



**EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA

**Instituto Tecnológico de Orizaba**

**División de Estudios de Posgrado e Investigación**

**Maestría en Sistemas Computacionales**

**Proyecto de Tesis**

**Título del proyecto: Desarrollo de un sistema de gestión de datos multimedia que utilice fragmentación horizontal para la optimización de consultas basadas en contenido**

**Ing. Ind. Marcos Joaquín Rodríguez Arauz M13010500**

**PARA OBTENER EL GRADO DE:**

**Maestro en Sistemas Computacionales**

**DIRECTOR DEL PROYECTO DE TESIS:**

**Dra. Lisbeth Rodríguez Mazahua**



Instituto Tecnológico de Orizaba  
División de Estudios de Posgrado e Investigación

Orizaba, Ver., **11/junio/2021**  
Dependencia: **División de Estudios de  
Posgrado e Investigación**  
Asunto: **Autorización de Impresión**  
OPCION: I

**C. MARCOS JOAQUÍN RODRÍGUEZ ARAUZ**

Candidato a Grado de Maestro en:  
**SISTEMAS COMPUTACIONALES**  
**P R E S E N T E.-**

De acuerdo con el Reglamento de Titulación vigente de los Centros de Enseñanza Técnica Superior, dependiente de la Dirección General de Institutos Tecnológicos de la Secretaría de Educación Pública y habiendo cumplido con todas las indicaciones que la Comisión Revisora le hizo respecto a su Trabajo Profesional titulado:

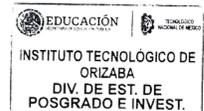
**" Desarrollo de un Sistema de gestión de datos multimedia que utilice fragmentación horizontal para la optimización de consultas basadas en contenido"**

Comunico a Usted que este Departamento concede su autorización para que proceda a la impresión del mismo.

**A T E N T A M E N T E**

Excelencia en Educación Tecnológica®  
**CIENCIA – TÉCNICA - CULTURA®**

**DR. MARIO LEONCIO ARRIJOA RODRÍGUEZ**  
**JEFE DE LA DIVISIÓN DE ESTUDIOS**  
**DE POSGRADO E INVESTIGACIÓN**



Avenida Oriente 9 No. 852  
Col. Emiliano Zapata, C.P. 94320  
Orizaba, Veracruz, México.  
Teléfono: 272-110-53-60  
Email: [cyd\\_orizaba@tecnm.mx](mailto:cyd_orizaba@tecnm.mx)  
[www.orizaba.tecnm.mx](http://www.orizaba.tecnm.mx)





Instituto Tecnológico de Orizaba  
División de Estudios de Posgrado e Investigación

Orizaba, Ver., 19/febrero/2021

Asunto: Revisión del Trabajo Escrito

**C. MARIO LEONCIO ARRIJOA RODRIGUEZ**  
**JEFE DE LA DIVISION DE ESTUDIOS**  
**DE POSGRADO E INVESTIGACION**  
**P R E S E N T E.**

Los que suscriben miembros del jurado, han realizado la revisión de la Tesis del (la) C.:

**MARCOS JOAQUÍN RODRÍGUEZ ARAUZ**  
**No. DE CONTROL: M13010500**

la cual lleva el título de:

**“Desarrollo de un Sistema de gestión de datos multimedia que utilice fragmentación horizontal para la optimización de consultas basadas en contenido”**

y concluyen que se acepta.

**A T E N T A M E N T E**

PRESIDENTE: DRA. LISBETH RODRÍGUEZ MAZAHUA

SECRETARIO: DR. MARIO L. ARRIJOA RODRÍGUEZ

VOCAL: M.C. MA. ANTONIETA ABUD FIGUEROA

VOCAL SUP.: DR. JOSÉ LUIS SÁNCHEZ CERVANTES

\_\_\_\_\_  
  
\_\_\_\_\_  
*ma. Antonieta Abud F.*  
\_\_\_\_\_  
  
\_\_\_\_\_

EGRESADO (A) DE LA MAESTRIA EN SISTEMAS COMPUTACIONALES

OPCION: I Tesis



Avenida Oriente 9 No. 852  
Col. Emiliano Zapata, C.P. 94320  
Orizaba, Veracruz, México.  
Teléfono: 272-110-53-60  
Email: depi\_orizaba@tecnm.mx  
www.orizaba.tecnm.mx



# Índice

|             |  |    |
|-------------|--|----|
| Capítulo 1. | Antecedentes .....   | 11 |
| 1.1         | Marco teórico.....   | 11 |
| 1.1.1       | Bases de datos multimedia .....                                  | 11 |
| 1.1.2       | Consultas basadas en contenido.....                              | 11 |
| 1.1.3       | Consultas rango.....   | 12 |
| 1.1.4       | Consultas kNN .....  | 12 |
| 1.1.5       | Fragmentación .....  | 12 |
| 1.1.6       | Tipos de fragmentación.....                                      | 13 |
| 1.2         | Situación tecnológica, económica y operativa de la empresa ..... | 14 |
| 1.3         | Planteamiento del problema .....                                 | 14 |
| 1.4         | Objetivos general y específicos .....                            | 14 |
| 1.4.1       | Objetivo general.....  | 14 |
| 1.4.2       | Objetivos específicos.....                                       | 15 |
| 1.4.3       | Justificación .....  | 15 |
| Capítulo 2. | Estado de la práctica .....                                      | 16 |
| 2.1         | Trabajos relacionados .....                                      | 16 |
| 2.2         | Análisis comparativo .....                                       | 20 |
| 2.3         | Propuesta de la solución .....                                   | 25 |
| Capítulo 3. | Aplicación de la Metodología.....                                | 26 |
| 3.1         | Análisis.....  | 26 |
| 3.2         | Selección.....   | 29 |
| 3.3         | Desarrollo del sistema.....                                      | 29 |
| 3.3.1       | Análisis de requisitos.....                                      | 30 |
| 3.3.2       | Modelo Conceptual .....  | 41 |
| 3.3.3       | Modelo de Navegación .....                                       | 45 |
| 3.3.4       | Modelo de presentación .....                                     | 47 |
| 3.3.5       | Modelo de procesos.....  | 52 |
| Capítulo 4. | Resultados .....   | 63 |
| 4.1         | Resultados del análisis .....                                    | 63 |
| 4.2         | Fragmentación Horizontal.....                                    | 67 |

|             |  |    |
|-------------|--|----|
| 4.3         | Sistema de Gestión de Datos Multimedia que utiliza un método de fragmentación horizontal para optimizar consultas basadas en contenido ..... | 68 |
| 4.4         | Evaluación de resultados.....  | 81 |
| Capítulo 5. | Conclusiones y recomendaciones .....   | 84 |
| 5.1         | Conclusiones.....  | 84 |
| 5.2         | Recomendaciones .....  | 85 |
|             | Productos Académicos.....  | 85 |
|             | Referencias.....   | 86 |
|             | Anexo.....   | 92 |

## Índice de tablas

|   |    |
|---|----|
| Tabla 2.1 Análisis comparativo de los trabajos relacionados ..... | 21 |
| Tabla 3.1 Análisis de Artículos. ....                             | 28 |
| Tabla 3.2 Actores de la aplicación.....                           | 30 |
| Tabla 4.1 Predicados usados por consultas.....                    | 82 |
| Tabla 4.2 Fragmentos de la base de datos multimedia.....          | 82 |
| Tabla 4.3 Matrix de Uso de Predicados.....                        | 82 |
| Tabla 4.4 Matrix de Uso de Fragmento-Predicado.....               | 83 |

## Índice de figuras

|   |    |
|---|----|
| Figura 3.1 Metodología de búsqueda y evaluación de los trabajos relacionados. ....  | 26 |
| Figura 3.2 Estructura de la solución de la propuesta .....                          | 30 |
| Figura 3.3 Diagrama de casos de uso.....  | 31 |
| Figura 3.4 Diagrama de caso de uso Iniciar sesión.....                              | 32 |
| Figura 3.5 Diagrama de caso de uso Cerrar sesión .....                              | 32 |
| Figura 3.6 Diagrama de caso de uso Registrar usuario .....                          | 33 |
| Figura 3.7 Diagrama de caso de uso Validar Información.....                         | 33 |
| Figura 3.8 Diagrama de caso de uso Proponer Contenido .....                         | 34 |
| Figura 3.9 Diagrama de caso de uso Alta Contenido .....                             | 35 |
| Figura 3.10 Diagrama de caso de uso Modificar Contenido.....                        | 36 |
| Figura 3.11 Diagrama de caso de uso Baja de Contenido .....                         | 36 |
| Figura 3.12 Diagrama de caso de uso Alta Usuario .....                              | 37 |
| Figura 3.13 Diagrama de caso de uso Baja de Usuario .....                           | 37 |
| Figura 3.14 Diagrama de caso de uso Modificar Usuario.....                          | 38 |
| Figura 3.15 Diagrama de caso de uso Consultar Usuario.....                          | 39 |
| Figura 3.16 Diagrama de caso de uso Consulta con Base en Texto .....                | 39 |
| Figura 3.17 Diagrama de caso de uso Consulta con base en contenido .....            | 40 |
| <i>Figura 3.18</i> Diagrama conceptual.....   | 41 |
| Figura 3.19 Diagrama Lógico .....   | 42 |
| Figura 3.20 Diagrama Físico .....   | 43 |
| Figura 3.21 Diagrama de Navegación.....   | 46 |
| Figura 3.22 Página de Inicio del modelo de presentación.....                        | 47 |
| Figura 3.23 Página Iniciar Sesión del modelo de presentación.....                   | 48 |
| Figura 3.24 Página Contenidos del modelo de presentación.....                       | 49 |
| Figura 3.25 Página Proponer Contenido del modelo de presentación.....               | 49 |
| Figura 3.26 Página Búsqueda de Contenido del modelo de presentación.....            | 50 |
| Figura 3.27 Página Gestionar Usuarios del modelo de presentación.....               | 51 |
| Figura 3.28 Página Gestionar Contenido del modelo de presentación.....              | 51 |
| Figura 3.29 Diagrama de proceso del caso de uso Iniciar Sesión.....                 | 52 |
| Figura 3.30 Diagrama de proceso del caso de uso Cerrar Sesión.....                  | 53 |
| Figura 3.31 Diagrama de proceso del caso de uso Registrar Usuario.....              | 53 |
| Figura 3.32 Diagrama de proceso del caso de uso Validar Información.....            | 54 |
| Figura 3.33 Diagrama de proceso del caso de uso Proponer Contenido.....             | 55 |
| Figura 3.34 Diagrama de proceso del caso de uso Alta Contenido.....                 | 56 |
| Figura 3.35 Diagrama de proceso del caso de uso Modificar Contenido.....            | 56 |
| Figura 3.36 Diagrama de proceso del caso de uso Baja de Contenido.....              | 57 |
| Figura 3.37 Diagrama de proceso del caso de uso Alta de Usuario.....                | 58 |
| Figura 3.38 Diagrama de proceso del caso de uso Baja de Usuario.....                | 58 |
| Figura 3.39 Diagrama de proceso del caso de uso Modificar Usuario.....              | 59 |
| Figura 3.40 Diagrama de proceso del caso de uso Consultar Usuario.....              | 60 |
| Figura 3.41 Diagrama de proceso del caso de uso Consulta con base en texto.....     | 61 |
| Figura 3.42 Diagrama de proceso del caso de uso Consulta con base en contenido..... | 62 |

## Resumen

El objetivo del presente trabajo es desarrollar un sistema gestor de datos multimedia que utilice un método de fragmentación horizontal, para optimizar las consultas basadas en contenido. El sistema se encargará de almacenar y gestionar los datos multimedia históricos del Instituto Tecnológico de Orizaba.

La importancia de este sistema radica en que ayudará a reducir el porcentaje de personas que ignoran gran parte de la historia del Instituto; también, permitirá que cualquier persona agregue información histórica a la base de datos multimedia del sistema, siempre y cuando pase por un proceso de supervisión para disminuir la posibilidad de agregar información falsa o inadecuada; y finalmente, logrará optimizar consultas basadas en contenido.

Las tecnologías seleccionadas para el desarrollo de este trabajo son: como marco de trabajo JavaServer Faces, el sistema gestor de bases de datos (SGDB) MongoDB y como IDE (*Integrated Development Environment*, Entorno de Desarrollo Integrado) NetBeans, ya que éste es el entorno que mejor se adecúa para el uso de estas tecnologías.

## **Abstract**

The objective of this work is to develop a multimedia data management system that uses a horizontal fragmentation method to optimize content-based queries. The system will be in charge of storing and managing the historical multimedia data of the Orizaba Technological Institute.

The importance of this system is that it will help reduce the percentage of people who ignore a large part of the Institute's history; Also, it will allow anyone to add historical information to the system's multimedia database, as long as it goes through a monitoring process to reduce the possibility of adding false or inappropriate information; and finally, it will be able to optimize content-based queries.

The technologies selected for the development of this work are: as for the framework uses JavaServer Faces, the database management system (SGDB) MongoDB and as IDE (Integrated Development Environment) NetBeans, since this is the environment that is best suited for the use of these technologies.

## **Introducción**

En el año 2000, se realizó un proyecto en el Instituto Tecnológico de Orizaba, donde se recuperaron una gran cantidad de datos multimedia. Al finalizar el proyecto, se detectó que al aumentar el contenido multimedia se volvía cada vez más difícil gestionar estos datos de forma óptima.

Este trabajo busca resolver la problemática antes mencionada, ya que se desarrolló un sistema para la gestión de los datos históricos del Instituto, el cual utiliza fragmentación horizontal para optimizar consultas basadas en contenido. Para lograr esto es necesario hacer un análisis de las técnicas de fragmentación horizontal existentes, elegir una para su posterior implementación y validación dentro del sistema que se desarrollará. De esta manera se mejorará la gestión de los datos históricos y se reducirá el tiempo de respuesta y el costo de ejecución de las consultas basadas en contenido.

Este trabajo se organiza en cinco capítulos principales: en el capítulo uno se abarca el marco teórico, planteamiento del problema, objetivo general y específicos y la justificación; el capítulo dos comprende el estado de la práctica y un análisis comparativo; el capítulo tres abarca un análisis de los métodos de fragmentación horizontal y el diseño de la aplicación Web para aplicar una técnica seleccionada del análisis; el capítulo cuatro describe los resultados obtenidos por la aplicación Web desarrollada y la descripción un caso de estudio para validarla; por último, en el capítulo cinco se presentan las conclusiones y recomendaciones.

# Capítulo 1. Antecedentes

En este capítulo se presentan algunos aspectos primordiales de la investigación como lo son: el marco teórico, el problema a resolver, los objetivos generales y específicos, así como la justificación de lo que se realizará.

## 1.1 Marco teórico

A continuación, se definen algunos términos relevantes para el trabajo de investigación.

### 1.1.1 Bases de datos multimedia

Los datos multimedia, como imágenes, audio y video, hoy en día casi siempre se almacenan fuera de las bases de datos, como en sistemas de archivos. Este tipo de almacenamiento no es un problema cuando la cantidad de objetos multimedia es relativamente pequeña, ya que las características proporcionadas por las bases de datos generalmente no son necesarias. Sin embargo, las características de la base de datos se vuelven importantes cuando el número de objetos multimedia almacenados aumenta. Las cuestiones como actualizaciones transaccionales, facilidades de consulta e indexación se vuelven importantes. Los objetos multimedia a menudo tienen atributos descriptivos, como los que indican cuándo se crearon, quién los creó y a qué categoría pertenecen. Un enfoque para construir una base de datos para tales objetos multimedia es usar bases de datos para almacenar los atributos descriptivos y para realizar un seguimiento de los archivos en los que se almacenan los objetos multimedia [1].

### 1.1.2 Consultas basadas en contenido

Los sistemas CBIR (*Content-Based Image Retrieval*, Recuperación de imágenes Basada en Contenido) emplean un conjunto de técnicas para administrar imágenes digitales basadas en su contenido visual. Para representar los datos, un sistema CBIR envía las imágenes a un extractor de características, que genera un vector que representa características específicas del contenido visual de las imágenes. Los ejemplos incluyen extractores de características basados en color, textura y forma. Los vectores de características generalmente no se comparan adecuadamente a través de operadores relacionales habituales ( $=$ ,  $<$ ,  $\leq$ ,  $>$  y  $\geq$ ) porque carecen de una relación de orden, es decir, la capacidad de determinar cuál

de los dos vectores de características precede al otro. Por lo tanto, los vectores de características normalmente se comparan por similitud. Dado un par de vectores de características, se emplea una función de distancia para determinar el grado de similitud entre ellos [2].

### **1.1.3 Consultas rango**

Se representan por  $R_q(s_q, \xi)$ , donde  $s_q$  es el vector de características del elemento de consulta y  $\xi$  es el radio, estas consultas recuperan todos los elementos dentro de una distancia  $\xi$  de  $s_q$ . Por ejemplo: devuelve los mamogramas que difieren del que se está analizando en hasta una unidad. Este tipo de consulta se aplica a escenarios específicos, donde un especialista conoce un valor apropiado de  $\xi$  para el dominio de datos, considerando los extractores de características y las funciones de distancia que se utilizan [2].

### **1.1.4 Consultas kNN**

Representado como  $kNN_q(s_q, k)$ , donde  $s_q$  es el vector de características del elemento de consulta y  $k$  es el número de elementos que se devolverán, estas consultas recuperan los  $k$  elementos más similares a  $s_q$ , generalmente ordenados desde el más cercano (primer elemento) al más lejano ( $k$ -ésimo elemento). Teniendo en cuenta un escenario en el que un médico necesita dar un diagnóstico, es posible que desee analizar una serie de casos similares para aumentar su confianza en el proceso de toma de decisiones. En este escenario, ese número sería el parámetro  $k$  [2].

### **1.1.5 Fragmentación**

La fragmentación es la separación de una base de datos en particiones llamadas fragmentos. La importancia de la fragmentación radica en que permite reducir el número de accesos a datos irrelevantes por lo que logra reducir el tiempo de respuesta y el costo de ejecución de las consultas. Hay dos tipos generales de alternativas de fragmentación [2].

## **1.1.6 Tipos de fragmentación**

### **1.1.6.1 Fragmentación horizontal**

Cada fragmento tiene un subconjunto de las tuplas de la relación. Hay dos tipos de fragmentación horizontal: primaria y derivada. La fragmentación horizontal primaria de una relación se realiza utilizando predicados que están definidos en esa relación. La fragmentación horizontal derivada, por otro lado, es la partición de una relación que resulta de la definición de predicados en otra relación [3].

### **1.1.6.2 Fragmentación vertical**

La fragmentación vertical de una relación  $R$  produce fragmentos  $R_1, R_2, \dots, R_n$ , cada uno de los cuales contiene un subconjunto de atributos de  $R$ , así como la clave primaria de  $R$ . El objetivo de la fragmentación vertical es dividir una relación en un conjunto de relaciones más pequeñas para que muchas de las aplicaciones de usuario se ejecuten en un solo fragmento. En este contexto, una fragmentación "óptima" es aquella que produce un esquema de fragmentación que minimiza el tiempo de ejecución de las aplicaciones de usuario que se ejecutan en estos fragmentos. La fragmentación vertical ha sido investigada dentro del contexto de sistemas de bases de datos centralizados, así como de sistemas distribuidos. Su motivación dentro del contexto centralizado es debido a que se considera una herramienta de diseño, que permite que el usuario consulte relaciones más pequeñas, lo que provoca un menor número de accesos a las páginas [3].

### **1.1.6.3 Fragmentación híbrida**

En la mayoría de los casos, una simple fragmentación horizontal o vertical de un esquema de base de datos no será suficiente para satisfacer los requisitos de las aplicaciones de uso. En este caso, la fragmentación vertical puede ser seguida por una horizontal, o viceversa, produciendo una partición estructurada en árbol. Dado que los dos tipos de estrategias de partición se aplican una tras otra, esta alternativa se llama fragmentación híbrida. También se le llama fragmentación mixta o fragmentación anidada [3].

## **1.2 Situación tecnológica, económica y operativa de la empresa**

El Instituto Tecnológico de Orizaba (ITO) es una institución que pertenece al Tecnológico Nacional de México, que se encuentra ubicado en Oriente 9, Colonia Emiliano Zapata, en la ciudad de Orizaba, Veracruz. Esta institución ofrece carreras de licenciatura, maestría y doctorado. En el área de maestrías ofrece las carreras de Maestría en Ingeniería Electrónica, Maestría en Ingeniería Industrial, Maestría en Ciencias en Ingeniería Química, Maestría en Ingeniería Administrativa y la Maestría en Sistemas Computacionales.

## **1.3 Planteamiento del problema**

Para mantener la memoria histórica del Instituto Tecnológico de Orizaba se llevó a cabo un proyecto de recopilación, digitalización y sistematización del material histórico del Instituto Tecnológico de Orizaba [4], el cual permitió recolectar una gran cantidad de datos multimedia. Sin embargo, al finalizar el proyecto se determinó que conforme el volumen de información disponible aumentaba, era más difícil la gestión de dichos datos. Además, en una tesis anterior [5], se desarrolló un sistema que permite realizar consultas basadas en contenido en una base de datos multimedia, y se encontró que uno de los problemas con dichas consultas es que conforme aumenta el número de objetos multimedia almacenados en la base, las consultas se vuelven muchos más lentas. Por lo tanto, en este caso, se desarrolló un sistema que permita gestionar los datos históricos del Instituto y que además aplica un método de fragmentación horizontal para la optimización de consultas basadas en contenido.

## **1.4 Objetivos general y específicos**

A continuación, se muestran el objetivo general y los específicos.

### **1.4.1 Objetivo general**

Desarrollar un sistema para la gestión de datos históricos del Instituto Tecnológico de Orizaba que aplique un método de fragmentación horizontal para optimizar consultas basadas en contenido.

### **1.4.2 Objetivos específicos**

1. Estudiar los conceptos principales de bases de datos multimedia, consultas basadas en contenido y fragmentación para entender el alcance de la investigación.
2. Realizar un análisis comparativo de los métodos de fragmentación horizontal para bases de datos multimedia para elegir el que se utilizará en el sistema.
3. Seleccionar las técnicas y tecnologías que se usarán para el desarrollo del sistema.
4. Determinar los requerimientos del sistema.
5. Realizar el diseño conceptual, lógico y físico de la base de datos multimedia para gestionar la información requerida.
6. Diseñar el sistema con base en la metodología seleccionada.
7. Implementar el sistema.
8. Validar el sistema por medio de un caso de estudio.

### **1.4.3 Justificación**

Debido a las necesidades presentadas es necesario mostrar en este trabajo un enfoque de fragmentación horizontal para mejorar el desempeño de la base de datos multimedia. El sistema que se desarrolló ayudará a reducir el porcentaje de personas que ignoran gran parte de la historia del Instituto; también, permitirá que cualquier persona agregue información histórica a la base de datos multimedia del sistema, siempre y cuando pase por un proceso de supervisión para disminuir la posibilidad de agregar información falsa o inadecuada; y finalmente, logrará reducir el tiempo de respuesta y costo de ejecución de consultas basadas en contenido por medio de la fragmentación horizontal. Cabe resaltar que, para evitar riesgos de seguridad personal, se mostrará información general a todos los usuarios, pero solo los usuarios registrados podrán acceder a su información personal. Los principales beneficiados con el desarrollo del sistema serán los estudiantes y profesores del Instituto Tecnológico de Orizaba, ya que serán capaces de acceder eficientemente a información histórica del Instituto.

## Capítulo 2. Estado de la práctica

Hoy en día los sistemas de recuperación de imágenes basada en contenido son una necesidad para el uso óptimo de bases de datos multimedia. También es muy común observar trabajos propuestos con el uso de diferentes métodos de fragmentación, es por eso que para proponer un enfoque relacionado con estas tecnologías es necesario primero analizar trabajos similares. En este capítulo se describen un conjunto de trabajos meticulosamente escogidos, los cuales se enfocan en el uso de sistemas CBIR usando algoritmos para optimizar y mejorar su rendimiento, así como en el desarrollo de métodos de fragmentación para bases de datos multimedia, posteriormente se realizó una síntesis por medio de una tabla y finalmente se analizó la propuesta exhibida en este trabajo.

### 2.1 Trabajos relacionados

La mayoría de los algoritmos de fragmentación vertical para bases de datos multimedia, son estáticos, lo que significa: que si se intenta optimizar un VPS (*Vertical Partitioning Scheme*, Esquema de Partición Vertical), de acuerdo a una carga de trabajo y ésta experimenta cambios, el VPS podría ser degradado, lo cual representaría largos tiempos de búsqueda. En [6] se presentó un sistema llamado DYMOND (*DYnamic Multmedia ONline Distribution*, Distribución dinámica multimedia en línea) el cual usa un grupo de reglas activas para ejecutar particiones verticales dinámicas en bases de datos multimedia. Como resultado se encontró que DYMOND mejoró el rendimiento de consultas en bases de datos multimedia.

La empresa AGROMAQ, que se dedica a la venta de maquinaria para la agricultura y repuestos, utiliza personal altamente capacitado para ayudar al cliente a identificar los productos, lo que provoca un incremento considerable de tiempo a la atención al cliente no solo en horas, sino también en días, debido a que los clientes no poseen el conocimiento necesario para identificar el tipo de maquinaria que necesitan. Rojas et al. [7] propusieron un sistema CBIR para el reconocimiento de maquinaria agrícola y repuestos, utilizando dos descriptores: SURF (*Speeded Up Robust Features*, Características Robustas Aceleradas) Y SIFT (*Scale Invariant Feature*

*Transform*, Transformación de Características Invariante de Escala). Los resultados demostraron que el descriptor SURF es más veloz y obtiene mayor precisión en comparación con el descriptor SIFT.

La fragmentación híbrida es una técnica utilizada para la optimización de consultas tanto en bases de datos relacionales como en bases de datos orientadas a objetos. En los últimos años, la creciente disponibilidad de aplicaciones multimedia se convirtió en objetivo de interés para utilizar técnicas de fragmentación para reducir el número de páginas requeridas para responder una consulta. Rodríguez-Mazahua et al. [8] propusieron un método de fragmentación híbrida para bases de datos multimedia, MHYP (*Multimedia Hybrid Partitioning Algorithm*, Algoritmo de fragmentación híbrida multimedia), junto con un modelo de costos. Los análisis experimentales demostraron que el algoritmo propuesto para este trabajo supera en rendimiento tanto a la fragmentación horizontal como vertical en la mayoría de los casos.

La necesidad de una recuperación eficiente de imágenes incrementa rápidamente y por eso, para mejorar el rendimiento y reducir el margen entre características visuales, se usa el proceso CBIR. En [9], se propuso un sistema CBIR basado en SURF con algoritmos genéticos para su optimización. Los resultados mostraron un aumento de rendimiento, después de las simulaciones se obtuvo una precisión de más del 98%, este sistema se implementó utilizando la caja de herramientas de procesamiento de imágenes del software MATLAB.

En [10] se presentó un sistema CBIR que usa técnicas de computación suave como: Redes Neuronales Artificiales, Lógica Difusa y SVM (*Support Vector Machine*, Máquina de soporte vectorial). El sistema propuesto fue evaluado pragmáticamente en términos de precisión y exactitud. Los resultados mostraron un enorme incremento en el rendimiento de recuperación de imágenes.

Con el gran crecimiento de dispositivos multimedia se generan grandes cantidades de imágenes cada día. El rápido acceso a estas enormes colecciones de imágenes significan retos cada vez más grandes y necesidad de algoritmos muy eficientes.

Ouhda et al. [11] aplicaron un nuevo método basado la técnica de agrupamiento *K-Means*, para proveer resultados exactos con menor tiempo de computación. Para la validación, se aplicó este sistema en dos bases de datos multimedia y obtuvo los resultados que los autores de este trabajo requerían.

Las técnicas CBIR se usan para recuperar imágenes de bases de datos masivas. En [12] se mostró un sistema CBIR con dos métodos de extracción de características: SIFT y ORB (*Oriented Fast Rotated and BRIEF*, Girado Rápido Orientado y BRIEF). ORB usa FAST (*Features from Accelerated Segment Test*, Características de Pruebas de Segmentos Acelerados), el cual es un detector de puntos clave y BRIEF (*Binary Robust Independent Elementary Features*, Características Elementales Independientes Robustas Binarias), como descriptor de una imagen. El método de agrupamiento *K-Means* se usó para analizar los datos, el cual genera grupos usando el vector descriptor. Se logró como resultado un porcentaje de precisión de 88.9%.

En [13] se propuso un sistema CBIR usando ANN (*Artificial Neural Network*, Red Neuronal Artificial), el cual se compone de tres fases principales: extracción de características, clasificación ANN y estrategia de correspondencia. En la primera fase se extraen las características locales y globales y se usa un filtro Gabor para mejorar las imágenes y lograr que mejores resultados sean obtenidos. La ANN sirve como clasificador para que las características de las imágenes que están siendo usadas sirvan para obtener mayores similitudes en la búsqueda de la imagen. Por último, la estrategia de correspondencia se usa para recuperar imágenes de la base de datos como resultado. Los resultados en este trabajo usando el sistema propuesto son precisos y toman menos tiempo de computación.

Buscar una imagen en volúmenes masivos de bases de datos de imágenes con la semántica deseada y con gran precisión se ha vuelto un problema. Esto ocurre cada vez más con la evolución de nuevas tecnologías del internet. Grandes sistemas de búsqueda han intentado proveer una solución a este problema. Skar et al. [14] analizaron un eficiente marco de trabajo CBIR. Hadoop MapReduce se propuso para operar establemente con gran rendimiento teniendo como objetivo grandes

repositorios de información. Las pruebas empíricas demostraron superar en rendimiento a las técnicas comparadas en el estado del arte.

Hoy en día, el reto de las búsquedas de imágenes basadas en contenido a grandes escalas ha tenido grandes acercamientos por numerosos trabajos prometedores. En el trabajo de Kamel et al. [15] se mostró un nuevo enfoque que conjuntamente optimiza la precisión de búsqueda y el tiempo de ejecución para la recuperación de imágenes basada en contenido a gran escala. Esto se logró usando descriptores de imágenes binarios locales, tal como BRIEF y métodos binarios de *hashing*, como SH (*Spherical Hashing, Hashing Esférico*). Como resultado mejoró la precisión comparándolo con las técnicas mencionadas en el estado del arte.

En [16], se propuso un sistema CBIR que usa extracción de características de imagen con momento de color, *wavelet* e histograma de borde. La combinación de características de color, textura y forma lleva a un resultado mucho más preciso. Se utilizó SVM para clasificar las imágenes en diferentes clases y la medida de similitud se tomó solo con imágenes de la misma clase. El sistema mostró un resultado mejor con el uso de la combinación de las características que con el uso individual de las mismas.

Un sistema CBIR que emplea técnicas de compresión y *clustering* sin supervisión, se enfoca en la recuperación rápida de las imágenes deseadas con una gran exactitud. Devi y Mathew [17] demostraron en su trabajo el uso de características extraídas de EDBTC (*Error Diffusion Block Truncation Coding, Codificación de Truncamiento de Bloque de Difusión de Error*). Un nuevo marco de trabajo de CBIR con *clustering* sin supervisión se utilizó para reducir significativamente el tiempo para comparar las imágenes objetivo y consulta.

Diagnósticos tempranos de cáncer de mama pueden mejorar el porcentaje de supervivencia. Kulshreshtha et al. [18] exhibieron un eficiente sistema de recuperación de mamografías basado en computación, el cual ayuda a diagnosticar el cáncer de mama al comparar el caso actual con casos anteriores. Los pasos propuestos incluyen cortes de mamografías, extracción de características usando

LBP (*Local Binary Pattern*, Patrón Binario Local) y *K-Means clustering*. Experimentos usando el punto de referencia de la base de datos de la MIAS (*Mammography Image Analysis Society*, Sociedad de Análisis de Imágenes Mamográficas), confirmaron la efectividad de este trabajo.

Hay cada vez más y más contenido valioso en archivos digitales y esto representa un gran esfuerzo en la digitalización y almacenamiento de este contenido. Para obtener un sistema de almacenamiento digital eficiente, Wang y Chang [19] exhibieron un enfoque de indexación distribuida para la recuperación multimedia basada en contenido colaborativo entre colecciones en archivos digitales. Para reducir los tiempos de carga de cada archivo, distribuyeron dinámicamente las tareas de rastreo, indexación y procesamiento de consultas según el tiempo de respuesta. Resultados experimentales mostraron el potencial en el enfoque de balanceo de cargas con distribución apropiada.

Se requiere una gran cantidad de trabajo para anotar manualmente grandes cantidades de materiales patrimoniales, y los enfoques computacionales automatizados dejan mucho que desear con respecto al nivel de conocimiento y riqueza semántica que proporcionan actualmente. Por este motivo Foley et al. [20] propusieron un algoritmo basado en metadatos para ayudar y aumentar el proceso de anotación. Este algoritmo llamado SAGA (*Semantically-Annotated Graph Analysis*, Análisis Gráfico Anotado Semánticamente) permite una anotación semiautomática, la cual balancea el valor de las contribuciones de anotadores humanos con los beneficios de reducción de tiempo y esfuerzo de un proceso automático. Se utiliza en el contexto de una infraestructura basada en web llamada SAGE (*Semantic Annotation by Group Exploration*, Anotación semántica por exploración de grupo). SAGA y SAGE se evaluaron desde un enfoque de exactitud. Como resultado demostraron un mayor rendimiento y aceptación del usuario.

## **2.2 Análisis comparativo**

La Tabla 2.1 muestra un análisis comparativo de los artículos relacionados con este trabajo.

Tabla 2.1 Análisis comparativo de los trabajos relacionados.

| Artículo                     | Problema  | Contribución  | Tecnología   | Resultados  | Estado               |
|------------------------------|---|---|--|---|----------------------|
| Rodríguez-Mazahua et al. [6] | La fragmentación vertical estática provoca grandes tiempos de búsqueda en bases de datos multimedia.  | DYMOND: sistema para fragmentación vertical de bases de datos multimedia.                                 | PostgreSQL, pgAgent, Fragmentación vertical, disparadores.   | Reducción del costo de ejecución y tiempo de respuesta de las consultas.  | Finalizado           |
| Rojas et al. [7]             | Tiempos de servicio al cliente ineficientes en la empresa AGROMAQ.                                    | Sistema CBIR para el reconocimiento de maquinaria para la agricultura.                                    | Sistema CBIR, BoofCV, SURF, Java, JavaServer Faces, PostgreSQL.  | Reducción y mejora de tiempos de servicio al cliente en la empresa AGROMAQ y se concluye que utilizar SURF para el sistema CBIR muestra mejor eficiencia. | En proceso de mejora |
| Rodríguez-Mazahua et al. [8] | Propuesta de un método de fragmentación híbrida para bases de datos multimedia.                       | Algoritmo de partición híbrida MHYP.<br><br>Modelo de costos para bases de datos multimedia distribuidas. | MMDB ( <i>Multimedia Data Base</i> , Base de datos multimedia), Fragmentación Híbrida, Fragmentación Vertical, Fragmentación Horizontal. | Optimización del costo de ejecución de consultas.   | Finalizado           |
| Prinka y Wasson [9]          | Propuesta de un sistema CBIR haciendo énfasis en técnicas de descripción de características visuales. | Sistema CBIR usando SURF con algoritmos genéticos.  | MATLAB, SURF, CBIR, algoritmos genéticos, procesamiento de imágenes.   | Precisión de más de 98%.  | Finalizado           |

Tabla 2.1 Continuación del análisis comparativo de los trabajos relacionados.

| Artículo                 | Problema   | Contribución  | Tecnología  | Resultados  | Estado     |
|--------------------------|--|---|---|---|------------|
| Khodaskar y Ladhake [10] | Análisis de un sistema CBIR usando técnicas de computación suaves.   | Marco de trabajo para CBIR utilizando técnicas de computación suaves. | CBIR, Redes Neuronales Artificiales, Lógica Difusa, Máquina de Soporte Vectorial. | Mejóro el rendimiento de recuperación de imágenes en términos de exactitud y precisión. | Finalizado |
| Ouhda et al. [11]        | Análisis de un sistema CBIR con técnicas de agrupamiento para proporcionar resultados exactos con menos tiempo de computación. | Sistema CBIR usando K-Means como técnica de agrupamiento.             | CBIR, MMDB, MATLAB, K-Means.  | Mejóro el tiempo de ejecución y la exactitud.   | Finalizado |
| Munish Kumar et al. [12] | Análisis de un sistema CBIR con dos métodos de extracción llamados SIFT y ORB.   | Sistema CBIR basado en SIFT y ORB.                                    | SIFT, ORB, CBIR, BRIEF, BayesNet, K-NN.   | Se logró un máximo de precisión de 88.9%.   | Finalizado |
| Singh et al. [13]        | Análisis de un sistema CBIR utilizando técnicas de computación suaves.   | Sistema CBIR usando Redes Neuronales Artificiales.                    | Redes Neuronales Artificiales, CBIR.  | Toma menos tiempo de computación y provee resultados precisos.                          | Finalizado |

Tabla 2.1 Continuación del análisis comparativo de los trabajos relacionados.

| Artículo                 | Problema   | Contribución  | Tecnología  | Resultados  | Estado     |
|--------------------------|--|---|---|---|------------|
| Sakr et al. [14]         | Construcción de un sistema CBIR para Big Data que opere eficientemente con datos en tiempo real. | Sistema CBIR basado en Hadoop MapReduce.  | CBIR, BOVW ( <i>Bag of Visual Words</i> , Bolsa de palabras visuales), Hadoop, MapReduce. | Mostró una mejora palpable en los factores de costos y tiempos.   | Finalizado |
| Kamel et al. [15]        | Un nuevo enfoque para la optimización de búsquedas de imágenes basadas en contenido.             | Sistema CBIR usando los descriptores binarios BRIEF o BRISK ( <i>Binary Robust Invariant Scalable Keypoints</i> , Puntos clave escalables invariables robustos binarios). | CBIR, Búsqueda Multi Binaria, Descriptores Binarios.                                      | Mejora significativa en la precisión de recuperación sobre los métodos originales de <i>hashing</i> .             | Finalizado |
| Naveena y Narayanan [16] | Análisis de un sistema CBIR para la recuperación de imágenes.                                    | Sistema CBIR basado en momentos de color, <i>wavelet</i> y un descriptor de borde.  | Momento de color, <i>Wavelet</i> , Histograma de borde, SVM, CBIR.                        | Las características combinadas exhiben una mejor recuperación de imágenes que las características individuales.   | Finalizado |
| Devi y Mathew [17]       | Análisis de un nuevo marco de trabajo para un sistema CBIR.                                      | Sistema CBIR con <i>clustering</i> sin supervisión.   | BTC, CBIR, EDBTC, <i>K-Means</i> , Técnica de <i>Clustering</i> , Error de Kernel.        | El método propuesto logra mejor rendimiento en comparación con un esquema de recuperación sin <i>clustering</i> . | Finalizado |

Tabla 2.1 Continuación del análisis comparativo de los trabajos relacionados.

| Artículo                 | Problema   | Contribución  | Tecnología   | Resultados  | Estado     |
|--------------------------|--|---|--|---|------------|
| Kulshreshtha et al. [18] | Detección de cáncer de mama en etapas tempranas.                                     | Sistema CBIR con LBP y <i>K-Means clustering</i> .  | LBP, SURF, SIFT, <i>k-means clustering</i> , CBIR.   | En comparación con otros descriptores LBP demuestra ser mejor para la correcta clasificación y mejora de rendimiento. | Finalizado |
| Wang y Chang [19]        | Grandes problemas en la digitalización de archivos y su respectivo almacenamiento.   | Enfoque de indexación distribuida para la recuperación multimedia basada en contenido colaborativo entre colecciones en archivos digitales. | Archivado distribuido, Rastreador distribuido, CBIR, Repositorio colaborativo, Computación <i>Peer-To-Peer</i> . | Mostraron el potencial del enfoque propuesto en la carga balanceada con distribución de tareas adecuadas.             | Finalizado |
| Foley et al. [20]        | Gran trabajo para anotar manualmente grandes cantidades de materiales patrimoniales. | Algoritmo basado en metadatos para ayudar y aumentar el proceso de anotación.   | SAGA, SAGE.  | Mejora de rendimiento y exactitud, también una mayor aceptación por usuarios.   | Finalizado |

La cantidad de contenido multimedia que crece exponencialmente día a día, presenta uno de los retos más grandes en el control y gestión de bases de datos para la búsqueda de imágenes basada en contenido, es por eso que los temas analizados anteriormente abordan diversos ángulos para la resolución de los retos que puedan surgir. Se observa en la Tabla 2.1 las diversas tecnologías propuestas para la optimización de sistemas CBIR y los métodos que utilizan para la optimización de estos sistemas. En este trabajo se propone un sistema de gestión de datos multimedia y que también utilice un método de fragmentación horizontal para optimizar las consultas basada en contenido. El sistema propuesto se diferencia claramente de los demás, ya que es el único que a pesar de exhibir un sistema de búsqueda de imágenes basada en contenido también utiliza un método de fragmentación horizontal.

### **2.3 Propuesta de la solución**

Se seleccionó MongoDB por su velocidad y sencillo sistema de consulta de los contenidos de la base de datos. Se optó por JavaServer Faces (JSF) debido a la flexibilidad para crear aplicaciones con el lenguaje puro de Java, su facilidad de desarrollo y que también es gratuito. Se eligió NetBeans como IDE por la facilidad para desarrollar aplicaciones con el marco de trabajo seleccionado. Se decide usar la metodología de desarrollo UWE debido a que se especializa en el desarrollo de aplicaciones web y multimedia. Por último se seleccionó BoofCV porque es una biblioteca de Java de código abierto y que tiene facilidad de uso y alto rendimiento.

## Capítulo 3. Aplicación de la Metodología

Para llevar a cabo la aplicación de la metodología, este capítulo se divide en dos partes. La primera realiza una búsqueda y un análisis profundo de los métodos de fragmentación horizontal y búsquedas basadas en contenido para seleccionar una técnica que cumpla con ciertos criterios. La segunda describe el diseño de la aplicación Web para realizar la ejecución de la técnica seleccionada. En este capítulo se aborda también la especificación de la técnica seleccionada.

### 3.1 Análisis

Conforme los objetivos general y específicos, se realizó un análisis profundo de los trabajos relacionados con la fragmentación horizontal de bases de datos multimedia encontrados en algunas de las principales bibliotecas digitales de editoriales científicas: ACM, IEEE, Springer y Elsevier. Como resultado del análisis, bajo algunos criterios principales, se obtuvo el método de fragmentación más adecuado para aplicar en este trabajo. Para realizar el estudio de los trabajos relacionados se siguió la metodología descrita a continuación en la Figura 3.1.

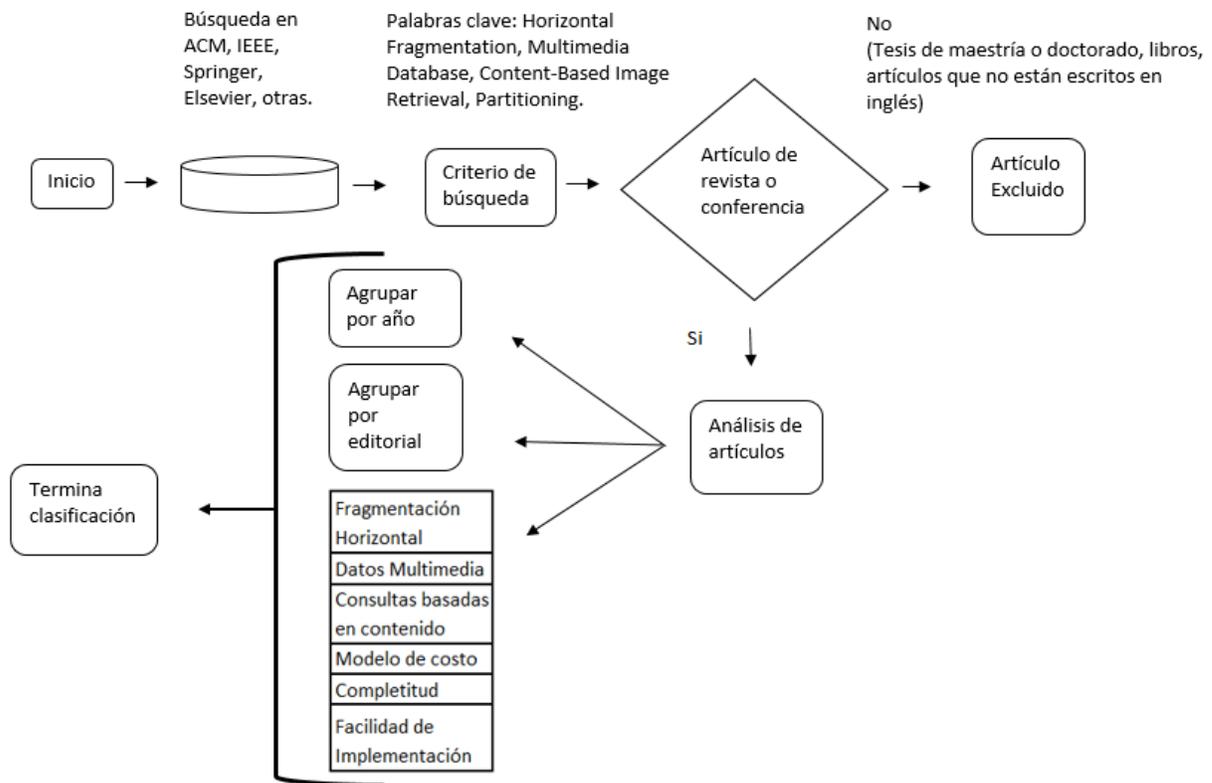


Figura 3.1 Metodología de búsqueda y evaluación de los trabajos relacionados.

En la Figura 3.1 se muestran todas las etapas de la metodología propuesta. Como ya se mencionó, se realizó la búsqueda de trabajos en las principales bibliotecas digitales de editoriales científicas, ACM, IEEE, Springer y Elsevier. Los trabajos encontrados que no son publicados por dichas editoriales se categorizan en “Otras”. La búsqueda consistió en encontrar todos los trabajos que contengan las siguientes palabras clave: *Horizontal Fragmentation* (Fragmentación Horizontal), *Multimedia Database* (Bases de datos multimedia), *Content-Based Image Retrieval* (Recuperación de Imágenes Basado en Contenido), y *Partitioning* (partición). Ya obtenidos todos los trabajos, se aplicó un filtro y se descartaron todos los trabajos que sean tesis de maestría o doctorado, así como libros y artículos que no estén escritos en inglés. Los artículos resultantes se clasificaron por editorial y por año. En esta última etapa se analizó cada uno de ellos bajo seis rubros principales: Fragmentación Horizontal (FH), Datos multimedia, Consultas basadas en contenido, Modelo de costo, Completitud y Facilidad de implementación. De esta manera, sólo los trabajos que cumplieron con la mayoría de los rubros se seleccionaron.

El proceso de selección dio como resultado 35 artículos, los cuales fueron analizados por las seis características para registrar si cumplieron cada rubro.

La Tabla 3.1 describe el registro de cada artículo y su cumplimiento con cada una de las seis características descritas anteriormente, se encontrarán tablas con información más detallada sobre el análisis de cada uno de los artículos en el anexo al final de este trabajo.

Tabla 3.1 Análisis de Artículos.

| Artículo   | 1 | 2 | 3 | 4 | 5 | 6 |
|--|---|---|---|---|---|---|
| Saad et al. [21]   | x | x | x |   |   | x |
| Getahun et al. [22]  | x | x | x |   |   |   |
| Fasolin et al. [23]  | x | x | x |   | x | x |
| Rodríguez-Mazahua et al. [24]  | x | x |   | x | x | x |
| Habbal et al. [25]   |   | x | x |   |   |   |
| Kulshreshtha et al. [18]   |   | x | x |   | x |   |
| Devi y Mathew [17]   |   | x | x |   | x | x |
| Banda y Angryk, 2014 [26]  |   | x | x |   | x |   |
| Mettes et al. [27]   | x |   |   |   | x | x |
| Ma y Kirchberg [28]  | x |   |   | x | x | x |
| Getahun et al. [29]  | x | x | x |   | x |   |
| Hui Ma et al. [30]   | x |   |   | x | x | x |
| Khan et al. [31]   | x |   |   |   | x |   |
| Ma et al. [32]   | x |   |   | x | x | x |
| Chbeir y Laurent [33]  |   | x | x |   | x | x |
| Concolato et al. [34]  |   | x |   |   | x | x |
| Chbeir y Laurent [35]  |   | x | x |   | x |   |
| Xiapoing Hu et al. [36]  |   | x |   |   | x |   |
| Antaris et al. [37]  |   | x | x |   | x | x |
| Torjmen et al. [38]  |   |   |   |   | x |   |
| Mezzoudj et al. [39]   |   | x | x |   | x | x |
| Sakr et al. [14]   |   | x | x |   | x | x |
| Torjmen-Khemakhem et al. [40]  |   | x |   |   | x | x |
| Gu y Gao [41]  |   | x |   |   | x | x |
| Włodarczyk-Sielicka y Stateczny [42]   |   |   |   |   | x | x |
| Kumar et al. [43]  |   | x | x |   | x |   |
| Moghimian et al. [44]  |   | x | x |   | x |   |
| Srivastava et al. [45]   |   | x | x |   | x | x |
| Li et al. [46]   |   | x |   |   | x |   |
| Al-Jubouri y Du [47]   |   | x | x |   | x | x |
| Mahmoudi et al. [48]   |   | x | x |   | x |   |
| Rejito et al. [49]   |   | x | x |   | x |   |
| Kouahla et al. [50]  |   | x |   |   |   | x |
| Le et al. [51]   |   | x |   |   | x | x |
| Yamamoto [52]  |   | x |   |   | x | x |
| 1)Fragmentación Horizontal, 2)Datos multimedia, 3)Consultas basadas en contenido, 4)Modelo de costo, 5)Compleitud, 6)Facilidad de implementación |   |   |   |   |   |   |

### **3.2 Selección**

Después de analizar los 35 métodos de optimización de consultas, se encontró que ninguno cumplió con los seis criterios, por lo que se seleccionaron tres artículos ([21], [23] y [24]), ya que involucran fragmentación horizontal y bases de datos multimedia, además son fáciles de implementar; [21] y [23] también consideran consultas basadas en contenido, mientras que [23] y [24] tienen completitud; solo [24] incluye un modelo de costos. En [21], no se describe un modelo de costos y el tipo de fragmentación horizontal que se utiliza no es el más adecuado para el propósito de esta aplicación.

Los métodos elegidos para fragmentar la base de datos históricos y evaluar el esquema de fragmentación son [23] y [24], debido a que el modelo de costos presentado en [24] es adecuado para el análisis del costo de acceso a cada uno de los fragmentos por medio de consultas basadas en contenido, mientras que el método de fragmentación horizontal en [23] presenta una metodología deseada y una facilidad de implementación, ya que se propone que los fragmentos sean por el tipo de imagen, acorde al propósito de esta aplicación.

### **3.3 Desarrollo del sistema**

En esta sección se describe el diseño de la aplicación siguiendo la metodología de desarrollo UWE (*UML-BASED WEB ENGINEERING*, UML-Basado en Ingeniería Web). Se llevan a cabo todos los diagramas y pasos de la metodología elegida para el desarrollo de la aplicación Web.

La arquitectura del sistema se diseñó con base en el patrón arquitectónico MVC (Modelo Vista Controlador), el cual distribuye los componentes del sistema de tal forma que facilite su mantenimiento y se representa de manera abstracta en la Figura 3.2, donde: en la vista están las páginas que interactúan con el usuario para solicitar información sobre la base de datos y cómo acceder a ella. Los *beans* administrados de JSF (JavaServer Faces) que gestionan el flujo de información se encuentran en el controlador. En el modelo, se definen las reglas del negocio, que en este escenario particular será la gestión de los datos históricos del Instituto Tecnológico de Orizaba, contemplando consultas basadas en contenido.

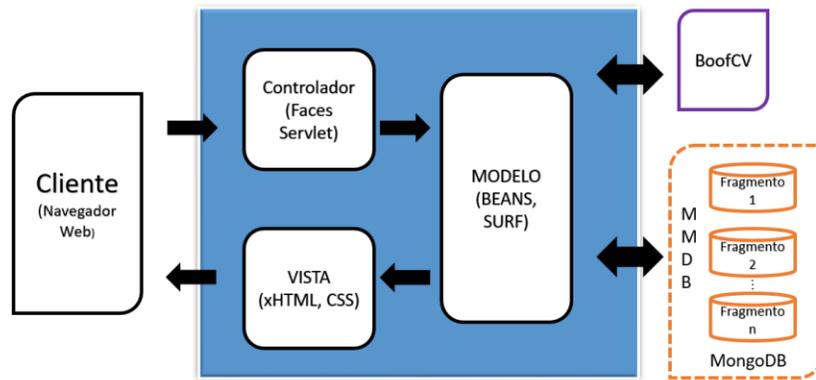


Figura 3.2 Estructura de la solución de la propuesta

### 3.3.1 Análisis de requisitos

El objetivo del análisis de requisitos es establecer los requisitos funcionales de la aplicación web, estos requisitos se representan mediante diagramas. El diagrama de casos de uso representa la relación actor-actividades, muestra las actividades a las que está relacionado cada actor.

En la Tabla 3.2 se definen los actores de la aplicación y una breve descripción del papel que desempeñan.

Tabla 3.2 Actores de la aplicación.

| Actor           | Descripción   |
|-----------------|---|
| Administrador   | Tiene control total sobre la base de datos y es el único que puede realizar eliminaciones de documentos o colecciones.  |
| Colaborador     | Es capaz de realizar capturas de información y tener acceso a la estructura de la base de datos pero de manera limitada, únicamente información referente a eventos y sus imágenes, así como validar información. |
| Docente         | Puede visualizar la información general del Instituto y también acceder a la información y/o imágenes donde aparece, únicamente información referente a él, y proponer contenido.                                 |
| Usuario General | Solo puede visualizar la información general del instituto, y proponer contenido.   |
| Usuario Anónimo | Solo puede visualizar la información general del instituto  |

Una vez ya definido los actores en la Tabla 3.2, en la Figura 3.3 se muestra el diagrama de casos de uso, el cual relaciona a los actores con sus actividades.

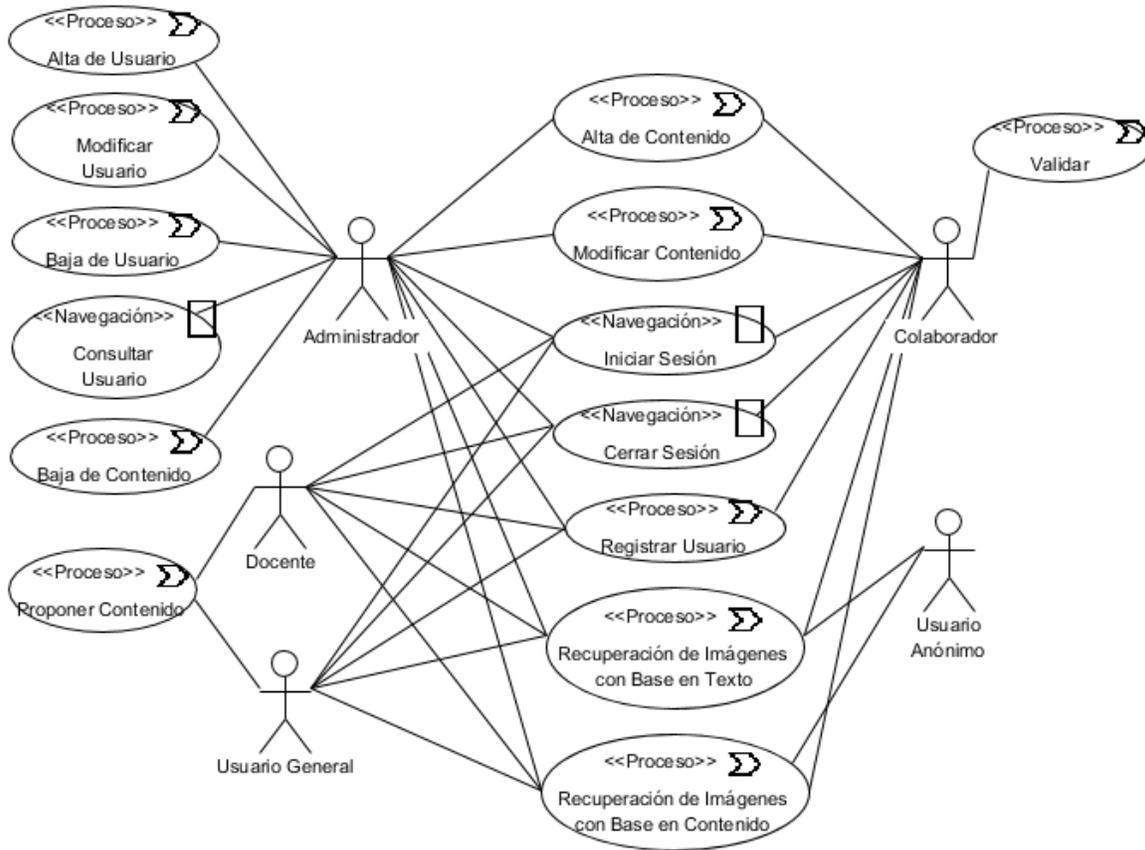


Figura 3.3 Diagrama de casos de uso

A continuación, en las Figuras 3.4 – 3.17 se presentan los diagramas de actividades para cada caso de uso que se muestra en la Figura 3.3

El caso de uso Iniciar Sesión se representa en el diagrama de actividad de la Figura 3.4. Éste indica que el proceso se inicia mostrando en pantalla un formulario mediante el cual el usuario proporciona sus credenciales de acceso para que el sistema inicie la sesión indicada de acuerdo a las mismas, lo cual da fin al referido caso de uso.

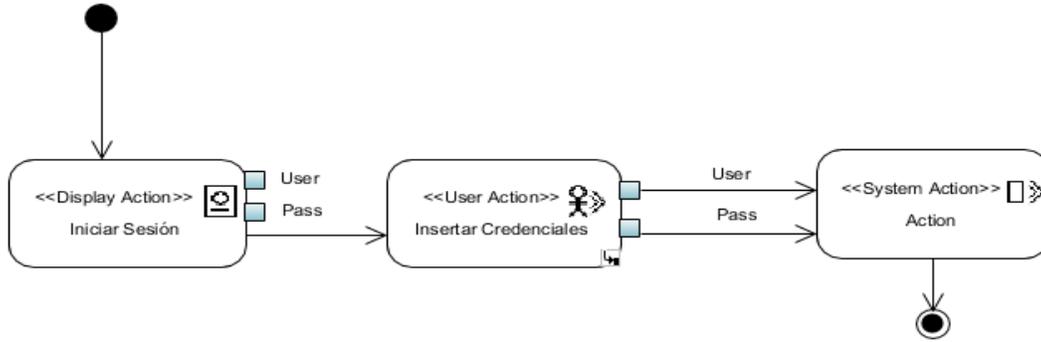


Figura 3.4 Diagrama de caso de uso Iniciar sesión

El diagrama de actividad del caso de uso contrapuesto, Cerrar Sesión, se presenta en la Figura 3.5. Éste indica que el caso inicia mostrando en pantalla un formulario mediante el cual el usuario selecciona la opción de cerrar sesión, para que el sistema cierre la sesión actual y finalice así el caso de uso.

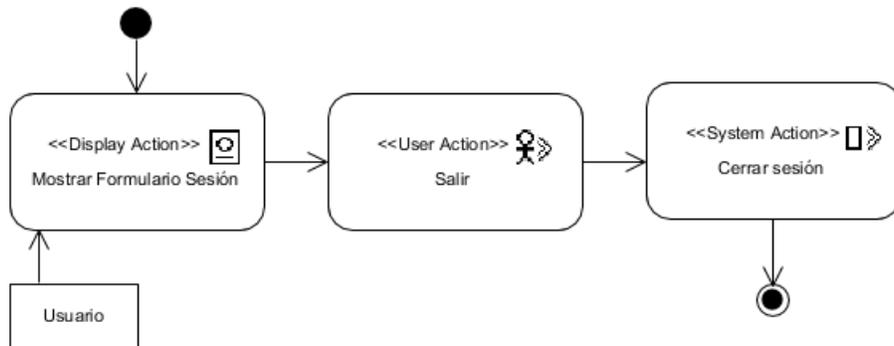


Figura 3.5 Diagrama de caso de uso Cerrar sesión

El diagrama de actividad para Registrar Usuario se visualiza en la Figura 3.6. Ésta indica que el caso de uso inicia mostrando en pantalla un formulario mediante el cual el usuario autorizado proporciona los datos necesarios, para que seguidamente el sistema guarde el nuevo usuario y dé fin al caso.

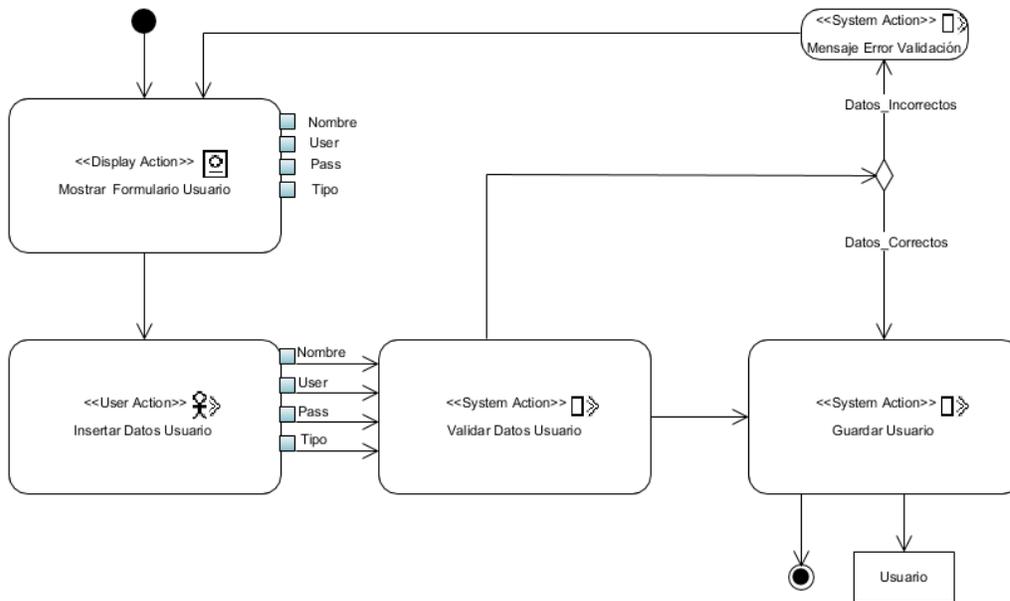


Figura 3.6 Diagrama de caso de uso Registrar usuario

El diagrama de actividad para el caso de uso Validar Información se observa en la Figura 3.7. Comienza mostrando en pantalla el listado de contenidos registrados en la aplicación y un formulario mediante el cual el usuario autorizado proporciona el identificador del contenido que desea validar. Hecho esto el sistema proporcionará los datos del mismo y en este punto el usuario indica que desea validar la información, para que el sistema proceda a realizar la operación. El caso de uso termina con la validación de contenido.

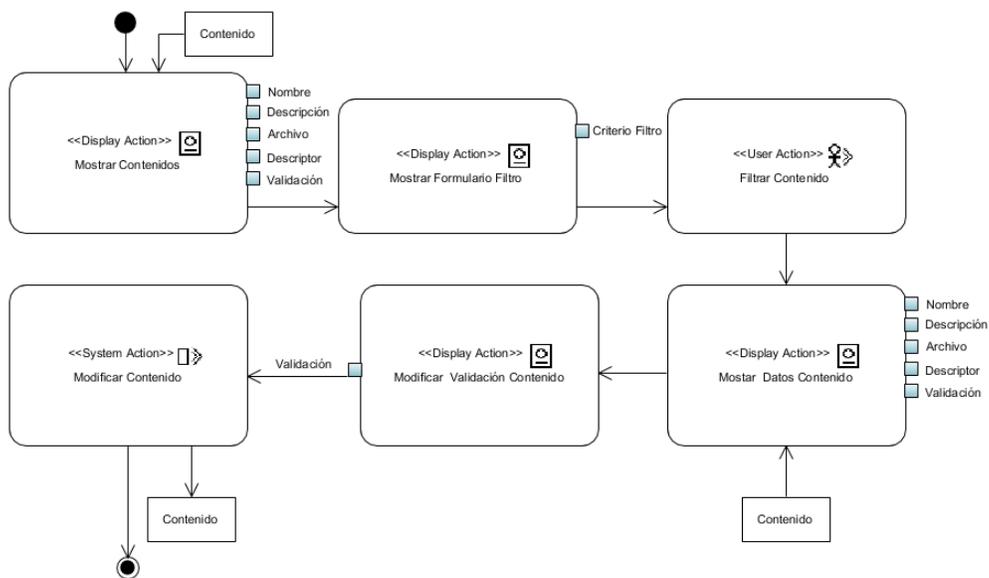


Figura 3.7 Diagrama de caso de uso Validar Información

En la Figura 3.8 se presenta el diagrama de actividad para el caso de uso Proponer Contenido. El caso de uso inicia mostrando en pantalla un formulario mediante el cual el usuario inserta el contenido y lo describe, para que seguidamente el sistema valide y guarde la nueva propuesta de contenido, para posteriormente poder ser validada por algún usuario colaborador. Cuando el contenido queda registrado en la aplicación el caso de uso se termina.

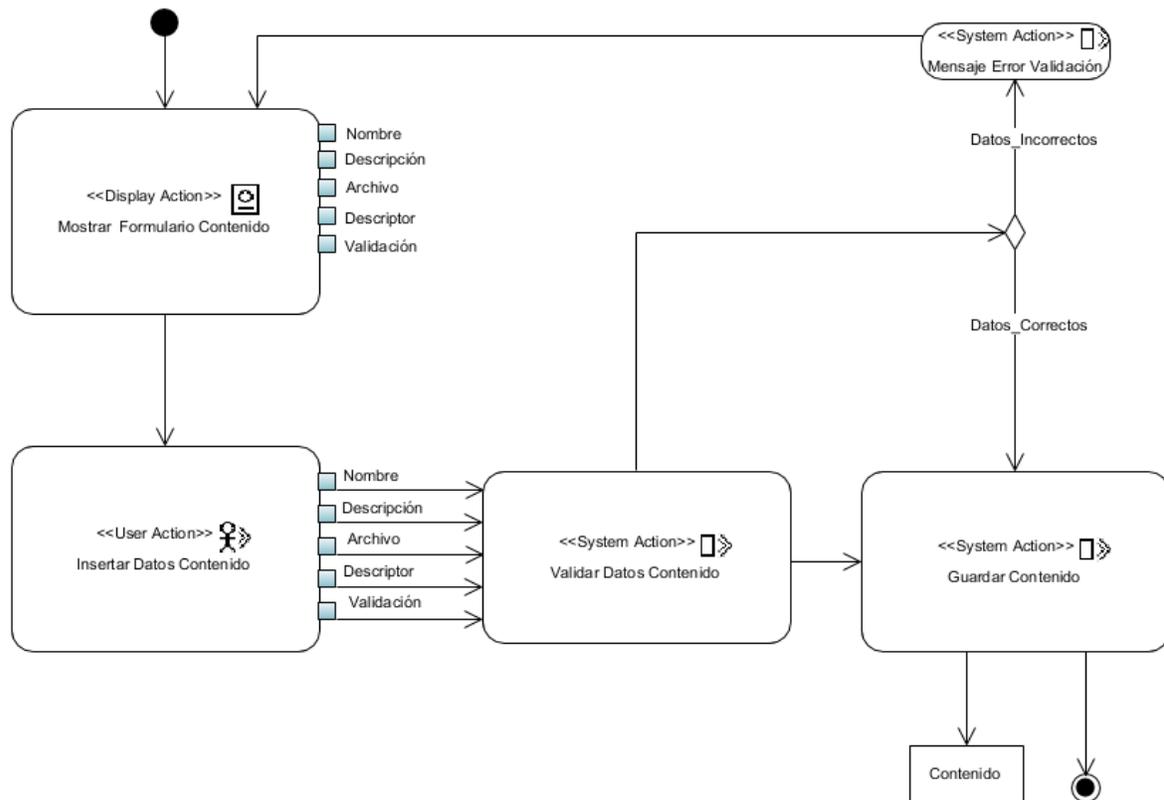


Figura 3.8 Diagrama de caso de uso Proponer Contenido

El diagrama de actividad para Alta de Contenido se observa en la Figura 3.9. Ésta indica que el caso de uso inicia mostrando en pantalla un formulario mediante el cual el usuario autorizado proporciona los datos necesarios, para que seguidamente el sistema guarde el nuevo contenido y dé fin al caso.

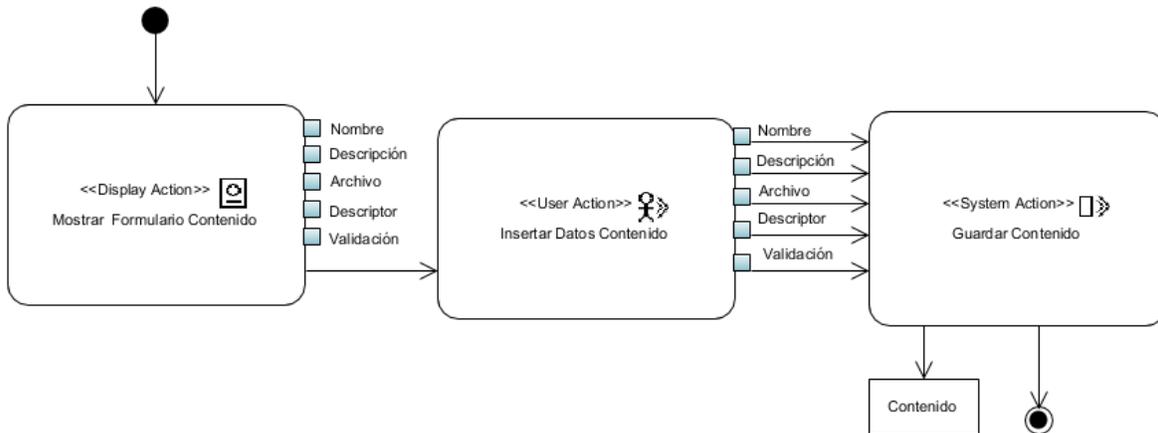


Figura 3.9 Diagrama de caso de uso Alta Contenido

El diagrama de actividad para el caso de uso Modificar Contenido se presenta en la Figura 3.10. Comienza mostrando en pantalla el listado de contenidos registrados en la aplicación y un formulario mediante el cual el usuario autorizado ingresa el identificador del contenido que desea modificar. Hecho esto el sistema proporcionará los datos del mismo y en este punto el usuario modifica el contenido que desee. El caso de uso termina con la modificación del contenido.

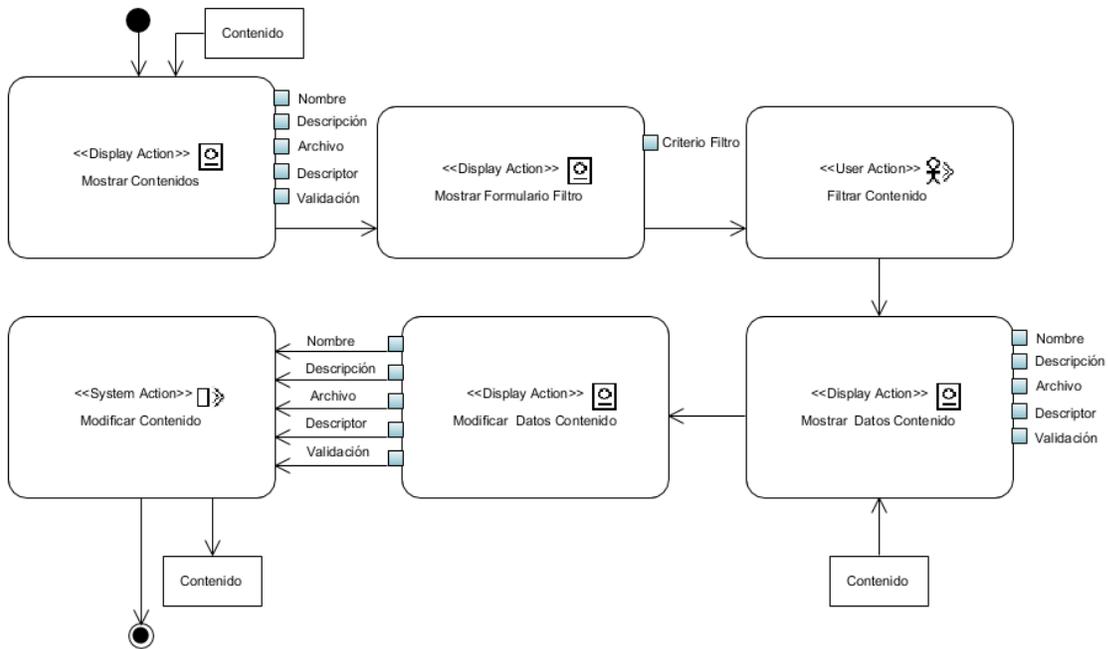


Figura 3.10 Diagrama de caso de uso Modificar Contenido

El diagrama de actividad para el caso de uso Baja de contenido se visualiza en la Figura 3.11. Comienza mostrando en pantalla el listado de contenidos registrados en la aplicación y un formulario mediante el cual el usuario autorizado ingresa el identificador del contenido que desea eliminar. Hecho esto el sistema proporcionará los datos del mismo y en este punto el usuario indica que desea dar de baja un contenido, para que el sistema proceda a realizar la operación. El caso de uso termina con la baja del contenido.

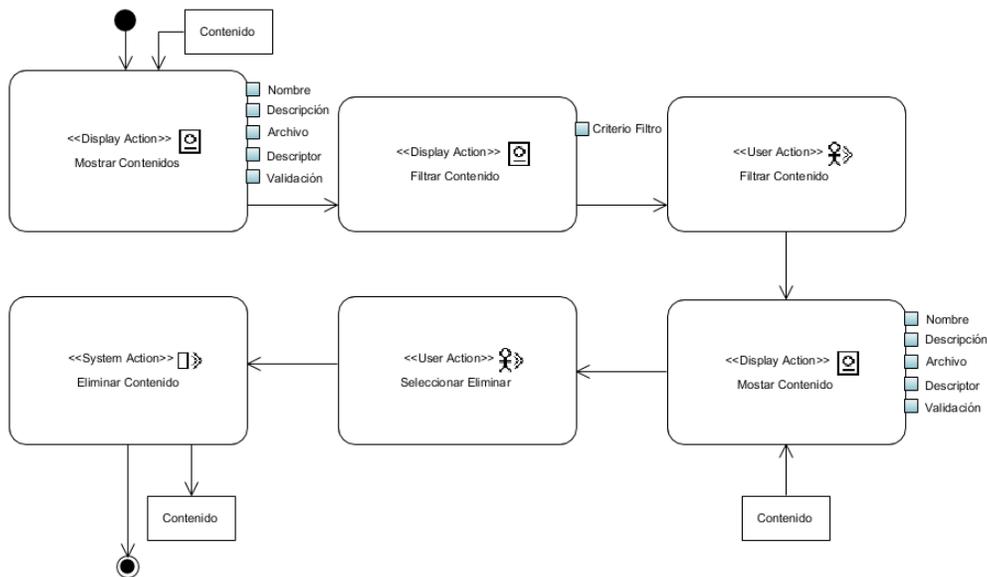


Figura 3.11 Diagrama de caso de uso Baja de Contenido

El diagrama de actividad para Alta Usuario se presenta en la Figura 3.12. Ésta indica que el caso de uso inicia mostrando en pantalla un formulario mediante el cual el usuario autorizado proporciona los datos necesarios, para que seguidamente el sistema guarde el nuevo usuario y dé fin al caso.

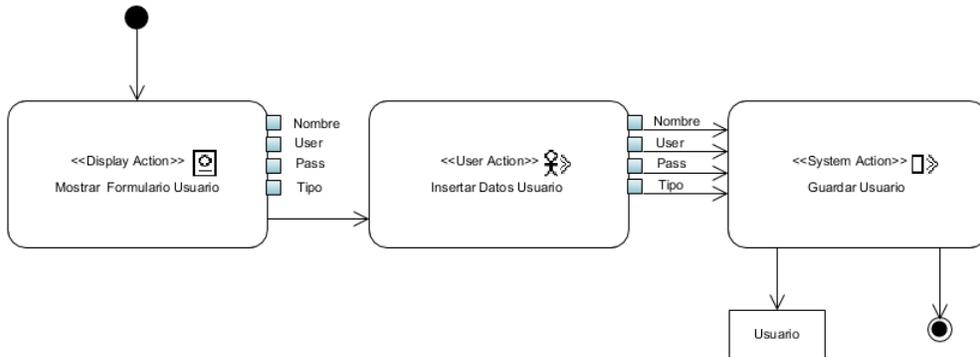


Figura 3.12 Diagrama de caso de uso Alta Usuario

El diagrama de actividad para el caso de uso Baja Usuario se observa en la Figura 3.13. Comienza mostrando en pantalla el listado de usuarios registrados en la aplicación y un formulario mediante el cual el usuario autorizado ingresa el identificador del usuario que desea eliminar. Hecho esto el sistema proporcionará los datos del mismo y en este punto el usuario indica que desea dar de baja un usuario, para que el sistema proceda a realizar la operación. El caso de uso termina con la baja del usuario.

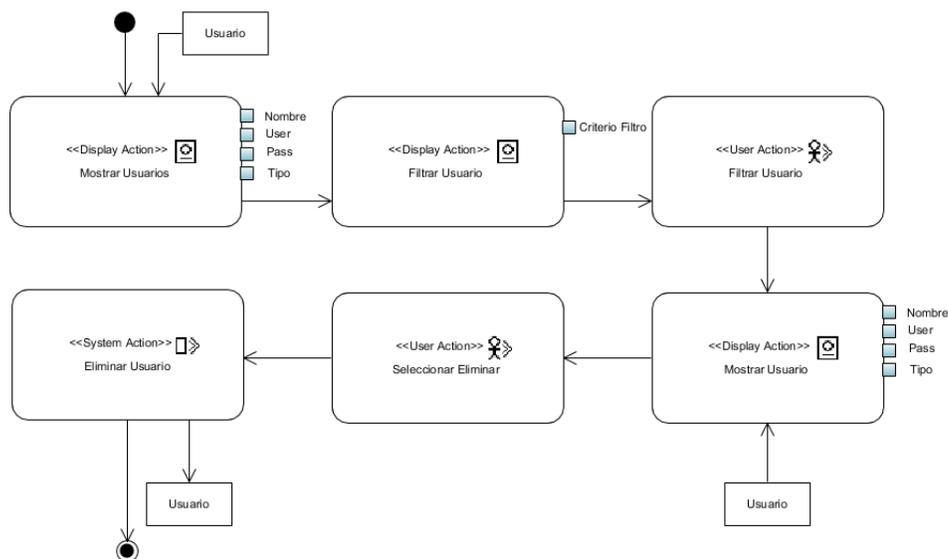


Figura 3.13 Diagrama de caso de uso Baja de Usuario

El diagrama de actividad para el caso de uso Modificar Usuario se presenta en la Figura 3.14. Comienza mostrando en pantalla el listado de usuarios registrados en la aplicación y un formulario mediante el cual el usuario autorizado ingresa el identificador del usuario que desea modificar. Hecho esto el sistema proporcionará los datos del mismo y en este punto el usuario modifica los datos que desee. El caso de uso termina con la modificación del usuario.

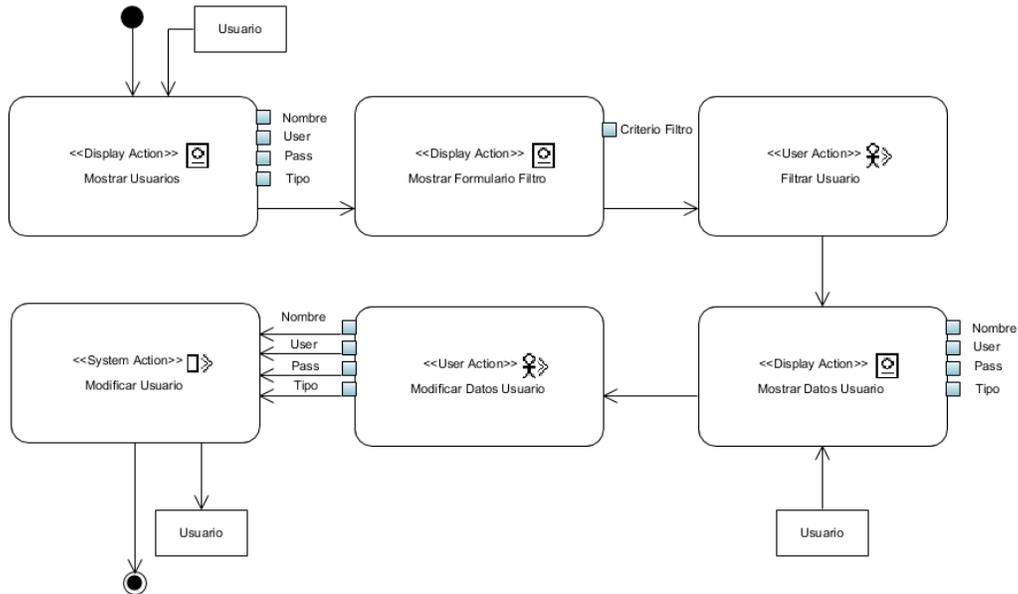


Figura 3.14 Diagrama de caso de uso Modificar Usuario

El caso de uso Consultar Usuario, como explica su diagrama de actividad en la Figura 3.15, inicia con la muestra en pantalla del listado de usuarios registrados en la aplicación y un formulario mediante el cual el usuario autorizado filtra los resultados. Para ello proporciona el identificador del usuario que se quiere consultar, a lo que el sistema responderá consecuentemente y dará fin al caso de uso.

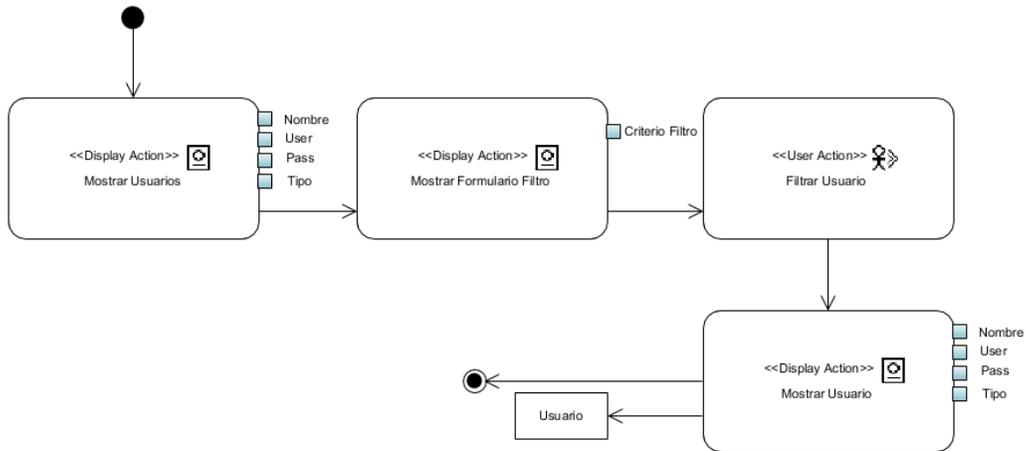


Figura 3.15 Diagrama de caso de uso Consultar Usuario

El caso de uso Consulta con Base en Texto, como explica su diagrama de actividad en la Figura 3.16, inicia con la muestra en pantalla de un formulario mediante el cual el usuario autorizado proporciona los datos necesarios para posteriormente el sistema busque los contenidos con los criterios de búsqueda. Se mostrarán los contenidos encontrados, y se dará fin al caso de uso.

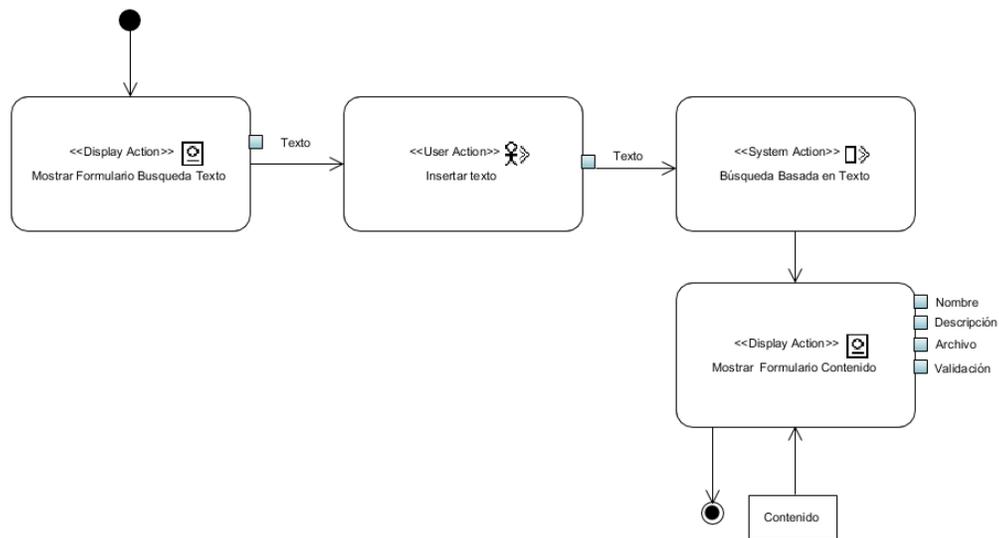


Figura 3.16 Diagrama de caso de uso Consulta con Base en Texto

El caso de uso Consulta con Base en Contenido, como explica su diagrama de actividad en la Figura 3.17, inicia con la muestra en pantalla de un formulario mediante el cual el usuario autorizado proporciona los datos multimedia para posteriormente el sistema realice la consulta basada en contenido. Se mostrarán los contenidos encontrados, y se dará fin al caso de uso.

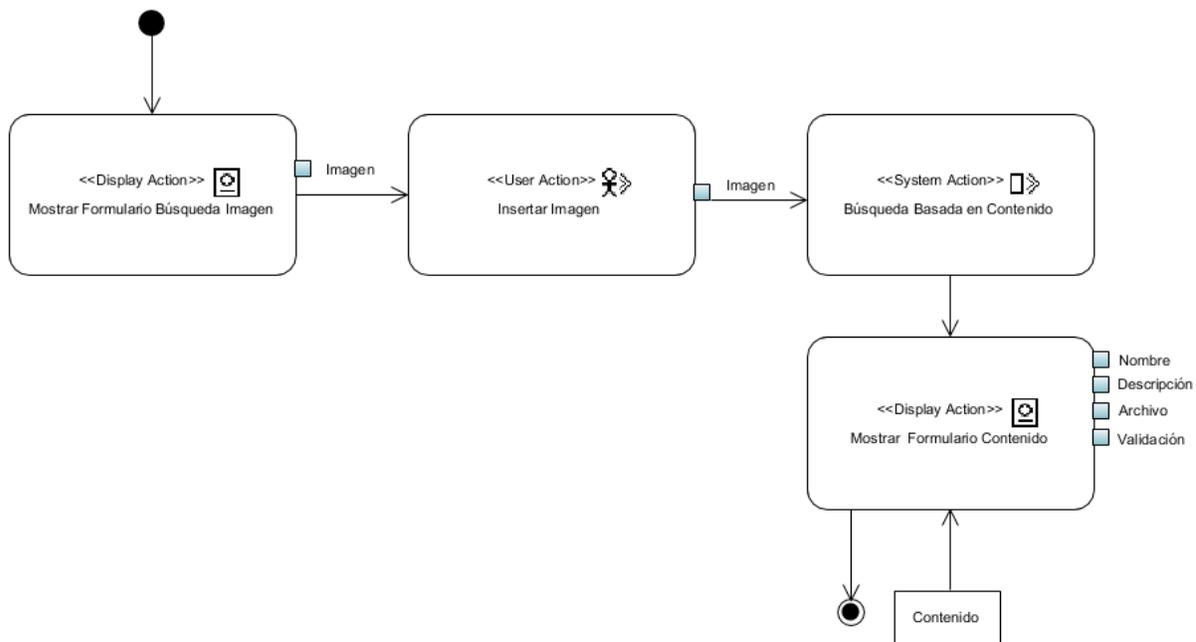


Figura 3.17 Diagrama de caso de uso Consulta con base en contenido







## **Tablas de la base de datos**

**Usuarios:** Esta tabla contiene la información de los usuarios que se han registrado en el sistema, misma que será necesaria para identificarse en el sistema.

**Personal:** Aquí se encuentran almacenados los datos del personal del Instituto. Esta tabla permite conocer fechas de defunción y egreso para poder conocer la situación actual del personal.

### **Grupo Puestos**

**Puestos, Personal[]:** Estas tablas y sus relaciones permiten clasificar los puestos y su información con relación al personal del Instituto.

### **Grupo Hechos**

**Hechos Destacados ,Reconocidos[]:** Conjunto de tablas que almacenan los datos correspondientes a los hechos destacados dentro del Instituto.

**Programas:** Tabla que almacena el nombre de los programas educacionales que se imparten dentro del Instituto.

### **Grupo departamento**

**Departamento, Programas, Oficinas:** Conjunto de tablas que almacenan la información de los departamentos y oficinas del Instituto.

### **Grupo Alumnos**

**Alumnos, Carrera:** Conjunto de tablas que almacenan la información de los alumnos del Instituto y su relación con respecto a la carrera a la que pertenecieron.

### **Grupo Imágenes**

**Imágenes, Relaciones:** Contiene los datos multimedia históricos del instituto tecnológico de Orizaba, donde solo se encuentran imágenes. Esta tabla será el principal objetivo para la aplicar un método de fragmentación horizontal para optimizar las consultas basadas en contenido.

**Aportaciones:** Tabla que almacena la información correspondiente a las aportaciones hechas sobre el Instituto Tecnológico de Orizaba.

### **Grupo Eventos:**

**Eventos, Actividades[], Participantes[]:** Conjunto de tablas que contienen la información correspondiente a los eventos que realiza el Instituto, así como su relación con los participantes a estos eventos o actividades.

### **Grupo Edificios**

**Edificios, Área, Secciones, Equipo:** Conjunto de tablas que contiene la información de todos los edificios dentro del instituto, así como su relación con respecto al área y secciones donde se localizan, así como la información del equipo que se encuentra en estos edificios.

### **3.3.3 Modelo de Navegación**

Mediante el modelo de navegación se conocen todos los caminos posibles dentro de la aplicación para la navegación de los usuarios. Este modelo propone crear mapas de navegación mediante diagramas de clase estereotipados, por medio de estos diagramas se describen todas las rutas posibles por las que se mueve un usuario dentro de un sitio o de una aplicación. En la Figura 3.21 se muestra el modelo de navegación de la aplicación.

En el mapa de navegación se ven todos los caminos posibles que indican cómo un usuario se mueve de un sitio a otro dentro de la aplicación. Partiendo de la página de inicio, desde el Menú Principal un usuario elige ir a registrarse como un nuevo usuario o iniciar sesión. Si decide seguir por la última opción, dependiendo el tipo de usuario tendrá hasta cinco opciones para continuar.

La primera de estas últimas muestra la posibilidad desde el Menú Acciones de navegar hasta la página para consultar contenido y realizar una búsqueda, cualquier usuario registrado cuenta con esta capacidad dentro del sistema.

La segunda ofrece al usuario autorizado la opción de proponer contenido, mostrando un formulario donde se tendrán que llenar los datos necesarios para su posterior guardado. En la tercera, se ofrece al usuario la opción de cerrar su sesión en la aplicación. La cuarta opción permite navegar al Usuario Autorizado hasta la página de Usuarios, donde es capaz de insertar un nuevo usuario, seleccionar uno específico y de acuerdo al Menú Usuario, es posible ver los datos del seleccionado, eliminarlo o modificarlo.

Por último, la quinta opción permite al usuario autorizado navegar a la página de Contenidos, donde podría insertar nuevo contenido, seleccionar uno específico y podría ver los datos del contenido seleccionado, eliminarlo o modificarlo.

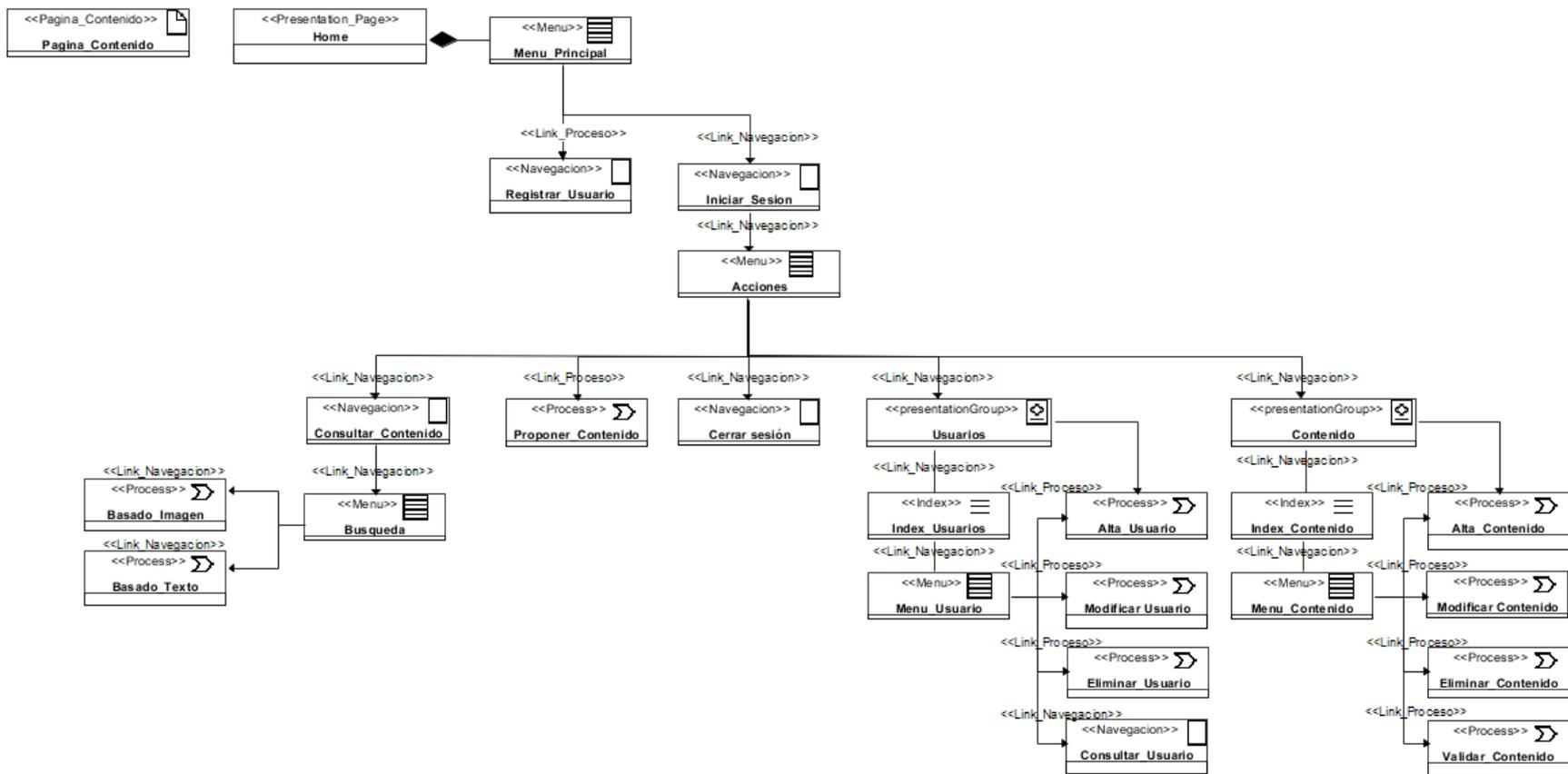


Figura 3.21 Diagrama de Navegación.

### 3.3.4 Modelo de presentación

El modelo de presentación define las páginas de la aplicación, la estructura y los componentes de las mismas. La aplicación tiene seis páginas, que se describen a continuación.

En la Figura 3.22 se muestra que la página Inicio da la bienvenida a los usuarios que acceden al sistema. En esta página se presenta información del Tecnológico Nacional de México, Campus Orizaba. También contiene un enlace para iniciar sesión dentro del sistema y acceder al contenido.

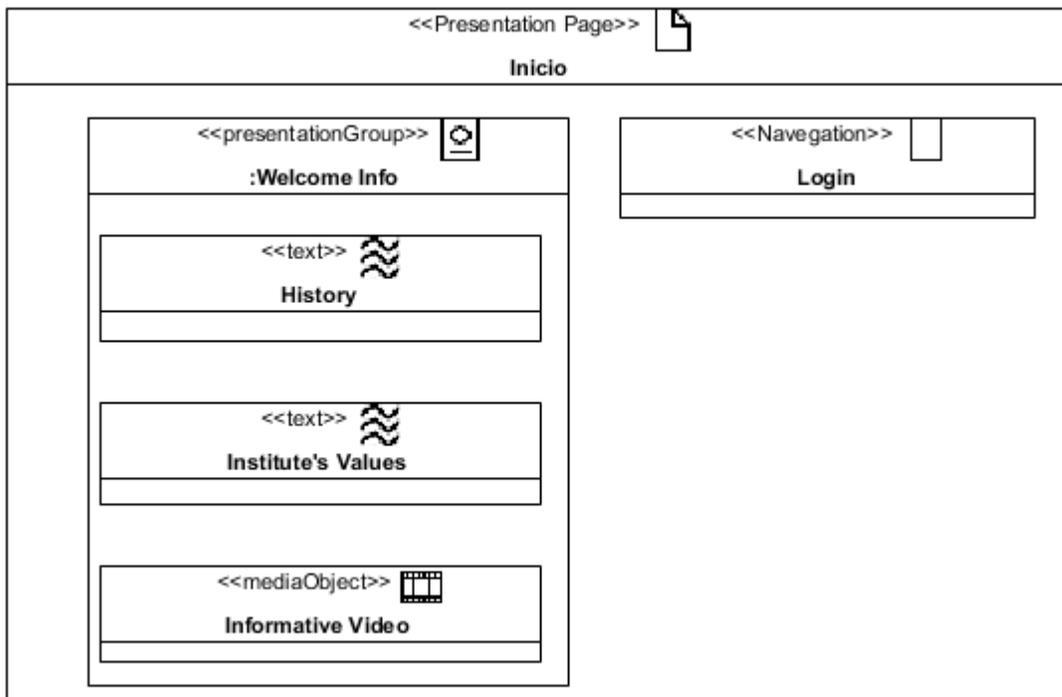


Figura 3.22 Página de Inicio del modelo de presentación.

La página *Iniciar Sesión*, ver Figura 3.23, muestra el formulario necesario para introducir las credenciales necesarias para acceder al sistema. En la parte inferior del formulario se encuentra la opción para crear una cuenta, la cual actualiza el formulario para registrar los datos de un nuevo usuario.

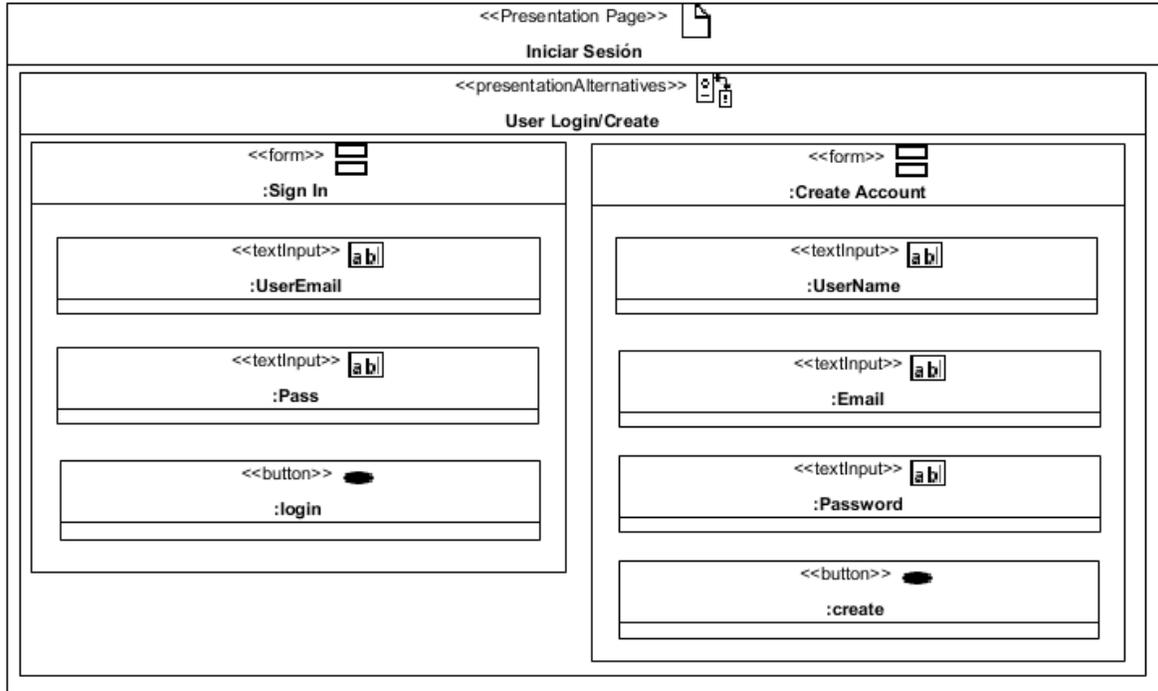


Figura 3.23 Página Iniciar Sesión del modelo de presentación.

En la primera pestaña de izquierda a derecha de la página *Contenidos*, ver en la Figura 3.24, se muestra información sobre la historia del Instituto Tecnológico de Orizaba. En las pestañas restantes se muestra información detallada del contenido correspondiente a los diferentes tipos de imágenes: edificios, eventos, equipamiento y personal. En la parte superior se encuentra un menú que dependiendo del tipo de usuario muestra opciones para salir de la sesión o gestionar usuarios o un tipo de contenido. También existe la opción de proponer contenido mediante un botón el cual re-direcciona al usuario a la página *Proponer Contenido*, ver Figura 3.25, donde se muestra el formulario necesario para la captura de los datos para proponer nuevo contenido. La opción para hacer búsquedas basadas en contenido o en texto está presente mediante un botón que re-direcciona al usuario a la página *Búsqueda de Contenido*, ver Figura 3.26, donde se presenta el formulario para la captura de los datos requeridos para realizar el tipo de búsqueda especificada.

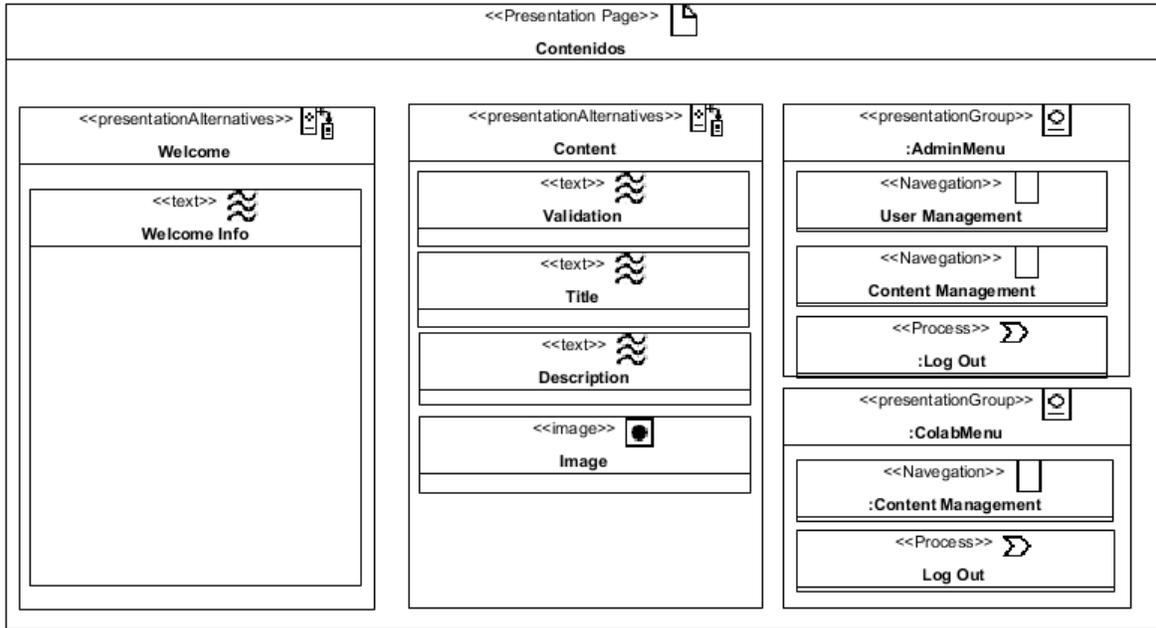


Figura 3.24 Página Contenidos del modelo de presentación.

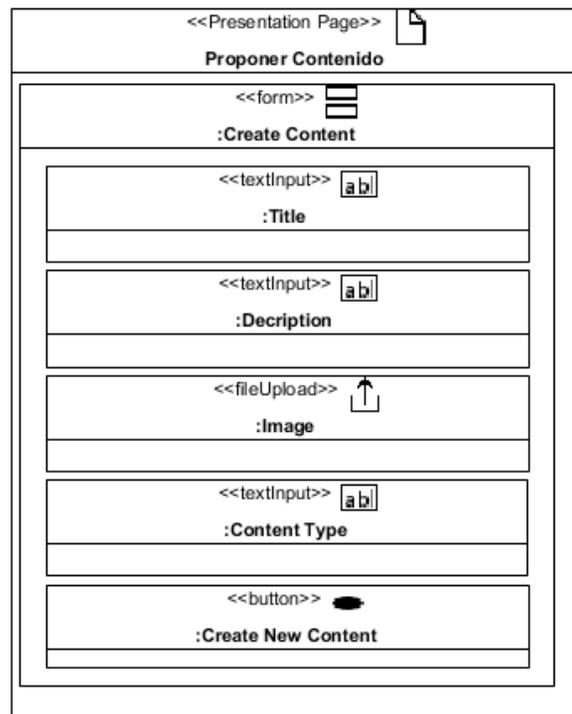


Figura 3.25 Página Proponer Contenido del modelo de presentación.

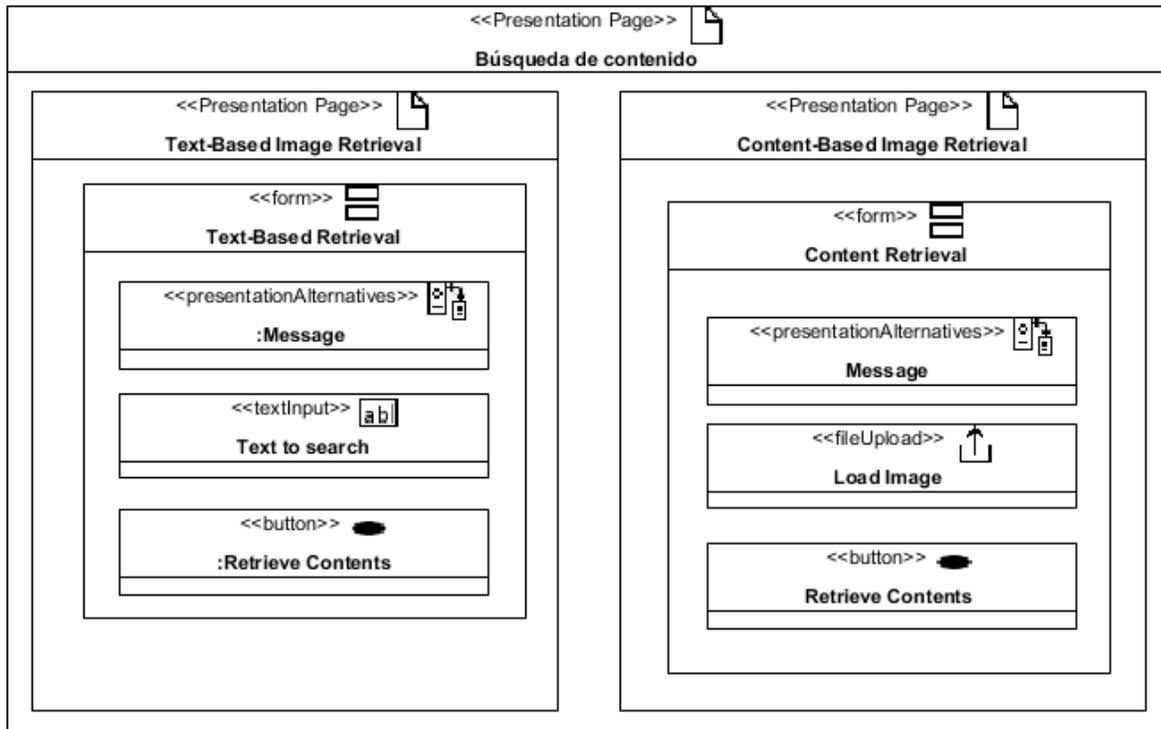


Figura 3.26 Página Búsqueda de Contenido del modelo de presentación.

En la Figura 3.27 se presenta la página *Gestionar Usuarios*, desde la cual un administrador gestiona las cuentas de los usuarios registrados dentro del sistema. Se muestra una tabla con los datos de los usuarios y las operaciones de actualizar y eliminar vinculadas a cada uno de ellos. Mediante el filtro de nombre de usuario y correo electrónico es posible buscar un usuario determinado. La opción de crear un nuevo usuario está presente mediante un botón que, al ser seleccionado, muestra un formulario emergente para la captura de los datos necesarios para hacer el nuevo registro.

En la Figura 3.28 se observa la página *Gestionar Contenido*, desde la cual un administrador gestiona el contenido de cada uno de los tipos de imagen que existe: edificios, equipamiento, eventos y personal. Se muestra una tabla con los datos del contenido deseado y las operaciones actualizar y eliminar vinculadas a cada uno de ellos. Mediante el filtro de nombre de contenido se busca un contenido determinado. La opción de crear un nuevo contenido está presente mediante un botón que, al ser seleccionado, muestra un formulario emergente para la captura de los datos necesarios para hacer el nuevo registro de contenido.

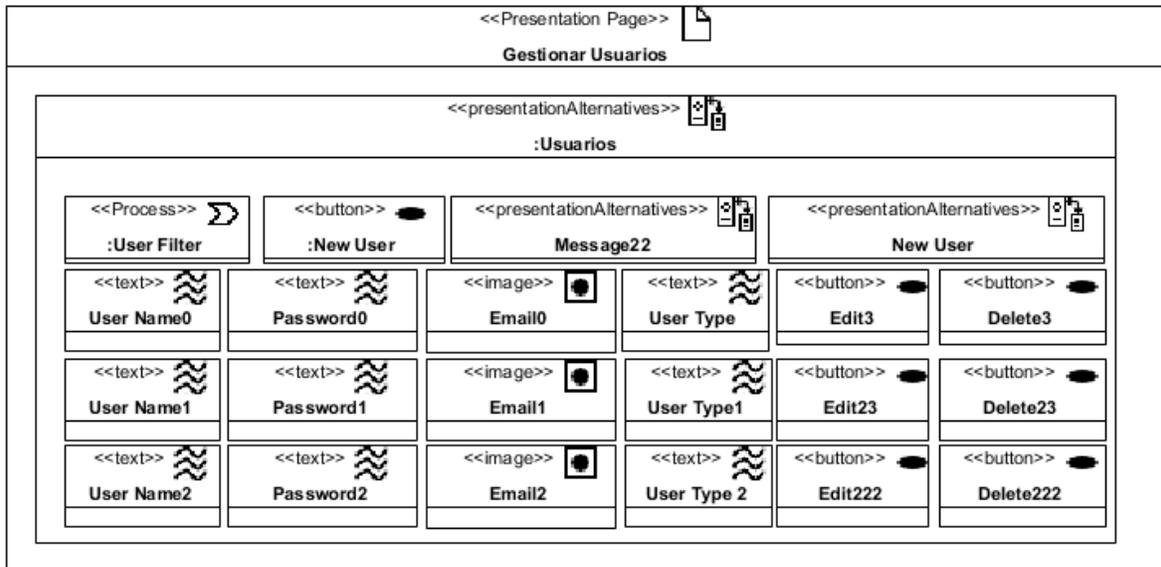


Figura 3.27 Página Gestionar Usuarios del modelo de presentación.

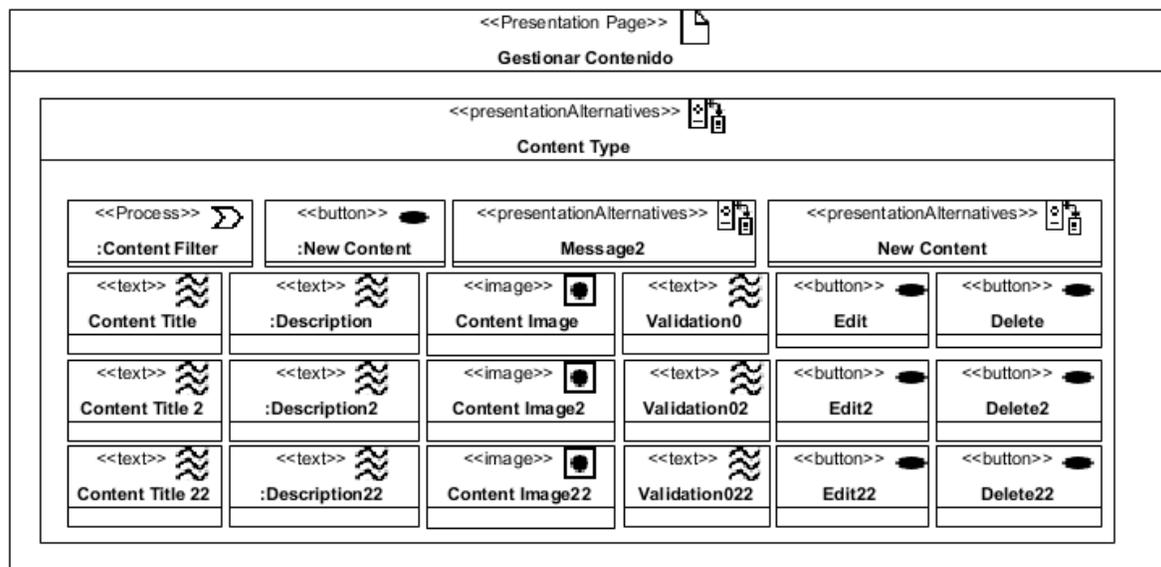


Figura 3.28 Página Gestionar Contenido del modelo de presentación.

### 3.3.5 Modelo de procesos

El diagrama de procesos se basa en el diagrama de actividades, agregando diversos controles que enriquecen la especificación de la aplicación.

En la Figura 3.29 se presenta el diagrama de proceso del caso de uso Iniciar sesión, agregando diversos componentes como los nodos de decisión y la representación de los objetos dentro del flujo de actividades. Este proceso inicia mostrando en pantalla un formulario mediante el cual el usuario proporciona sus credenciales de acceso para que el sistema inicie la sesión indicada de acuerdo a las mismas, lo cual da fin al referido caso de uso.

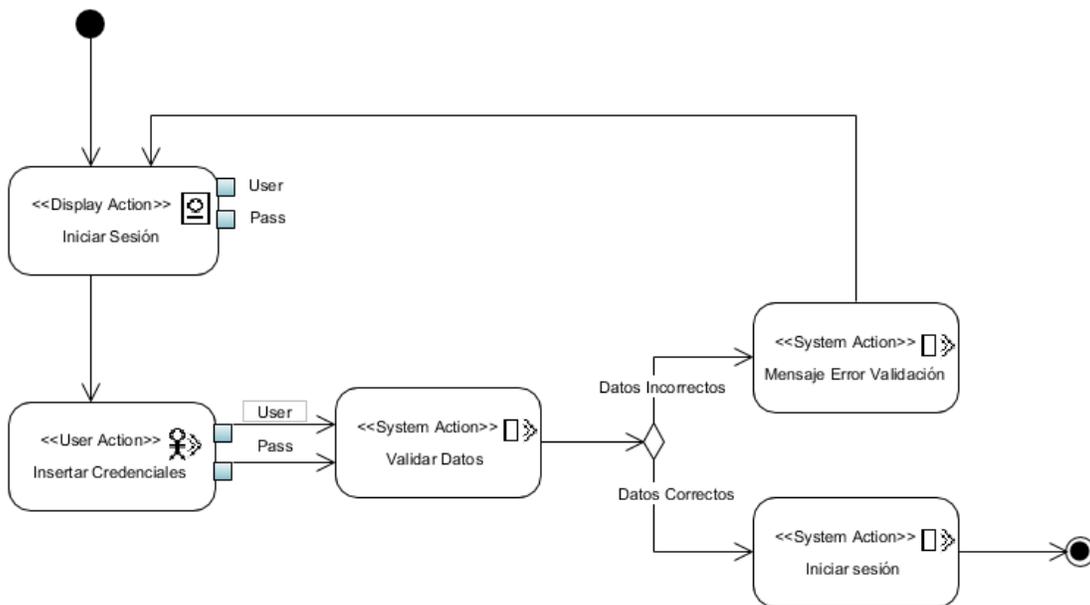


Figura 3.29 Diagrama de proceso del caso de uso Iniciar Sesión.

El diagrama de proceso del caso de uso Cerrar Sesión se presenta en la Figura 3.30. Éste indica que el proceso inicia mostrando en pantalla un formulario mediante el cual el usuario selecciona la opción de cerrar sesión, la cual al ser seleccionada muestra un breve mensaje emergente para confirmar la decisión, para que el sistema cierre la sesión actual y finalice así el caso de uso.

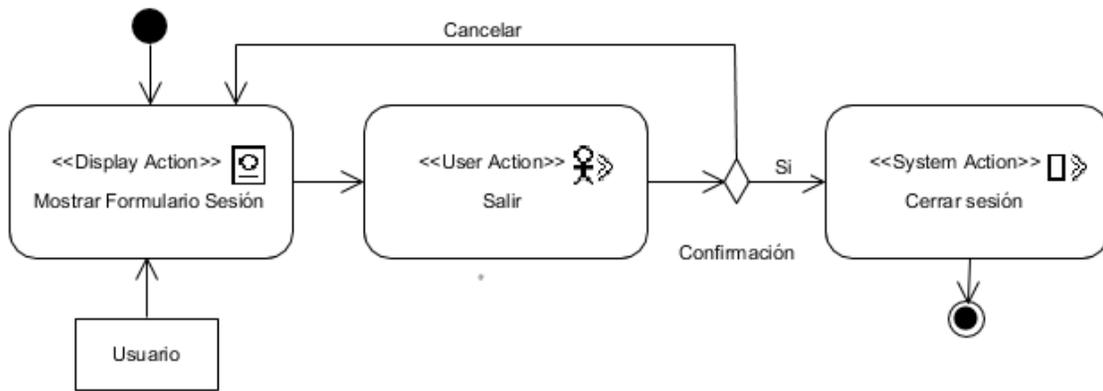


Figura 3.30 Diagrama de proceso del caso de uso Cerrar Sesión.

El diagrama de proceso para el caso de uso Registrar Usuario se visualiza en la Figura 3.31. Ésta indica que el proceso inicia mostrando en pantalla un formulario mediante el cual el usuario autorizado proporciona los datos necesarios, para que seguidamente el sistema guarde el nuevo usuario y dé fin al caso, no sin antes validar los datos previamente introducidos, de tal manera que si los datos son correctos, el nuevo usuario será registrado satisfactoriamente.

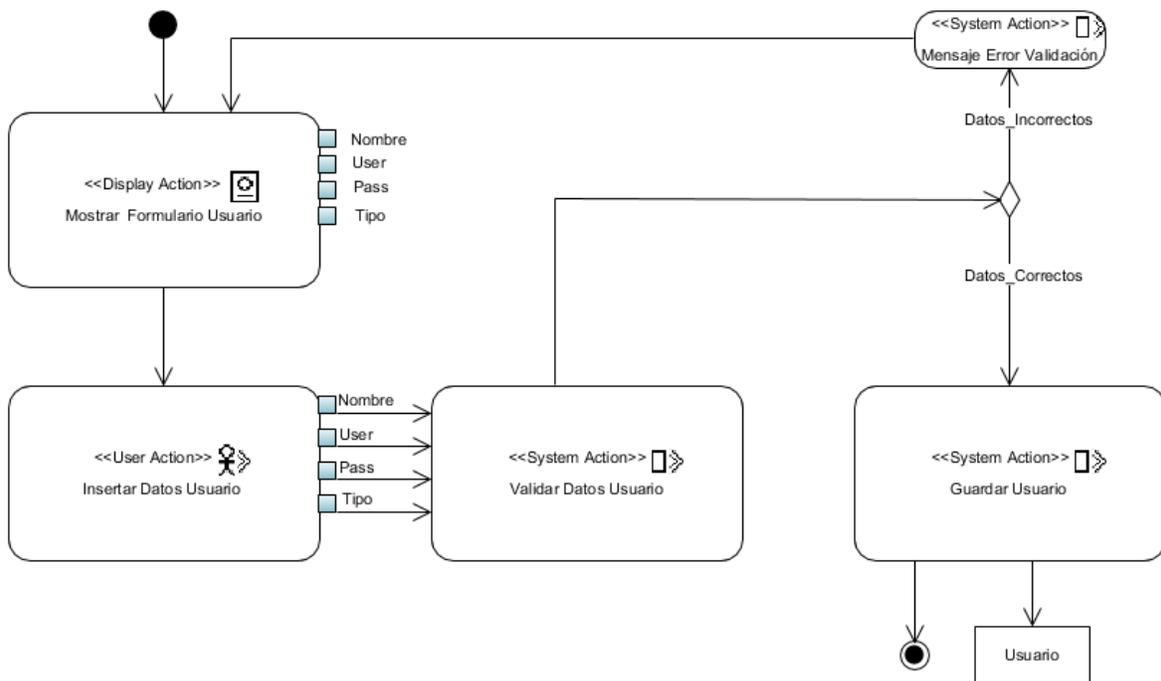


Figura 3.31 Diagrama de proceso del caso de uso Registrar Usuario.

El diagrama de procesos para el caso de uso Validar Información se presenta en la Figura 3.32. El proceso comienza mostrando en pantalla el listado de contenidos registrados en la aplicación y un formulario mediante el cual el usuario autorizado ingresa el identificador del contenido que desea validar. Hecho esto el sistema proporcionará los datos del mismo y en este punto el usuario indica que desea validar la información, para que el sistema proceda a realizar la operación. El caso de uso termina con la validación de contenido.

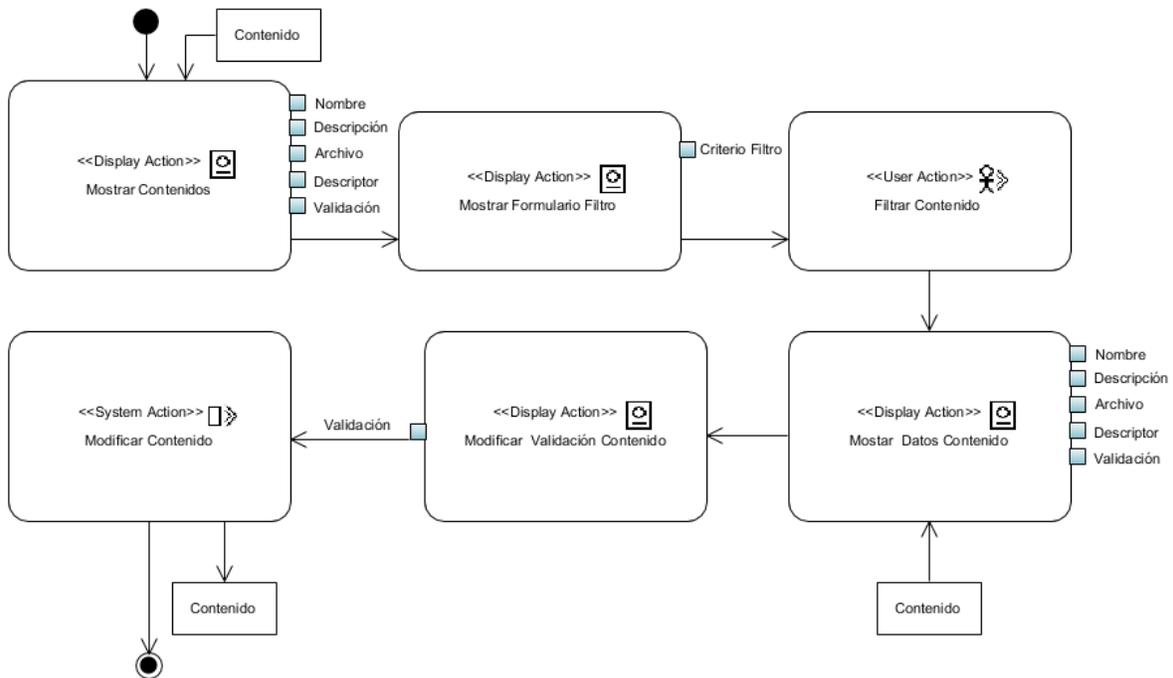


Figura 3.32 Diagrama de proceso del caso de uso Validar Información.

En la Figura 3.33 se observa el diagrama de procesos para el caso de uso Proponer Contenido. El proceso inicia mostrando en pantalla un formulario mediante el cual el usuario inserta el contenido y lo describe, para que seguidamente el sistema valide y guarde la nueva propuesta de contenido, para posteriormente poder ser validada por algún usuario colaborador. Cuando el contenido queda registrado en la aplicación el proceso finaliza.

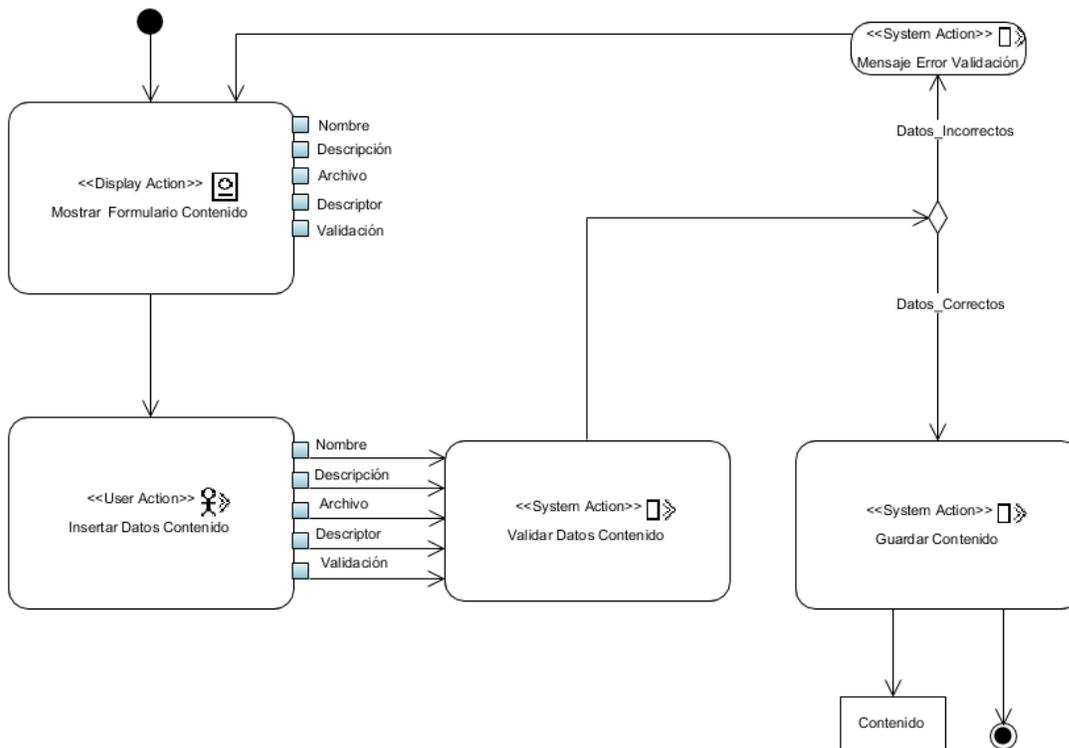


Figura 3.33 Diagrama de proceso del caso de uso Proponer Contenido.

El diagrama de procesos para el caso de uso Alta de Contenido se presenta en la Figura 3.34. Este proceso inicia mostrando en pantalla un formulario mediante el cual el usuario autorizado proporciona los datos necesarios, para que seguidamente el sistema guarde el nuevo contenido y dé fin al proceso, no sin antes haber validado los datos previamente introducidos.

El diagrama de procesos para el caso de uso Modificar Contenido se visualiza en la Figura 3.35. El proceso comienza mostrando en pantalla el listado de contenidos registrados en la aplicación y un formulario mediante el cual el usuario autorizado ingresa el identificador del contenido que desea modificar. Hecho esto el sistema proporcionará los datos del mismo y en este punto el usuario modifica el contenido que desee. El proceso termina con la modificación del contenido.

El diagrama de procesos para el caso de uso Baja de contenido se observa en la Figura 3.36. Comienza mostrando en pantalla el listado de contenidos registrados en la aplicación y un formulario mediante el cual el usuario autorizado ingresa el identificador del contenido que desea eliminar. Hecho esto el sistema proporcionará los datos del mismo y en este punto el usuario indica que desea dar de baja un contenido, para que el sistema proceda a realizar la operación. El proceso termina con la baja del contenido.

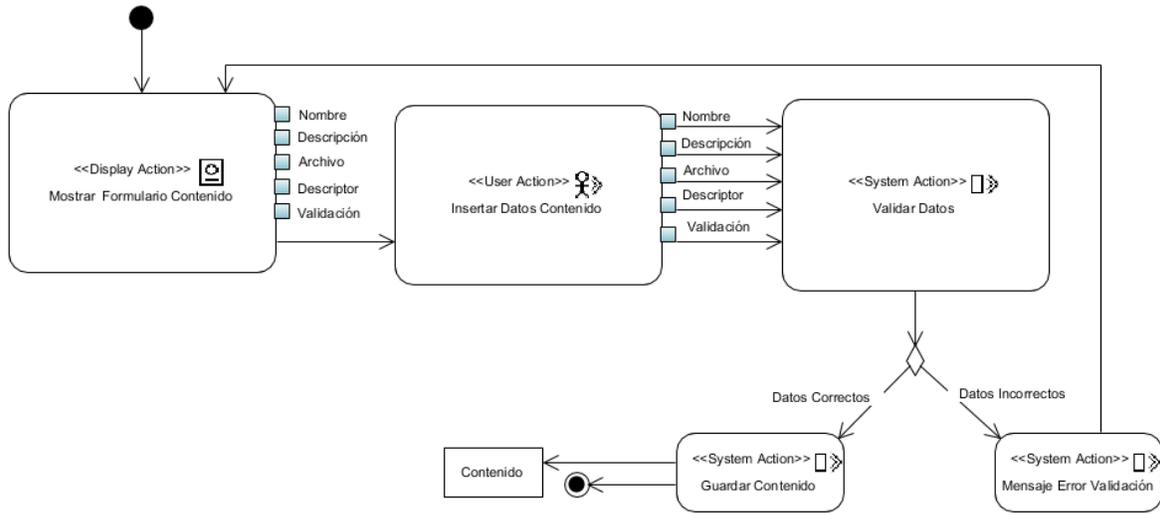


Figura 3.34 Diagrama de proceso del caso de uso Alta Contenido.

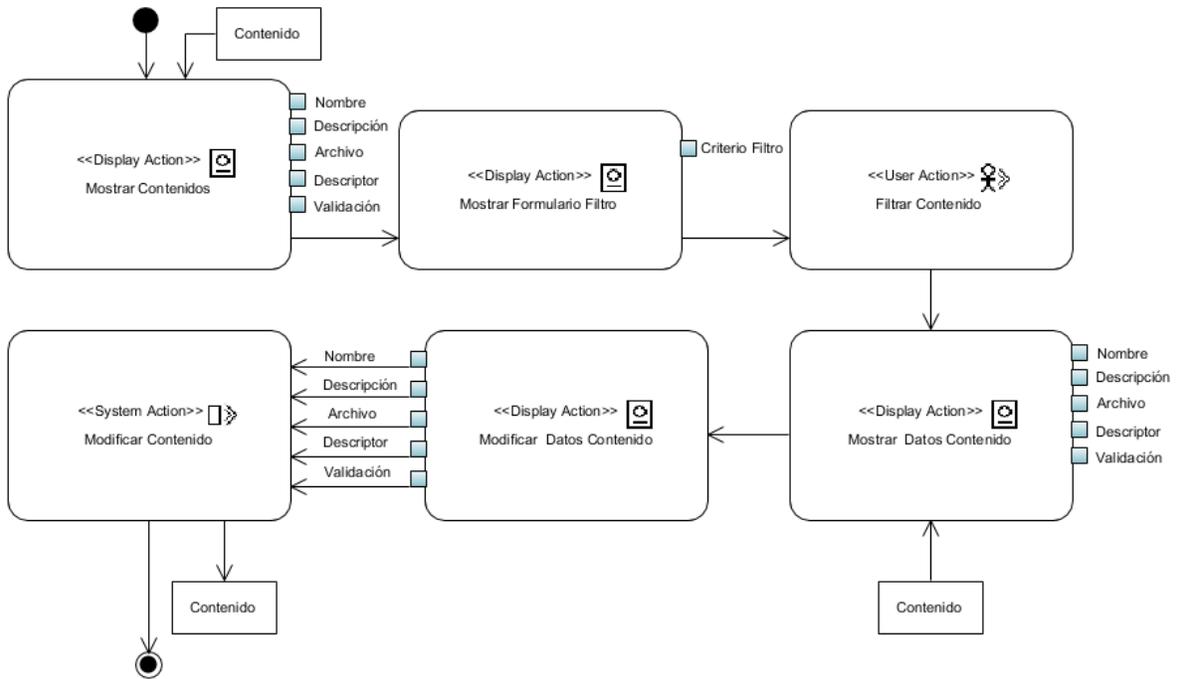


Figura 3.35 Diagrama de proceso del caso de uso Modificar Contenido.

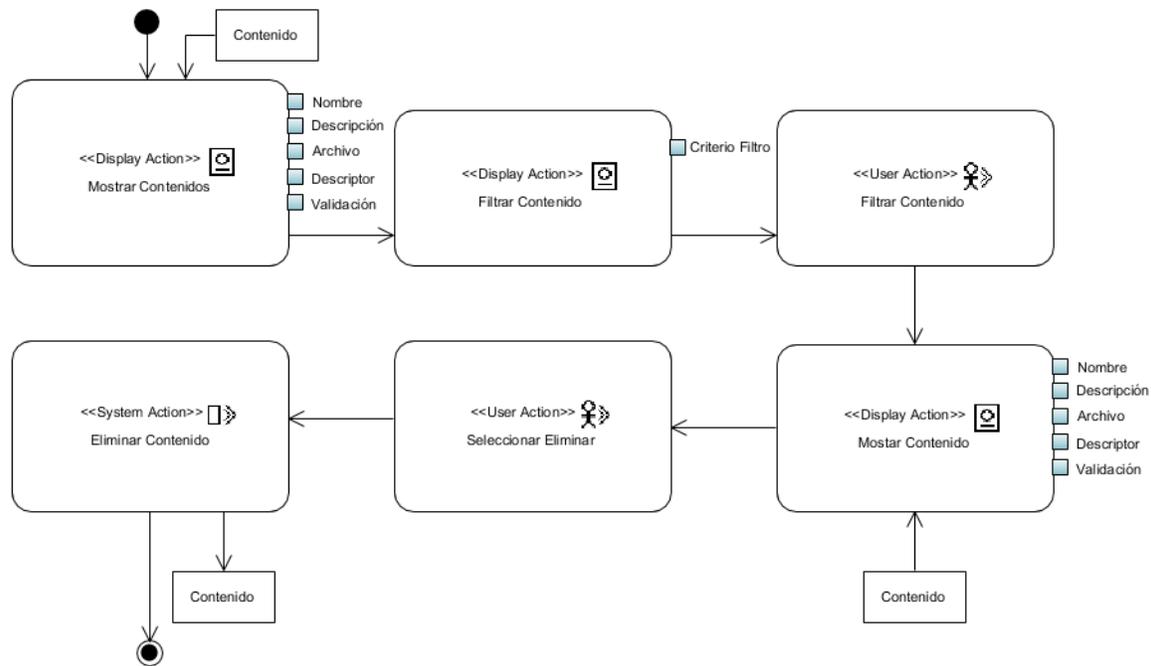


Figura 3.36 Diagrama de proceso del caso de uso Baja de Contenido.

El diagrama de procesos para el caso de uso Alta de Usuario se presenta en la Figura 3.37. Ésta indica que el proceso inicia mostrando en pantalla un formulario mediante el cual el usuario autorizado proporciona los datos necesarios, para que seguidamente el sistema guarde el nuevo usuario y dé fin al proceso.

El diagrama de procesos para el caso de uso Baja de Usuario se visualiza en la Figura 3.38. Comienza mostrando en pantalla el listado de usuarios registrados en el sistema y un formulario mediante el cual el usuario autorizado ingresa el identificador del usuario que desea eliminar. Hecho esto el sistema proporcionará los datos del mismo y en este punto el usuario indica que desea dar de baja un usuario, para que el sistema proceda a realizar la operación. El proceso termina con la baja del usuario. El diagrama de proceso para el caso de uso Modificar Usuario se observa en la Figura 3.39. El proceso inicia mostrando en pantalla el listado de usuarios registrados en la aplicación y un formulario mediante el cual el usuario autorizado ingresa el identificador del usuario que desea modificar. Hecho esto el sistema proporcionará los datos del mismo y en este punto el usuario modifica el contenido que desee. El proceso termina con la modificación del usuario.

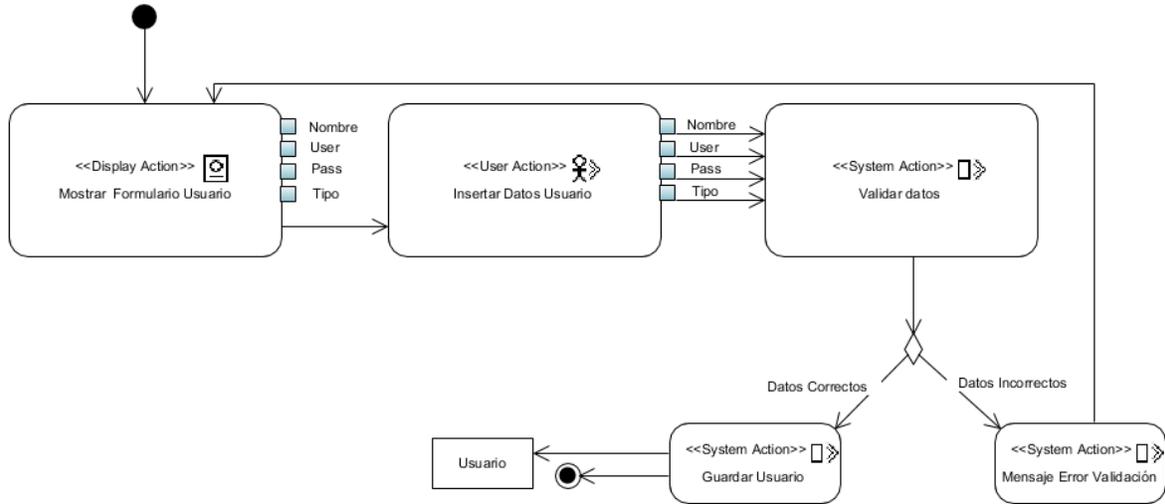


Figura 3.37 Diagrama de proceso del caso de uso Alta de Usuario.

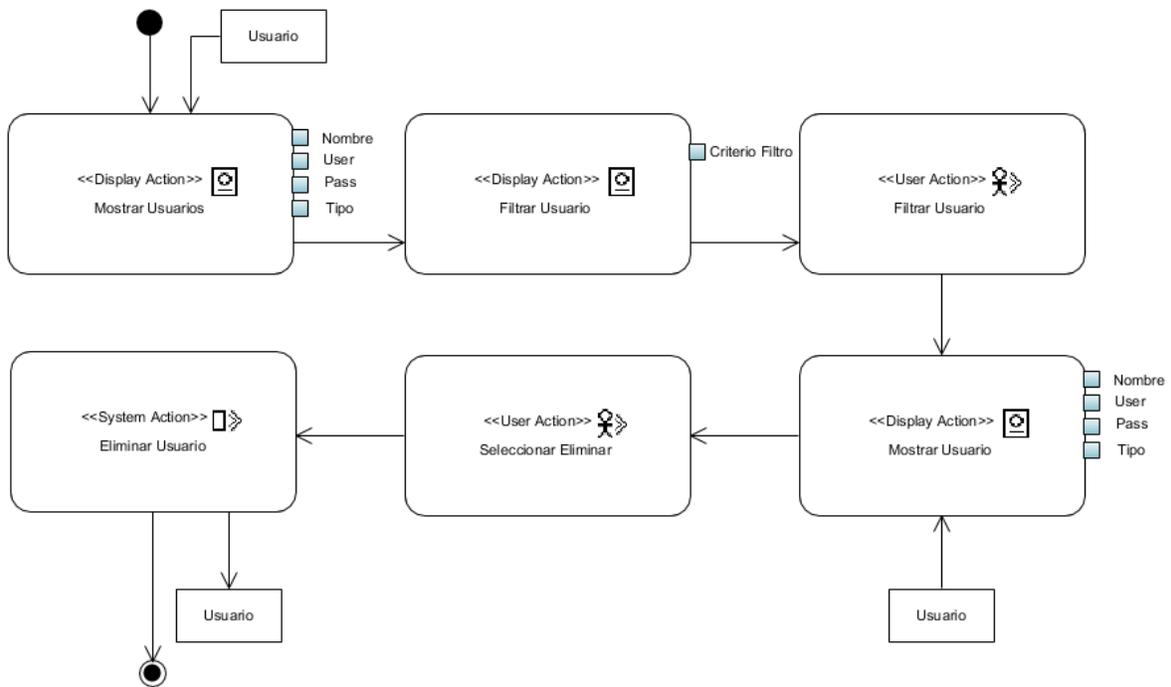


Figura 3.38 Diagrama de proceso del caso de uso Baja de Usuario.

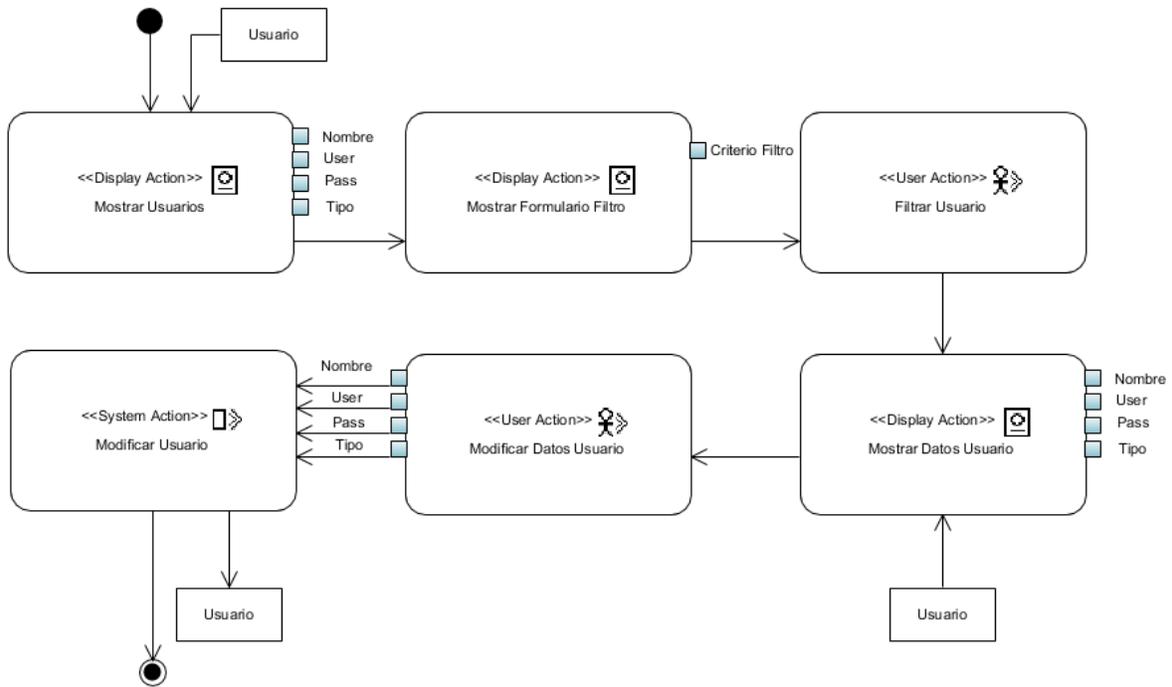


Figura 3.39 Diagrama de proceso del caso de uso Modificar Usuario.

El caso de uso Consultar Usuario, como explica su diagrama de proceso en la Figura 3.40, inicia con la muestra en pantalla del listado de usuarios registrados en la aplicación y un formulario mediante el cual el usuario autorizado filtra los resultados. Para ello proporciona el identificador del usuario que se quiere consultar, a lo que el sistema responderá consecuentemente y dará fin al proceso.

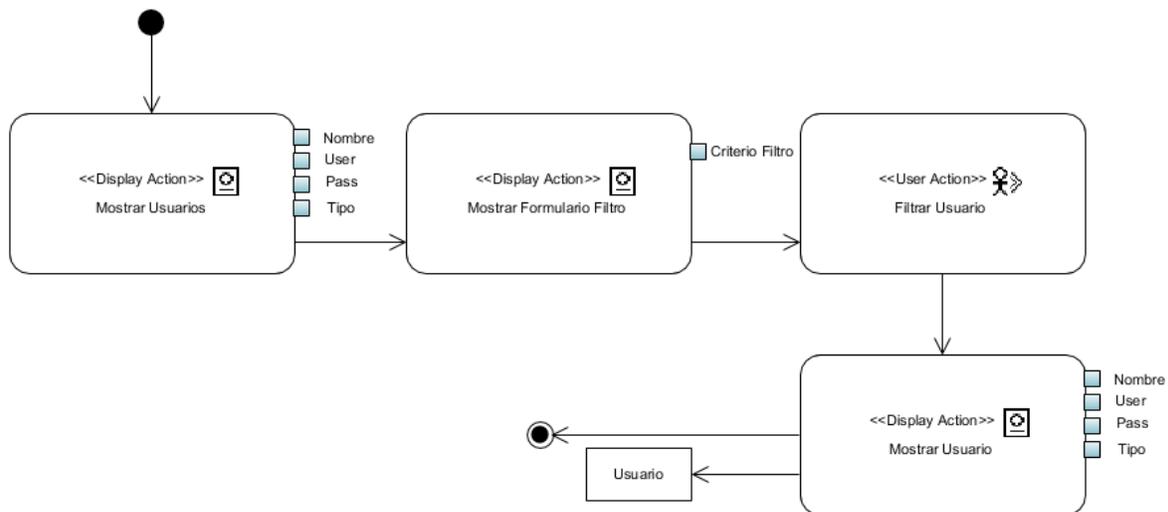


Figura 3.40 Diagrama de proceso del caso de uso Consultar Usuario.

El caso de uso Consulta con Base en Texto, como explica su diagrama de proceso en la Figura 3.41, inicia con la muestra en pantalla de un formulario mediante el cual el usuario autorizado proporciona los datos necesarios para que posteriormente el sistema busque los contenidos con los criterios de búsqueda. Se mostrarán los contenidos encontrados, y se dará fin al proceso.

El caso de uso Consulta con Base en Contenido, como explica su diagrama de proceso en la Figura 3.42, inicia con la muestra en pantalla de un formulario mediante el cual el usuario autorizado proporciona los datos multimedia para posteriormente el sistema realice la consulta basada en contenido. Se mostrarán los contenidos encontrados, y terminará el proceso.

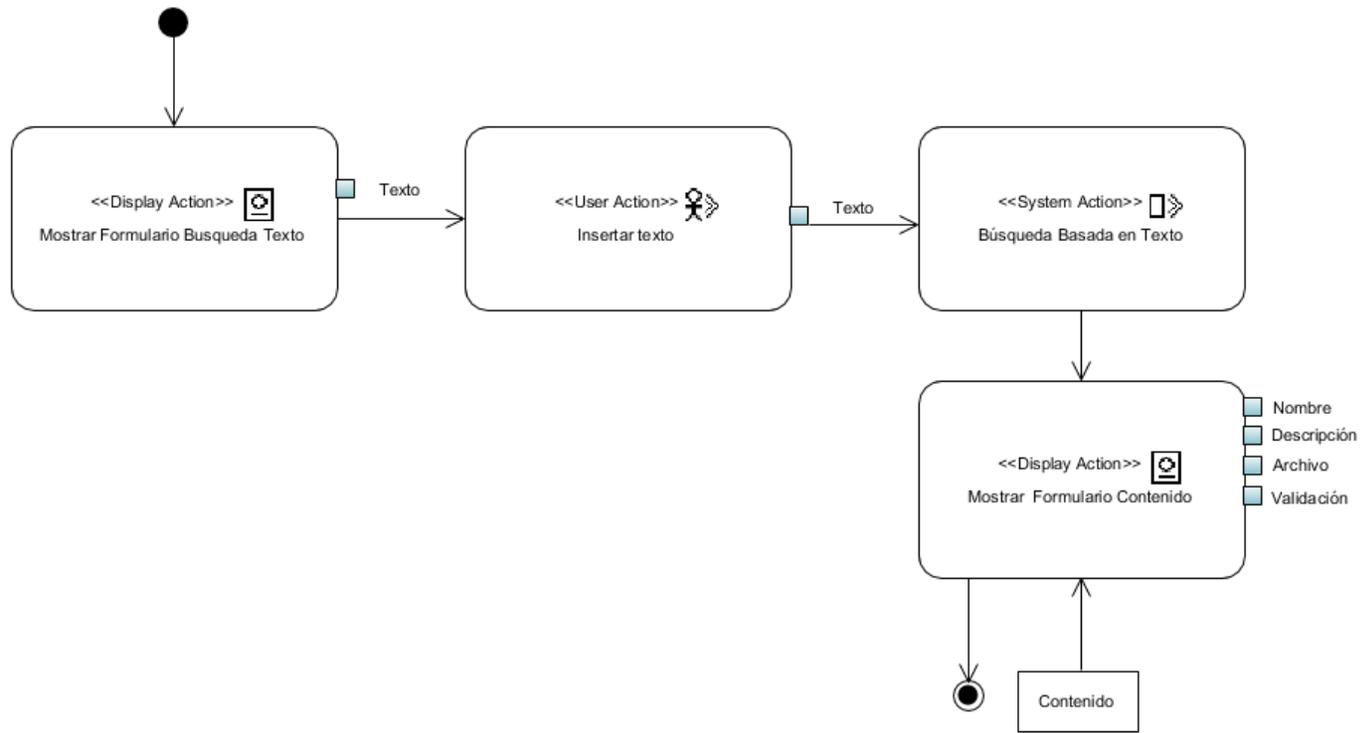


Figura 3.41 Diagrama de proceso del caso de uso Consulta con base en texto.

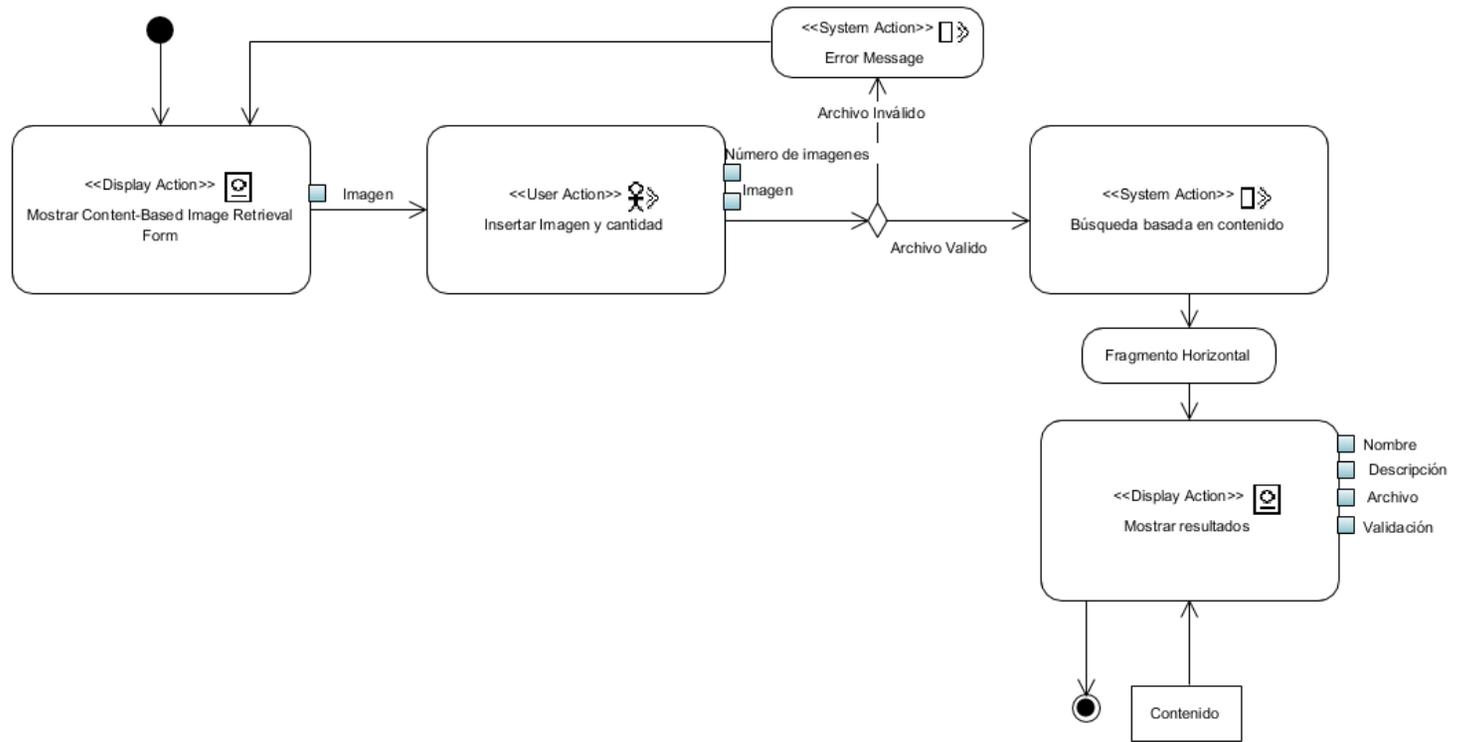


Figura 3.42 Diagrama de proceso del caso de uso Consulta con base en contenido.

## Capítulo 4. Resultados

En este capítulo se muestra el resultado del análisis, se describe el funcionamiento de la aplicación Web y por último se muestran los resultados obtenidos con base en el caso de estudio.

### 4.1 Resultados del análisis

Para finalizar la etapa del análisis, en esta sección se muestran los resultados mediante gráficas y tablas. De esta manera se observa con mayor claridad otro enfoque de los datos mostrados en la Tabla 3.1.

La Figura 4.1 muestra un gráfico de barras que relaciona el número de artículos con la editorial; la mayor cantidad de artículos se concentra en otras editoriales. Los artículos del IEEE abordaron principalmente la recuperación de imágenes basada en contenido ([17], [18], [23], [25], [26]), 5 de 8 artículos. Solo 10 artículos consideraron fragmentación horizontal ([21], [22], [23], [24] y [27] - [32]).

La distribución de los artículos analizados según su año de publicación se muestra en la Figura 4.2. Se observa que la mayoría de los artículos son de 2013 a 2019. En este último año, los trabajos se enfocan principalmente en optimizar el almacenamiento y procesamiento de consultas de *Digital Imaging and Communication in Medicine* (DICOM) [51] y métodos para sistemas CBIR eficientes [39], [44], [45], [48].

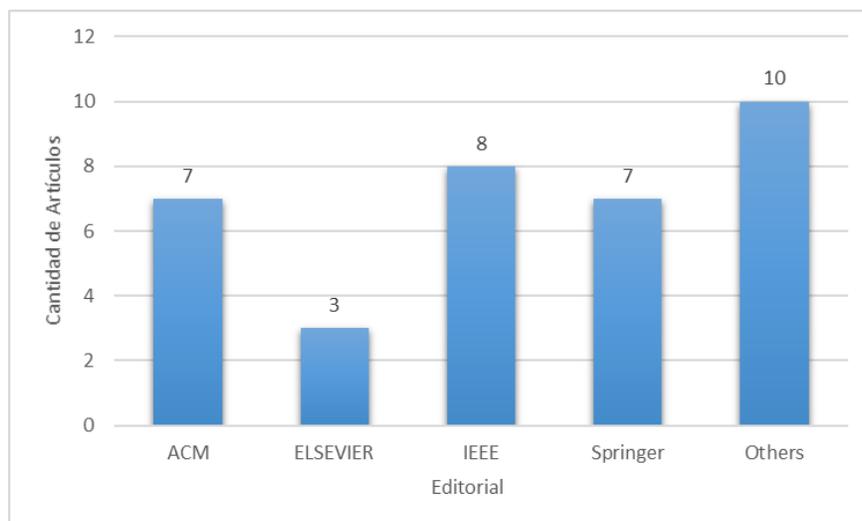


Figura 4.1 Cantidad de artículos por editorial

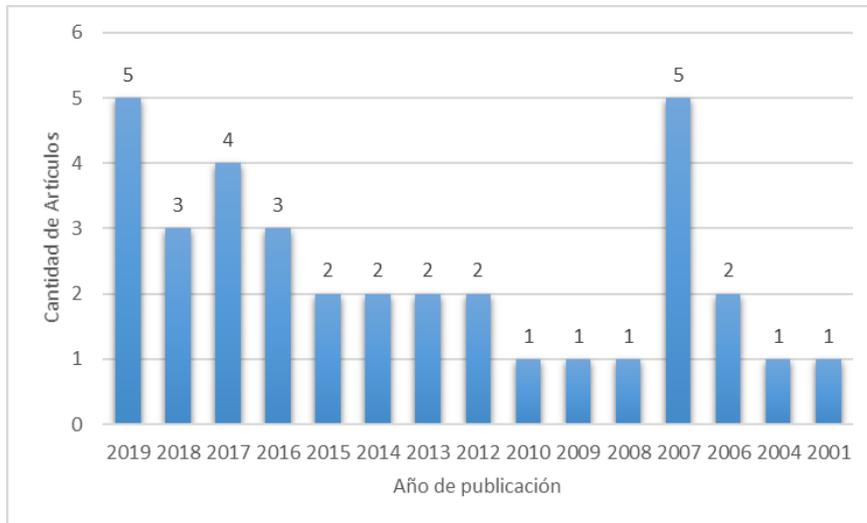


Figura 4.2 Cantidad de artículos por año

En la Figura 4.3 se compara la cantidad de trabajos que consideran fragmentación horizontal ([21], [22], [23], [24] y [27] - [32]) y el número de trabajos que no toman en cuenta fragmentación horizontal ([14], [17], [18] [25], [26], [33]-[52]). En la Figura 4.3, se observa que la fragmentación horizontal es un tema menos común entre los 35 artículos. Sin embargo, hay artículos que incluyen otro tipo de fragmentación ([33] - [36], [38], [40], [42], [50], [51]).

La Figura 4.4 muestra un gráfico de barras que relaciona el número de artículos por tipo de fragmentación horizontal. El número de artículos que incluyen la fragmentación horizontal primaria es 9, y solo 1 artículo considera fragmentación horizontal derivada [32]. En [33] y [35], los autores realizaron fragmentación mixta multimedia.

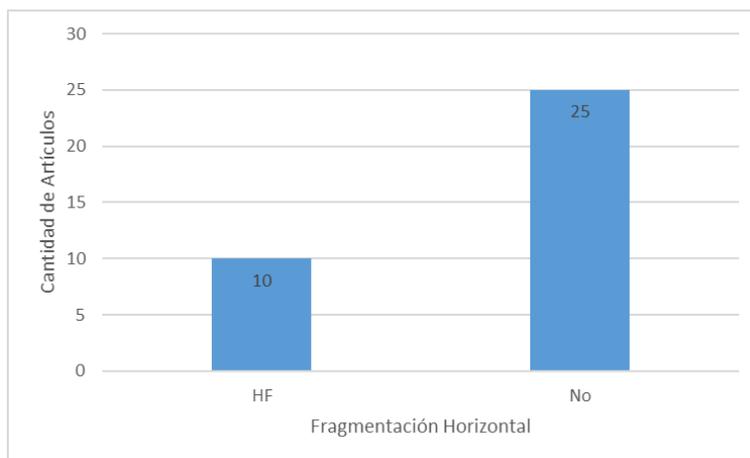


Figura 4.3 Cantidad de artículos por fragmentación horizontal

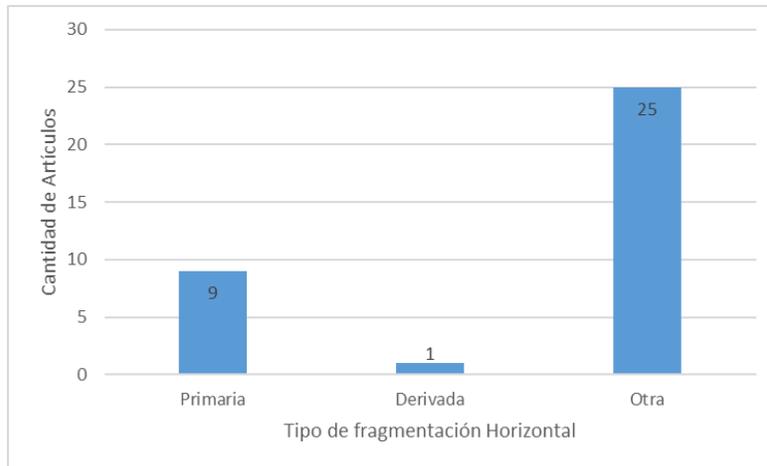


Figura 4.4 Cantidad de artículos por tipo de fragmentación horizontal

La Figura 4.5 muestra un gráfico de barras que relaciona el número de artículos que tienen en cuenta la recuperación basada en contenido; 19 de 35 propusieron un método CBIR. Varios trabajos consideraron la agrupación de las características de imagen [17], [18], [42], [43], [45], [47], [49] o el uso de Map-Reduce [14], [39], [41], [52] para optimizar los sistemas CBIR, minimizando el tiempo de respuesta de las consultas. Las consultas de similitud más usadas fueron las consultas de rango y  $k$ -NN [21], [22], [23], [29], [33]. 16 artículos no consideraron consultas basadas en contenido ([24], [27], [28], [30]-[32], [34], [36], [38], [40]-[42], [46] y [50]-[52]).

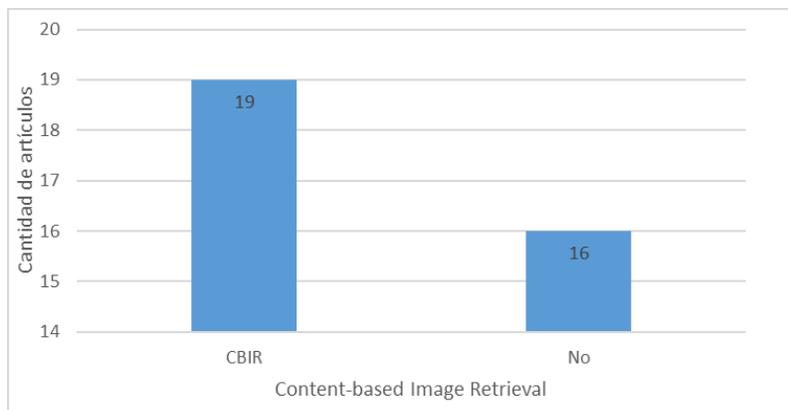


Figura 4.5 Cantidad de artículos por content-based image retrieval

La Figura 4.6 muestra un gráfico de barras que relaciona el número de artículos que tienen en cuenta un modelo de costos; 4 de 35 métodos utilizaron un modelo de costos, [24] y [28] de Springer, [30] y [32] de otras editoriales. En [24], el costo de un esquema de fragmentación horizontal se compone de dos partes: costo de acceso a datos irrelevantes y costo de transporte. Por el contrario, los artículos [28], [30] y [32] se centraron en el análisis de los costos de consulta, que consisten en costos de almacenamiento y costos de transporte. Por lo tanto, 31 artículos no consideraron un modelo de costos ([14], [17], [18], [21], [22], [23], [25], [26], [27], [29], [31], [33]-[52]).

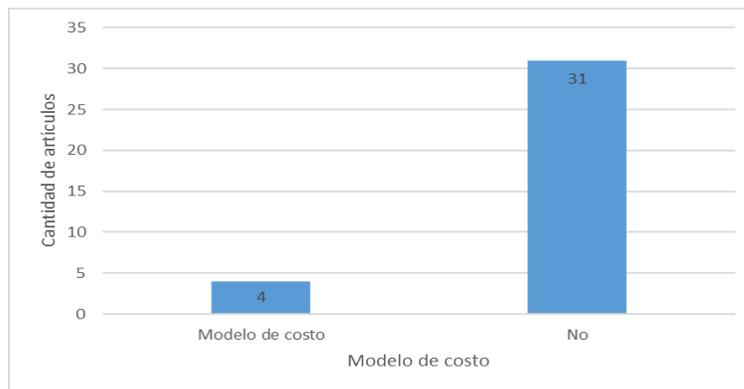


Figura 4.6 Cantidad de artículos por modelo de costo

La Figura 4.7 muestra el número de artículos que tienen una alta facilidad de implementación, la cual se mide por la cantidad de conocimientos específicos necesarios para implementar las técnicas propuestas por cada artículo; se considera que 21 de 35 artículos tienen una gran facilidad de implementación ([4], [14], [17], [23], [24], [27], [28], [30], [32]-[34], [35], [39]-[42], [45], [47], [50]-[52]). La mayoría de estos métodos son fáciles de implementar porque son completos, es decir, presentan toda la información necesaria para reproducirlos. Se consideró que 14 artículos tenían poca facilidad de implementación ([18], [22], [25], [26], [29], [31], [35], [36], [38], [43], [44], [46], [48], [49]).

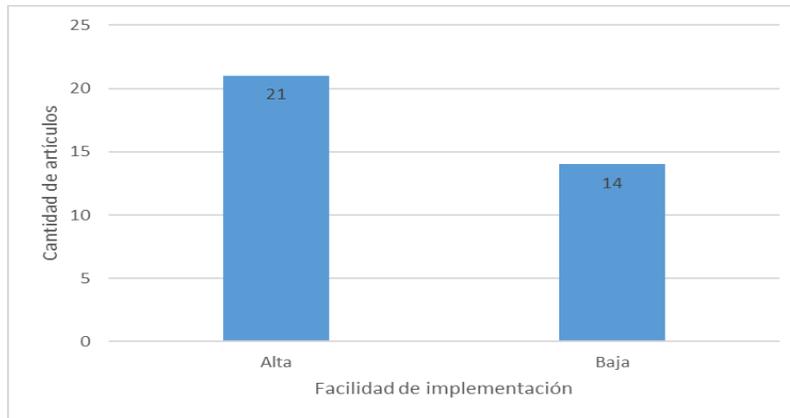


Figura 4.7 Cantidad de artículos por facilidad de implementación

## 4.2 Fragmentación Horizontal

La colección “imagenes” de la base de datos HITO (Historia del Instituto Tecnológico de Orizaba) está fragmentada considerando el atributo tipo (el cual tiene cuatro valores: edificio, equipo, evento o persona) según el método de Fasolin et al. [8], que da como resultado los fragmentos horizontales denominados: *con\_buildings*, *con\_equipment*, *con\_events* y *con\_personal*. Esto se aprecia en la Figura 4.8.

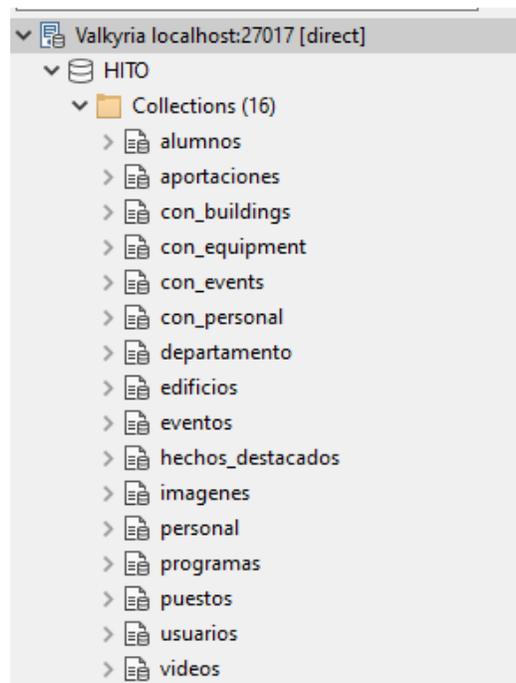


Figura 4.8 Fragmentación horizontal aplicada a la colección “imagenes” de la base de datos HITO

### 4.3 Sistema de Gestión de Datos Multimedia que utiliza un método de fragmentación horizontal para optimizar consultas basadas en contenido

La primer pantalla del sistema muestra información sobre la historia del Instituto Tecnológico de Orizaba, así como su misión y valores. En la parte inferior de la página también se encuentran ciertos enlaces a las redes sociales del instituto. En la parte superior se tiene un enlace para poder iniciar sesión dentro del sistema (ver Figuras 4.9 y 4.10).



Figura 4.9 Pagina de bienvenida del sistema 1

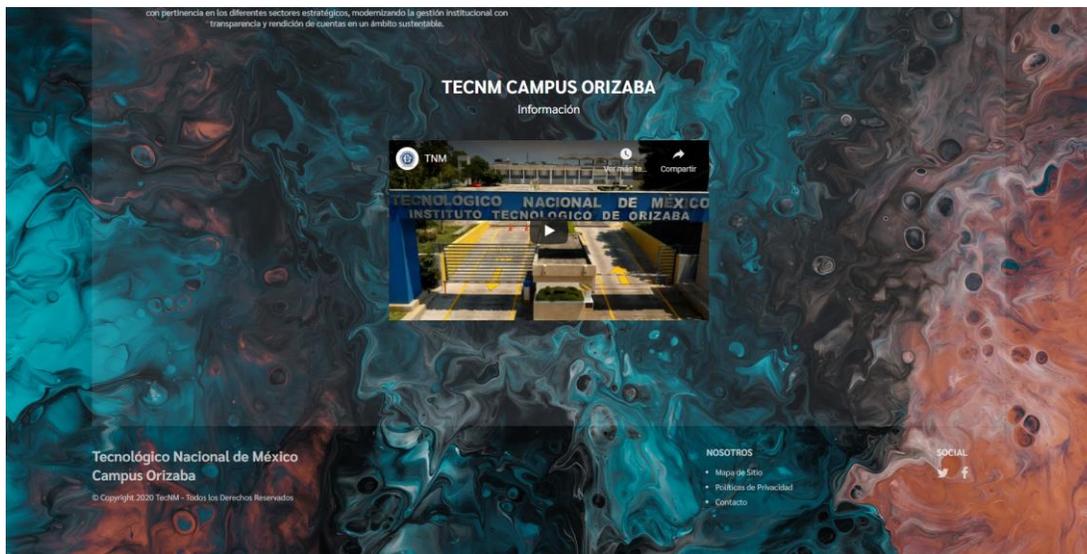


Figura 4.10 Pagina de bienvenida del sistema 2

Antes de acceder al sistema es necesario iniciar sesión ya sea como un administrador, colaborador, profesor o usuario general. En la Figura 4.11 se muestra la interfaz para el inicio de sesión, dentro de la misma interfaz se encuentra la opción para crear un usuario nuevo. La Figura 4.12 muestra la interfaz para crear un usuario nuevo.

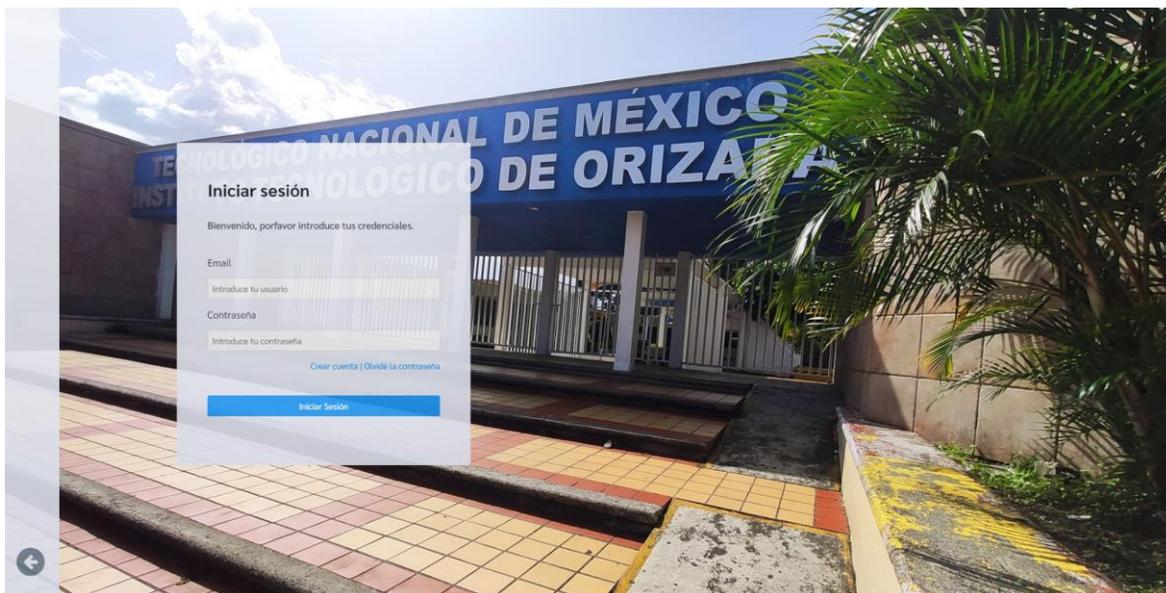


Figura 4.11 Página de Inicio de sesión 1

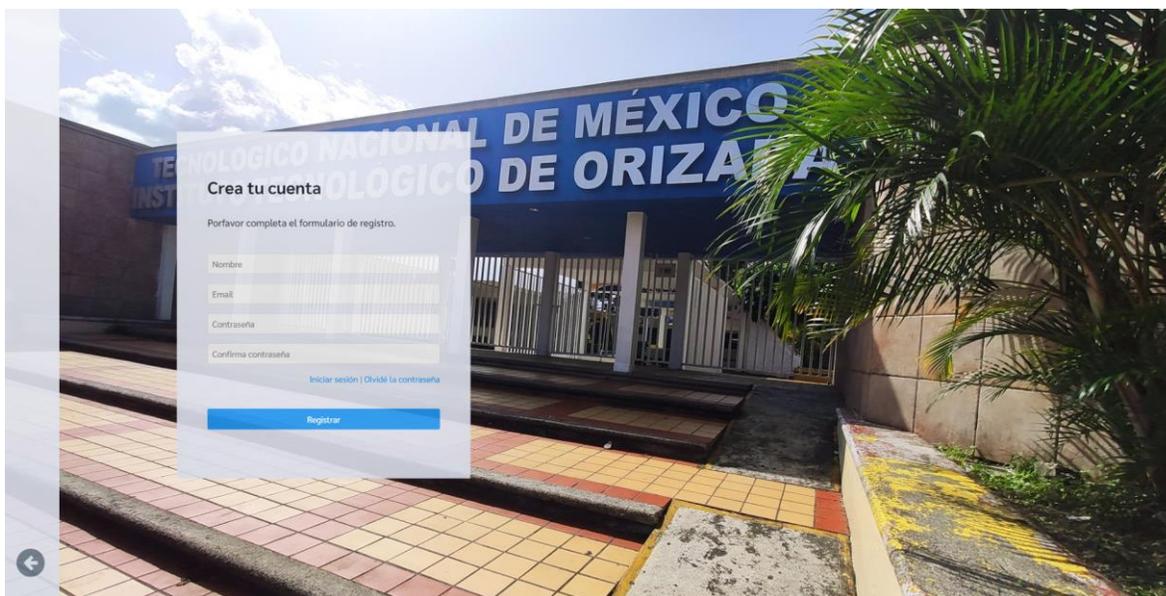


Figura 4.12 Página de Inicio de sesión - Crear cuenta

La Figura 4.13 muestra la página de bienvenida una vez se haya iniciado sesión dentro del sistema. Al inicio de la página de bienvenida se encuentran los datos del usuario y un mensaje confirmando una conexión exitosa al sistema, así como un breve espacio para anuncios o información correspondiente al Instituto Tecnológico de Orizaba.

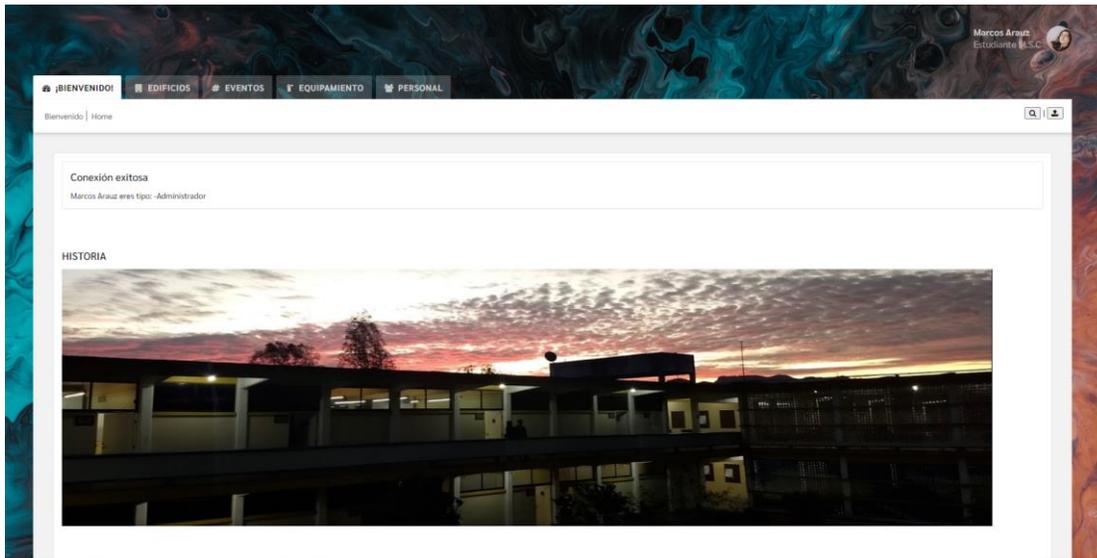


Figura 4.13 Página de bienvenida al sistema una vez se inicia sesión

La Figura 4.14 muestra la primera de cuatro páginas similares que contienen los datos multimedia de los fragmentos horizontales obtenidos de la colección "imágenes" de la base de datos. En este caso se muestra la colección de imágenes del fragmento horizontal llamado: *con\_buildings*.

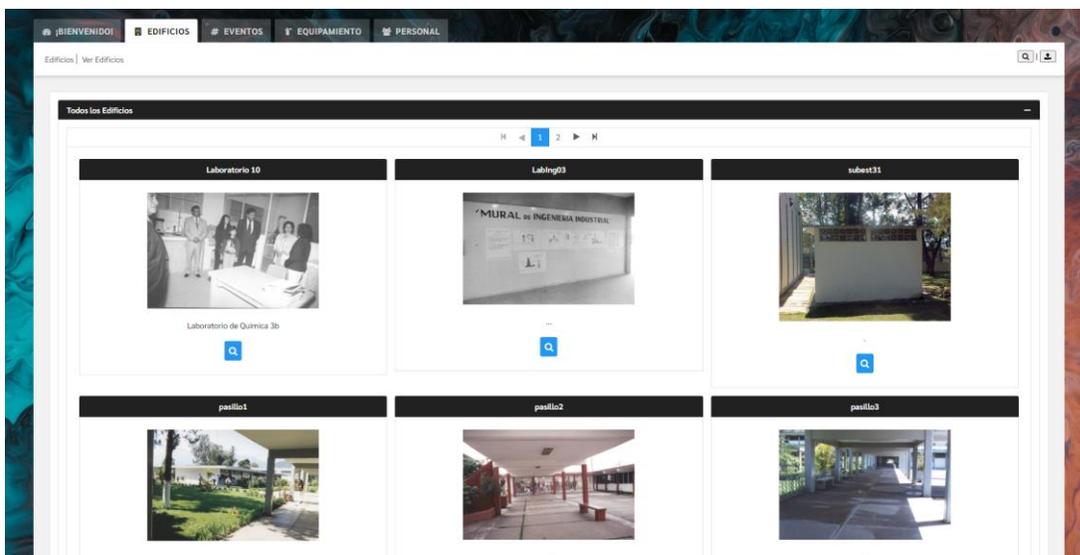


Figura 4.14 Página de contenido de Edificios 1

Cada una de las imágenes que se muestran en pantalla tienen debajo un botón para ver los detalles de la imagen. Esto se observa en la Figura 4.15.

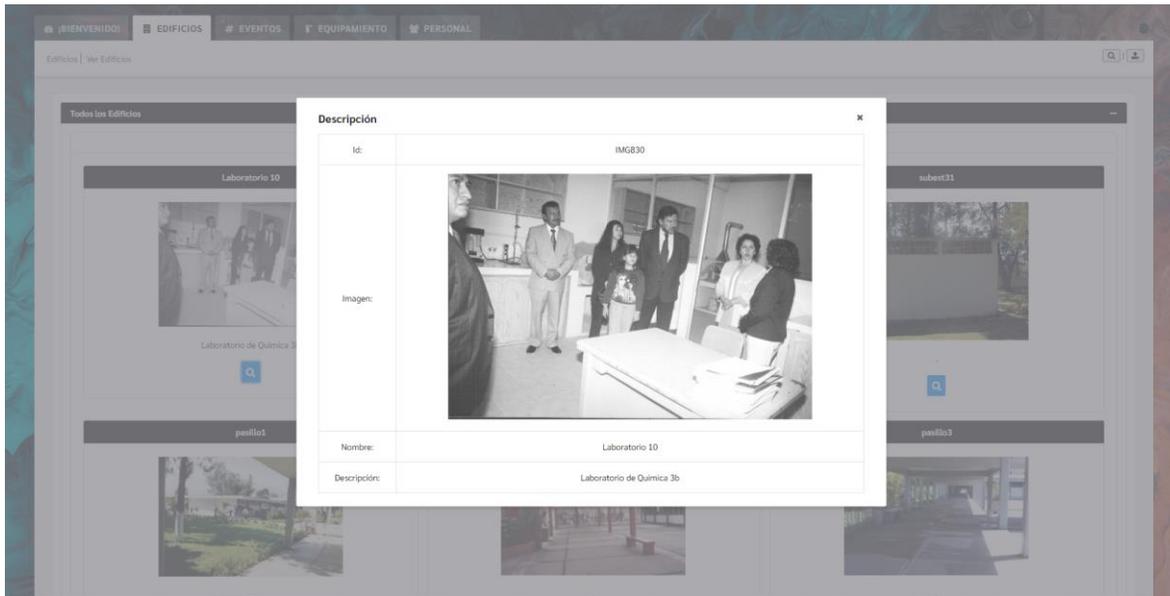


Figura 4.15 Página de contenido de Edificios 2

En la Figura 4.16 se muestra la página que contiene las imágenes del fragmento horizontal: *con\_events*.

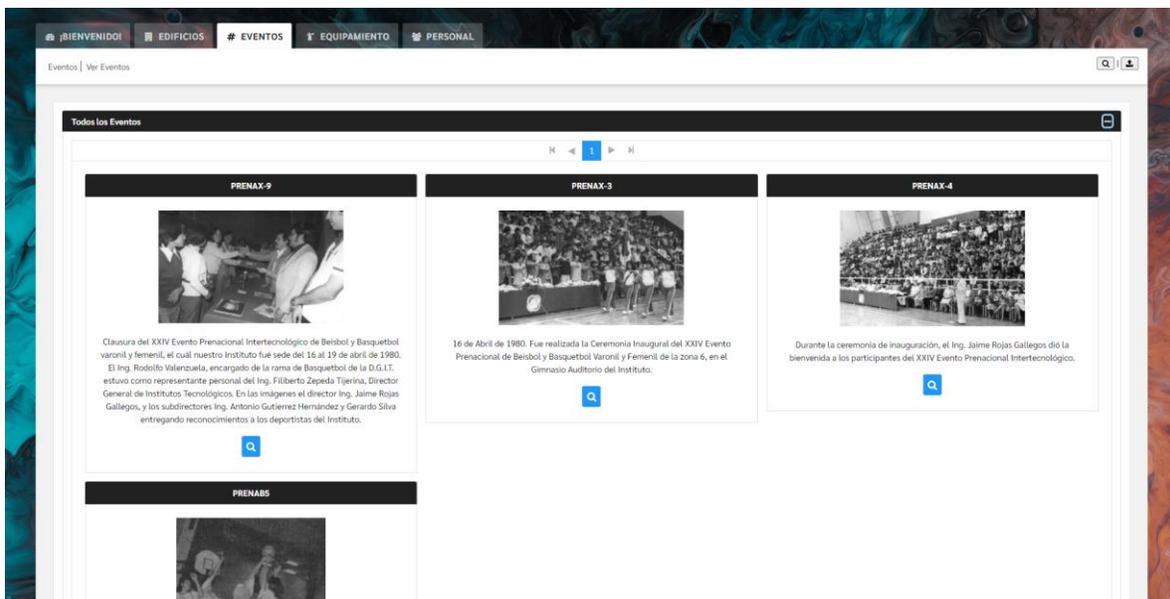


Figura 4.16 Página de contenido de Eventos

En la Figura 4.17 se muestra la página que contiene las imágenes del fragmento horizontal: *con\_equipment*.

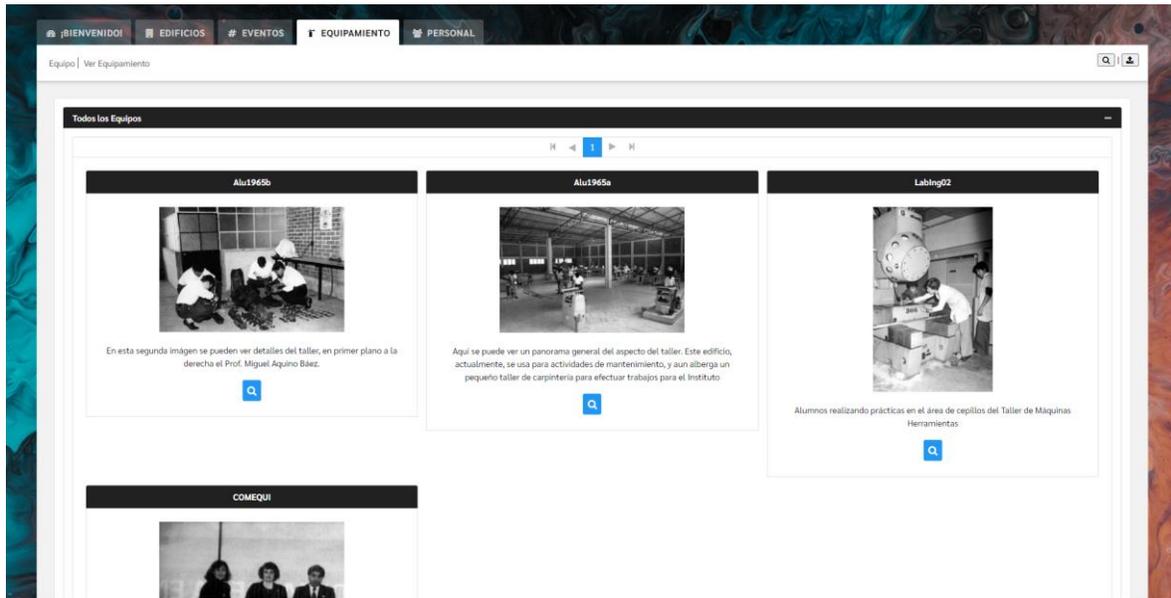


Figura 4.17 Página de contenido de Equipo

En la Figura 4.18 se muestra la página que contiene las imágenes del fragmento horizontal: *con\_personal*.

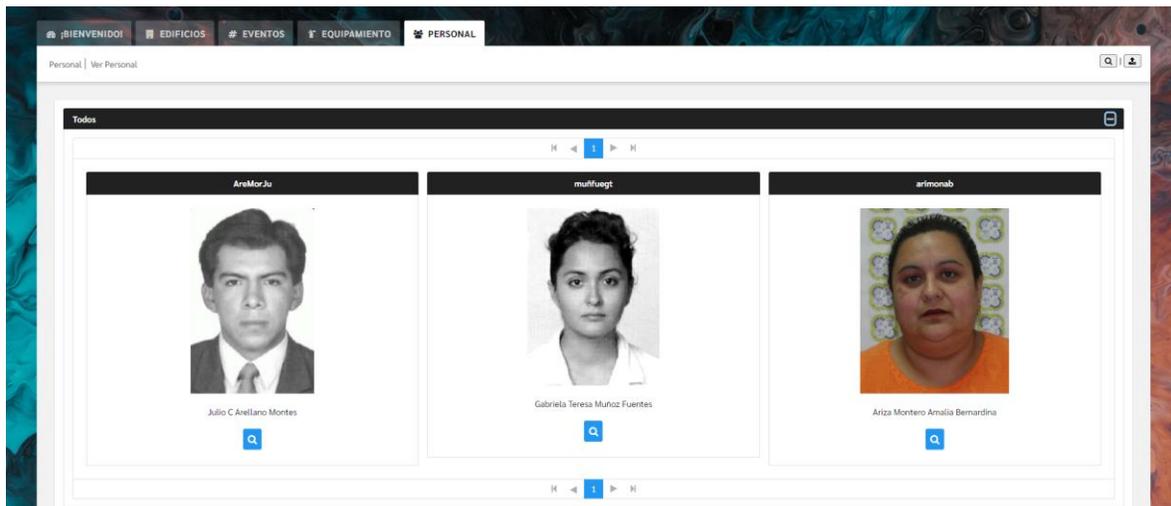


Figura 4.18 Página de contenido de Personal

Los usuarios de tipo administrador tienen la posibilidad de acceder a todas las características de gestión de información dentro del sistema. Los usuarios de tipo colaborador solo tienen acceso a las opciones de gestión de información de contenido, como lo son las opciones de: Edificios, Eventos, Equipo y Personal. Mientras que los usuarios tipo profesor y usuario general, no tienen acceso a ninguna de estas opciones. Esto se representa en la Figura 4.19, donde también se observa la página que permite la gestión de la información de los usuarios. En esta página se muestra una tabla de datos, que contiene la información de los usuarios y donde cada fila tiene dos botones, uno para editar y otro para eliminar la fila correspondiente.

| Id                                   | Usuario        | Contraseña | Email               | Tipo          |
|--------------------------------------|----------------|------------|---------------------|---------------|
| PERS4                                | Marcos Arauz   | 123        | marcos@hotmail.com  | Administrador |
| 8ce54d25-53cc-4c86-8904-5249278a000f | Marcos Dos     | 123        | marcos@aa.com       | Profesor      |
| d8815d26                             | Joaquin        | 123        | joaquin@hotmail.com | Profesor      |
| bf00849f                             | Marcos Arauz   | 123456     | m@gmail.com         | Colaborador   |
| 84db003c                             | Marcos         | 123        | marcos@gmail.com    | Administrador |
| 96819d49                             | Joacho         | 134        | joachogmail.com     | Profesor      |
| 1e070a1e                             | Joacho         | 134        | joa@gmail.com       | Administrador |
| 33d9fb5f                             | Carlos Alberto | cesar123   | carlos@gmail.com    | Colaborador   |
| 07d32ba2                             | Juanito Perez  | mejor123   | juan@gmail.com      | Profesor      |
| a4682c0c                             | Maria          | 123        | maria@gmail.com     | Administrador |

Figura 4.19 Página de gestión de usuarios 1

Cuando se selecciona el botón para editar cualquier fila, la misma fila cambia y muestra los espacios editables para su posterior modificación, esto se visualiza en la Figura 4.20. Si se seleccionó una fila para ser editada pero ningún campo fue modificado o si se cancela la edición, se procederá a cambiar el modo de edición de la fila, a un modo de solo lectura, también el sistema notifica que la fila seleccionada no fue editada, como se observa en la Figura 4.21. La Figura 4.22 representa el caso en el que una fila haya sido modificada satisfactoriamente, entonces el sistema notifica que la fila seleccionada fue modificada exitosamente. Mientras que en la Figura 4.23 se puede observar que al seleccionar el botón para eliminar una fila, aparecerá un diálogo requiriendo la confirmación del administrador, que en caso de proceder, eliminará el usuario, caso contrario se cerrará el diálogo sin cambios en la tabla de datos.

| Id                                   | Usuario        | Contraseña | Email               | Tipo          |
|--------------------------------------|----------------|------------|---------------------|---------------|
| PERS4                                | Marcos Arauz   | 123        | marcos@hotmail.com  | Administrador |
| 8ce54d25-53cc-4c86-8904-5249278a000f | Marcos Dos     | 123        | marcos@aa.com       | Profesor      |
| d815d26                              | Joaquín        | 123        | joaquin@hotmail.com | Profesor      |
| bf00849f                             | Marcos Arauz   | 123456     | m@gmail.com         | Colaborador   |
| 84c003c                              | Marcos         | 123        | marcos@gmail.com    | Administrador |
| 968104d9                             | Joacho         | 134        | joacho@gmail.com    | Profesor      |
| 1e070a1e                             | Joacho         | 134        | joa@gmail.com       | Administrador |
| 33df65f                              | Carlos Alberto | cesar123   | carlos@gmail.com    | Colaborador   |
| 07d32be2                             | Juanito Perez  | mejor123   | juan@gmail.com      | Profesor      |
| a4682c0c                             | Maria          | 123        | maria@gmail.com     | Administrador |

Figura 4.20 Página de gestión de usuarios 2

Edición cancelada PERS4

| Id                                   | Usuario        | Contraseña | Email               | Tipo          |
|--------------------------------------|----------------|------------|---------------------|---------------|
| PERS4                                | Marcos Arauz   | 123        | marcos@hotmail.com  | Administrador |
| 8ce54d25-53cc-4c86-8904-5249278a000f | Marcos Dos     | 123        | marcos@aa.com       | Profesor      |
| d815d26                              | Joaquín        | 123        | joaquin@hotmail.com | Profesor      |
| bf00849f                             | Marcos Arauz   | 123456     | m@gmail.com         | Colaborador   |
| 84c003c                              | Marcos         | 123        | marcos@gmail.com    | Administrador |
| 968104d9                             | Joacho         | 134        | joacho@gmail.com    | Profesor      |
| 1e070a1e                             | Joacho         | 134        | joa@gmail.com       | Administrador |
| 33df65f                              | Carlos Alberto | cesar123   | carlos@gmail.com    | Colaborador   |
| 07d32be2                             | Juanito Perez  | mejor123   | juan@gmail.com      | Profesor      |
| a4682c0c                             | Maria          | 123        | maria@gmail.com     | Administrador |

Figura 4.21 Página de gestión de usuarios 3

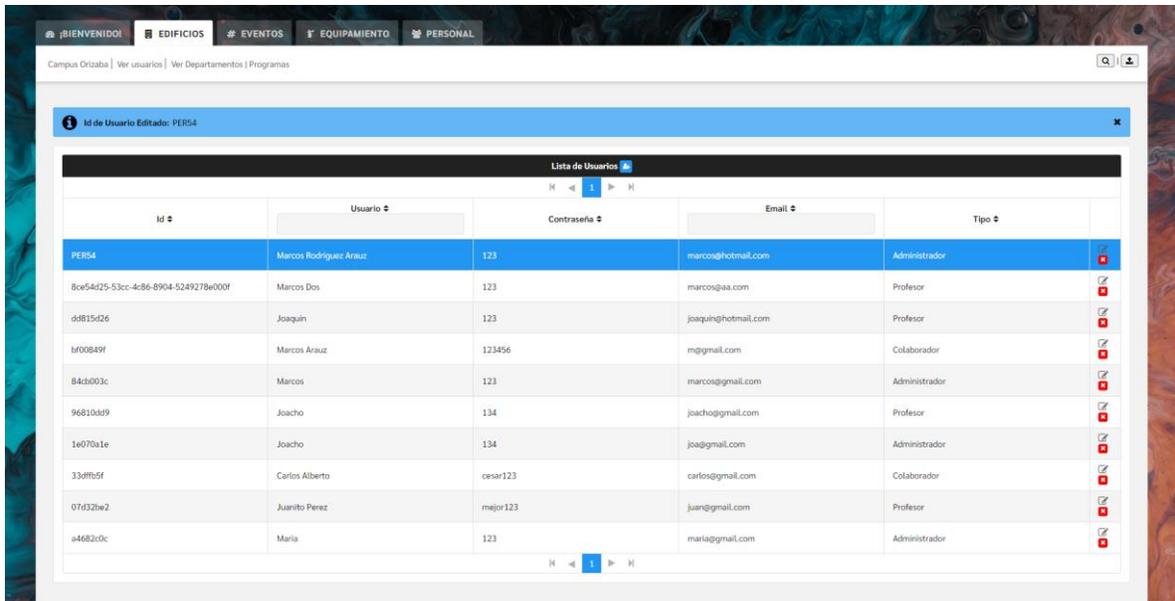


Figura 4.22 Página de gestión de usuarios 4

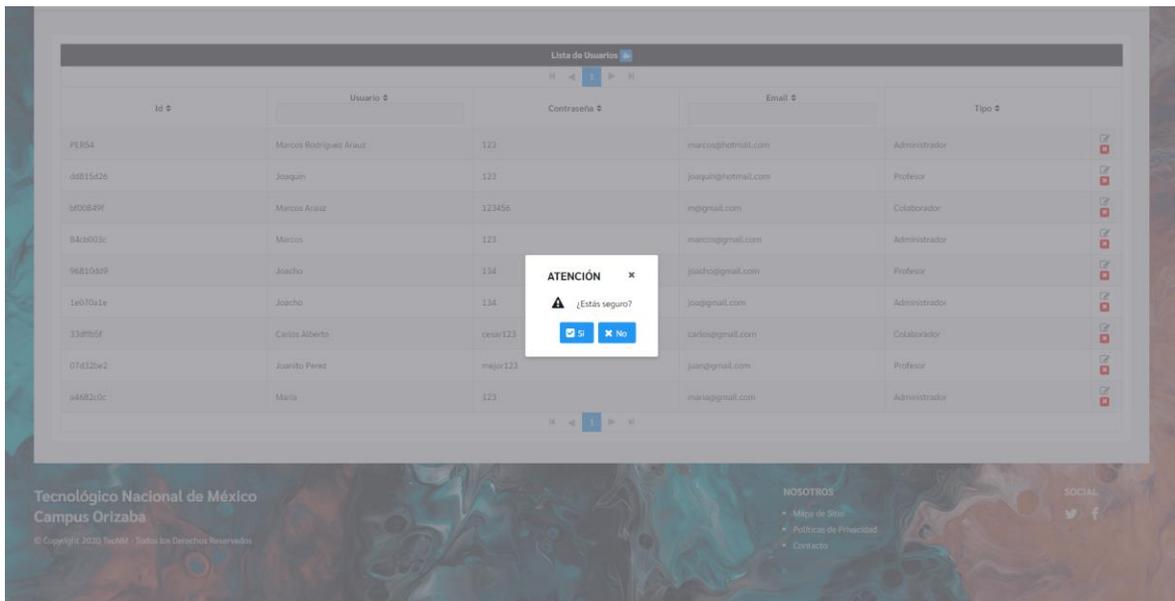


Figura 4.23 Página de gestión de usuarios 5

En la Figura 4.24 se observa la interfaz que aparece cuando se selecciona el botón localizado a la derecha del título de la tabla de datos, la cual contiene el formulario necesario para registrar un nuevo usuario y también la posibilidad de escoger el tipo de usuario que se creará (Administrador, Colaborador, Profesor o Usuario General).

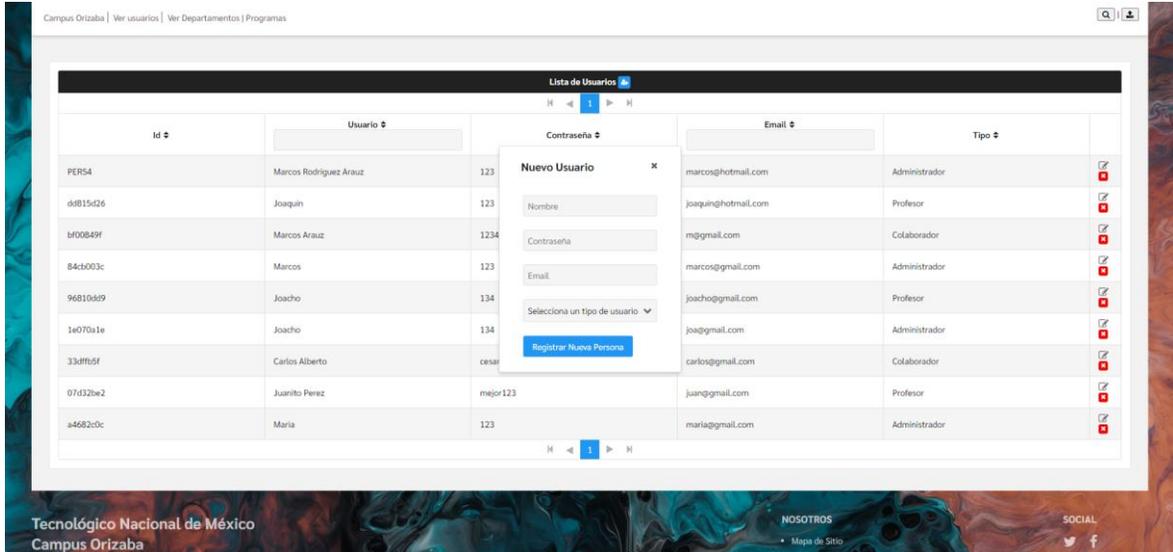


Figura 4.24 Página de gestión de usuarios 6

En la Figura 4.25 se presenta la página para la gestión de información del fragmento horizontal: *con\_buildings*, la cual posee una tabla de datos semejante a la de gestión de información de usuarios, con la diferencia de que en esta tabla se muestran imágenes previamente almacenadas en el fragmento correspondiente. La única diferencia con la página de gestión de información de usuarios, es que la opción para crear un nuevo contenido que se localiza a lado del título de la tabla de datos, ver Figura 4.26, muestra un formulario para la correspondiente captura del contenido, registrando también las nuevas imágenes a la base de datos, no sin antes analizarlas y extraer descriptores de las imágenes para su posterior uso en el módulo de búsquedas basadas en contenido.

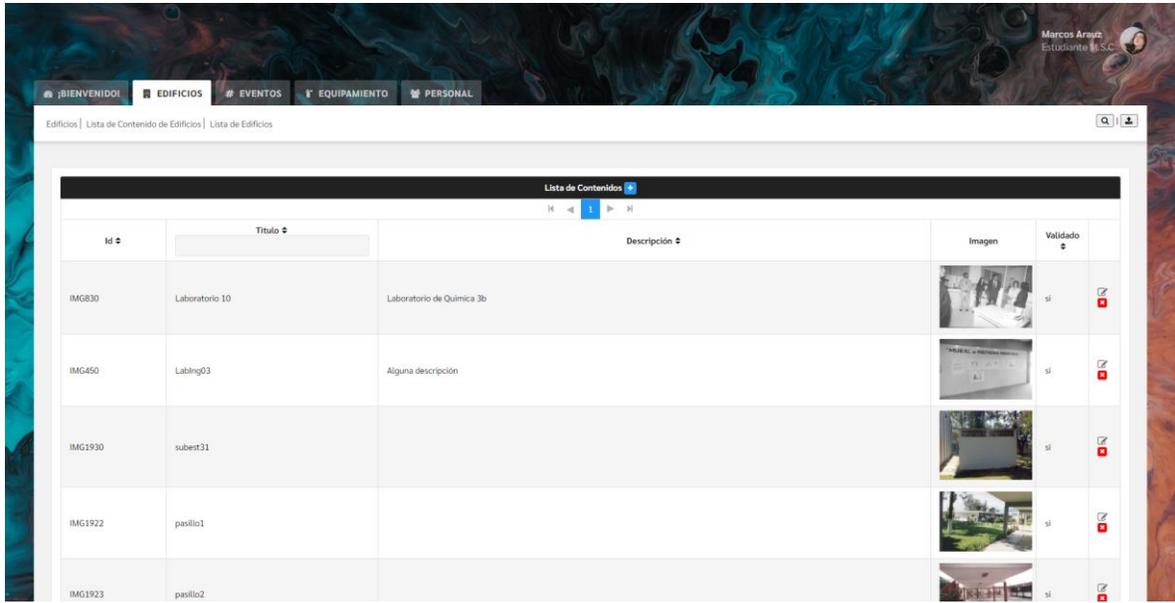


Figura 4.25 Página de gestión de edificios 1

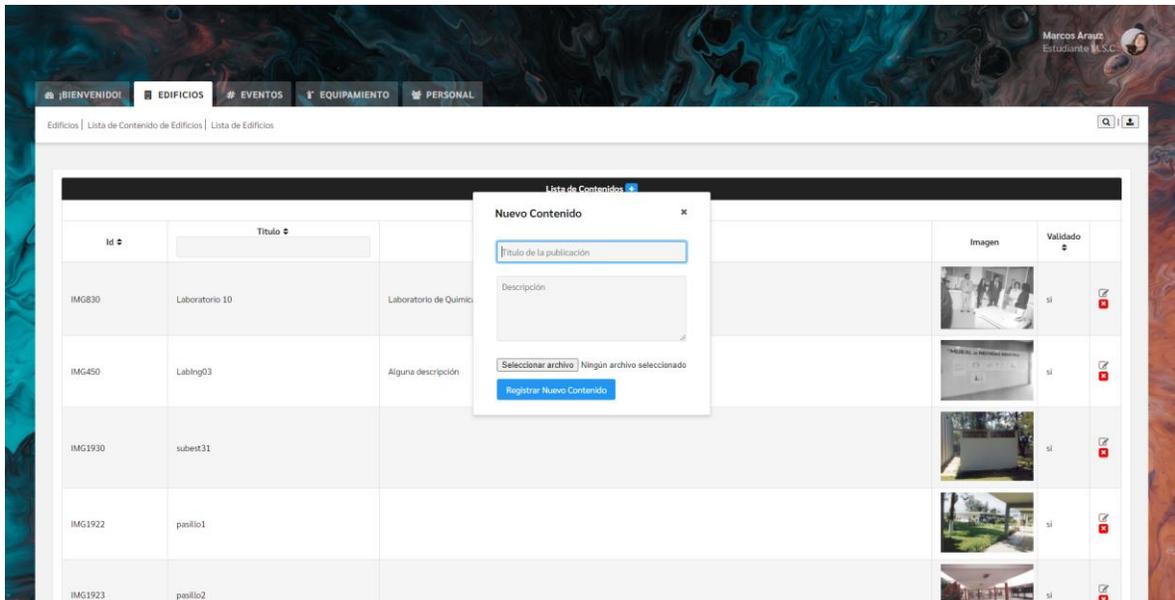


Figura 4.26 Página de gestión de edificios 2

En la Figura 4.27 se observa la página para la gestión de información del fragmento horizontal: *con\_events*.

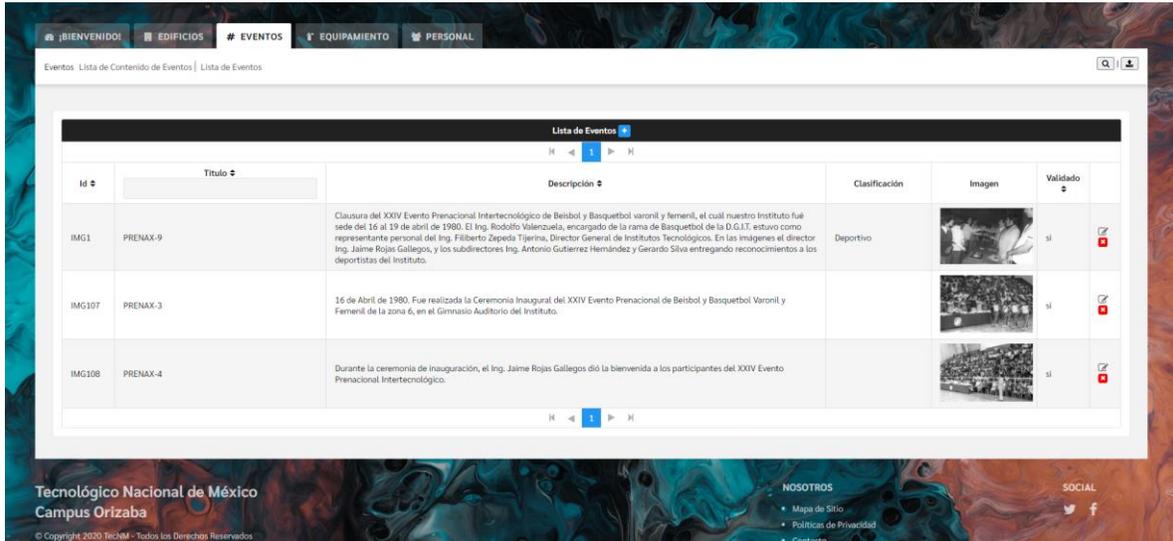


Figura 4.27 Página de gestión de eventos

En la Figura 4.28 se muestra la página para la gestión de información del fragmento horizontal: *con\_equipment*.

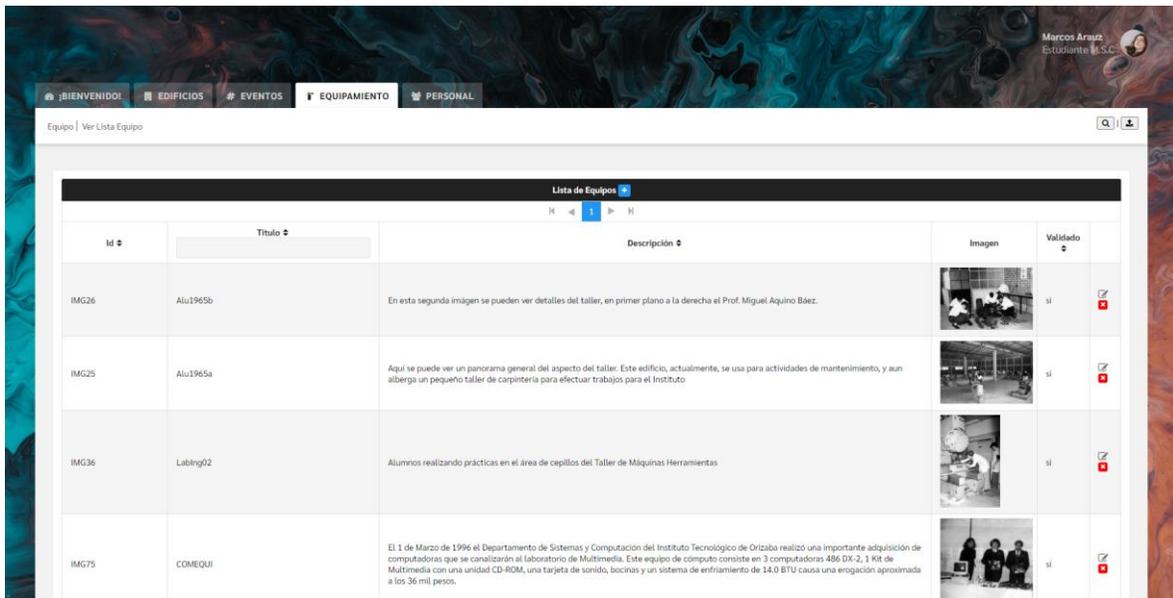


Figura 4.28 Página de gestión de equipo

En la Figura 4.29 se visualiza la página para la gestión de información del fragmento horizontal: *con\_personal*.

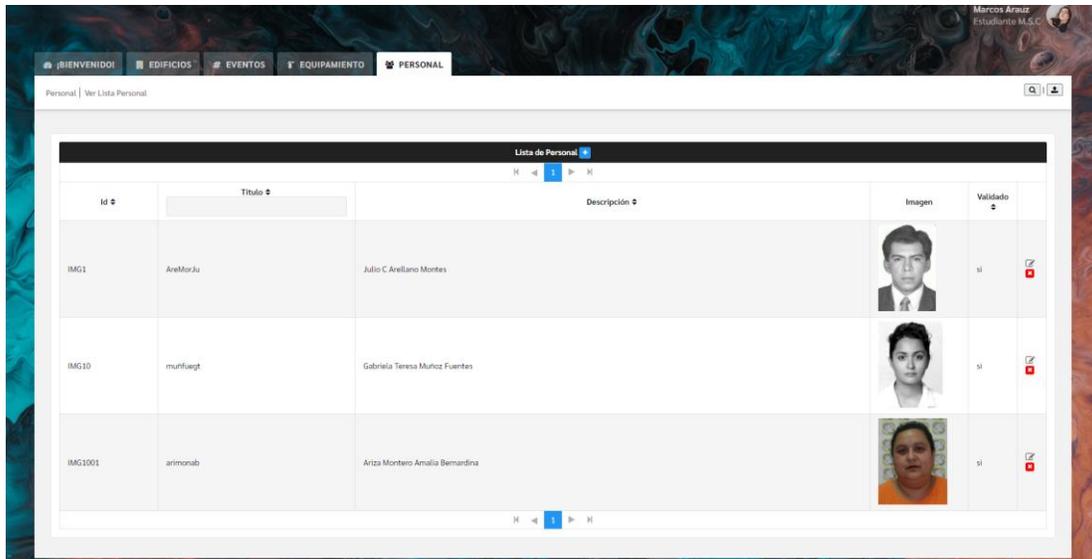


Figura 4.29 Página de gestión de personal

En la Figura 4.30 se observa la página para proponer contenido, función que cualquier tipo de usuario puede utilizar dentro del sistema. La página muestra el formulario requerido para registrar la propuesta de contenido, la cual será registrada dentro de la base de datos con el atributo “validado” con un valor negativo, ya que necesitará la previa aprobación de un colaborador o administrador para que este contenido sea mostrado en las páginas de contenido respectivas, también se extraerán los descriptores a las imágenes para su uso.

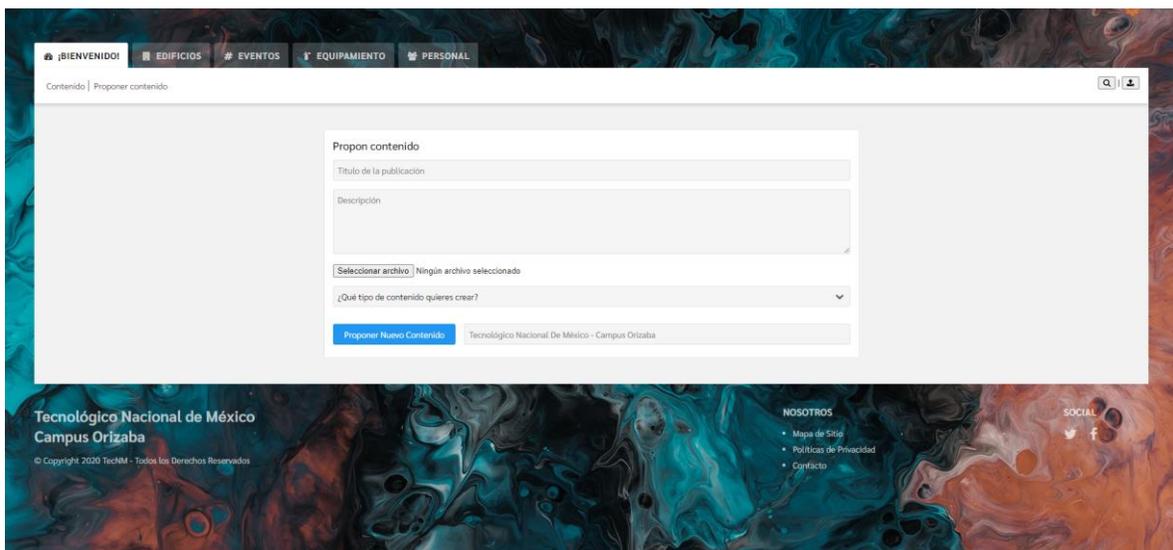


Figura 4.30 Página para proponer contenido

La Figura 4.31 presenta la página para realizar búsquedas basadas en contenido dentro del sistema. El formulario correspondiente muestra la opción para poder elegir el tipo de contenido que se quiere buscar: edificios, eventos, equipo o

personal. Una vez seleccionado el tipo de contenido a buscar se procede a cargar la imagen para la obtención de las imágenes más similares encontradas en el fragmento del tipo de contenido. Una vez obtenida la información necesaria el sistema procede a comparar los descriptores de las imágenes encontradas en el tipo de fragmento seleccionado y de la imagen de búsqueda para buscar la similitud en los puntos de interés de cada una de las imágenes. Una vez terminado el proceso anterior, se ordenan las imágenes más similares y se obtiene una lista con el resultado que se muestra en la página.

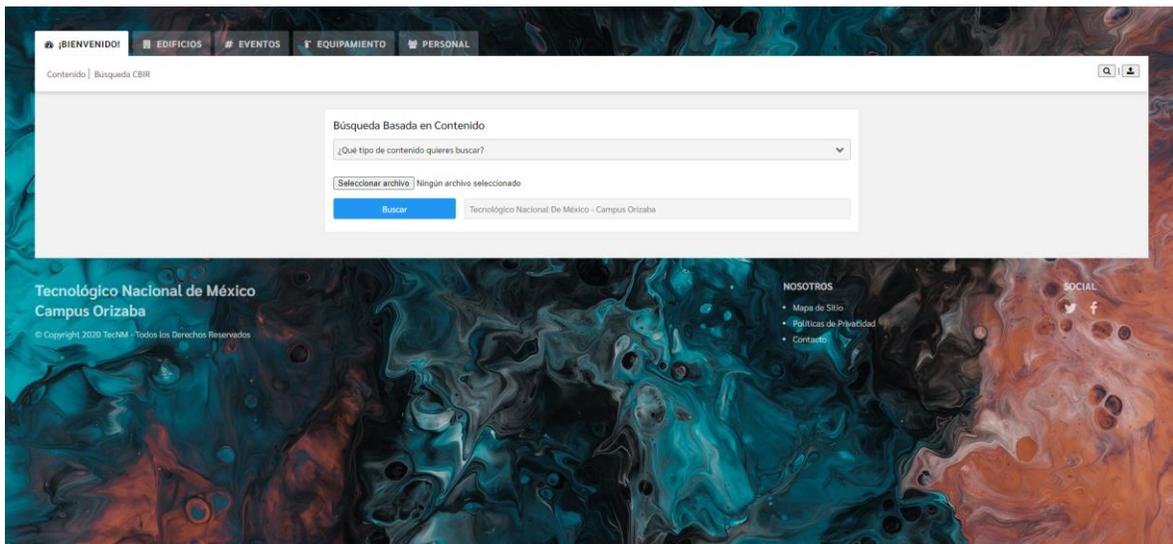


Figura 4.31 Página para búsquedas basadas en contenido

#### 4.4 Evaluación de resultados

En esta sección se evalúa la eficiencia del método de fragmentación horizontal usado en la base de datos HITO. El primer paso es el análisis de los datos de la base de datos multimedia del Instituto Tecnológico de Orizaba. En este primer paso, la colección de imágenes se fragmenta considerando el atributo tipo [23], que da como resultado los fragmentos horizontales denominados: edificios, eventos, personas y equipos. Estos cuatro fragmentos optimizan las consultas basadas en contenido, ya que minimizan su costo de ejecución al reducir las imágenes irrelevantes a las que acceden.

El segundo paso es evaluar las consultas y sus frecuencias utilizadas en la base de datos, ver Tabla 4.1, así como obtener el número de imágenes por fragmento, ver Tabla 4.2. En cuanto al tercer y cuarto paso, es necesaria la construcción de la Matriz de Uso de Predicados (PUM), ver Tabla 4.3, y la Matriz de Uso de Fragmentos-Predicados (FPUM), ver Tabla 4.4, para obtener el Costo de Acceso de Tuplas Irrelevantes (ITAC) descrito en [24]. El quinto paso es obtener una tabla con los resultados finales del ITAC, donde se muestra para cada fragmento, la comparación del costo de ejecución de las consultas. El sexto paso muestra el resultado de este flujo de trabajo, ver Figura 4.32, que es un esquema de fragmentación horizontal con cuatro fragmentos.

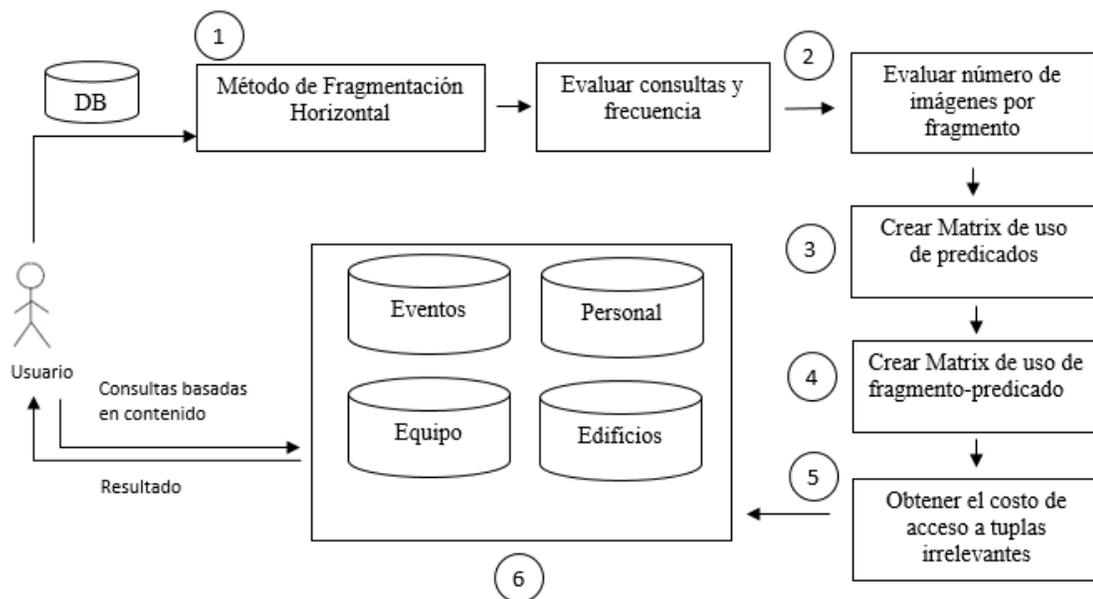


Figura 4.32 Flujo de trabajo del método de fragmentación horizontal y su evaluación.

Tabla 4.1 Predicados usados por consultas

| Q              | Pr                                   |
|----------------|--------------------------------------|
| q <sub>1</sub> | p <sub>1</sub> =5NN(event_w.png)     |
| q <sub>2</sub> | p <sub>2</sub> =5NN(equipment_x.png) |
| q <sub>3</sub> | p <sub>3</sub> =5NN(person_y.png)    |
| q <sub>4</sub> | p <sub>4</sub> =5NN(building_z.png)  |

Tabla 4.2 Fragmentos de la base de datos multimedia

| Fragmento | Número de imágenes |
|-----------|--------------------|
| Evento    | 756                |
| Equipo    | 63                 |
| Personal  | 886                |
| Edificios | 158                |

Tabla 4.3 Matrix de Uso de Predicados

| Q/Pr             | p <sub>1</sub> | p <sub>2</sub> | p <sub>3</sub> | p <sub>4</sub> | f <sub>i</sub> |
|------------------|----------------|----------------|----------------|----------------|----------------|
| q <sub>1</sub>   | 1              | 0              | 0              | 0              | 15             |
| q <sub>2</sub>   | 0              | 1              | 0              | 0              | 5              |
| q <sub>3</sub>   | 0              | 0              | 1              | 0              | 20             |
| q <sub>4</sub>   | 0              | 0              | 0              | 1              | 10             |
| sel <sub>j</sub> | 5              | 5              | 5              | 5              |                |

Tabla 4.4 Matrix de Uso de Fragmento-Predicado

| Fragmento       | p <sub>1</sub> | p <sub>2</sub> | p <sub>3</sub> | p <sub>4</sub> | card <sub>k</sub> |
|-----------------|----------------|----------------|----------------|----------------|-------------------|
| fr <sub>1</sub> | 1              | 0              | 0              | 0              | 756               |
| fr <sub>2</sub> | 0              | 1              | 0              | 0              | 63                |
| fr <sub>3</sub> | 0              | 0              | 1              | 0              | 886               |
| fr <sub>4</sub> | 0              | 0              | 0              | 1              | 158               |

Para obtener el costo de acceso a tuplas irrelevantes de cada consulta, se utilizan las siguientes ecuaciones, donde  $Pr_j$  tiene los predicados usados por una consulta  $q_j$  y ubicados en el fragmento  $fr_k$ , y  $n_p$  es el número de predicados en  $Pr_j$ . La Figura 4.33 muestra la comparación del costo de ejecución de las consultas de un base de datos que usa fragmentación horizontal contra la misma base de datos sin un método de fragmentación.

$$ITAC(q_j) = \begin{cases} f_j (card_k - \sum_{p_t \in Pr_j} sel_t) & \text{if } n_p \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$$Pr_j = \{p_t | PUM(q_j, p_t) = 1 \wedge FPUM(fr_k, p_t) = 1\} \quad (2)$$

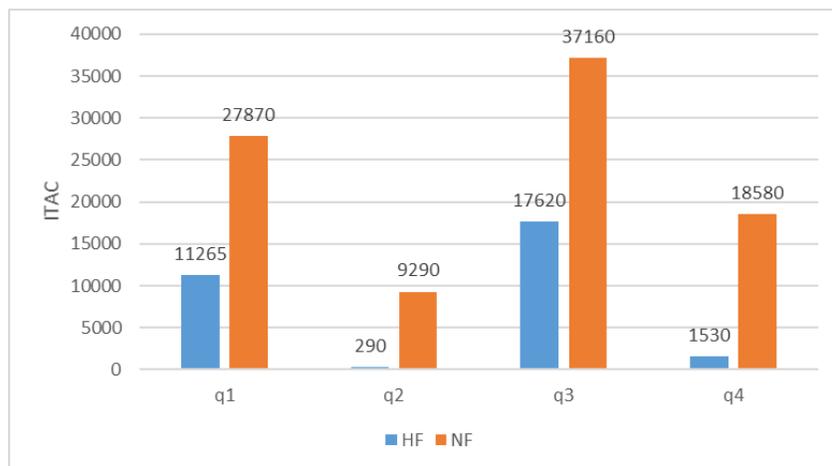


Figura 4.33 Costo de ejecución de consultas basadas en contenido

## Capítulo 5. Conclusiones y recomendaciones

### 5.1 Conclusiones

La fragmentación de datos es un objeto de estudio muy popular hoy en día y se utiliza constantemente en la industria. Un número grande de técnicas de fragmentación se utilizan actualmente, incluyendo fragmentación horizontal, que es la técnica que se propuso para ser utilizada en el desarrollo de este trabajo. Esta técnica ayuda no solo a mejorar la gestión de bases de datos multimedia, sino también para optimizar la eficiencia y la velocidad de acceso a la información que se necesita.

Este trabajo presentó un sistema para la gestión de datos multimedia, en el cual se utilizó MongoDB como gestor de base de datos por su velocidad y flexibilidad, y se diseñó el sistema con base en el marco de trabajo JavaServer Faces haciendo uso del patrón arquitectónico MVC dentro del ambiente de desarrollo integrado NetBeans, por último se utilizó BoofCV para la construcción del módulo de búsquedas basadas en contenido. El sistema utiliza fragmentación horizontal para mejorar el costo de ejecución de consultas basadas en contenido. Los usuarios del Instituto Tecnológico de Orizaba se beneficiarán enormemente, provocando un impacto social, que pretende crear conciencia sobre la historia del Instituto. El sistema también generará un impacto económico, ya que al utilizar un método de fragmentación, se observan ventajas, como costos de ejecución más bajos o tiempos de recuperación reducidos.

Dos métodos fueron seleccionados después del análisis de trabajos relacionados, ([23] y [24]) ya que cumplieron cinco criterios del análisis comparativo. Este trabajo combina el método de fragmentación para bases de datos multimedia de [23] con el modelo de costos propuesto en [24].

## **5.2 Recomendaciones**

A pesar de que este trabajo cumplió con el objetivo planteado, una de las limitaciones de esta arquitectura es que utiliza un tipo de fragmentación estática, como resultado, el esquema de fragmentación no cambiará cuando los patrones de acceso varíen, por lo tanto se planea utilizar un método de fragmentación dinámica que también implemente un método de fragmentación vertical para una optimización aún mejor de las consultas basadas en contenido. Además, tipos de consultas complejas conjuntivas se incluirán para la evaluación del esquema de fragmentación horizontal.

## Productos Académicos



M. J. Rodríguez-Arauz, L. Rodríguez-Mazahua, M. L. Arrijoja-Rodríguez, A. Abud-Figueroa, S. G. Peláez-Camarena “*Horizontal Fragmentation of Multimedia Databases to Optimize Content-based Queries: A Review*”, en 2020 17th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE) Mexico City, Mexico, 2020.

Estado: Publicado.



M. J. Rodríguez-Arauz, L. Rodríguez-Mazahua, M. L. Arrijoja-Rodríguez, A. Abud-Figueroa, S. G. Peláez-Camarena, and L. C. Martínez Méndez, “*Design of a Multimedia Data Management System that Uses Horizontal Fragmentation to Optimize Content-based Queries*”, International Conference on Advances in Information Mining and Management (IMMM) 2020, pp. 15-21.

Estado: Publicado.

## Referencias

- [1] A. Silberschatz, H. F. Korth, y S. Sudarshan, *Database system concepts*, 6th ed. New York: McGraw-Hill, 2011.
- [2] P. H. Oliveira *et al.*, «Employing Domain Indexes to Efficiently Query Medical Data from Multiple Repositories», *IEEE J. Biomed. Health Inform.*, vol. 23, no. 6, pp. 2220-2229, 2018.
- [3] M. T. Özsu y P. Valduriez, *Principles of distributed database systems*, 3rd ed. New York: Springer Science+Business Media, 2011.
- [4] M. L. Arrijoa Rodríguez, «Recopilación, Digitalización y Sistematización de Material Histórico del Instituto Tecnológico de Orizaba», Instituto Tecnológico de Orizaba, Departamento de Sistemas y Computación, Orizaba, Veracruz, México, Reporte de Proyecto de Investigación, 2000.
- [5] R. Rojas Ruiz, «Desarrollo de un sistema que permita realizar consultas basadas en contenido en una base de datos multimedia», Tesis de Maestría, Instituto Tecnológico de Orizaba, Orizaba, Veracruz, México, 2018.
- [6] L. Rodríguez-Mazahua, G. Alor-Hernández, X. Li, J. Cervantes, y A. López-Chau, «Active rule base development for dynamic vertical partitioning of multimedia databases», *J. Intell. Inf. Syst.*, vol. 48, n.º 2, pp. 421-451, abr. 2017.
- [7] R. R. Ruiz, L. Rodríguez-Mazahua, A. López-Chau, S. G. Peláez-Camarena, M. A. Abud-Figueroa, y I. Machorro-Cano, «A CBIR System for the Recognition of Agricultural Machinery», *J. Res. Comput. Sci.*, vol. 3, pp. 9-16, 2018.
- [8] L. Rodríguez-Mazahua, G. Alor-Hernández, J. Cervantes, A. López-Chau, y J. L. Sánchez-Cervantes, «Un método de fragmentación híbrida para bases de datos multimedia», *DYNA*, vol. 83, n.º 198, p. 59, sep. 2016.
- [9] Prinka y V. Wasson, «An efficient content based image retrieval based on speeded up robust features (SURF) with optimization technique», en *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, 2017, pp. 730-735.
- [10] A. A. Khodaskar y S. A. Ladhake, «A novel approach for content based image retrieval in context of combination S C techniques», en *2015 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2015, pp. 1-6.

- [11] M. Ouhda, K. El Asnaoui, M. Ouanan, y B. Aksasse, «A Content Based Image Retrieval Method Based on K-Means Clustering Technique»:., *J. Electron. Commer. Organ.*, vol. 16, n.º 1, pp. 82-96, ene. 2018.
- [12] M. Kumar, P. Chhabra, y N. K. Garg, «An efficient content based image retrieval system using BayesNet and K-NN», *Multimed. Tools Appl.*, vol. 77, n.º 16, pp. 21557-21570, ago. 2018.
- [13] M. Singh, S. Chopra, y J. Kaur, «An efficient content based image retrieval system using soft computing technique», *J. Comput. Sci. Mob. Comput.*, vol. 4, pp. 148-157, 2015.
- [14] N. A. Sakr, Ali. I. ELdesouky, y H. Arafat, «An efficient fast-response content-based image retrieval framework for big data», *Comput. Electr. Eng.*, vol. 54, pp. 522-538, ago. 2016.
- [15] A. Kamel, Y. B. Mahdy, y K. F. Hussain, «Multi-Bin search: improved large-scale content-based image retrieval», *Int. J. Multimed. Inf. Retr.*, vol. 4, n.º 3, pp. 205-216, sep. 2015.
- [16] A. K. Naveena y N. K. Narayanan, «Image retrieval using combination of color, texture and shape descriptor», en *2016 International Conference on Next Generation Intelligent Systems (ICNGIS)*, Kottayam, India, 2016, pp. 1-5.
- [17] S. Devi y A. Mathew, «Fast image retrieval using error diffusion block truncation coding and unsupervised clustering», en *2016 International Conference on Emerging Technological Trends (ICETT)*, Kollam, India, 2016, pp. 1-6.
- [18] D. Kulshreshtha, V. P. Singh, A. Shrivastava, A. Chaudhary, y R. Srivastava, «Content-based mammogram retrieval using k-means clustering and local binary pattern», en *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, Chengdu, China, 2017, pp. 634-638.
- [19] J.-H. Wang y H.-C. Chang, «CoBITS: a distributed indexing approach to collaborative content-based multimedia retrieval across digital archives», *Multimed. Tools Appl.*, vol. 74, n.º 8, pp. 2639-2658, abr. 2015.
- [20] J. Foley, P. Kwan, y M. Welch, «A Web-Based Infrastructure for the Assisted Annotation of Heritage Collections», *J. Comput. Cult. Herit.*, vol. 10, n.º 3, pp. 1-25, jul. 2017.

- [21] S. Saad, J. Tekli, R. Chbeir, and K. Yetongnon, "Towards Multimedia Fragmentation", in *Advances in Databases and Information Systems*, vol. 4152, Y. Manolopoulos, J. Pokorný, y T. K. Sellis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 415-429.
- [22] F. Getahun, J. Tekli, S. Atnafu, and R. Chbeir, "Towards Efficient Horizontal Multimedia Database Fragmentation using Semantic-based Predicates Implication", in *XXII Simpósio Brasileiro de Banco de Dados*, 15-19 de Outubro, João Pessoa, Paraíba, Brasil, Anais, 2007, pp. 68–82.
- [23] K. Fasolin, R. Fileto, M. Kruger, D. S. Kaster, M. R. P. Ferreira, R. L. F. Cordeiro, A. J. M. Traina, C Traina Jr., "Efficient Execution of Conjunctive Complex Queries on Big Multimedia Databases", in *2013 IEEE International Symposium on Multimedia*, Anaheim, CA, USA, 2013, pp. 536-543.
- [24] L. Rodríguez-Mazahua, G. Alor-Hernández, Ma. A. Abud-Figueroa, and S. G. Peláez-Camarena, "Horizontal Partitioning of Multimedia Databases Using Hierarchical Agglomerative Clustering", in *Nature-Inspired Computation and Machine Learning*, vol. 8857, A. Gelbukh, F. C. Espinoza, y S. N. Galicia-Haro, Eds. Cham: Springer International Publishing, 2014, pp. 296-309.
- [25] A. Habbal, F. B. Adamu, S. Hassan, R. L. Cottrell, B. White, M. Kaiiali, A. S. Wazan, "BIND: An indexing strategy for big data processing", in *TENCON 2017 - 2017 IEEE Region 10 Conference*, Penang, 2017, pp. 645-650.
- [26] J. M. Banda and R. A. Angryk, "Scalable solar image Retrieval with Lucene", in *2014 IEEE International Conference on Big Data (Big Data)*, Washington, DC, USA, 2014, pp. 11-17, doi: 10.1109/BigData.2014.7004399.
- [27] P. Mettes, J. C. van Gemert, S. Cappallo, T. Mensink, and C. G. M. Snoek, "Bag-of-Fragments: Selecting and Encoding Video Fragments for Event Detection and Recounting", in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval - ICMR '15*, Shanghai, China, 2015, pp. 427-434, doi: 10.1145/2671188.2749404.
- [28] H. Ma and M. Kirchberg, "Cost-Based Fragmentation for Distributed Complex Value Databases", in *Conceptual Modeling - ER 2007*, vol. 4801, C. Parent, K.-D. Schewe, V. C. Storey, y B. Thalheim, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 72-86.

- [29] F. Getahun, J. Tekli, S. Atnafu, and R. Chbeir, "The use of semantic-based predicates implication to improve horizontal multimedia database fragmentation", in Workshop on multimedia information retrieval on The many faces of multimedia semantics - MS '07, Augsburg, Bavaria, Germany, 2007, pp. 29-38, doi: 10.1145/1290067.1290073.
- [30] H. Ma, K.-D. Schewe, and Q. Wang, "A Heuristic Approach to Cost-Efficient Fragmentation and Allocation of Complex Value Databases", in Seventeenth Australasian Database Conference, Hobart, Australia, 2006, pp. 183-192.
- [31] S. I. Khan, "Efficient Partitioning of Large Databases without Query Statistics", *Database Systems Journal*, vol. VII, 2016. 34-53.
- [32] H. Ma, K. Schewe, and Q. Wang, "A Heuristic Approach to Cost-Efficient Derived Horizontal Fragmentation of Complex Value Databases", in Eighteenth Australasian Database Conference, 2007, vol. 63, pp. 103-111.
- [33] R. Chbeir and D. Laurent, "Towards a novel approach to multimedia data mixed fragmentation", in Proceedings of the International Conference on Management of Emergent Digital EcoSystems - MEDES '09, France, 2009, pp. 200-204.
- [34] C. Concolato, J. Le Feuvre, and J.-C. Moissinac, "Timed-fragmentation of SVG documents to control the playback memory usage", in Proceedings of the 2007 ACM symposium on Document engineering - DocEng '07, Winnipeg, Manitoba, Canada, 2007, pp. 121-124, doi: 10.1145/1284420.1284453.
- [35] R. Chbeir and D. Laurent, "Enhancing Multimedia Data Fragmentation", *Journal of Multimedia Processing and Technologies*, vol. 1, no. 2, pp. 112-131, 2010.
- [36] Xiaoping Hu, U. Y. Ogras, N. H. Zamora, and R. Marculescu, "Data partitioning techniques for pervasive multimedia platforms", in 2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No.04TH8763), Taipei, Taiwan, 2004, pp. 1035-1038.
- [37] S. Antaris and D. Rafailidis, "Similarity Search over the Cloud Based on Image Descriptors' Dimensions Value Cardinalities", *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 11, no. 4, pp. 1-23, Jun. 2015, doi: 10.1145/2716315.
- [38] M. Torjmen, K. Pinel-Sauvagnat, and M. Boughanem, "Towards a structure-based multimedia retrieval model", in Proceeding of the 1st ACM international conference on Multimedia information retrieval - MIR '08, Vancouver, British Columbia, Canada, 2008, pp. 350-357.

- [39] S. Mezzoudj, A. Behloul, R. Seghir, and Y. Saadna, "A parallel content-based image retrieval system using spark and tachyon frameworks", *Journal of King Saud University - Computer and Information Sciences*, Jan. 2019.
- [40] M. Torjmen-Khemakhem, K. Pinel-Sauvagnat, y M. Boughanem, "Investigating the document structure as a source of evidence for multimedia fragment retrieval", *Information Processing & Management*, vol. 49, no. 6, pp. 1281-1300, Nov. 2013, doi: 10.1016/j.ipm.2013.06.001.
- [41] C. Gu and Y. Gao, "A Content-Based Image Retrieval System Based on Hadoop and Lucene", in 2012 Second International Conference on Cloud and Green Computing, Xiangtan, Hunan, China, 2012, pp. 684-687, doi: 10.1109/CGC.2012.33.
- [42] M. Wlodarczyk-Sielicka and A. Stateczny, "Fragmentation of Hydrographic Big Data Into Subsets During Reduction Process", in 2017 Baltic Geodetic Congress (BGC Geomatics), Gdansk, Poland, 2017, pp. 193-198.
- [43] M. Kumar, P. Chhabra, and N. K. Garg, "An efficient content based image retrieval system using BayesNet and K-NN", *Multimed Tools Appl*, vol. 77, no. 16, pp. 21557-21570, Aug. 2018, doi: 10.1007/s11042-017-5587-8.
- [44] Moghimian, M. Mansoorizadeh, y M. Dezfoulian, "Content based image retrieval using fusion of multilevel bag of visual words", *SN Appl. Sci.*, vol. 1, no. 12, p. 1735, Dec. 2019.
- [45] D. Srivastava, R. Bakthula, y S. Agarwal, "Image classification using SURF and bag of LBP features constructed by clustering with fixed centers", *Multimed Tools Appl*, vol. 78, no. 11, pp. 14129-14153, Jun. 2019, doi: 10.1007/s11042-018-6793-8.
- [46] W.-S. Li, K. S. Candan, K. Hirata, y. Hara, "Supporting efficient multimedia database exploration", *The VLDB Journal*, vol. 9, no. 4, pp. 312-326, Apr. 2001
- [47] H. Al-Jubouri y H. Du, "A Content-Based Image Retrieval Method by Exploiting Cluster Shapes", *Iraqi Journal for Electrical and Electronic Engineering*, vol. 14, no. 2-2018, pp. 90-102, 2018.
- [48] S. A. Mahmoudi, M. A. Belarbi, E. W. Dadi, S. Mahmoudi, y M. Benjelloun, "Cloud-Based Image Retrieval Using GPU Platforms", *Computers*, vol. 8, no. 2, Jun. 2019, doi: 10.3390/computers8020048.
- [49] J. Rejito, A. S. Abdullahi, Akmal, D. Setiana, y B. N. Ruchjana, "Image indexing using color histogram and k-means clustering for optim

ization CBIR in image database”, The Asian Mathematical Conference, Oct. 2017, doi: 10.1088/1742-6596/893/1/012055.

[50] Z. Kouahla, A. Anjum, y H. Seridi, “Indexing through separable partitioning for complex data sharing in P2P system”, *Journal of Intelligent & Fuzzy Systems*, vol. 35, no. 6, pp. 6469-6478, Dec. 2018, doi: 10.3233/JIFS-18398.

[51] T.-D. Le, V. Kantere, y L. Orazio, “Optimizing DICOM data management with NSGA-G”, in International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data, 2019.

[52] M. Yamamoto, “Parallel Image Database Processing with Mapreduce and Performance Evaluation in Pseudo Distributed Mode”, *International Journal of Electronic Commerce Studies*, vol. 3, no. 2, pp. 211-228, Dec. 2012, doi: 10.

## Anexo

A continuación se muestra un análisis más detallado de la Tabla 3.1 Análisis de Artículos.

Tabla 1 Análisis detallado de artículos.

| Artículo                      | FH   | Datos multimedia   | Consultas basadas en contenido | Modelo de costo | Complejidad   | Facilidad de implementación | SGBD                                  |
|-------------------------------|--|--|--------------------------------|-----------------|---|-----------------------------|---------------------------------------|
| Saad et al. [21]              | Sí   | Predicados Multimedia. Video, audio e Imagen.  | Consultas Rango y KNN          | No              | No se muestra la arquitectura del prototipo, solo los resultados de las simulaciones.   | 7/10                        | Oracle 9i                             |
| Getahun et al. [22]           | Sí   | Predicados Multimedia basados en texto. Predicados basados en semántica.                         | Consultas de rango y KNN       | No              | En teoría mantiene un desarrollo entendible, sin embargo, es teórico y no hay simulaciones.   | 7/10                        | Oracle 9i                             |
| Fasolin et al. [23]           | Sí (Criterio de fragmentación basado en la frecuencia de las consultas más usadas) | Imágenes   | Sí                             | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto.  | 6/10                        | Oracle DB11g                          |
| Rodríguez-Mazahua et al. [24] | Sí (Basado en un algoritmo jerárquico de clustering)                               | Imágenes, graficas, audio y video (base de datos de equipo en un empresa de venta de maquinaria) | No                             | Si              | Plantea correctamente la arquitectura para el desarrollo del proyecto.  | 7/10                        | No                                    |
| Habbal et al. [25]            | No   | No (Big Data) (PingER data)  | No                             | No              | Se habla superficialmente del desarrollo del algoritmo usado, sin mencionar como se aplicó Java como lenguaje de programación para hacer las pruebas. | 7/10                        | HDFS (Hadoop Distributed File System) |
| Kulshreshtha et al. [18]      | No   | Imágenes (MIAS database benchmark)   | Si. (LBP y K-Means Clustering) | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto.  | 6/10                        | ROI DB (Region of Interest)           |

Tabla 1 Análisis detallado de artículos.

| Artículo                  | FH                            | Datos multimedia   | Consultas basadas en contenido  | Modelo de costo  | Complejidad  | Facilidad de implementación | SGBD                                       |
|---------------------------|-------------------------------|--|---|--|--|-----------------------------|--|
| Devi y Mathew [17]        | No                            | Imágenes (Corel)   | Si (usando técnica de compresión y un enfoque de <i>unsupervised clustering</i> ) | No   | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 7/10                        | No   |
| Banda y Angryk, 2014 [26] | No                            | Imágenes (Imágenes Solares) (Atmospheric Imagin Assembly, AIA)   | Sí (basado en Lucene)   | No   | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 6/10                        | PostgreSQL 9.2 & Apache Lucene Version 4.0 |
| Mettes et al. [27]        | No                            | Video (THUMOS 14 temporal localization dataset)  | No  | No   | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 9/10                        | No   |
| Ma y Kirchberg [28]       | Sí (y fragmentación vertical) | No (orientado a objetos, Objetos Relacionales y XML ( <i>eXtensible Markup Language</i> , Lenguaje de marcado extensible)) | No  | Si (cuanto menor sea el valor del costo de asignar un atributo a un sitio, menores serán los costos totales para acceder a él) | Mantiene una buena complejidad debido a su modelo de costo.            | 8/10                        | No   |

Tabla 1 Análisis detallado de artículos.

| Artículo            | FH                                     | Datos multimedia  | Consultas basadas en contenido | Modelo de costo   | Compleitud   | Facilidad de implementación | SGBD       |
|---------------------|--|---|--------------------------------|---|--|-----------------------------|------------|
| Getahun et al. [29] | Sí                                     | Predicados Multimedia basados en texto.   | Consultas rango y KNN          | No  | No muestra la arquitectura del prototipo, ni cómo fue construido, ni sus requerimientos. | 7/10                        | Oracle 9i  |
| Hui Ma et al. [30]  | Sí                                     | No (Orientado a objetos, Objetos Relacionales y XML)  | No                             | Sí (enfoque heurístico para minimizar los costos de consulta para los casos de fragmentación horizontal). | El modelo de costo lo vuelve un trabajo mejor desarrollado.                              | 8/10                        | No.        |
| Khan et al. [31]    | Sí (fragmentación Basada en la Matrix) | No (Base de datos de banca distribuida en el laboratorio de pos-graduados de BUET: Bangladesh University of Engineering and Technology) | No                             | No  | Plantea correctamente la arquitectura para el desarrollo del proyecto.                   | 6/10                        | Oracle 10g |
| Ma et al. [32]      | Sí (derivada)                          | No (Datos complejos)  | No                             | Sí  | Plantea correctamente la arquitectura para el desarrollo del proyecto.                   | 8/10                        | No         |

Tabla 1 Análisis detallado de artículos.

| Artículo                | FH                            | Datos multimedia   | Consultas basadas en contenido              | Modelo de costo | Complejidad  | Facilidad de implementación | SGBD |
|-------------------------|-------------------------------|--|---|-----------------|--|-----------------------------|------|
| Chbeir y Laurent [33]   | Híbrido (Fragmentación Mixta) | Implicación de predicado Audio, Video, Imagen.   | No  | No              | No muestran una implementación del método propuesto.                 | 8/10                        | No   |
| Concolato et al. [34]   | No (Timed Fragmentation)      | Scalable Vector Graphics (SVG). Audio, video, vector graphics, interactivity, animation. | No  | No              | Buen desarrollo sobre los métodos para complementar herramientas SVG | 8/10                        | No   |
| Chbeir y Laurent [35]   | No (Fragmentación Mixta)      | Selección de predicados atómicos. Vinculación de predicados.                             | No  | No              | Se desarrolla adecuadamente el método propuesto.                     | 7/10                        | No   |
| Xiapoing Hu et al. [36] | No (Fragmentación Mixta)      | Video  | No (Content-Bases power management schemes) | No              | Mantiene un buen desarrollo sobre cómo lograr el objetivo propuesto. | 7/10                        | No   |

Tabla 1 Análisis detallado de artículos.

| Artículo             | FH | Datos multimedia   | Consultas basadas en contenido                   | Modelo de costo | Complejidad  | Facilidad de implementación | SGBD                       |
|----------------------|----|--|--|-----------------|--|-----------------------------|----------------------------|
| Antaris et al. [37]  | No | Conjunto de datos de vectores de descriptores de imágenes. (The ImageClef Wikipedia Retrieval 2010 Collection) | Sí   | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 9/10                        | No (simulación en la nube) |
| Torjmen et al. [38]  | No | No (Documentos XML ) (INEX framework, INEX Multimedia Fragment task 2006)                                      | No (Recuperación multimedia contextual)          | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 7/10                        | No                         |
| Mezzoudj et al. [39] | No | Imágenes de gran escala. (ImageCLEFphoto, ImageNet)  | Sí (k-NN)  | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 8/10                        | HDFS & Taychon framework   |
| Sakr et al. [14]     | No | Imágenes en Big Data con gran dimensión (NUS-WIDE, Holiday dataset)  | Sí (Marco de trabajo basado en Hadoop MapReduce) | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 8/10                        | Hadoop Yarn 2.3.0          |

Tabla 1 Análisis detallado de artículos.

| Artículo                             | FH   | Datos multimedia   | Consultas basadas en contenido              | Modelo de costo | Complejidad   | Facilidad de implementación | SGBD          |
|--------------------------------------|--|--|---|-----------------|---|-----------------------------|---------------|
| Torjmen-Khemakhem et al. [40]        | No   | Imágenes (Objetos multimedia descritos en documentos estructurados como XML) (Wikipedia XML, INEX) | No  | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto.              | 7/10                        | No            |
| Gu y Gao [41]                        | No   | Imágenes   | Sí (Sistema CBIR basado en Hadoop y Lucene) | No              | La arquitectura no es óptima y aún tiene problemas para el desarrollo del proyecto. | 7/10                        | HDFS, HBase   |
| Włodarczyk-Sielicka y Stateczny [42] | No. (Fragmentación de Datos Hidrográficos) | No (Big Data batimétrica)  | No  | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto.              | 8/10                        | GRID Database |
| Kumar et al. [43]                    | No   | Imágenes (Wang Dataset)  | Sí (Sistema CBIR basado en SIFT y ORB)      | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto.              | 7/10                        | Wang Database |

Tabla 1 Análisis detallado de artículos.

| Artículo               | FH | Datos multimedia  | Consultas basadas en contenido  | Modelo de costo | Complejidad  | Facilidad de implementación | SGBD          |
|------------------------|----|---|---|-----------------|--|-----------------------------|---------------|
| Moghimian et al. [44]  | No | Imágenes (Wang Dataset, Corel9C Dataset, Corel5K Dataset)                     | Sí (Enfoque de características de fusión)   | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 7/10                        | Wang Database |
| Srivastava et al. [45] | No | Imágenes (Datasets: Caltech101, ORL Face, Bangla Signature y Hindi Signature) | Si (A través de ROI usando SURF, Extracción de LBP Y agrupamiento de características LBP) | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 8/10                        | No            |
| Li et al. [46]         | No | Imágenes (SEMCOG Multimedia Database)   | No  | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 7/10                        | No            |
| Al-Jubouri y Du [47]   | No | Imágenes. (Dataset: Wang y Caltech6)  | Sí (Basado en agrupamiento de formas)   | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 8/10                        | No            |

Tabla 1 Análisis detallado de artículos.

| Artículo             | FH | Datos multimedia   | Consultas basadas en contenido       | Modelo de costo | Complejidad  | Facilidad de implementación | SGBD           |
|----------------------|----|--|--------------------------------------|-----------------|--|-----------------------------|----------------|
| Mahmoudi et al. [48] | No | Imágenes (Dataset: Wan, Corel-10k, GHIM-10K)                             | Sí (uso de combinado de SIFT y SURF) | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 6/10                        | No             |
| Rejito et al. [49]   | No | Imágenes. (Base de datos Wang)   | Sí (usando un histograma de color)   | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto  | 7/10                        | Delphi Program |
| Kouahla et al. [50]  | No | Imágenes. (Ciudades de Francia, Objetos complejos de un subset MPEG-7)   | No                                   | No              | La arquitectura para el desarrollo del proyecto no ha sido corregida.  | 8/10                        | No             |
| Le et al. [51]       | No | Imágenes ( <i>Digital Imaging and Communication in Medicine</i> , DICOM) | No                                   | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 8/10                        | No. (NSGA-G)   |
| Yamamoto [52]        | No | Video.   | No                                   | No              | Plantea correctamente la arquitectura para el desarrollo del proyecto. | 8/10                        | HDFS           |