistema de afinación distinto, produciendo la sensación de desafinación discordante para la música occidental. Esto quiere lograr una sonoridad cercana a la música indígena, donde los criterios de afinación occidentales están ausentes por completo. Las flautas de esta obra utilizan Temperamento Pelog & Slendro, Temperamento Mesotónico y Temperamento Pitagórico [15].

En el proceso de masterización, el compositor agregó a las flautas efectos de reverberación en graduaciones ligeras, un ligero efecto de distorsión para los instrumentos de percusión y otros efectos acústicos que se mezclaron con los sonidos pregrabados de selva. Estos sonidos fueron manipulados para que el canto de los pájaros coincidiera con ciertas secciones de la percusión y las flautas, especialmente en el primer y el tercer *track*: ello llevó al compositor a desplazar los sonidos, subir el volumen e incluso manipular la velocidad de las ondas de los sonidos pregrabados.

Del compositor, también se incluye también la obra *Espíritus del Orinoco*. Esta obra está inspirada en el Río Orinoco, el gran afluente que recorre gran parte de su superficie y atraviesa diversos paisajes del territorio venezolano. *Espíritus del Orinoco* busca evocar el fluir del río a través del timbre de las flautas, cuya relación contrapuntística produce efectos acústicos que emulan el sonido del agua; a la vez, es una pieza que persigue un efecto de misticismo y ritualidad a partir de la repetición de fragmentos pequeños de música, para así captar el espíritu mágico del río.

V. EXPERIMENTACIÓN

Con el fin de analizar los resultados obtenidos de una experiencia lúdica en los jugadores, se diseñó un estudio cualitativo basado en una encuesta que aborda los aspectos más resaltantes del juego. El objetivo es evaluar el impacto que se logra transmitir en cuanto a la difusión de las culturas indígenas. La encuesta está compuesta por las preguntas que se muestran en la Tabla II.

Para el diseño de la encuesta no se utilizó algún modelo preexistente, o alguna fundamentación basada en estudios psicológicos. Se crearon preguntas que permitan abarcar los aspectos del diseño de juego (como parte del proceso creativo) que pueden ofrecer al jugador alguna forma de aprendizaje o la adquisición de conocimientos. Estamos conscientes que existen otras preguntas que están enfocadas a diversas percepciones particulares del juego.

La encuesta fue aplicada a ocho personas, obteniendo diversas respuestas. En función de los resultados se pueden obtener algunas conclusiones:

- El adjetivo para describir el videojuego de todos los encuestados fue: **difícil**.
- El conocimiento de la etnia por parte del jugador era casi nulo.

Tabla II: Conjunto de Preguntas Realizadas como Parte de la Encuestas

Pregunta	Opciones
¿Con qué adjetivo describiría mejor su experiencia de juego?	Divertido, aburrido, fácil, difícil
¿Conocía la etnia con la que jugó?	Nada, poco, algo, mucho, todo
¿Tenía conocimiento de la presencia de la etnia con la que jugó en la región geográfica escogida?	Sí, no, pensaba que habitaban en otro lugar
¿Conocía los recursos naturales y especies vegetales de la región geográfica en donde jugó?	Ninguno, pocos, algunos, muchos, todos
¿Conocía los alimentos que consumían los indígenas de la etnia con que jugó?	Ninguno, pocos, algunos, muchos, todos
¿Conocía los productos que manufacturaban los indígenas de la etnia con la que jugó?	Ninguno, pocos, algunos, muchos, todos
¿Conocía el tipo de viviendas donde habitaban los indígenas de la etnia con que jugó?	Sí, no, pensaba que era otro tipo de vivienda
¿Sabía que los indígenas de la etnia con que jugó construían otro tipo de infraestructuras además de las viviendas?	Sí, no, no aplica
¿Reconoce en la cultura venezolana actual elementos de las culturas indígenas exploradas en el juego, producto de la herencia cultural de esos pueblos?	N/A
¿Considera usted que adquirió un nuevo conocimiento y aprendió acerca de las culturas indígenas?	No aprendí nada, aprendí poco, aprendí mucho
¿Qué impresión tiene acerca de las culturas indígenas venezolanas luego de jugar Einalu'u?	Rechazo, indiferencia, curiosidad, interés,
admiración, otra (indique cuál); (se puede marcar más de una opción)	
¿Logró completar el escenario de juego?	En caso de ser negativa, indique qué le
impidió terminar	
¿Cuántos escenarios jugó? ¿Volvería a jugar? ¿Siente curiosidad por probar otro escenario de juego?	N/A

- Por referencia, los jugadores conocían los recursos naturales, especies vegetales, alimentos y productos de la región geográfica.
- El 75% pensaba que las etnias solamente construían viviendas.
- El nuevo conocimiento adquirido por el juego va desde poco a mucho.
- Las palabras empleadas al momento de capturar la impresión de los jugadores sobre las culturas indígenas venezolanas fueron: **curiosidad, interesante e histórico**.
- Dada la dificultad del juego, la mayoría no pudo culminar el juego (solo 2 jugadores pudieron culminarlo con éxito), sin embargo, exploraron diversos escenarios.

Toda esta información supone para el jugador una oportunidad para la exploración y el aprendizaje. Sin embargo, la densidad del contenido resultó agobiante para los jugadores que tuvieron dificultades para comprender el juego y gestionar correctamente las actividades y alcanzar el objetivo final del juego. Si bien es cierto que la presencia de elementos desconocidos en el juego despierta la curiosidad del jugador, la gran cantidad de dichos elementos dificulta la capacidad de relacionarlos entre sí. Esta relación es clave dado que permite crear estrategias de juego para desarrollar una buena gestión de las actividades de extracción y producción.

De hecho, el videojuego permite mostrar más información opcional asociada a cada etnia: descripción de cada una, zonas

que habitaron, recursos que había en su medio y el contexto adecuado para la comprensión del jugador. Sin embargo, esta información forma parte de la Ayuda al jugador, por lo cual no esta obligado a revisar dicha información considerada opcional.

Para minimizar este resultado, los elementos de simulación que forman parte de la cultura de una etnia son introducidos paulatinamente conforme el jugador aumenta su población y sube de nivel, en lugar de ser presentados desde el inicio de la partida. Los elementos de simulación cuentan con un ícono que permite distinguirlos e identificarlos gráficamente, facilitando así el proceso de asociación.

VI. DISCUSIÓN

La investigación subyacente para Einalu'u abarca aspectos etnográficos de gran parte de las etnias que hicieron vida en Venezuela antes de la llegada de los europeos. Entre los aspectos investigados más significativos se resaltan: la distribución geográfica de la etnia, el tipo de alimentos que consumían, sus hábitos de caza, pesca, recolección y agricultura, los productos que manufacturaban. Igualmente, el tipo de infraestructuras que construían y cómo se distribuían sus habitantes en ellas, la cantidad de población promedio en las aldeas, la fisionomía, costumbres al vestir y características generales del *modus vivendi* de cada pueblo. Todos estos elementos fueron adaptados a un diseño de un videojuego del género de simulación de construcción y gestión.

Para una partida del videojuego, el jugador elige un escenario, producto de la combinación de una etnia y una ecorregión seleccionada, y en cada escenario el jugador debe alcanzar una cantidad determinada de población para poder ganar la partida. Las diferentes configuraciones de ecorregión ofrecen al jugador la posibilidad de identificar características biomáticas de su selección. El escenario presentado dentro del mapa de juego muestra una gran diversidad de especies vegetales que el jugador, basado en las encuestas, pudo identificar y reconocer en algunos casos.

El simulador, ente capaz de gestionar diferentes escenarios de juego, permite la construcción de una arquitectura modular que aprovecha el enfoque basado en componentes que utiliza Unity, lo cual favorece la construcción de una arquitectura de juego robusta. Con el fin de reconocer los elementos de la simulación relativos a cada cultura, cada etnia es representada por un dibujo de un indígena ataviado con sus ornamentos típicos y las ecorregiones son representadas por un paisaje característico del bioma. Todos los elementos de simulación tienen un ícono que los identifica, y las infraestructuras se construyen sobre el mapa de juego. Todos los mapas tienen un arte 2D dimétrico que son representados sobre un sistema de *tiles*.

Con el fin de complementar la experiencia lúdica y dar más contexto indígena al juego, se utilizaron obras indigenistas únicas de compositores venezolanos como banda sonora, bajo previa solicitud de permisos y acuerdos de derechos de autor. También se introdujeron animaciones en forma de tutorial, para introducir a los jugadores a dar sus primeros pasos con más facilidad, y guiarlos a medida que van avanzando dentro de la simulación.

Es posible deducir un interés en la exploración de culturas desconocidas que puede tener un jugador al abordar un juego de este género, sin embargo, muchos jugadores encontraron el juego difícil de abordar. Aspectos como la experiencia en el tipo de juego, donde la paciencia y el tiempo para generar objetivos a mediano o largo plazo son destacables, podría perjudicar una calificación general de Einalu'u. Sumado a ello, algunos elementos de arte podrían perjudicar el equilibrio de juego o su fluidez, pero para ello se requiere de un proceso posterior como jornadas de *playtesting* para perfeccionar la configuración de los parámetros de juego.

Para reducir el impacto del denso contenido etnográfico, se plantea agregar descripciones a los elementos de la simulación para complementar el proceso de aprendizaje dentro del juego. Igualmente, se considera introducir más información de los tiempos de producción, las cantidades de elementos producidos y de personas requeridas, el nivel donde se habilita cada elemento y la curva de crecimiento poblacional.

Con las recomendaciones mencionadas se espera mejorar la fluidez del juego, acelerar su proceso de aprendizaje y de comprensión de la dinámica, potenciar en el jugador la capacidad de crear estrategias viables y hacer aún más coherente y enfático el contenido etnográfico que se espera comunicar a los jugadores.

Como parte del proceso de liberación al mercado de un videojuego, contar con una población diversa de jugadores para obtener un primera retroalimentación, es un punto clave en el éxito de un videojuego. Al mismo tiempo, contar con una cantidad aproximada de entre 30 a 50 personas nos daría un mayor grado de confianza en nuestro resultado. Sin embargo, no dudamos el alcance efectivo y eficaz de Einalu'u como un videojuego de simulación para construcción y gestión centrado en las culturas indígenas prehispánicas de Venezuela.

AGRADECIMIENTOS

Los autores quieren agradecer al coro de cámara de voces claras Aequalis Aurea de la Fundación Aequalis. Del mismo modo, queremos agradecer a Ana María Raga y Verónica Sosa por la grabación del audio empleado en Einalu'u.

REFERENCIAS

- [1] Instituto Nacional de Estadística, *Resultados Población Indígena. Censo 2011*, <http://www.ine.gov.ve/documentos/Demografia/CensodePoblacionyVivienda/pdf/ResultadosBasicos>, 2011.
- [2] *El Canto del Autana*, https://es.wikipedia.org/wiki/El_canto_del_Autana.
- [3] *Prensa TeleSur, Venezolanos Crearon Videojuego sobre Simón Bolívar*, <http://goo.gl/XLL4Q0>, 2014.
- [4] L. Lima, D. Torres and E. Ramírez, *Un Juego Serio para la Preservación de la Fauna Silvestre en Peligro de Extinción en Venezuela*, 3ra Conferencia Nacional de Computación, Informática y Sistemas (CoNCISA 2015), pp. 50-60, 2015.
- [5] D. M. Captain y L. Captain, *Diccionario Básico Ilustrado; Wayuunaiki-Español; Español-Wayuunaiki*, 1st ed., Editorial Fundación para el Desarrollo de los Pueblos Marginado, 2005.
- [6] *Anno 1800, Ubisoft*, <https://www.ubisoft.com/en-us/game/anno-1800>.
- [7] *City Building Series, Sierra Entertainment*, [https://en.wikipedia.org/wiki/City_Building_\(series\)](https://en.wikipedia.org/wiki/City_Building_(series)).
- [8] Entertainment Software Association, *ESRB Ratings*, http://www.esrb.org/ratings/ratings_guide.jsp, 2019.
- [9] Ecoregions, *World Wildlife Fund*, <https://www.worldwildlife.org/biomes>.
- [10] J. Bose, *Creating Isometric Worlds: A Primer for Game Developers*, <https://tinyurl.com/y5v1cjhs>, 27 May 2013.
- [11] Unity, *Tilemap Documentation*, <https://docs.unity3d.com/Manual/class-Tilemap.html>, Febrero 2020.
- [12] S. Cadenas and A. J. Rojas, *Clase Abierta de Petroglifos y Pintura Lítica*, USB Sede Litoral, 2017.
- [13] A. Rojas, *Petroglifo Carmen de Uria: Una Lectura Etnográfica*, El Tarmaño, 2014.
- [14] S. Padilla, *De los Petroglifos y otras Expresiones Primitivas de América*, Leopoldo Kripacz Koch, <https://books.google.es/books?id=q1ezPgAACAAJ>, 1957.
- [15] J. Goldáraz Gainza, *Afinación y Temperamentos Históricos*, Grupo Anaya Publicaciones Generales: Alianza Música, 2004.

Manual-design of Blocks: Una Herramienta para Gestionar Segmentaciones Manuales de Páginas Web

Jean Garcia¹, Andres Sanoja¹
jeangarcialunardi95@gmail.com, andres.sanoja@ciens.ucv.ve

¹ Escuela de Computación, Universidad Central de Venezuela, Caracas, Venezuela

Resumen: La segmentación es una parte importante en el análisis de páginas Web. El objetivo es dividir una página en bloques, cada uno representando una parte (o segmento) coherente del contenido. En el presente trabajo describimos el desarrollo de la herramienta Manual-design of Blocks (MoB). Se incluyen detalles de la investigación, donde se comprueba la importancia de MoB en la evaluación de algoritmos de segmentación. El objetivo de nuestro trabajo es el desarrollo de la herramienta MoB. Al mismo tiempo describir los mecanismos para la obtención de una “base referencial” de segmentaciones manuales sobre una misma página y para la posterior obtención de “la mejor segmentación manual”. La mejor segmentación manual es definida en base a nuestra experiencia y los datos obtenidos, en este trabajo se define una forma de su obtención pero no se considera que sea la única forma de obtener. La mejor segmentación obtenida queda disponible para la evaluación de un algoritmo de segmentación usando el framework Block-o-Matic. Mediante el uso de MoB se soporta el proceso de conformación de la base referencial, apoyado con una interfaz de programación de aplicaciones Web (API) y un repositorio para su manejo y consulta. Se presentan resultados de las pruebas de aceptación.

Palabras Clave: Segmentación; Página web; Segmentación Manual; Bom; Block-o-matic.

Abstract: Web page segmentation is an important task in Web page analysis. The objective is to divide a Web page into blocks, each one representing a coherent part (or segment) of the content. In this work we describe the development of the Manual-design of Blocks (MoB). At the same time we describe how to get a ground truth of segmentations and how to compute the “best manual segmentation”. The best manual segmentation is defined based on our experience and the data obtained, in this investigation we define one way to obtain it, but we do not consider there’s only one way to achieve this. The best segmentation is then available to be used on the evaluation process of segmentation algorithm using the Block-o-Matic framework. Also, a Web API and a Web repository for managing the data. Acceptance test results are presented in this document.

Keywords: Segmentation; Web Page; Manual Segmentation; Bom; Block-o-matic.

I. INTRODUCCIÓN

La página Web es un documento digital de información accesible mediante un navegador de Internet. Esta información se presenta generalmente en formato HTML. Está compuesta por un conjunto de elementos ordenados en una estructura de árbol (el árbol DOM), generado por el navegador a partir del código fuente HTML [1]. En el caso del presente trabajo, la segmentación de una página Web es la acción de dividir una página Web en fragmentos coherentes (*i.e* cada fragmento debe tener un sentido para un usuario) llamados bloques [2]. Cada bloque representa distintos elementos de información en la página. Un algoritmo de segmentación define las reglas para la selección de dichos segmentos. En general, estas reglas se

definen en base al árbol DOM y las pistas visuales, *e.g.* tamaño de la letra, separación, espaciado o colores. Otros enfoques se definen en base a *screenshots* o el texto de una página. La segmentación puede ser aplicada en diferentes áreas como por ejemplo:

- **Procesos de SEO o Search Engine Optimization.** La segmentación de la página Web permite realizar un análisis del contenido de la página para que esta pueda ser calificada y ubicada en un *ranking*.
- **Migración de Formatos.** La segmentación de la página Web puede ser utilizada para ayudar en el proceso de conversión del formato de un documento a otro diferente. Por ejemplo, se puede migrar una página Web de la

versión del lenguaje HTML4 a HTML5. Se convierten los bloques de la página en HTML4 a elementos semánticos, los cuales se utilizan para la creación de la nueva versión de la página usando el lenguaje HTML5 (e.g. section, article).

- **Archivado de la Web o Web Archiving.** La segmentación permite comparar dos versiones de la misma página Web, la versión que actualmente se tiene almacenada y la versión que se planea almacenar. El encontrar las diferencias entre ellas permite detectar si resulta eficiente descargar y almacenar la nueva versión y sus dependencias.
- **Bloqueo de Contenido o Content Blocking.** La segmentación de la página Web permite identificar los segmentos dentro de la página que poseen contenido no deseado para ciertas audiencias. De esta manera bloquear solamente una porción de la página sin afectarla como a un todo.

Es importante considerar las necesidades de cada área, dependiendo de si se desea un algoritmo de segmentación genérico o un algoritmo de segmentación específico. Se entiende que mientras más genérico sea, podrá ser usado en mayor cantidad de páginas Web pero con menos precisión. Para poder precisar esto el algoritmo de segmentación debe ser evaluado [3]. Esta evaluación se enfoca en comparar la segmentación obtenida con un algoritmo contra una segmentación denominada como “la mejor segmentación manual”. Así mismo con la conformación de una “base referencial”. En ella residen las páginas previamente segmentadas y anotadas. Se utiliza para apoyar tanto la evaluación como la construcción de la mejor segmentación manual. Definir un método de evaluación formal, en el alcance de nuestro conocimiento, no es práctico en base a la naturaleza de los estándares (ie reglas heurísticas) W3C. Consideramos una segmentación manual como la mejor opción para este caso. Obviamente, esta técnica no sustituye un enfoque formal, ya que puede resultar subjetivo dependiendo de la visualización que posea el usuario [4]. Buscamos la mejor segmentación manual, la cual es una composición basada en el consenso entre las diferentes segmentaciones realizadas por varios usuarios sobre una misma página Web y con la misma granularidad. El principio es que un bloque estará en la mejor segmentación si el área de la página ha sido seleccionada por la mayoría de los usuarios como un bloque. Ambas segmentaciones deben estar ajustadas a una misma granularidad, la cual define el tamaño general de los bloques dentro de una segmentación. Segmentar una página Web manualmente puede ser una tarea compleja y propensa a errores. Un usuario debe tener conocimientos avanzados del lenguaje HTML y sus dependencias.

Contribución: El presente trabajo tiene como objetivo ofrecer una solución para la obtención de la mejor segmentación manual mediante el desarrollo de una herramienta que permite a cualquier usuario realizar segmentaciones manuales de páginas Web. La solución incluye tres componentes: una extensión del navegador para la segmentación manual (MoB extension), una aplicación Web como *endpoint* que recibe las segmen-

taciones manuales, construye la mejor segmentación manual y gestiona los datos (MoB API). Finalmente, un repositorio como archivo Web (MoB Repository). Están disponible MoB Repository y descargar la extensión de MoB desde el sitio Web¹ o descargar el código fuente desde Github².

Organización: En la Sección II presentamos los antecedentes. En la Sección III describimos la herramienta de segmentación manual y consideraciones sobre la granularidad. En la Sección IV presentamos la API, el repositorio y la propuesta de como obtener la mejor segmentación manual. En la Sección V describimos la verificación funcional de la herramienta. En la Sección VI presentamos las conclusiones y trabajos a futuro.

II. ANTECEDENTES

A diferencia de su contraparte en el área de Computación Gráfica, las herramientas de segmentación manual para la Web no han sido muy populares. Lo que si se encuentra, y es mas común, es realizar anotaciones dentro del código HTML. En principio esta técnica nos permitiría indicar cuales elementos pertenecen a un bloque y asignarle una etiqueta. Sin embargo, es una actividad tediosa, larga y propensa a errores. Requiere conocimientos avanzado de HTML y sobre navegadores Web. Algunas soluciones utilizan la segmentación manual como anotaciones y son muy enfocadas en medir las métricas de precisión y recall, por ejemplo en [5][6]. Otros se enfocan en considerar bloques como nodos de un grafo para así anotar páginas con pesos para aplicar métricas de clustering, como es el caso de [7][8].

Es de nuestro interés contar con un mecanismo que permita de manera genérica, segmentar manualmente cualquier página para ser utilizado en un proceso de evaluación y de manera gráfica. Kiesel *et al.* (2017) [9] presentan su versión para anotación de etiquetas de segmentos. Dada una segmentación (en este caso los bloques son oraciones en un texto), un experto aporta las etiquetas a cada uno de los segmentos especificados. Aun cuando en su trabajo se realiza una segmentación de texto (realizada sobre el texto de una página directamente), la herramienta WAT-SL, desarrollada a lo largo de su investigación, es en esencia similar a a la herramienta MoB, ya que también es usada para segmentar y anotar dichos segmentos con etiquetas. Contiene apoyo gráfico y no está restringida a un conjunto de datos de prueba en particular.

Otros investigadores han trabajado también en la evaluación de segmentaciones de páginas Web, construyendo segmentaciones manuales para probar los evaluadores [10][11]. Estos últimos generaron *datasets* para pruebas de sus tesis doctorales, que luego quedaron disponibles en línea. Utilizan elementos de CSS para denotar un bloque, mediante el uso de clases CSS. Aun cuando es un enfoque interesante, deja de lado los aspectos geométricos inherentes de los elementos de una página Web. Siendo estos un elemento clave, común en

¹<https://mob.ciens.ucv.ve>

²<https://github.com/JeanGarcia/MoB>

todos los algoritmos de segmentación y así poder compararlos. Nuestro trabajo forma parte de la evaluación de la segmentación de páginas Web. Fue descrito como un trabajo futuro propuesto por Sanoja y Gançarski (2015) [2][3]. En el trabajo se presenta el uso del framework de segmentación de páginas Web Block-o-Matic (BoM), para migrar documentos en formato HTML4 a formato HTML5 para evitar la obsolescencia en el contexto de archivos Web. BoM es la herramienta basada en el algoritmo homónimo [12], para la segmentación de una página Web de forma automática. *MoB-prototype* (Prototipo de Manual-design of Blocks) es la herramienta de segmentación manual, cuya versión debe ser mejorada. Se utilizó como un prototipo de baja fidelidad implementado como una extensión para el navegador Chrome, llevando el proceso de manera semi-automática. Esta herramienta está orientada a usuarios expertos.

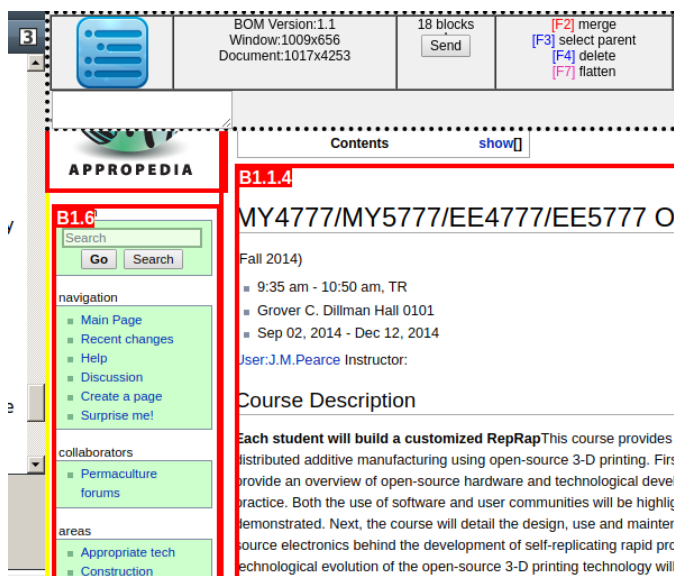


Figura 1: Versión Prototipo de *MoB*

En la Figura 1 se puede observar un ejemplo de como el usuario crea bloques dependiendo de los elementos del DOM. Se obtiene un grafo de bloques (respetando la jerarquía en el DOM) o simplemente una segmentación plana (sólo bloques terminales). Ambas segmentaciones se materializan en un documento XML. Produce también un conjunto de rectángulos presentados de manera visual.

En la Figura 1 se observa que *MoB-prototype* posee un panel en la parte superior. En el panel se muestra la leyenda de los comandos a usar para realizar la segmentación manual. Se le presenta al usuario una segmentación realizada previamente por BoM. Se le propone al usuario aceptar o modificar la segmentación propuesta. Si el usuario desea modificar la segmentación, por ejemplo agregar un bloque, debe presionar F9 y hacer clic en el elemento que desea segmentar. Mediante el uso de las flechas del teclado el usuario puede navegar la jerarquía del DOM.

Una vez se hayan completado los pasos anteriormente men-

cionados, y el usuario escoja el elemento DOM que desea segmentar, se resaltaré dicho elemento indicando que se ha segmentado. Este proceso resulta tedioso, complicado y propenso a errores, incluso para usuarios expertos. Es recomendable realizar una mejora en la usabilidad de *MoB-prototype*. Por ejemplo, resaltar tentativamente el elemento que se desea segmentar permitiéndole al usuario una retroalimentación activa y se crea también un sistema de puntajes, el cual motiva a los usuarios a realizar segmentaciones, entre otros aspectos a considerar. Proponer una segmentación usando BoM al inicio puede no ser justo con otros algoritmos, por lo que se recomienda no utilizarlo.

En las siguientes secciones se presentaran las mejoras realizadas a *MoB-prototype* en su nueva versión.

III. EXTENSIÓN MOB

La extensión de MoB es la herramienta de segmentación manual desarrollada por nosotros³. Se utiliza Javascript, HTML5 y CSS3, como una extensión para navegadores Chrome y Chromium.

A. Funcionalidades de la Extensión MoB

- **Consultar información:** sección de información donde el usuario puede aprender más de la herramienta y sus diferentes acciones.
- **Cambiar idioma:** permite al usuario cambiar el idioma de la interfaz escogiendo entre: inglés, francés o español.
- **Consultar puntuaciones:** permite a un usuario autenticado en el sistema consultar sus puntuaciones personales sobre una determinada página Web o las puntuaciones globales.
- **Registro:** permite al usuario registrarse dentro del sistema MoB.
- **Autenticación:** una vez el usuario está registrado, el sistema le permite autenticarse mediante un inicio de sesión.
- **Segmentación manual:** el sistema le permite al usuario identificado inicializar la herramienta de segmentación manual.

B. Acciones de la Herramienta de Segmentación

A continuación describimos las acciones que ofrece la extensión MoB. Al inicializar la herramienta de segmentación manual, aparecerá un menú con las siguientes acciones:

- **Agregar Nuevo Bloque:** agregar un nuevo bloque a la segmentación. Al ser seleccionada, el usuario puede recorrer los elementos del DOM con el ratón y estos serán iluminados. Una vez el usuario haga clic sobre alguno de estos elementos del DOM, se insertará en el HTML un rectángulo representando el bloque segmentado. Este rectángulo posee entre sus atributos los datos del bloque segmentado (etiqueta, ancho, alto, posición horizontal,

³Disponible en <https://mob.ciens.ucv.ve>

posición vertical y el área). Estos datos son almacenados dentro de un arreglo en Javascript. Este arreglo es utilizado para realizar las comparaciones necesarias entre los diferentes bloques para aplicar las acciones. Al realizar esta acción se activa un subproceso que verifica si hay otros bloques dentro del que está a punto de crearse. De ser así se eliminan dichos bloques y se conserva únicamente el nuevo que ha sido creado. En caso de que existan bloques que se intersecten parcialmente con otros bloques, es necesario la acción de “cortar”, *i.e.* ajustar los rectángulos.

- **Eliminar Bloque:** una vez activada, se iluminarán aquellos bloques al que el usuario señale con el ratón. Al hacer clic sobre alguno de estos bloques se disparará una función de Javascript la cual removerá el elemento DOM que representa el rectángulo del bloque. Será removido del arreglo de bloques.
- **Unir Bloques:** al ser seleccionada, esta acción permite al usuario seleccionar dos bloques los cuales se unirán. Esta acción elimina los dos bloques seleccionados y crea uno nuevo que comparte los límites superiores e inferiores máximos de los bloques anteriores. Después se comprueba que no cubra otros bloques dentro del nuevo, en cuyo caso se eliminarían los internos.
- **Cortar Bloques:** esta acción permite seleccionar dos bloques que se solapan (A y B) y realizar un corte entre los dos. El orden de selección importa, pues A será el bloque que predominará (se mantendrá intacto) y B será el bloque que se recortará y ajustará.
- **Etiquetar Bloque:** al hacer clic sobre un bloque se mostrará una modal con una lista donde el usuario podrá escoger la etiqueta que mejor se adapte al bloque.
- **Seleccionar Bloque:** esta acción permite al usuario seleccionar cualquier bloque y obtener una ventana de información con los datos de dicho bloque.
- **Panel de Información:** esta acción despliega un panel informativo con metadatos de la segmentación para el usuario. Ofrece la opción de cambiar la granularidad de la segmentación y muestra todas las alertas que puede presentar la segmentación.
- **Enviar Segmentación:** esta acción activa los procesos necesarios para la recolección de datos y envío hacia la API de MoB. Comienza por comprobar el estado de la segmentación en busca de errores o advertencia, en caso de presentar errores, la segmentación no será enviada y los procesos de recolección de datos se cancelan. En caso contrario, se crea una estructura de JSON para enviar todos los datos necesarios al API. Se almacena el HTML renderizado de la página en un string, esto vendría siendo el HTML de la página versión MoB. A la vez se hace una búsqueda entre los elementos DOM de la página para identificar cuales son los elementos DOM que están presente en cada bloque segmentado, esta información se almacena en un arreglo junto con los bloques de la segmentación. También se incluyen otros datos como el título de la página, la dirección URL, la categoría,

colección y dimensiones.

C. Ajustando Bloques a la Granularidad

Para poder realizar la evaluación se requiere de la mejor segmentación manual y la segmentación realizada por el algoritmo, ambas con una misma granularidad. Se puede ver el nivel de granularidad como el nivel de detalle de la segmentación, a menor granularidad mayor nivel de detalle. Se presenta a continuación un análisis que se desprende del descrito en [2], pero considerando aspectos para su implementación.

1) Consideraciones Sobre la Granularidad:

- La granularidad se escribe como la relación existente entre el área de los bloques y el área total del documento.
- Se considera la cantidad de bloques en la segmentación. A menor número de bloques, mayor granularidad.
- La granularidad puede variar de un documento a otro en términos absolutos. Debe mantenerse la relación en términos relativos. Se debe acotar la cantidad de bloques relativo al parámetro de granularidad.

2) *Calculando la Granularidad:* La granularidad se considera como un valor G que se lee “Granularidad del documento a un nivel G ”. G es un parámetro de la segmentación. G toma valores enteros entre 0 y 10 (incluyendo el 0 y e 10). Para poder garantizar una cantidad máxima de bloques segmentados con respecto a la granularidad seleccionada, se consideran los casos descritos en la Tabla I.

Tabla I: Cantidades Máximas de Bloques

G	Cantidad Máxima de Bloques
0	Se puede tener tantos Bloques como elementos tenga el DOM.
1	40
2	36
3	31
4	27
5	22
6	18
7	13
8	8
9	3 bloques (eg. header,content, footer)
10	Se tiene 1 solo Bloque que coincide con el documento

Se escribe $G(b)$ como el valor de granularidad de un bloque b cualquiera. En aras de simplificar la notación, se asume el valor de granularidad del documento, $G(document)$, como simplemente G .

Esto representa el número máximo de segmentos en los que se puede dividir el documento, lo que sigue es considerar un análisis similar para el área. Se considera el área mínima dado un valor de G . Pueden existir menos bloques de los esperados, pero no más. Se establece una correspondencia entre el área de un bloque y su valor de granularidad. Así un bloque b con granularidad $G(b)$ debe tener su área dentro de un rango de valores predefinido. En la fórmula (1) se muestra como seleccionar el área mínima que debe tener cada bloque en un determinado nivel de granularidad. $area(b_i)$ representa el área del rectángulo asociado a un bloque y $area(doc)$ representa el área del rectángulo del documento.

$$area(b_i) \geq \frac{area(doc)}{G - 1} \quad (1)$$

En una situación ideal la segmentación está constituida con bloques que posean todos la misma granularidad. Sin embargo, ese no siempre es el caso. Dado que esto puede resultar muy restrictivo, se decide dar un rango de tolerancia de una unidad (1) en el valor G. En otras palabras, el bloque b es aceptado si $G - 1 \leq G(b) \leq G + 1$.

D. Diferencias Entre el Prototipo y la Nueva Versión

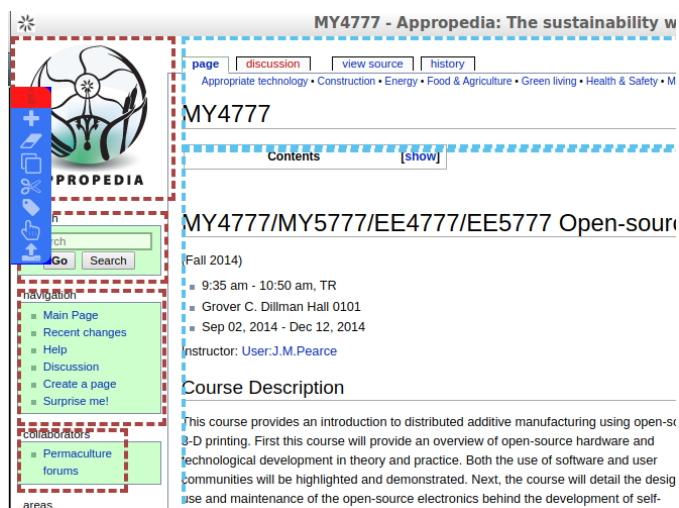


Figura 2: Nueva Versión de MoB

Como se mencionó anteriormente en la Sección II, en el trabajo de Sanoja y Gancarski (2015) [2][3] se presenta una versión prototipo de la herramienta de segmentación manual la cual se busca mejorar en este presente trabajo. En la Figura 2 se puede observar que la nueva versión de la herramienta posee las siguientes características a diferencia de la antigua versión:

- El color de los bloques refleja el estado de los mismos y no el nivel de la segmentación, pues la segmentación es plana. En este ejemplo azul significa que está correcto y rojo que necesita atención. Se considera correcto verificando la granularidad, si tiene etiqueta, si no hay superposición de bloques, entre otras consideraciones menores.
- Se restringe al usuario a realizar una segmentación plana, lo cual se considera deseable, a diferencia de la versión anterior la cual es propensa a segmentar en múltiples niveles.
- El panel de información ofrece mayor información sobre la granularidad presente y los bloques, además de los posibles errores o advertencias que puedan ocurrir.
- La caja de herramienta ocupa menos espacio al estar conformada únicamente de metáforas de las acciones, también es semi-transparente para poder observar el contenido que existe detrás.

- Todas las acciones son llevadas a cabo únicamente con el ratón, sin tener que usar el teclado.
- Se ofrece la acción de “cortar” para poder separar bloques que se intersectan, y la acción “seleccionar” que permite obtener toda la información de un bloque en específico.

En la Figura 3 se pueden observar los elementos que conforman el prototipo anterior. Se concibió como una extensión simple de Chrome. Un usuario interactúa con la página modificada por la extensión y puede obtener los resultados como un archivo separado por comas (CSV). Listando datos básicos, por ejemplo el *id* del bloque, los puntos del rectángulo o la cantidad de palabras contenidas en el bloque. En la

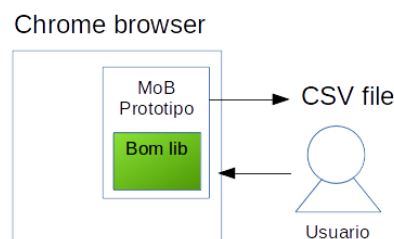


Figura 3: Elementos del Prototipo Anterior de MoB

Figura 4 se observa los elementos de la nueva versión. Aquí se plantea la segmentación manual como una solución. Se consideran varios usuarios y elementos de referencia como la *ground truth*. La extensión de Chrome se replantea, tal como se describe en este documento. Se incluyen elementos fuera del alcance de este trabajo, solo deben considerarse los elementos delimitados. Se consideran los datos básicos de la versión anterior, más una lista de imágenes contenidas, el texto, así como un *hash* de todo el contenido (c.f Sección C).

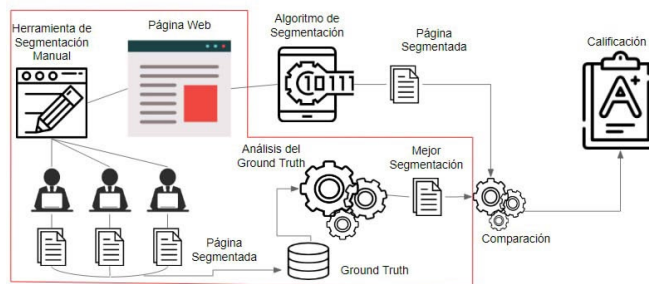


Figura 4: Elementos de la Nueva Versión de MoB

IV. MOB API Y MOB REPOSITORY

El MoB API y el MoB Repository están estrechamente relacionados, ya que, el MoB API no sólo ofrece servicios RESTful sino también actúa como backend del MoB Repository. A continuación se hace una descripción de las características más resaltantes de ambos y su finalidad en el sistema.

A. MoB API

El desarrollo de la API se realizó con el lenguaje de programación Python v.3.5, apoyado con el microframework Flask

v.0.12.2. Para la creación de la base de datos que va conectada a la API se utilizó el manejador de base de datos PostgreSQL v.10.1 junto con un componente llamado Postgis 2.4 para realizar las operaciones entre tablas.

En general, la API de MoB se divide en dos partes: todos los servicios RESTful que pueden ser ofrecidos a la Extensión MoB o a terceros y todas aquellas funciones que manejan el backend del Repositorio MoB. A continuación se describen los servicios RESTful que son ofrecidos por la API, mientras que las funciones que manejan el backend del Repositorio de MoB se explicarán en la siguiente subsección. Los servicios ofrecidos son los siguientes:

- **Registrar usuario:** Permite registrar a un usuario en el sistema. Para completar el registro se le envía al usuario un enlace de activación a su correo.
- **Iniciar sesión:** Permite al usuario registrado (y activado) iniciar sesión en el sistema para hacer uso de sus funcionalidades.
- **Cerrar sesión:** Borra las *cookies* de sesión existentes en el navegador y la sesión existente en el API.
- **Recuperar contraseña:** Permite al usuario recuperar su contraseña en caso de extravío. El sistema envía una combinación aleatoria de caracteres como contraseña temporal. Por medidas de seguridad las contraseñas se encuentran encriptadas por *hash* en la base de datos.
- **Obtener colecciones:** Permite obtener una lista con los nombres de las colecciones y categorías de estas existentes en la base de datos del sistema.
- **Obtener etiquetas:** Permite obtener una lista con los nombres de las etiquetas existentes en la base de datos del sistema.
- **Obtener puntajes globales:** Permite obtener una lista con los mejores puntajes en cada una de las granularidades de una página Web específica.
- **Obtener puntajes del usuario:** Permite obtener una lista con los puntajes en cada una de las granularidades de una página Web específica para un usuario determinado.
- **Cargar segmentación:** Este es uno de los servicios más importante del API pues representa la base de todo el sistema. Permite cargar los resultados de una segmentación a la base de datos y los datos de la página Web, en caso de que sea la primera vez que se segmenta. Formando así lo que denominamos anteriormente como base de la verdad.
- **Vista previa de segmentación:** Devuelve un *canvas* con las figuras y etiquetas de los bloques segmentados para una segmentación en específica.
- **Obtener segmentación en formato JSON:** Retorna un JSON con todos los datos de una segmentación en específica.
- **Obtener segmentación en formato V-PRIMA:** Retorna todos los datos de una segmentación en formato V-PRIMA. El formato V-PRIMA es un XML donde se especifican los bloques existentes en la segmentación y los enlaces, imágenes y textos existentes dentro de estos.

Es una adaptación del formato V-DIFF a nuestro contexto [13].

- **Obtener segmentación en formato MoB HTML:** Retorna el HTML original con la información de la segmentación incluida. Se usan *tags* personalizados, eg. `< block >`.
- **Obtener página Web en formato WARC:** Devuelve la información de una página Web en formato WARC (Web ARChive). El formato WARC permite la concatenación de múltiples objetos de datos o recursos en un sólo archivo. Es utilizado para almacenar la información de páginas Web junto con sus recursos y metadata.

B. Mob Repository

El sitio Web MoB Repository se desarrolló haciendo uso del lenguaje de marcado HTML5, CSS3 y el lenguaje de programación Javascript. Usamos también el *framework JQuery*, principalmente para el comportamiento de las páginas y control de eventos. Todo esto del lado del cliente. Del lado del servidor está apoyado por el MoB API (Python y Flask) y la base de datos conectada a este (PostgreSQL y Postgis). La finalidad del MoB Repository es ofrecerle a los usuarios del sistema una interfaz para que puedan visualizar las colecciones de segmentaciones manuales almacenadas en la base de datos. Ver las mejores segmentaciones manuales de cada página Web segmentada. A los usuarios administradores les permite administrar las etiquetas utilizadas en la segmentación, las colecciones y sus categorías, así como los roles de otros usuarios.

C. Creación de la Mejor Segmentación

La mejor segmentación manual es aquella que contiene los bloques más populares en un conjunto de segmentaciones manuales. Si la mayoría de usuarios han marcado un área de la página como un bloque, este se considerará popular. Entonces tomamos los bloques más populares, creamos una nueva segmentación que los contenga, esta será considerada la mejor.

Al llevar a cabo la comparación, los bloques recibirán una serie de puntajes dependiendo de sus similitudes geométricas, con respecto a los otros bloques. Los puntajes de todos los bloques correspondientes a una misma segmentación podrán ser sumados dando la puntuación total de la segmentación. Esto quiere decir que mientras mayor sea el puntaje obtenido, mayor es el número de segmentaciones que la respaldan. Varios de estos atributos se nombran durante el proceso de comparación, por lo que aquí se presenta una leyenda de los mismos:

- **Identificador:** Un identificador arbitrario del bloque.
- **Geometría:** La geometría del bloque (área, ubicación y coordenadas del rectángulo asociado)
- **Puntaje Geométrico:** La puntuación obtenida por similitudes geométricas.
- **Etiqueta:** La etiqueta que le fue asignada al bloque.
- **Puntaje Semántico:** La puntuación obtenida por similitudes semánticas (*ie.* labels).

El proceso de creación de la mejor segmentación consta de las siguientes 3 etapas:

- **Identificación de los Bloques:** principalmente se deben identificar los bloques de la nueva segmentación a ser creada S_{n+1} . Se seleccionan en base a una comparación utilizando las n segmentaciones ya almacenadas, S_1, \dots, S_n . El proceso se puede describir en dos pasos:
 - se fija un bloque $b_k^{n+1} \in S_{n+1}$,
 - se compara contra todos los bloques en todas las segmentaciones pertinentes, $\forall b_j^i \in S_i$ con $i \in [1, n]$ en las demás segmentaciones. Se considera una tolerancia basada en la Distancia Hausdorff ⁴ entre b_k^{n+1} y un b_j^i . Si la distancia es menor o igual a T ⁵ se considera que ambos bloques son similares y el b_j^i tendría el mismo identificador que b_k^{n+1} .
- **Contabilización de Puntos:** para complementar el paso anterior, se contabilizan todos los bloques bajo un mismo identificador, para obtener el *puntaje geométrico*. Después, dentro del mismo *pool* de bloques con el mismo identificador se contabilizan todos los que posean las mismas etiquetas, de esta forma obtener el *puntaje semántico*. En la Figura 5 se observa el cálculo del *puntaje geométrico*. El bloque b_1^1 tiene equivalentes en las tres segmentaciones restantes. Esto contabiliza 4 ocurrencias en el conjunto de segmentaciones. Para el bloque b_1^4 tiene correspondencia en dos de las segmentaciones restantes. Finalmente se suman las ocurrencias y se obtiene el puntaje geométrico. Este cálculo se realiza cada vez que se incorpora una nueva segmentación de una página.

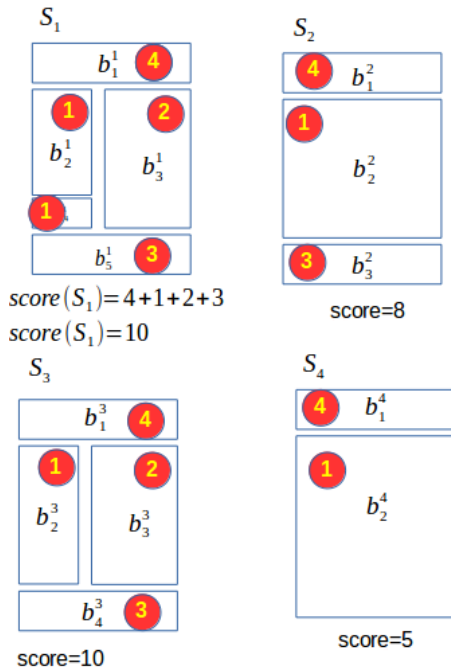


Figura 5: Ejemplo de Puntajes Geométricos

En la Figura 6 se muestra un ejemplo para el cálculo del *puntaje semántico*. Se basa en la coincidencia geométrica y adicionalmente que la etiqueta de los bloques coincidan. En el ejemplo se puede observar para la segmentación S_1 las siguientes etiquetas: HDADF (corto para *header, aside, article, aside* y *footer*, respectivamente). La segmentación S_1 tiene el mayor *puntaje semántico*. Es un fuerte candidato para ser incorporado

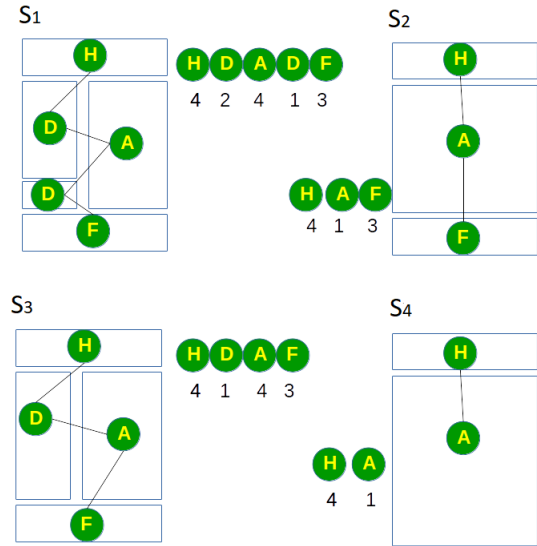


Figura 6: Ejemplo de Puntajes Semánticos

- **Creación de la Mejor Segmentación:** para crear la mejor segmentación se incluyen todos aquellos bloques (uno por cada identificador) cuyo *puntaje geométrico* sea mayor que el 50% del número de segmentaciones realizadas, esto garantiza que la mayoría de los usuarios opina que ese bloque debe existir. Dicho bloque poseerá la etiqueta más utilizada para ese bloque, es decir, se busca entre los de un mismo identificador la etiqueta que tenga el mayor *puntaje semántico*. En el ejemplo mostrado en la Figura 6 el bloque b_1^1 tendría asociada la etiqueta H la cual es la más popular para esa área de la página.

V. VERIFICACIÓN DEL SISTEMA

Para verificar el sistema desarrollado, se llevaron a cabo dos pruebas de aceptación, una funcional y otra no funcional, las cuales se describirán a continuación.

A. Prueba Funcional

Para comprobar la funcionalidad del sistema, se realiza una prueba de caja negra. Se busca comprobar si el sistema se comporta como es esperado según las funciones que debe realizar. Tal es el caso en la MoB Extension de insertar, modificar y eliminar bloques. Al modificar se verifica la unión y separación de bloques. En el MoB API la gestión de

⁴La distancia Hausdorff es la mayor de todas las distancias existentes desde un punto en un conjunto hasta el punto más cercano en otro conjunto.

⁵generalmente $T = 30px$

segmentaciones (eg agregar, visualizar), así como la creación de la mejor segmentación. Y para el MoB Repository observar que las segmentaciones y páginas estén completas y con todas sus dependencias. La técnica utilizada es principalmente la observación, dada la experiencia con los datos manejados, de manera satisfactoria.

B. Prueba no Funcional

Las pruebas no funcionales se enfocaron en asegurar si la herramienta es usable o no. Específicamente la herramienta MoB Extension, para esto se evalúa la reacción de 5 individuos ante el sistema. Se observó las reacciones de los participantes mientras completan 2 objetivos planteados y finalmente se les dio un cuestionario para responder según su experiencia. Cabe destacar que la herramienta está orientada a usuarios de 13 años en adelante, con inexistentes, bajos, intermedio o avanzados conocimientos en segmentación de páginas Web. En la Figura 7 se muestra un resultado de ejemplo del formulario. Dada la extensión del formato de los datos, no se muestran en este documento los gráficos completos, pero pueden ser consultados en [14]. En la misma figura puede observarse la pregunta del cuestionario: ¿cuánto tiempo se tomó en realizar el segundo objetivo? teniendo que el 40% lo tomó entre 5 y 10 minutos, mientras que el 60% lo finalizó en menos de 5 minutos. Respuestas como estas nos permiten tener la confianza que se ha mejorado la usabilidad substancialmente. Tomando en consideración que los usuarios de la versión anterior debían tener un *background* muy avanzado para poder usar la herramienta.

¿Cuánto tiempo se tomó en realizar el segundo objetivo?

5 respuestas

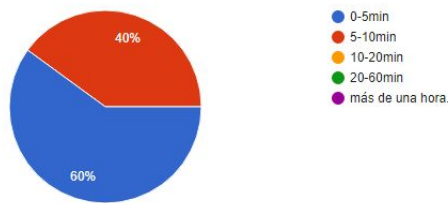


Figura 7: Prueba No Funcional: Pregunta 5 del Cuestionario

- Realizar una segmentación manual sobre la página Web: <https://wiki.apache.org/httpd/RedirectSSL>.
- Visitar el Repositorio de MoB y observar la segmentación realizada.

En conclusiones generales, basándose en las respuestas obtenidas del cuestionario, los comentarios hechos por los participantes y el comportamiento observado de los mismos, se tiene que:

- El sistema en general presenta un aspecto estético agradable para los usuarios.
- La herramienta de segmentación permite a los usuarios inexpertos realizar segmentaciones rápidamente sobre una página Web, de una forma sencilla.

- La navegación general del Repositorio MoB es entendible, sin embargo cuando se debe profundizar, buscar segmentaciones específicas, el usuario debe invertir un poco de tiempo en entender la lógica de la navegación.

Se puede considerar que el sistema MoB es usable, sin embargo, algunos usuarios requieren una breve inducción.

VI. CONCLUSIONES Y TRABAJOS A FUTURO

Durante el desarrollo de la herramienta de segmentación manual (Extensión de MoB), se presentó uno de los retos más grandes del proyecto. Se requirió no sólo que cumpliera con su misión de segmentar manualmente la página Web, sino mejorar significativamente la usabilidad. Las pruebas de caja negra y usabilidad permitieron validar los objetivos iniciales de la herramienta. Se logró reducir la complejidad de una tarea tediosa y propensa a errores mediante la automatización.

En cuanto al repositorio de MoB, el mayor reto fue lograr una forma comprensible de ordenar toda la información sobre las páginas Web y sus segmentaciones. Después de realizar las pruebas de caja negra se evidenció que el repositorio funcionaba como se esperaba. Aun algunos usuarios encontraban difícil la navegación en el sitio. Esta información recopilada puede ser tomada en cuenta para las mejoras a futuro que se vayan a realizar sobre el sistema.

Entre los retos presentados en el desarrollo del API de MoB se encontraba: el poder desarrollar todos los servicios pertinentes de una forma modular. Presentar estructuras de respuesta que fuesen fáciles de comprender y manejar, en especial la estructura de bloques que se debe usar para la carga de los datos de la segmentación. Se incluye una parte de la implementación en background para poder detectar una nueva segmentación e invocar el proceso de cálculo de la mejor segmentación manual. Reducir el tiempo de respuesta de este proceso fue un reto, y que dio respuesta en un tiempo aceptable. Esto fue resuelto sin contratiempos incluyendo una base de datos con extensiones geográficas como *Postgis*. De esta forma el análisis de comparar los diferentes bloques se realizó de manera rápida y eficaz.

Este proyecto comenzó con la idea de desarrollar una herramienta para ayudar en el desarrollo de otros proyectos. Sin embargo, a lo largo del desarrollo se fueron extendiendo las funcionalidades principales y añadiendo funcionalidades adicionales a los elementos principales del sistema, reforzando la funcionalidad y permitiendo la evolución del mismo. Como resultado tenemos un sistema bastante completo en donde se pueden realizar segmentaciones manuales, incluir segmentaciones hechas por algoritmos y ser almacenadas. La información de las segmentaciones puede ser mostrada a través de una interfaz Web (Repositorio MoB), y en el caso de las segmentaciones manuales, pueden ser analizadas y obtener ese elemento que representa la razón principal por la cual se realiza este trabajo, la mejor segmentación manual.

Para finalizar, la solución presentada permite incrementar la cantidad de páginas existentes en el repositorio, que hasta el momento posee cerca de 60 páginas realizadas con el prototipo de MoB. Durante la realización de las pruebas de la nueva versión pudimos observar que agregar una nueva página manualmente segmentada promete ser menos complejo. Esto ayuda al equipo de investigación a tener un dataset cada vez más completo. Puede visitar el repositorio en <http://bom.ciens.ucv.ve> y en <https://mob.ciens.ucv.ve>

A. Trabajos Futuros

La inclusión de otros formatos de exportación para las segmentaciones sería una buena actualización. Según los resultados arrojados por la prueba de usabilidad, el Repositorio MoB presenta una navegación que puede resultar algo confusa para algunos usuarios, es por eso que se propone como un trabajo a futuro la implementación de una navegación más intuitiva. Este tipo de retroalimentación nos permitirá acelerar la conformación de la base referencial. Además, es importante destacar que la extensión actual funciona únicamente para el navegador Chrome/Chromium es por eso que se plantea un trabajo a futuro donde se adapte dicha extensión para funcionar en una mayor variedad de navegadores (Firefox, Safari, Opera, entre otros). Esta nueva versión permite incrementar la cantidad de usuarios participantes. Esto implica a corto plazo que se debe mejorar el MoB API para trabajar de manera concurrente en un ambiente de alto desempeño.

REFERENCIAS

- [1] Y. Zhu, *W3C dom 4.1* W3C working draft, W3C, Feb. 2018. <https://www.w3.org/TR/2018/WD-dom41-20180201/>.
- [2] A. Sanoja, *Web Page Segmentation, Evaluation and Applications*. PhD thesis, Université Pierre et Marie Curie-Paris VI, <https://hal.inria.fr/tel-01128002/>, 2015.
- [3] A. Sanoja and S. Gañçarski, *Web Page Segmentation Evaluation*, in Proceedings of the 30th Annual ACM Symposium on Applied Computing, pp. 753–760, ACM, 2015.
- [4] D. Cai, S. Yu, J.-R. Wen, and W.-Y. Ma, *Extracting Content Structure for Web Pages Based on Visual Representation*, in Proceedings of the 5th Asia-Pacific Web Conference on Web Technologies and Applications, APWeb'03, (Xian, China), pp. 406–417, Springer-Verlag, 2003.
- [5] X. Liu, H. Lin, and Y. Tian, *Segmenting Webpage with Gomory-Hu Tree Based Clustering*, JSW, vol. 6, no. 12, pp. 2421–2425, 2011.
- [6] E. Uzun, H. V. Agun, and T. Yerlikaya, *A Hybrid Approach for Extracting Informative Content from Web Pages*, Information Processing & Management, vol. 49, no. 4, pp. 928–944, 2013.
- [7] D. Chakrabarti, R. Kumar, and K. Punera, *A Graph-Theoretic Approach to Webpage Segmentation*, in Proceedings of the 17th International Conference on World Wide Web, (Beijing, China), pp. 377–386, ACM, 2008.
- [8] Z. Jiang, H. Yin, Y. Wu, Y. Lyu, G. Min, and X. Zhang, *Constructing Novel Block Layouts for Webpage Analysis*, ACM Trans. Internet Technol., vol. 19, July 2019.
- [9] J. Kiesel, H. Wachsmuth, K. Al-Khatib, and B. Stein, *WAT-SL: A Customizable Web Annotation Tool for Segment Labeling*, in Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics, (Valencia, Spain), pp. 13–16, Association for Computational Linguistics, April 2017.
- [10] J. Blustein, N. R. Di Matteo, and D. Macrini, *Designing Experiments to Compare Web Page Segmenters*, in Proceedings of the 2nd International Workshop on Human Factors in Hypertext, pp. 27–29, 2019.
- [11] R. Kreuzer, J. Hage, and A. Feelders, *A Quantitative Comparison of Semantic Web Page Segmentation Approaches*, in International Conference on Web Engineering, pp. 374–391, Springer, 2015.
- [12] A. Sanoja and S. Gañçarski, *Block-o-matic: A Web Page Segmentation Framework*, in International Conference on Multimedia Computing and Systems (ICMCS), (Marrakesh, Morocco), pp. 595–600, April 2014.
- [13] Z. Pehlivan, M. Ben-Saad, and S. Gañçarski, *Vi-diff: Understanding Web Pages Changes*, in Proceedings of the 21st International Conference on Database and Expert Systems Applications: Part I, DEXA'10, (Berlin, Heidelberg), pp. 1–15, Springer-Verlag, 2010.
- [14] J. P. Garcia, *Desarrollo de una Herramienta Interactiva para la Construcción de un "Ground Truth" de Segmentaciones de Páginas Web*, T.E.G. Escuela de Computación. Universidad Central de Venezuela, 2018.

Sistema de Apoyo a la Terapia Física Basado en el Internet de las Cosas

Francisco José Moreno Alvarez¹, Walter Hernández¹
francisco.j.moreno@ucv.ve, walter.hernandez@ciens.ucv.ve

¹ Escuela de Computación, Universidad Central de Venezuela, Caracas, Venezuela

Resumen: Desde 2010, un alto porcentaje de la población (15%) a nivel mundial posee alguna discapacidad que afecta su desenvolvimiento en la sociedad, cifra que continúa en ascenso debido a diversos factores (envejecimiento de la población, proliferación de enfermedades crónicas, entre otros). El proceso tradicional de terapia posee un conjunto de problemas que involucran desde la infraestructura y personal en el centro médico, hasta el traslado del paciente. Dichos problemas, inciden en la calidad del servicio personalizado que requiere el paciente. Con la masificación del Internet y el surgimiento de paradigmas como el Internet de las Cosas (IoT por sus siglas en inglés), se manifiestan nuevas posibilidades para mejorar la atención en el área de la rehabilitación. En este trabajo se propone una arquitectura basada en IoT, que permite al paciente realizar su terapia desde la comodidad de su casa y al personal médico: seguir y adaptar la misma basándose en su progreso. La arquitectura propuesta, está compuesta por un conjunto de dispositivos *inteligentes* orientados a rehabilitación, una aplicación de Monitoreo de Actividades Terapéuticas (MAT) utilizada por el médico para asignar y controlar las sesiones de terapia, y un conjunto de servicios en la nube que coordina los demás componentes. Entre los aportes de esta solución se incluyen: un conjunto de mecanismos y reglas que permiten la integración y correcto funcionamiento de cualquier dispositivo *inteligente* en la arquitectura, y una herramienta que permite el diseño, asignación y monitoreo de actividades terapéuticas. La arquitectura propuesta fue sometida a un conjunto de pruebas de integración basadas en la simulación de escenarios, que validan su efectividad como herramienta computacional. Por último, se estima un incremento de la velocidad de atención en una sesión terapéutica de al menos unas 10 veces mediante el uso del sistema.

Palabras Clave: Arquitectura de Software; Internet de las Cosas; Terapia; IoT; Tele-Rehabilitación; Rehabilitación.

Abstract: Since 2010, a high percentage of the world's population (15%) has been established to have some disability which constrains its life in society. This percentage is still growing due to several factors such as population aging and chronic diseases. Traditional therapy has an inherent set of issues that range from infrastructure to the lack of staff in medical centers. These problems affect the quality of both the personal and customized service required by the patient. With the massification of the Internet and new paradigms such as the Internet of Things (IoT), new possibilities to improve the rehabilitation service arise. In this proposal, an architecture based on IoT which allows the patient to do therapy from home has been developed. It also provides tools which enables the therapist with the required information to track the patient's progress and adapt the therapy accordingly. The proposed architecture is compound by a set of *smart* devices for therapy, a software application used by the therapist to assign and monitor therapy sessions and cloud services to control all the components of the architecture. The main contributions are: a set of tools and rules to easily configure and integrate any *smart* device for therapy into the architecture, as well as an application to design, assign and monitor therapy activities. The proposed architecture has been tested through simulated scenarios that validate its computational effectiveness. Finally, with the use of the proposed architecture a patient has been estimated to be served 10 times faster than with traditional therapy.

Keywords: Software Architecture; Internet of Things; Therapy; IoT; Tele-Rehabilitation; Rehabilitation.

I. INTRODUCCIÓN

En el año 2011, la Organización Mundial para la Salud (OMS) y el Grupo del Banco Mundial presentó un informe en el que se estimó que más de mil millones de personas (aprox. 15% de la población mundial) vivían con algún tipo de discapacidad [1].

Dicha cifra resulta superior a la estimación del 10% realizada por el mismo organismo en el año 1970. Este crecimiento de la proporción de individuos con discapacidades, se ve principalmente impulsado por el envejecimiento de la población, el cual es un factor de relevancia debido a que el riesgo de

sufrir una lesión o desarrollar alguna enfermedad crónica que conlleve una discapacidad, aumenta con respecto a la edad de la persona. En el año 2002, la OMS [2] reportó que al menos cinco millones de personas sobreviven a un ataque cerebro vascular cada año y que la mitad de ellos permanecen con hemiparesia (debilidad o parálisis parcial en la mitad del cuerpo). En el año 2017, el Centro de Control y Prevención de Enfermedades (CDC por sus siglas en inglés), indica que el 25.6% de la población adulta en Estados Unidos sufre de algún tipo de discapacidad [3]. Por su parte, el Instituto Nacional de Estadísticas (INE) en el último censo realizado en 2011, indica que el 5.38% de la población venezolana sufre de alguna discapacidad, agrupándolas en tres tipos principales como: visual, cardiovascular y músculo esquelética, representando el 1.7%, 1.1% y 0.9% respectivamente [4].

En la Figura 1, se evidencia como una deficiencia visual, motora o cardiovascular, afecta actividades primordiales de la vida en sociedad tales como: el acceso a la educación, el acceso y desempeño en el trabajo y, la facilidad de aprendizaje. Adicionalmente, un gran esfuerzo y tiempo es necesario para cumplir con tareas esenciales como: aseo personal, alimentación, movilización e incluso socializar. En consecuencia, el paciente normalmente depende de asistencia, que de no obtenerla, podrían incrementar sus problemas de salud física y mental, desencadenando trastornos depresivos y finalmente deteriorando su calidad y la de su grupo familiar.

Al someterse a rehabilitación, el paciente puede erradicar o disminuir el grado de su discapacidad, hecho de vital importancia ya que evidentemente una discapacidad significa un riesgo directo e indirecto para su salud. Sin embargo, los métodos tradicionales de terapia involucran asistir a un centro médico, lo cual conlleva otro conjunto de problemas. Diversos factores deben ser considerados y planificados para poder asistir a la cita médica, por ejemplo: evaluar qué medios de transporte son los más adecuados para el paciente, cuáles son los mejores horarios para trasladarse, cómo satisfacer los requerimientos de un posible régimen especial de alimentación fuera de casa, entre otros. Además, se deben tener en cuenta las necesidades del acompañante del paciente, por ende, es necesario añadir: la alimentación y transporte del acompañante, tiempo fuera del trabajo, entre otros. Dichos factores, impactan económica y emocionalmente tanto al paciente como al acompañante y se traducen en asistencias irregulares a las sesiones de terapia que eventualmente conllevan a una posible deserción.

El problema anterior, es complementado con la falta de personal y espacio físico que existe en los centros hospitalarios a nivel mundial, razón que provoca el distanciamiento entre cada sesión de terapia o el deterioro de la calidad de las mismas, lo que afecta la velocidad de recuperación del paciente. Como solución, los pacientes son comúnmente forzados a cumplir gran parte de sus terapias en sus casas. Esto, opaca los problemas del paciente para asistir a la cita médica y los del hospital para prestarle el servicio adecuado; sin embargo, en casa y sin supervisión directa de un médico calificado, resulta imposible garantizar el cumplimiento de los ejercicios y su

correcta ejecución, lo cual incide en la correcta recuperación del paciente.

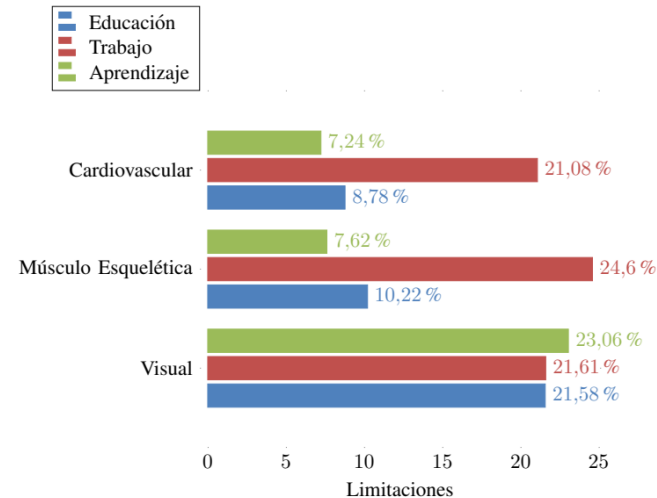


Figura 1: Principales Limitaciones de las Personas con Discapacidad, CENSO 2011 [4]

En el presente trabajo, se ofrece una solución que le permite al personal médico asignar y monitorear sesiones de terapia de manera remota a sus pacientes. Para el desarrollo de dicha aplicación, se propone una arquitectura basándose en una rama de estudio emergente conocida como: El Internet de las Cosas (IoT por sus siglas en inglés). En este escenario se plantea una red global de dispositivos que interactúan entre ellos [5][6] mediante la incorporación de la capacidad para: identificarse, percibir su entorno, procesar información e interconectarse.

Particularmente, se plantea una aplicación Web que el personal médico utiliza para: asignar sesiones de terapia a sus pacientes y monitorear el progreso de los mismos. De igual manera, se generan las estructuras de la arquitectura que permiten: la interconexión de la aplicación utilizada por el terapeuta y los dispositivos inteligentes que imparten la terapia al paciente. La arquitectura está diseñada de manera tal que facilite la integración de un conjunto heterogéneo de dispositivos capaces de trabajar con diversas patologías médicas. Por ende, se plantean mecanismos para: registrar un nuevo dispositivo a la red, suministrar los parámetros de configuración de terapia e indicar el tipo de medidores de desempeño que el dispositivo registra durante la terapia. En la propuesta actual se presume la existencia de dispositivos inteligentes cuya efectividad médica está comprobada y que son capaces de suministrar y evaluar por si mismos la actividad para la que fueron diseñados.

Mediante la simulación de dispositivos básicos integrados a la propuesta, se pudieron realizar un conjunto de pruebas para probar la efectividad del sistema como herramienta de agilización del proceso de terapia. Se pudo estimar que el tiempo de atención de un paciente disminuye al menos 10 veces con el uso de la aplicación, considerando únicamente el tiempo que el paciente invierte viajando a un centro hospitalario y el tiempo promedio de espera por atención *in situ*.

El resto del trabajo se presenta en 5 secciones. La Sección

II menciona un conjunto de investigaciones que a lo largo del tiempo comprueban la efectividad de la tele-rehabilitación como técnica de terapia. La Sección III describe brevemente algunos trabajos relacionados a la propuesta que se presenta en la Sección IV. La Sección V plantea las pruebas que se realizaron al sistema para evaluar su efectividad computacional y su habilidad para agilizar el proceso que involucra la terapia. Finalmente, la Sección VI expone las conclusiones del trabajo y trabajos futuros.

II. EFECTIVIDAD DE LA TELE-REHABILITACIÓN SOBRE EL MÉTODO TRADICIONAL

Los avances tecnológicos generan propuestas alternativas a la terapia tradicional, como la rehabilitación a través de realidad virtual, la cual fue inicialmente estudiada como un mecanismo para afrontar fobias [7][8] y se ha expandido hasta ser utilizada como una forma alternativa de terapia física [9][10][11]. Con los avances en las telecomunicaciones, la tele-rehabilitación prueba ser una solución factible para monitorear el desempeño del paciente fuera de la institución médica. Efectividad que se comprueba en el estudio [12], en donde se valida la tele-rehabilitación como técnica de terapia en pacientes que poseen problemas motores en los miembros superiores producto de una lesión cerebral. En dicho estudio, analizan el progreso que obtienen seis pacientes, luego de ser entrenados de manera remota durante seis semanas, con una frecuencia de cinco días por semana y setenta minutos por día; contando con la supervisión intermitente por parte del personal médico, la cual fue eliminada progresivamente por decisión de los pacientes a medida que manifestaban la capacidad de realizar los ejercicios de manera autónoma. Como resultado, todos los pacientes involucrados en el estudio mejoraron su rango de movilidad, confirmando la efectividad de la tele-rehabilitación como alternativa a la terapia tradicional.

Un estudio más reciente [13] realiza una extensa revisión bibliográfica en la que evalúa la efectividad médica de la tele-rehabilitación con respecto a los métodos tradicionales. En dicho estudio, abarcan cualquier tipo de patología (desde terapia física a psicológica) que requiera rehabilitación. Además, aíslan los pacientes que fueron rehabilitados mediante el proceso tradicional, de los pacientes que fueron tratados exclusivamente mediante tele-rehabilitación o siguiendo un enfoque híbrido. Los autores concluyen, que la tele-rehabilitación es tan efectiva como la terapia tradicional.

Otro estudio [14] analiza la perspectiva de los allegados que cuidan a la persona que sufre de una discapacidad. En este, se asigna un costo monetario al tiempo que dedican dichas personas durante los días que el paciente debe realizar terapia en el centro hospitalario. El costo es calculado en base a: los gastos en transporte, alimentación del día y el costo asociado a las horas que deben faltar al trabajo para acompañar al paciente. Estiman, que los cuidadores ahorran un promedio de 1024 dólares, luego de realizar 16 sesiones de manera remota. Las sesiones se suministraron a través de vídeo llamadas, permitiéndole al cuidador realizar otras tareas que antes no podían efectuar durante el transcurso de la misma. El estudio

concluye indicando el alto grado de aceptación por parte de los cuidadores y de los pacientes, apoyándose en el ahorro económico producto del tratamiento remoto, además de la cantidad de horas ahorradas en preparación para asistir a la terapia y el grado de estrés que representaba la misma.

III. TRABAJOS RELACIONADOS

Según Atzori et al. [15], las áreas de estudio de *IoT* se pueden agrupar en 4 dominios: Transporte y Logística, Ambientes inteligentes, Personal y Social y Salud.

Específicamente en el contexto de la salud, *IoT* se ha explorado como herramienta de prevención de reacciones adversas a medicamentos suministrados a pacientes [16]. Vigilancia de pacientes en ambientes cerrados con necesidades especiales y sistemas de alerta temprana [17][18][19]. Control de signos vitales de pacientes en áreas rurales remotas con baja inversión monetaria [20]. Predicción de diagnóstico de enfermedades basado en signos vitales [21]. Atención y monitoreo de un gran número de pacientes mediante el uso de agentes especialistas en tipo de signo vital o patología [22].

Para la asistencia de personas con discapacidades Domingo et al. [23] propone una arquitectura para *IoT*, compuesta de tres capas:

- **Capa de Percepción:** Compuesta por sensores, efectores y estaciones de monitoreo, cuyo objetivo es recolectar información.
- **Capa de Redes:** Consiste de las redes alámbricas e inalámbricas utilizadas para transmitir la información recolectada por la capa de percepción.
- **Capa de Aplicación:** Es un conjunto de aplicaciones que utilizan la información recolectada, para satisfacer la necesidad del paciente.

Cabe destacar que en dicho trabajo se hace hincapié en la importancia de la “Capa de Percepción”, como herramienta para minimizar el efecto de la discapacidad sobre la vida de la persona, nunca se propone un mecanismo de rehabilitación para mejorar la condición del paciente.

Por otra parte Bui et al. [24] y Dohr et al. [25] proponen un esquema similar para *IoT* enfocado en el cuidado de la salud. En dichos trabajos exponen un escenario donde el paciente debe cumplir con una rutina diaria compuesta de una medición de signos vitales y el consumo de algunos medicamentos prescritos. Para garantizar el cumplimiento de la rutina, se integran sensores al equipo médico utilizado por el paciente (capa de *hardware*) y los datos que estos recolectan son capturados por un teléfono inteligente y son enviados a un sistema central en el centro médico (capa de *middleware*). Dichos datos, son dispuestos al personal médico y familiares con el fin de permitirles evaluar la condición del paciente y tomar acciones (capa de servicios). Finalmente, se complementa con un sistema de alertas en caso de poseer valores anómalos en los datos recolectados o en caso de que el paciente olvide realizar el registro de los signos vitales.

Vukićević et al. [26] presenta un sistema de tele-rehabilitación basado en *IoT*, que aplica efectivamente terapia física de

manera remota, eliminando la intervención presencial del personal médico durante su práctica y procurando mantener motivado al paciente a realizar diariamente el mismo grupo de ejercicios.

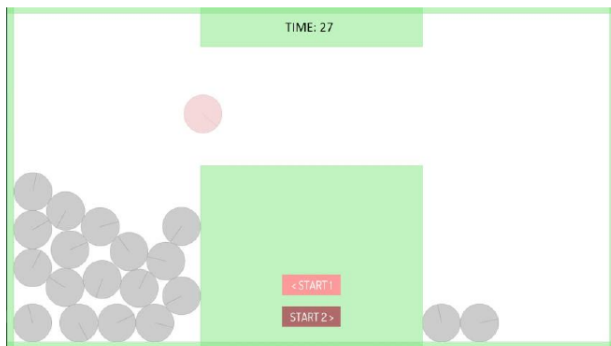


Figura 2: Juego Serio Desarrollado en [26]

Los ejercicios de rehabilitación son presentados al paciente en forma de un juego serio con el objeto de proponerle al paciente: un reto, un objetivo claro y una forma divertida de realizar los ejercicios repetitivos que componen una sesión de terapia. Los ejercicios propuestos, son diseñados para tratar pacientes con movilidad reducida en los miembros superiores. Durante el desarrollo de la actividad, el desempeño del paciente es evaluado mediante un conjunto de *hardware* especializado. Precisamente, se observa la postura del brazo trabajado mediante un *Microsoft Kinect* (*hardware* diseñado para hacer seguimiento del cuerpo del usuario). Por otra parte, se evalúa la contracción del bíceps, generada al abrir o cerrar la mano, mediante un sensor que supervisa la actividad eléctrica producida en los músculos, el cual se encuentra conectado a un *Arduino* (computador compacto en una sola pieza) que procesa la señal.

El juego consiste de un conjunto de círculos, que el paciente debe mover de un extremo de la pantalla al otro sobrepasando un obstáculo (ver Figura 2). Para mover los círculos, el paciente debe mover el brazo hasta alcanzar uno y cerrar la mano para recogerlo; análogamente, deberá mover el brazo para colocarlo en el extremo contrario de la pantalla y abrir la mano para depositarlo. El objetivo es evaluar la velocidad con la que el paciente es capaz de mover todos los círculos de la posición inicial a la final.

Los autores proponen una arquitectura que se puede observar en la Figura 3. En la misma, se dispone de una aplicación para el terapeuta mediante la cual puede observar el progreso del paciente, una aplicación que contiene el juego serio, una capa de software para almacenar y procesar los datos recolectados durante la terapia, un conjunto de herramientas de análisis de los datos para uso del terapeuta y la integración del registro médico del paciente para facilitar el diagnóstico y evaluación del progreso de la patología del mismo.

Finalmente, demuestran la factibilidad del sistema como herramienta de terapia probándolo con un paciente de 60 años, el cual es tratado durante un mes exclusivamente mediante la aplicación de manera remota. Los datos recolectados por la

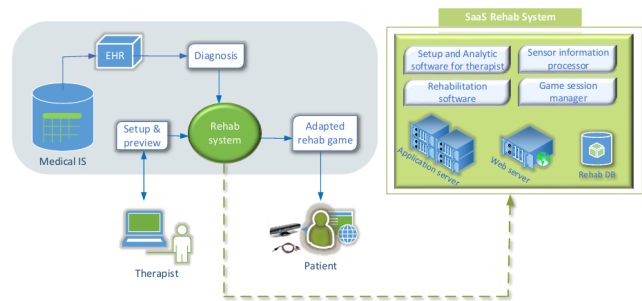


Figura 3: Arquitectura Propuesta en [26]

aplicación dan como resultado la Figura 4, donde se percibe la mejora en movilidad y precisión del paciente en el transcurso de los días.

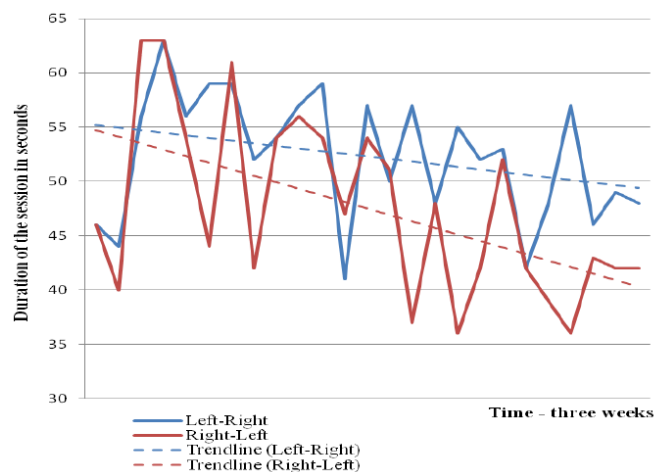


Figura 4: Progreso Obtenido por el Paciente en el Periodo de Prueba [26]

Por último, Ferreira et al. [27] investiga como *ludificar* los aparatos que se utilizan tradicionalmente en la terapia de rehabilitación de pacientes con problemas motores en los miembros superiores. Para ello proponen un conjunto de juegos serios (ver Figura 5), que con la ayuda de hardware especializado (*Razer Hydra Motion Controllers* y *Oculus Rift*), son capaces de suministrarle la terapia al paciente siguiendo los parámetros de la terapia tradicional. La terapia es suministrada de manera remota, por lo que la solución se complementa con un módulo que le permite al terapeuta visualizar los resultados capturados por el juego serio y configurar parámetros para la siguiente terapia.

La propuesta de este documento nace con la intención de dar continuidad a la investigación de Moreno et al. [28] donde se propone un *framework* basado en realidad virtual semi-inmersiva, que tiene el objetivo de actuar como herramienta de rehabilitación física para pacientes con discapacidades en los miembros superiores, cuyas edades oscilan entre los 6 y 12 años. Dicho trabajo tuvo como principal objetivo mantener motivado a los pacientes a cumplir con la terapia (disminuir el índice de deserción) y proveer datos de progreso al personal médico. Para cumplir con el mismo, las actividades



Figura 5: Paciente Haciendo Uso del Sistema Propuesto en [27]

terapéuticas se diseñaron en la forma de juegos serios, en los que la motricidad gruesa y fina del paciente eran evaluadas por *hardware* especializado, los cuales proveen al médico de información del desempeño del paciente, durante y luego de la culminación de la actividad.

El problema que posee la investigación previa es la imposibilidad de utilizar la herramienta fuera del centro médico, ya que la misma requiere de la intervención presencial de personal médico. Problema que será abordado mediante la incorporación de técnicas de *IoT*, planteando una arquitectura enfocada en la rehabilitación de pacientes con problemas motores, tomando en cuenta los esquemas de desarrollo y consideraciones expuestas por [24][25][15].

IV. PROPUESTA

El Internet de las Cosas es definido como una red ubicua compuesta por un grupo heterogéneo de dispositivos “inteligentes”. Un dispositivo es considerado “inteligente” si es capaz de mantener una comunicación permanente con otros objetos y ser consciente de su entorno, esto implica que debe percibir y procesar de manera constante todo estímulo proveniente del ambiente. Un estímulo puede originarse como producto de la interacción del usuario, el análisis de algún área de interés (medición de un sensor, análisis de una imagen, entre otros) o como resultado de los mensajes transmitidos a través de la red.

En el contexto de la terapia de rehabilitación, *IoT* puede aplicarse a la asignación, ejecución y evaluación de una actividad terapéutica mediante el uso de dispositivos “inteligentes”.

El escenario propuesto inicia con el acceso del terapeuta a una aplicación *Web* en donde asigna una sesión de terapia al paciente, hecho que desencadena una notificación que recibe el paciente a través de un dispositivo “inteligente”. Dicha notificación, le indica la actividad a realizar y la fecha en que debe cumplir con la misma. El día de la terapia, el dispositivo le recuerda al paciente la sesión pautada, este realiza las actividades propuestas, siguiendo las instrucciones suministradas por el dispositivo. Una vez que concluye, continúa con sus actividades del día. Los datos recolectados por el dispositivo, son enviados de vuelta al médico para que analice el desempeño del paciente y diseñe la nueva sesión de terapia que mejor se ajuste al progreso demostrado.

En el escenario anterior, los estímulos provenientes de la red vienen en la forma de notificaciones:

- Nueva sesión terapéutica asignada.
- Recordatorios para el cumplimiento de la terapia asignada.
- Actividades completadas.
- Sesión inconclusa (caso especial en que el paciente no realiza la terapia a tiempo).

De igual forma, los estímulos que recibe el dispositivo son una combinación de los estímulos de la red más la interacción con el usuario:

- Usuario detectado en el área de interés.
- Interacción específica del usuario.
- Registro de los datos correspondientes al efectuar la actividad propuesta.

En la Figura 6 se presenta la arquitectura basada en *IoT* propuesta en este trabajo, en la misma se pueden observar un conjunto de elementos claramente separados:

- **Dispositivos terapéuticos** (color rojo): Conjunto heterogéneo de dispositivos “inteligentes” diseñados para evaluar un conjunto de ejercicios específicos que tratan una patología médica.
- **Servicios en la nube** (color verde): Que tiene como ente central el *API* (*Application Programming Interface* - Interfaz de Programación de Aplicaciones), que tiene como objetivo principal coordinar el resto de los componentes. Almacenes de datos. Un gestor de mensajes que mantiene una comunicación permanente con los dispositivos terapéuticos. Y finalmente, un conjunto de programas que atienden las peticiones realizadas por los dispositivos terapéuticos (Trabajadores, en color amarillo).
- **Capa de aplicación** (color azul): Contiene una aplicación denominada *MAT* (*Monitor de Actividades Terapéuticas*) que es utilizada por el médico y los pacientes para monitorear el progreso de la rehabilitación.

A. Servicios en la Nube

En esta sección se describen cada uno de los servicios dispuestos en la nube que, en conjunto, mantienen en perma-

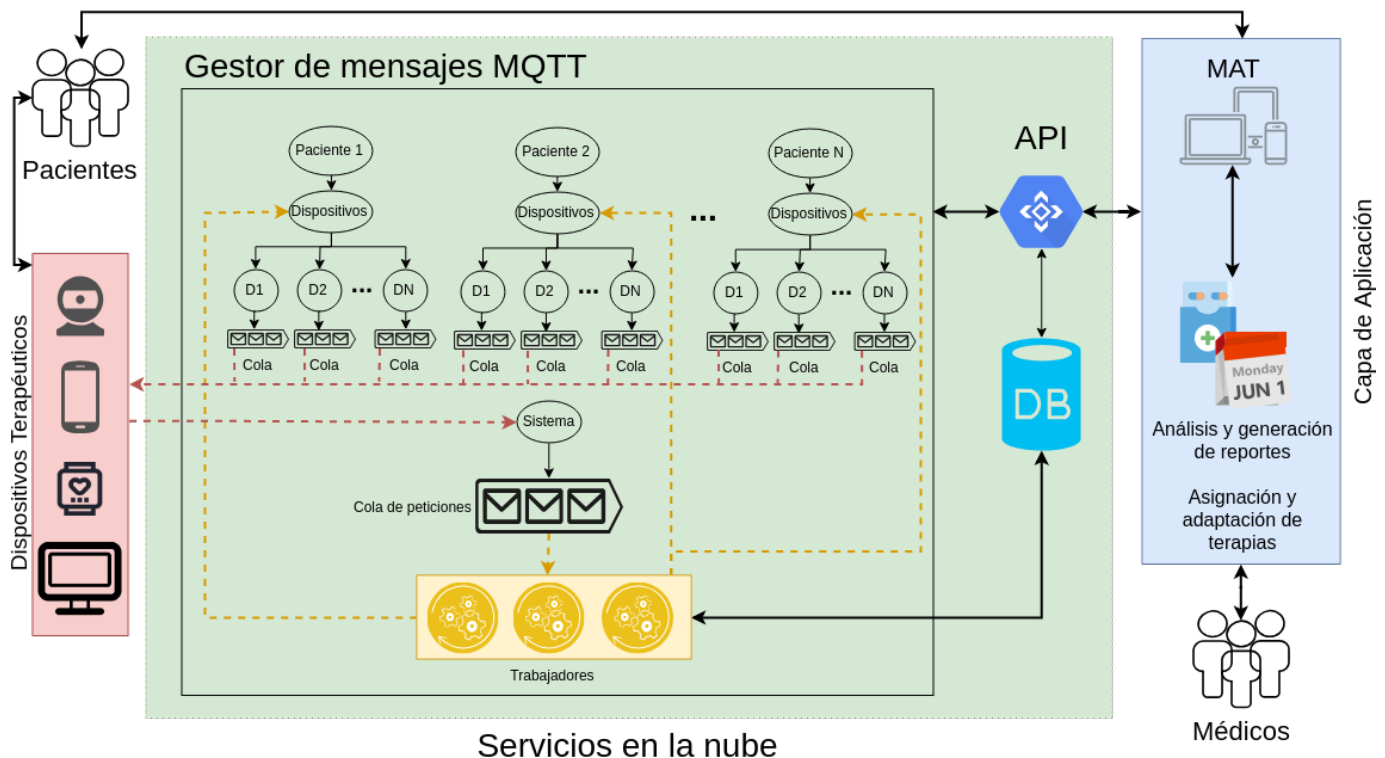


Figura 6: Arquitectura Basada en IoT para Apoyar la Terapia de Rehabilitación

nente comunicación a todos los componentes del sistema y almacenan todos los datos registrados por cada uno de ellos.

1) *API (Application Programming Interface - Interfaz de Programación de Aplicaciones)*: Es un *API Web* que tiene como objetivo principal registrar los datos asociados a las sesiones terapéuticas, los cuales son de naturaleza heterogénea y motivan el uso de bases de datos no relacionales (*NoSQL*), ya que las mismas poseen la capacidad de manejar datos con estructuras cambiantes y diversas. Adicionalmente, la aplicación generará un gran contenido de información rápidamente, lo que explota la característica de escalamiento horizontal que presentan las bases de datos *NoSQL* [29]. Particularmente para este trabajo, se utilizan bases de datos orientadas a documentos, el cual es el tipo de base de datos *NoSQL* más adaptable a una vasta gama de aplicaciones [30]. Específicamente se emplea como manejador *MongoDB*¹, ya que posee una gran madurez al ser extensamente utilizado en el mercado competitivo.

Con el objeto de permitir trabajos futuros donde se deseen realizar búsquedas complejas sobre la base datos, por ejemplo aplicar procesos de minería de datos, se decidió implementar el *API* utilizando el lenguaje de consulta (*Query*) *GraphQL* [31] diseñado por la compañía Facebook. *GraphQL* flexibiliza el acceso a la información provista por el *API*, permitiéndole al cliente diseñar sus propias consultas especializadas sobre los datos que requiera. Como servidor que implementa *GraphQL* se utilizó *Apollo Server*².

Además de los datos asociados a las sesiones terapéuticas, el *API* provee funciones administrativas como:

- Registrar usuarios.
- Registrar tipos de dispositivos.
- Registrar los tipos de actividades soportadas por dispositivo.
- Asignar pacientes a un terapeuta.
- Asignar dispositivos a un paciente.

Al momento de registrar un nuevo tipo de dispositivo, el usuario (en este caso el desarrollador del dispositivo) debe indicar la lista de parámetros generales que requiere que el *API* le suministre al mismo para iniciar una actividad terapéutica, así como los tipos de actividades terapéuticas que soporta y los parámetros que requiere. Para esto, se proveen de un conjunto de herramientas con las que puede especificar: el tipo de cada parámetro, los valores válidos para el mismo, el tipo de unidad (en caso de que aplique) y si es obligatorio u opcional; de igual manera permite especificar tipos de datos estructurados complejos y listas. Dichas funciones son provistas al desarrollador del dispositivo a través del *API*, con la intención de maximizar la cantidad de dispositivos y terapias que pueda manejar la arquitectura.

2) *Gestor de Mensajes*: Tiene como objetivo gestionar la comunicación en tiempo real entre los dispositivos “inteligentes”, el *API* y la capa de aplicaciones; por lo tanto, es necesario contar con un protocolo de comunicación de buena calidad. *MQTT*³ (*Message Queue Telemetry Transport - Transporte de Telemetría de Cola de Mensajes*), es un

¹<https://www.mongodb.com/es>

²<https://www.apollographql.com/docs/apollo-server>

³<http://mqtt.org>

protocolo de mensajería ampliamente utilizado en el ámbito de *IoT* puesto que es eficiente, está probado ser de bajo consumo energético, puede ser cifrado y está desarrollado para trabajar en redes de baja velocidad [32]. *MQTT* trabaja bajo un esquema suscripción/publicación, en el que cada nodo se suscribe a un tópico del cual quiere recibir información. Cuando algún paquete dirigido a un tópico es recibido, un nodo central denominado *broker* se encarga de distribuirlo a todos los nodos que se encuentren suscritos a dicho tópico. La especificación de los tópicos puede ser jerárquica, permitiendo establecer esquemas de transmisión complejos, en los que se generan filtros de mensajes al suscribirse a nodos específicos del árbol que establece la jerarquía. Como *broker* se utiliza *RabbitMQ*⁴, el cual es políglota y tiene la capacidad de operar con múltiples protocolos de mensajería simultáneamente, característica que es explotada en la arquitectura propuesta.

En la Figura 6 se puede observar parte de la estructura de tópicos (representada por óvalos) que se emplea en el sistema. Se provee de un tópico por cada usuario registrado, el cual tiene múltiples hijos para canalizar distintos tipos de mensajes, entre ellos, los dirigidos a dispositivos que tenga registrado el paciente (“dispositivo”). Contar con dicho tópico general permite poder recibir los mensajes dirigidos a todos los dispositivos del paciente, al suscribirse al mismo. Cada dispositivo está suscrito al tópico específico que tiene como identificador su nombre, dicho tópico es un hijo en el árbol de jerarquía del paciente que posea el dispositivo.

La estructura descrita anteriormente es utilizada por los dispositivos para recibir información dirigida a ellos, cuando desee solicitar alguna operación, el mismo publica un mensaje en el tópico “sistema”, su solicitud será procesada por un grupo de procesos especializados (trabajadores) y estos publicarán el resultado de la misma en el tópico específico del dispositivo.

3) *Trabajadores*: Conjunto de procesos diseñados con la intención de permitirle a los dispositivos solicitar información a través del protocolo *MQTT*. El problema con dicho protocolo es que al ser basado en tópicos, resulta imposible implementar de manera directa una cola de trabajo que permita distribuir equitativamente un conjunto de trabajos entre un grupo de procesos. Como solución se explotan las capacidades de *RabbitMQ* para trabajar con múltiples protocolos simultáneamente. A través del protocolo de comunicación *AMQP* (*Advanced Message Queuing Protocol* - Protocolo Avanzado de Cola de Mensajes) [33] se crea una cola de trabajos específica que se asocia al tópico “Sistema”. Dicha cola se configura para que despache un trabajo a cualquier proceso registrado a la misma, de esta forma se garantiza que ningún trabajador este ocioso si quedan peticiones por atender, maximizando el tiempo de trabajo. Ya que la cola se encuentra asociada a un tópico, ambos protocolos de mensajería pueden coexistir. De igual forma, cuando el trabajador culmina su trabajo, puede comunicar el resultado del mismo publicando un mensaje en el canal específico del dispositivo que realizó la solicitud.

A fin de mejorar la robustez del sistema, los trabajadores

realizan las siguientes consideraciones:

- No eliminan un trabajo de la cola, hasta que el mismo no fue completamente procesado.
- En caso de requerir confirmación por parte del dispositivo, no aceptan otro trabajo hasta recibir la confirmación.
- Poseen un tiempo de espera por confirmación. En caso de no recibirla en el lapso establecido, reenvían el mensaje. Si sobrepasan un número de intentos, desechan el trabajo.

B. Dispositivos Terapéuticos

Compuesta por un conjunto de dispositivos “inteligentes” especializados en un tipo de ejercicios o terapia que tienen como objetivos:

- Notificar al usuario la recepción de una nueva terapia.
- Suministrar la terapia al paciente con los parámetros especificados por el médico.
- Registrar y evaluar la calidad de los ejercicios realizados por el paciente durante la terapia.
- Proveer de información al paciente sobre su desempeño mientras el mismo efectúa la terapia.
- Enviar los datos recolectados al gestor de mensajes para que registre la información en la base de datos.

Debido a la variabilidad del tipo de dispositivo y de su propósito, es responsabilidad del desarrollador de un nuevo dispositivo, registrar correctamente los parámetros que requiera para poder ejecutar las actividades terapéuticas que proponga. De igual manera, es responsable de enviar en conjunto con los resultados del ejercicio y algún tipo de métrica general con la que se pueda evaluar el desempeño del paciente.

El dispositivo debe tomar las siguientes consideraciones para mejorar la robustez del sistema:

- En caso de perder conexión con el servidor de mensajes, solicitar nuevas terapias a partir de la última recibida.
- Si se solicitó información al servidor de mensajes y no se recibe respuesta luego de un periodo de tiempo, volver a solicitar la información.
- Esperar la confirmación de la recepción de los datos de desempeño de una terapia concluida. Si no se recibe la confirmación, reenviar la misma.

Cabe destacar, que con el desarrollo de los trabajadores, se incentiva a que los dispositivos realicen la comunicación exclusivamente a través de *MQTT*. Sin embargo, en caso de ser necesario, el dispositivo puede realizar peticiones directamente al *API*.

Todo dispositivo “inteligente” registrado en el sistema, es responsable de la administración y evaluación de la terapia. En este trabajo, se asume la existencia de dispositivos cuya efectividad médica fue previamente validada. Dichos dispositivos, pueden ser adaptados para que cumplan con las consideraciones anteriormente expuestas y se puedan incluir de manera transparente a la solución.

⁴<https://www.rabbitmq.com>

C. Capa de Aplicación

Comprende las aplicaciones de consulta y análisis de los datos terapéuticos, así como aplicaciones administrativas.

1) *Monitor de Actividades Terapéuticas (MAT)*: El Monitor de Actividades Terapéuticas es la aplicación Web encargada de gestionar la asignación y configuración de las actividades terapéuticas, así como actividades administrativas que corresponden: a la creación de usuarios, asignación de pacientes a médicos, gestión de los datos del paciente, entre otros.

Esta aplicación le permite tanto al terapeuta como al paciente: observar el progreso obtenido a lo largo de la terapia, mediante con un conjunto de estadísticas que indican el progreso en general del paciente; así como la posibilidad de consultar su desempeño en una actividad en particular. Tiene como función proveer al terapeuta de una interfaz que le permita: seleccionar las actividades terapéuticas que realizará el paciente, adaptar los parámetros en base a los resultados de las últimas sesiones y asignar el día de la misma.

El desarrollo de la aplicación se llevó con la librería de amplio uso para la construcción de interfaces de usuario React⁵, utilizando como cliente de GraphQL Apollo Client⁶. Mediante las características que provee GraphQL, la aplicación es capaz de recibir notificaciones en tiempo real al conectarse al servidor de gestión de mensajes.

V. PRUEBAS

Con el objetivo de comprobar la eficacia computacional del sistema, se desarrollaron dos tipos de pruebas:

- **Pruebas caja blanca:** Este tipo de pruebas tienen conocimiento total de la estructura interna de cada componente, por ende, analiza en detalle el correcto funcionamiento de las clases y funciones que lo conforman. Están compuestas por un conjunto de pruebas unitarias y funcionales para cada componente.
- **Pruebas caja negra:** Con este método se considera el sistema completo como una caja negra [34], por lo cual no se tiene ningún conocimiento acerca de la estructura interna de ninguno de los componentes que lo conforman. Corresponden a un conjunto de pruebas de integración, en las que se valida el correcto funcionamiento del sistema ante escenarios hipotéticos. Este tipo de pruebas son de mayor complejidad ya que consideran la interacción entre cada elemento presente en la arquitectura.

Para las pruebas de caja negra se consideró el caso de uso principal del sistema. En el mismo, un terapeuta asigna una sesión terapéutica a través del MAT. El dispositivo "inteligente" del paciente recibe la terapia asignada y se lo informa. El paciente efectúa la terapia asignada y el resultado de la misma es devuelto al personal médico.

En dicho caso, se plantea una evaluación basada simulación [35]: en donde el comportamiento de los dispositivos "inteligentes", paciente y médicos son simulados. Ya que la

interacción del paciente es exclusiva con el dispositivo, ambos serán considerados como un solo ente en las simulaciones; dicha asunción también aplica para el médico que interactúa exclusivamente con el MAT.

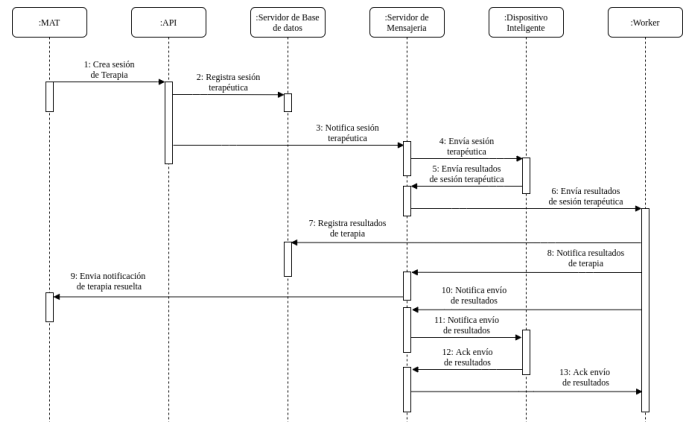


Figura 7: Caso Ideal

En la Figura 7, se puede observar como los mensajes fluyen a través de los componentes del sistema desde que se asigna una sesión de terapia al paciente, hasta que el mismo la realiza. En dicha Figura se expone el caso ideal, donde ningún componente falla y todos los mensajes se transmiten sin ningún tipo de inconveniente. En las pruebas realizadas, se simulan fallos de conexión en puntos claves del flujo de mensajes con el objeto de verificar la robustez del sistema. Es importante mencionar, que dichas pruebas no tienen la intención de validar la correcta transmisión de un mensaje entre dos componentes, ya que el protocolo de comunicación *MQTT* y los canales de transmisión la garantizan. Las pruebas, validan el envío de información entre los puntos finales del flujo, lo que involucra el paso de múltiples mensajes entre distintos componentes en el sistema.

A. Dispositivo Sin Conexión

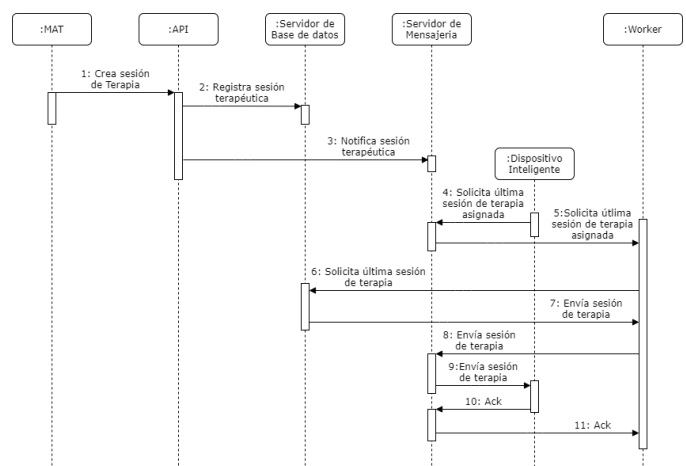


Figura 8: Dispositivo Sin Conexión

En la Figura 8 se muestra el flujo de mensajes que el sistema genera cuando el dispositivo "inteligente" al que se le asigna la sesión de terapia, no se encuentra conectado en el momento de

⁵<https://es.reactjs.org>

⁶<https://www.apollographql.com/docs/react>

dicha asignación. En este caso, como se puede observar en el paso 4, cuando el dispositivo restablece la conexión al sistema, el mismo solicita la última sesión de terapia que se le haya asignado. Dicha solicitud es atendida por un trabajador, el cual, le suministra la última sesión que le fuera asignada. Debido a la existencia de los trabajadores, el sistema puede continuar funcionando aunque el dispositivo no estuviese conectado en el momento que el terapeuta creó la sesión terapéutica.

B. Dispositivo Pierde Conexión Luego de Solicitud

Expande el caso planteado en la Figura 8, agregando un fallo de conexión al dispositivo luego de que este solicita la última sesión de terapia (paso 4). Dependiendo de la duración de la desconexión se presentan 4 posibles escenarios:

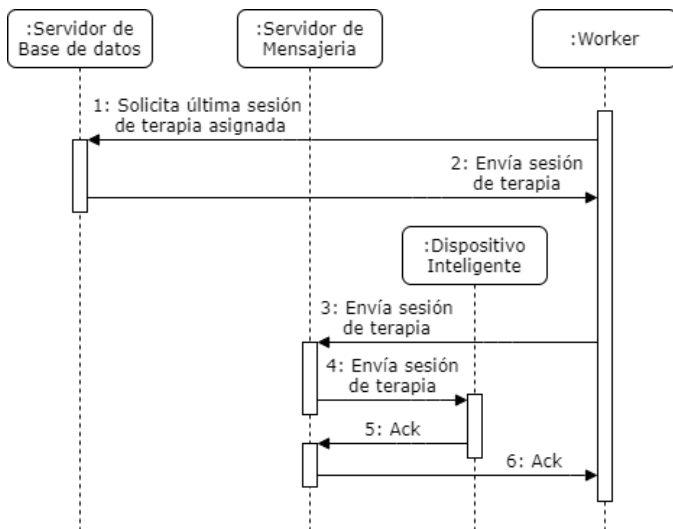


Figura 9: Reconexión Antes de Respuesta

1) *Restablecimiento de la Conexión Antes de Recibir la Respuesta:* El escenario presente en la Figura 9 corresponde con un periodo de desconexión muy corto. En el mismo, se puede observar que el dispositivo restablece la conexión antes de que el trabajador termine de procesar la solicitud. Debido a esto, el trabajador no se percata de la momentánea desconexión del dispositivo “inteligente” y el flujo de mensajes continúa de manera normal.

2) *Reintentos Fallidos:* En contraposición al caso anterior, en la Figura 10 se ejemplifica el funcionamiento del sistema cuando el dispositivo “inteligente” pierde la conexión con el sistema por un periodo de tiempo largo, luego de haber solicitado la última sesión de terapia. En este caso, el trabajador luego de esperar un tiempo por la confirmación de recepción (por parte del dispositivo), reenvía el resultado de la solicitud. Cuando el número de reintentos alcanza su límite, el trabajador descarta la solicitud. Cuando el dispositivo restablece la conexión, el mismo vuelve a solicitar la última sesión de terapia y el flujo continúa de manera normal como se establece en la Figura 8.

3) *Restablecimiento de la Conexión Durante Reintentos(1):* En la Figura 11 se plantea un tiempo de desconexión moderado

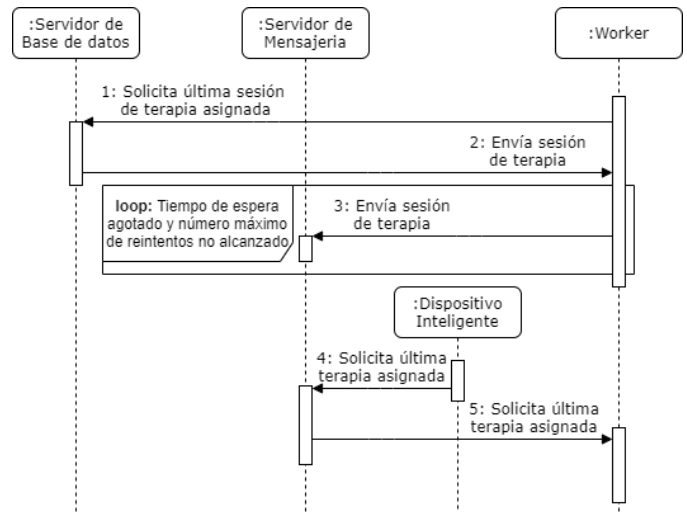


Figura 10: Reintentos Fallidos

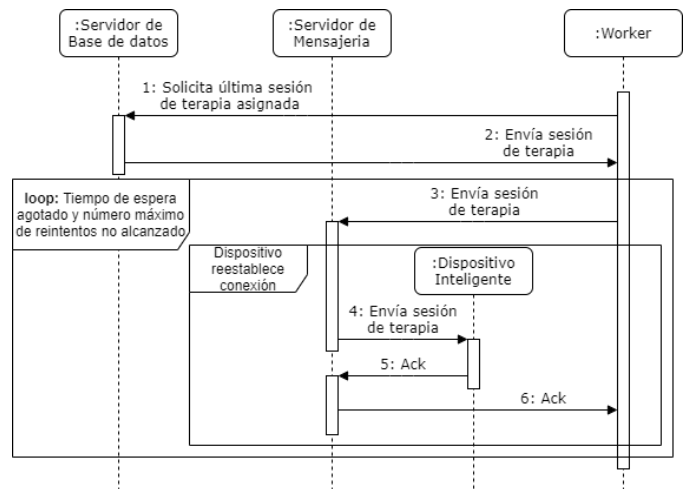


Figura 11: Reconexión Durante Reintentos(1)

luego de la solicitud. En este caso, el dispositivo restablece la conexión mientras el trabajador está reintentando enviar el resultado de la solicitud. Ya que el dispositivo se conecta durante el paso 3 y 4, recibe el resultado de la solicitud y el proceso continúa de manera normal.

4) *Restablecimiento de la Conexión Durante Reintentos(2):* Se expande el escenario anterior, incrementando el tiempo que el trabajador tarda en procesar la solicitud o disminuyendo el tiempo de espera por respuesta del dispositivo “inteligente”. En este caso (ver Figura 12), el dispositivo restablece la conexión durante el proceso de reintentos del dispositivo o durante el procesamiento de la solicitud. El dispositivo, espera un tiempo por la respuesta a su solicitud, pero al no recibirla, la reenvía (paso 4) y es atendida por otro trabajador (paso 5-9). Eventualmente, el dispositivo recibe la respuesta a la solicitud de ambos trabajadores y el flujo finaliza.

Aunque el sistema es capaz de trabajar ante el fallo, varios inconvenientes surgen del mismo:

- El dispositivo recibe la respuesta a su solicitud dos veces. Esto es resuelto por el mismo, ignorando una de las

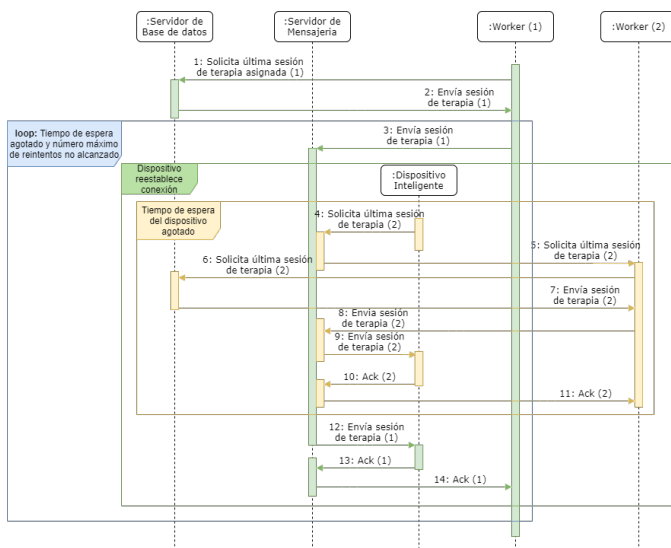


Figura 12: Reconexión Durante Reintentos(2)

respuestas.

- La solicitud es procesada dos veces. El segundo trabajador en atender la solicitud, verifica que no exista registro de los resultados a la terapia previamente indicada, lo que minimiza la posibilidad de generar incongruencias en la base de datos.
- Dependiendo de el tiempo de espera del dispositivo o de la resolución de las solicitudes, el dispositivo podría reiniciar múltiples veces la solicitud.

Estos problemas pueden solucionarse, agregando mecanismos que permitan al dispositivo verificar el estatus de la solicitud previa, antes de realizar una nueva solicitud.

C. Trabajador Falla Antes de Completar la Solicitud Asignada

El trabajador acepta una solicitud y falla antes de culminarla. En este caso, el gestor de mensajes resuelve el problema de manera automática. En el momento en el que el trabajador pierde la conexión, la solicitud que procesaba es devuelta a la cola de trabajos, donde otro trabajador se encarga de esta.

D. Fallo del Gestor de Mensajes

El gestor de mensajes representa el punto central de comunicación del sistema, mientras falle, todo el flujo se detiene. Si algún dispositivo se encontraba enviando alguna petición, el mismo continuará intentando realizarla hasta que el gestor de mensajes vuelva a funcionar.

Todos los mensajes que los trabajadores estuviesen atendiendo, son desechados, ya que los mismos no podrán recibir respuesta de confirmación por parte del solicitante. Sin embargo, cuando el gestor de mensajes reinicia, dichos trabajos son devueltos a la cola o son nuevamente solicitados por los dispositivos. Por ende, el sistema en las pruebas realizadas a baja escala, parece ser capaz de recuperarse a una falla de dicha magnitud.

Además de las pruebas realizadas para comprobar la eficacia computacional del sistema, se estimó que para transportarse

una distancia de 3.5 kilómetros en la ciudad de Caracas (Venezuela), utilizando medios de transporte público, una persona invierte un promedio de 2 horas. Si adicionalmente, se considera un promedio de 30 minutos de espera por atención en el centro médico y 30 minutos de ejecución de la terapia, un paciente invertiría un mínimo de 5 horas para asistir a la misma y regresar a su hogar. Del tiempo invertido, solo el 10% corresponde a la sesión terapéutica. Al realizar la terapia a través del sistema, los tiempos de traslado y espera son eliminados, lo que implica que el tiempo de atención se limita al de la terapia (30 minutos), dando como resultado una velocidad de atención 10 veces más rápida.

VI. CONCLUSIÓN Y TRABAJOS FUTUROS

En el presente trabajo se describió y probó una arquitectura basada en *IoT*, que sirve como herramienta de apoyo al proceso de recuperación de un paciente que requiere realizar terapia de rehabilitación. En base a los datos estimados, mediante el uso del sistema, el paciente puede incrementar la velocidad de atención en una sesión terapéutica al menos unas 10 veces, pudiendo incluso realizar la terapia desde la comodidad de su casa. Esto a su vez, propone una mejora para el personal médico, el cual podrá atender una mayor cantidad de pacientes y brindarles una atención más personalizada, lo que impactaría positivamente la velocidad de recuperación del paciente.

Aunque la arquitectura fue probada mediante un conjunto de pruebas unitarias y de integración, se considera necesario realizar pruebas a mayor escala como las propuestas en [35][36], con el objeto de someter a estrés al sistema y detectar una mayor cantidad de posibles fallos y comportamientos emergentes que permitan futuras optimizaciones.

De igual manera, es conveniente desarrollar numerosos dispositivos terapéuticos que abarquen distintos tipos de terapias y patologías médicas, utilizando variedad de sensores y datos entrada. En este sentido, actualmente se están desarrollando varios dispositivos y aplicaciones continuando el trabajo hecho en sistemas de detección de movimiento (*Kinect* y cámaras *web* de bajo costo) [28], además de otras implementaciones basadas en dispositivos móviles inteligentes (*Android*) y microcomputadores (*Raspberry Pi*). Estos desarrollos si bien no formaban parte de los objetivos de la presente propuesta, serán incluidos en futuras publicaciones donde se haga mayor énfasis en el funcionamiento y usabilidad desde la perspectiva de los clientes del sistema en la nube.

Por último, se propone desarrollar un sistema de recomendaciones de actividades terapéuticas, con el objeto de agilizar el proceso de diseño de la terapia de rehabilitación. El sistema de recomendación puede utilizar como insumo las actividades que le fueron asignadas previamente a pacientes con patologías similares, además de contar con un método de clasificación que pondere la efectividad de cada ejercicio, en base al progreso que obtuvo el paciente y al estado previo en el que se encontraba.

REFERENCIAS

- [1] World Health Organization and World Bank, *World Report on Disability*, World Health Organization, report, 2011.

- [2] World Health Organization, *Reducing Risk, Promoting Healthy Life*, World Health Organization, report, 2002.
- [3] Center for Disease Control and Prevention. *Disability and Health U.S. State Profile Data: Adults 18+ Years of Age*, <https://www.cdc.gov/nbddd/disabilityandhealth/impacts/index.html>
- [4] Instituto Nacional de Estadísticas, *Boletín de Mesa Técnica de Estadísticas de Discapacidad*, INE, reporte, 2014.
- [5] I. Lee and K. Lee, *The Internet of Things (IoT): Applications, Investments, and Challenges For Enterprises*, Business Horizons, vol. 58, no. 4, pp. 431–440, 2015.
- [6] A. Whitmore, A. Agarwal, and L. Xu, *The Internet of Things—a Survey of Topics and Trends*, Information Systems Frontiers, vol. 17, 2014.
- [7] S.M. North, M.M. North and J.R. Coble, *Virtual Reality Therapy: an Effective Treatment for Phobias*, Stud Health Technol Inform, no. 58, pp. 112–119, 1998.
- [8] S. Carlin, G. Hoffman and S. Weghorst, *Virtual Reality and Tactile Augmentation in the Treatment of Spider Phobia: a Case Report*, Behav Res Ther, vol. 2, no. 35, pp. 157–163, 1997.
- [9] G. Kim K. Yoo and J. Im, *A New Rehabilitation Training System for Postural Balance Control Using Virtual Reality Technology*, IEEE Trans Rehabil Eng, vol. 4, no. 7, pp. 482–485, 1999.
- [10] L. Piron, F. Cenni, P. Tonin and M. Dam, *Virtual Reality as an Assessment Tool for Arm Motor Deficits After Brain Lesions*, Stud Health Technol Inform, no. 81, pp. 386–392, 2001.
- [11] J. Fung et al., *Locomotor Rehabilitation in a Complex Virtual Environment*, Conf Proc IEEE Eng Med Biol Soc, vol. 4, no. 7, pp. 4859–4861, 2004.
- [12] J. Langan, K. DeLave, L. Phillips, P. Pangilinan, and H. Brown, *Home-Based Telerehabilitation Shows Improved Upper Limb Function in Adults With Chronic Stroke: a Pilot Study*, Journal of Rehabilitation Medicine, vol. 45, no. 2, 2013.
- [13] K.E. Laver, Z. Adey-Wakeling, M. Crotty, N.A. Lannin, S. George and C. Sherrington, *Telerehabilitation Services for Stroke*, Cochrane Database of Systematic Reviews, no. 1, 2020.
- [14] L. Tindall and R. Huebner, *The Impact of an Application of Telerehabilitation Technology on Caregiver Burden*, International Journal of Telerehabilitation, vol. 1, no. 1, 2009.
- [15] L. Atzori, A. Iera, and G. Morabito, *The Internet of Things: A Survey*, Computer Networks, vol. 54, no. 15, pp. 2787–2805, 2010.
- [16] A. Jara, F. Belchi, A. Alcolea, J. Santa, M. Zamora-Izquierdo, and A. Skarmeta, *A Pharmaceutical Intelligent Information System to Detect Allergies and Adverse Drugs Reactions Based on Internet of Things*, 2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops, PERCOM Workshops 2010, pp. 809–812, 2010.
- [17] H. Luo, S. Ci, D. Wu, N. Stergiou, and K.C. Siu, *A Remote Markerless Human Gait Tracking for E-Healthcare Based on Content-Aware Wireless Multimedia Communications*, IEEE Wireless Communication, vol. 17, pp. 44–50, 2010.
- [18] G. Nussbaum, *People With Disabilities: Assistive Homes and Environments*, in Computers Helping People with Special Needs, K. Miesenberger, J. Klaus, W. L. Zagler, and A. I. Karshmer, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 457–460, 2006.
- [19] G.H. Khan, A.P. Manwatkar, N.P. Khan, A.D. Bhotmange, and N. Meshram, *Real-time IoT-Based Health Care Monitoring for Prediction and Analysis*, Our Heritage, vol. 68, no. 30, pp. 217–220, 2020.
- [20] V.M. Rohokale, N.R. Prasad, and R. Prasad, *A Cooperative Internet of Things (IoT) for Rural Healthcare Monitoring and Control*, en 2011 2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace Electronic Systems Technology (Wireless VITAE), pp. 1–6, 2011.
- [21] P. Verma and S. Sood, *Cloud-Centric IoT Based Disease Diagnosis Healthcare Framework*, Journal of Parallel and Distributed Computing, 2017.
- [22] C. Fernandes and C. Lucena, *A Software Framework for Remote Patient Monitoring by Using Multi-Agent Systems Support*, JMIR Medical Informatics, vol. 5, p. e9, 2017.
- [23] M.C. Domingo, *An Overview of the Internet of Things for People With Disabilities*, Journal of Network and Computer Applications, vol. 35, no. 2, pp. 584–596, 2012.
- [24] N. Bui and M. Zorzi, *Health Care Applications: A Solution Based on the Internet of Things*, in Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, ISABEL '11. New York, NY, USA: ACM, pp. 131:1–131:5, 2011
- [25] A. Dohr, R. Modre-Osprian, M. Drobnic, D. Hayn, and G. Schreier, *The Internet of Things for Ambient Assisted Living*, Third International Conference on Information Technology: New Generations, pp. 804–809, 2010.
- [26] S. Vukicevic, Z. Stamenkovic, S. Murugesan, Z. Bogdanović, and B. Radenković, *A New Telerehabilitation System Based on Internet of Things*, Facta Universitatis Series Electronics and Energetics, vol. 29, pp. 395–405, 2016.
- [27] B. Ferreira and P. Menezes, *Gamifying Motor Rehabilitation Therapies: Challenges and Opportunities of Immersive Technologies*, Information, vol. 11, no. 02, 2020.
- [28] F. Moreno, J. Ojeda, E. Ramírez, C. Mena, O. Rodríguez, J. Rangel, y S. Álvarez, *Un Framework para la Rehabilitación Física en Miembros Superiores con Realidad Virtual*, Proceedings del I Congreso Nacional de Computación, Informática y Sistemas (CoNCISa), vol. 1, no. 1, pp. 77–84, 2013.
- [29] M. Team. *NoSQL Databases Explained*, <https://www.mongodb.com/nosql-explained>
- [30] M. Team. *Top 5 Considerations when Evaluating NoSQL Databases*, <https://www.mongodb.com/collateral/top-5-considerations-when-evaluating-nosql-databases>
- [31] Facebook. *GraphQL*, <https://graphql.github.io/graphql-spec>
- [32] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, *Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications*, IEEE Communication Surveys and Tutorials, vol. 17, no. 4, 2015.
- [33] S. Vinoski, *Advanced Message Queuing Protocol*, IEEE Internet Computing, vol. 10, no. 6, pp. 87–89, 2006.
- [34] B. Beizer, *Black-Box Testing: Techniques for Functional Testing of Software and Systems*. J. Wiley, 1995.
- [35] S. Bosmans, S. Mercelis, P. Hellinckx, and J. Denil, *Towards Evaluating Emergent Behavior of the Internet of Things Using Large Scale Simulation Techniques*, 2018.
- [36] S. Bosmans, S. Mercelis, J. Denil, and P. Hellinckx, *Testing IoT Systems Using a Hybrid Simulation Based Testing Approach*, Computing, 2018.

Índice de Autores

B

Briceño Arlán 1

G

García Jean 19

H

Hernández Walter 28

M

Moreno Francisco 28

N

Núñez Haydemar 1

R

Ramírez Esmitt 10

Rodríguez Aneriz 1

S

Sanoja Andres 19

U

Urdaneta Adriana 10

REVECOM

Sociedad Venezolana de Computación

La Sociedad Venezolana de Computación está comprometida con el impulso de una nueva generación académica y profesional en nuestra área de saber para el desarrollo del país.

Los conceptos y puntos de vista expresados en los trabajos publicados en este libro representan las opiniones personales de los autores y no reflejan el juicio de los editores o de la Sociedad Venezolana de Computación.

ISSN: 2244-7040



9 772244 704006

www.svc.net.ve/revecom

