



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Instituto Tecnológico de Orizaba

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

OPCIÓN I.- TESIS

TRABAJO PROFESIONAL

“Recuperación de videos con
base en contenido de la base
de datos multimedia HITO”

**QUE PARA OBTENER EL GRADO DE:
MAESTRO EN SISTEMAS
COMPUTACIONALES**

PRESENTA:

I.S.C. Gustavo Steve Luis Peña Riquer

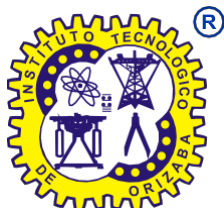
M10011279

DIRECTOR DE TESIS:

Dra. Lisbeth Rodríguez Mazahua

CODIRECTOR DE TESIS:

Dr. Mario Leoncio Arrioja Rodríguez





EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Instituto Tecnológico de Orizaba
División de Estudios de Posgrado e Investigación

Orizaba, Veracruz, **18/octubre/2023**
Dependencia: **División de Estudios de**
Posgrado e Investigación
Asunto: **Autorización de Impresión**
OPCION: I

C. GUSTAVO STEVE LUIS PEÑA RIQUEL
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:

" Recuperación de videos con base en contenido de la Base de Datos Multimedia HITO"

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

ATENTAMENTE

Excelencia en Educación Tecnológica®
CIENCIA - TÉCNICA - CULTURA®

Cuahtémoc Sánchez R.

DR. CUAHTÉMOC SÁNCHEZ RAMÍREZ
JEFE DE LA DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN



OG-13-F06



Av. Oriente 9 Núm.852, Colonia Emiliano Zapata. C.P. 94320 Orizaba, Veracruz.
Tel. 01 (272)1105360 e-mail: dir_orizaba@tecnm.mx tecnm.mx | orizaba.tecnm.mx



2023
Francisco
VILLA
AÑO DEL CENTENARIO DE SU NACIMIENTO

Orizaba, Veracruz, **11/septiembre/2023**
Asunto: **Revisión de trabajo escrito**

C. CUAUHTÉMOC SÁNCHEZ RAMÍREZ
JEFE DE LA DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN
P R E S E N T E.-

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

GUSTAVO STEVE LUIS PEÑA RIQUER

La cual lleva el título de:

Recuperación de videos con base en contenido de la Base de Datos Multimedia HITO

Y concluyen que se acepta.

ATENTAMENTE
Excelencia en Educación Tecnológica®
CIENCIA - TÉCNICA - CULTURA®

**PRESIDENTE: DRA. LISBETH RODRÍGUEZ
MAZAHUA**


FIRMA

**SECRETARIO: DR. MARIO LEONCIO ARRIJOJA
RODRÍGUEZ**


FIRMA

VOCAL: DR. ULISES JUÁREZ MARTÍNEZ


FIRMA

**VOCAL SUP.: M.C. MARÍA ANTONIETA ABUD
FIGUEROA**


FIRMA

TA-09-18



CARTA DE ORIGINALIDAD

En la ciudad de Orizaba, Veracruz, el día 25 del mes de Octubre del año 2023,
el (la) que suscribe ISC. Gustavo Steve Luis Peña Riquer,

alumno (a) del programa de Maestría en Sistemas Computacionales con número de control
M10011279, manifiesta que es autor (a) del trabajo de tesis titulado
“Recuperación de videos con base en contenido de la Base de Datos Multimedia HITO” y declara que
el trabajo es original, ya que su contenido es producto de su directa contribución intelectual.
Todos los datos y las referencias a materiales ya publicados están debidamente identificados
con su respectivo crédito e incluidos en las notas bibliográficas y en las citas que se destacan
como tal y, en los casos que así lo requieran, se tienen las debidas autorizaciones de quienes
poseen los derechos patrimoniales. Por lo tanto, se hace responsable de cualquier litigio o
reclamación relacionada con derechos de propiedad intelectual, exonerando de toda
responsabilidad al Tecnológico Nacional de México / Instituto Tecnológico de Orizaba.



Gustavo Steve Luis Peña Riquer
Nombre y firma autógrafa

CARTA DE CESIÓN DE DERECHOS

En la ciudad de Orizaba, Ver el día 25 del mes de Octubre
del año 2023, el (la) que suscribe ISC. Gustavo Steve Luis Peña Riquer

alumno (a) del programa de Maestría en Sistemas Computacionales con número de control
M10011279, manifiesta que es autor (a) intelectual del trabajo de tesis bajo la dirección de
Dra. Lisbeth Rodríguez Mazahua y ceden los derechos del trabajo
intitulado "Recuperación de videos con base en contenido de la Base de Datos Multimedia HITO"
al TecNM/Instituto Tecnológico de Orizaba para su difusión, con fines académicos y de
investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del
trabajo sin el permiso expreso del autor y del director del trabajo. Este puede ser obtenido
escribiendo a la siguiente dirección: msc@orizaba.tecnm.mx. Si el permiso se otorga, el
usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.


Gustavo Steve Luis Peña Riquer

Nombre y firma

Agradecimientos

En primer lugar, quiero expresar mi agradecimiento a mis padres, cuyo recuerdo y promesa han sido la fuente inagotable de mi motivación durante todo este viaje. Antes de que partieran de este mundo, les prometí que no me daría por vencido, que perseguiría mis sueños y metas con pasión y perseverancia. Su amor y confianza en mí me impulsaron a seguir adelante, incluso en los momentos más desafiantes de esta travesía académica. A pesar de que ya no están físicamente conmigo, su legado vive en cada página de este trabajo.

Agradezco a mi hermana por su apoyo constante, su paciencia y su aliento. Mi sobrina, con su inocencia y sonrisa, ha sido una fuente constante de inspiración, recordándome la importancia de dejar un legado positivo a las generaciones futuras.

Mi agradecimiento al Tecnológico Nacional de México por brindarme una educación de calidad y las herramientas necesarias para alcanzar mis metas académicas. La excelencia académica de esta institución y el compromiso de sus profesores han sido un faro que ha guiado mi aprendizaje y crecimiento.

Finalmente, agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por su generoso apoyo financiero, que me permitió concentrarme en mis estudios y dedicar el tiempo y los recursos necesarios para la investigación y la redacción de esta tesis.

Gracias a mi familia, amigos y todos los que creyeron en mí, por ser la brújula de mi travesía académica y por ayudarme a cumplir la promesa que hice a mis padres. Estoy eternamente agradecido y comprometido a honrar su apoyo y confianza en mí, trabajando para contribuir al avance del conocimiento y la sociedad.

Con sincero agradecimiento y aprecio.

Índice

ÍNDICE DE TABLAS	VIII
ÍNDICE DE CÓDIGOS.....	XII
RESUMEN.....	XIII
ABSTRACT	XIV
INTRODUCCIÓN	XV
CAPÍTULO 1. ANTECEDENTES.....	1
1.1 MARCO TEÓRICO.....	1
1.1.1 BASES DE DATOS MULTIMEDIA	1
1.1.2 RECUPERACIÓN DE VIDEO CON BASE EN CONTENIDO	1
1.1.2.1 CONSULTAS EN CBVR.....	2
1.1.2.2 DESCRIPTORES	3
1.1.2.3 REDES NEURONALES CONVOLUCIONALES.....	4
1.1.3 BASE DE DATOS HITO (HISTORIA DEL INSTITUTO TECNOLÓGICO DE ORIZABA)	5
1.2 SITUACIÓN TECNOLÓGICA, ECONÓMICA Y OPERATIVA DE LA EMPRESA.....	5
1.3 PLANTEAMIENTO DEL PROBLEMA	5
1.4 OBJETIVOS GENERALES Y ESPECÍFICOS	6
1.4.1 OBJETIVO GENERAL.....	6
1.4.2 OBJETIVOS ESPECÍFICOS.....	6
1.5 JUSTIFICACIÓN	7
CAPÍTULO 2. ESTADO DE LA PRÁCTICA.....	8
2.1 TRABAJOS RELACIONADOS.....	8
2.2 ANÁLISIS COMPARATIVO	14
2.3 PROPUESTA DE LA SOLUCIÓN	24
CAPÍTULO 3. APLICACIÓN DE LA METODOLOGÍA	25
3.1 ANÁLISIS.....	25
3.2 SELECCIÓN	43
3.3 DISEÑO DEL SISTEMA	43
3.3.1 ANÁLISIS DE REQUISITOS	44
3.3.2 MODELO CONCEPTUAL	55

3.3.3	MODELO DE NAVEGACIÓN	59
3.3.4	MODELO DE PRESENTACIÓN	65
3.3.5	MODELO DE PROCESOS	69
3.4	PROCESO CBVR	79
3.4.1	FLUJO DE TRABAJO.....	79
3.4.2	PROPUESTA DE VIDEO	79
3.4.3	BASE DE DATOS.....	83
3.4.4	ALGORITMO	84
CAPÍTULO 4. RESULTADOS		86
4.1	RESULTADOS DEL ANÁLISIS	86
4.2	SISTEMA DE RECUPERACIÓN DE VIDEOS CON BASE EN CONTENIDO DE LA BASE DE DATOS MULTIMEDIA HITO	90
CAPÍTULO 5. CONCLUSIONES Y RECOMENDACIONES.....		101
5.1	CONCLUSIONES.....	101
5.2	RECOMENDACIONES	102
PRODUCTOS ACADÉMICOS		103
REFERENCIAS.....		104
ANEXO		108

Índice de tablas

TABLA 2.1 ANÁLISIS COMPARATIVO DE ARTÍCULOS ASOCIADO	15
TABLA 3.1 COMPARACIÓN DE TRABAJOS.....	27
TABLA 3.2 ANÁLISIS DE ARTÍCULOS	28
TABLA 3.3 ACTORES DE LA APLICACIÓN	44

Índice de figuras

FIGURA 3.1 METODOLOGÍA DE BÚSQUEDA Y ANÁLISIS DE ARTÍCULOS RELACIONADOS.....	26
FIGURA 3.2 ARQUITECTURA DE LA APLICACIÓN WEB	44
FIGURA 3.3 DIAGRAMA DE CASOS DE USO.....	46
FIGURA 3.4 DIAGRAMA DEL CASO DE USO INICIAR SESIÓN.....	46
FIGURA 3.5 DIAGRAMA DEL CASO DE USO CERRAR SESIÓN	47
FIGURA 3.6 DIAGRAMA DEL CASO DE USO REGISTRAR USUARIO	47
FIGURA 3.7 DIAGRAMA DEL CASO DE USO VALIDAR INFORMACIÓN	48
FIGURA 3.8 DIAGRAMA DEL CASO DE USO PROPONER CONTENIDO.....	49
FIGURA 3.9 DIAGRAMA DEL CASO DE USO ALTA DE CONTENIDO	49
FIGURA 3.10 DIAGRAMA DEL CASO DE USO MODIFICAR CONTENIDO	50
FIGURA 3.11 DIAGRAMA DEL CASO DE USO BAJA DE CONTENIDO	51
FIGURA 3.12 DIAGRAMA DEL CASO DE USO ALTA DE USUARIO	51
FIGURA 3.13 DIAGRAMA DEL CASO DE USO BAJA DE USUARIO.....	52
FIGURA 3.14 DIAGRAMA DEL CASO DE USO MODIFICAR USUARIO	53
FIGURA 3.15 DIAGRAMA DEL CASO DE USO CONSULTAR USUARIO.....	53
FIGURA 3.16 DIAGRAMA DEL CASO DE USO RECUPERACIÓN CON BASE EN TEXTO	54
FIGURA 3.17 DIAGRAMA DEL CASO DE USO RECUPERACIÓN CON BASE EN CONTENIDO	54
FIGURA 3.18 DIAGRAMA CONCEPTUAL.....	55
FIGURA 3.19 DIAGRAMA LÓGICO	56
FIGURA 3.20 DIAGRAMA FÍSICO	57
FIGURA 3.21 DIAGRAMA DE NAVEGACIÓN DEL ADMINISTRADOR	61
FIGURA 3.22 DIAGRAMA DE NAVEGACIÓN DEL COLABORADOR	62
FIGURA 3.23 DIAGRAMA DE NAVEGACIÓN DEL DOCENTE/ USUARIO GENERAL	63
FIGURA 3.24 DIAGRAMA DE NAVEGACIÓN DEL USUARIO ANÓNIMO	64
FIGURA 3.25 PÁGINA INICIO DEL MODELO DE PRESENTACIÓN.....	65
FIGURA 3.26 PÁGINA INICIAR SESIÓN DEL MODELO DE PRESENTACIÓN	66

FIGURA 3.27 PÁGINA DE CONTENIDOS DEL MODELO DE PRESENTACIÓN	67
FIGURA 3.28 PÁGINA DE PROPONER CONTENIDO DEL MODELO DE PRESENTACIÓN	67
FIGURA 3.29 PÁGINA DE BÚSQUEDA DE CONTENIDO DEL MODELO DE PRESENTACIÓN	68
FIGURA 3.30 PÁGINA DE GESTIONAR USUARIOS DEL MODELO DE PRESENTACIÓN	69
FIGURA 3.31 PÁGINA DE GESTIONAR CONTENIDO DEL MODELO DE PRESENTACIÓN	69
FIGURA 3.32 DIAGRAMA PROCESO DEL CASO DE USO INICIAR SESIÓN.....	70
FIGURA 3.33 DIAGRAMA PROCESO DEL CASO DE USO CERRAR SESIÓN	70
FIGURA 3.34 DIAGRAMA DE PROCESO DEL CASO DE USO REGISTRAR USUARIO	71
FIGURA 3.35 DIAGRAMA DE PROCESO DEL CASO DE USO VALIDAR INFORMACIÓN	72
FIGURA 3.36 DIAGRAMA DE PROCESO DEL CASO DE USO PROPONER CONTENIDO.....	73
FIGURA 3.37 DIAGRAMA DE PROCESO DEL CASO DE USO ALTA DE CONTENIDO	73
FIGURA 3.38 DIAGRAMA DE PROCESO DEL CASO DE USO MODIFICAR CONTENIDO	74
FIGURA 3.39 DIAGRAMA DE PROCESO DEL CASO DE USO BAJA DE CONTENIDO	74
FIGURA 3.40 DIAGRAMA DE PROCESO DEL CASO DE USO ALTA DE USUARIO	75
FIGURA 3.41 DIAGRAMA DE PROCESO DEL CASO DE USO BAJA DE USUARIO	76
FIGURA 3.42 DIAGRAMA DE PROCESO DEL CASO DE USO MODIFICAR USUARIO	76
FIGURA 3.43 DIAGRAMA DE PROCESO DEL CASO DE USO CONSULTAR USUARIO.....	77
FIGURA 3.44 DIAGRAMA DE PROCESO DEL CASO DE USO RECUPERACIÓN CON BASE EN TEXTO	78
FIGURA 3.45 DIAGRAMA DE PROCESO DEL CASO DE USO RECUPERACIÓN CON BASE EN CONTENIDO	78
FIGURA 3.46 FLUJO DE TRABAJO	79
FIGURA 3.47 PROPONER VIDEO.....	80
FIGURA 3.48 DIAGRAMA OBTENCIÓN DE CUADROS CLAVE	81
FIGURA 3.49 CUADROS CLAVE.....	83
FIGURA 3.50 COLECCIÓN VIDEOS MONGODB.....	84
FIGURA 3.51 COLECCIÓN FRAMES MONGODB.	84
FIGURA 4.1 TIPO DE RECUPERACIÓN	86
FIGURA 4.2 NÚMERO DE ARTÍCULOS POR EDITORIAL.....	87

FIGURA 4.3 CANTIDAD DE ARTÍCULOS POR AÑO	88
FIGURA 4.4 TIPO DE CONSULTA.....	88
FIGURA 4.5 TECNOLOGÍAS MÁS UTILIZADAS.....	89
FIGURA 4.6 CONJUNTO DE DATOS.....	90
FIGURA 4.7 PROPONER VIDEO	91
FIGURA 4.8 COLECCIÓN DE VIDEOS	91
FIGURA 4.9 OPCIONES PARA GESTIONAR CONTENIDO	92
FIGURA 4.10 FORMULARIO PARA INSERCIÓN DE VIDEO DESDE LA PÁGINA DE GESTIÓN.....	92
FIGURA 4.11 EDITAR VIDEO.....	93
FIGURA 4.12 ELIMINAR VIDEO.....	94
FIGURA 4.13 VIDEO ELIMINADO	94
FIGURA 4.14 PESTAÑA VIDEOS.....	95
FIGURA 4.15 TODOS LOS VIDEOS.....	95
FIGURA 4.16 REPRODUCTOR DE VIDEO.....	96
FIGURA 4.17 BÚSQUEDA BASADA EN CONTENIDO	96
FIGURA 4.18 RESULTADO DE BÚSQUEDA BASADA EN CONTENIDO 1	97
FIGURA 4.19 RESULTADO DE BÚSQUEDA BASADA EN CONTENIDO 2	97
FIGURA 4.20 REPRODUCTOR DE VIDEO EN RESULTADO.....	98
FIGURA 4.21 PÁGINA BÚSQUEDA BASADA EN TEXTO.....	98
FIGURA 4.22 SELECCIÓN BÚSQUEDA BASADA EN TEXTO	99
FIGURA 4.23 BÚSQUEDA BASADA EN TEXTO.....	99
FIGURA 4.24 RESULTADO BÚSQUEDA BASADA EN TEXTO.....	100
FIGURA 4.25 REPRODUCCIÓN RESULTADO BASADA EN TEXTO	100

Índice de Códigos

CÓDIGO 3.1 CÓDIGO INSERTAR VIDEO 1.....	81
CÓDIGO 3.2 CÓDIGO INSERTAR VIDEO 2.....	82
CÓDIGO 3.3 CÓDIGO INSERTAR VIDEO 3.....	83

Resumen

El objetivo del presente trabajo fue extender el sistema para administrar los datos históricos del Instituto Tecnológico de Orizaba desarrollado en una tesis anterior con la agregación de un módulo de recuperación de videos con base en contenido de la base de datos multimedia HITO (Historia del Instituto Tecnológico de Orizaba), que permite la consulta de videos similares a una imagen proporcionada por el usuario. El módulo facilita la administración de los videos en la base de datos y el acceso efectivo y eficiente a videos históricos del Instituto Tecnológico de Orizaba.

La trascendencia de este módulo radica en que es una herramienta para disminuir el promedio de personas que desconocen gran parte de la historia del Instituto; además, incrementa la efectividad de las consultas basadas en texto al permitir la recuperación de videos más similares a una imagen de búsqueda.

En este proyecto se eligieron diversas tecnologías para su desarrollo. Como marco de trabajo, se utilizó JavaServer Faces. Para la gestión de la base de datos, se empleó Mongo DB. Además, se utilizó la biblioteca de visión artificial OpenCV. Tomando en cuenta el IDE (*Integrated Development Environment*, Entorno Integrado para el Desarrollo), se optó por NetBeans, ya que ofrece un entorno óptimo para aprovechar al máximo estas tecnologías.

Abstract

The aim of this work was to extend the system for the administration of the historical data of the Technological Institute of Orizaba developed in a previous thesis with the addition of a module of retrieval of videos based on content of the multimedia database HITO (History of the Instituto Tecnológico de Orizaba), which allows the consultation of videos similar to an image provided by the user. The module facilitates the administration of the videos in the database and the effective and efficient access to historical videos of the Technological Institute of Orizaba.

This module is relevant because it is a tool to reduce the average number of people who are unaware of a considerable part of the Institute's history; moreover, it increases the effectiveness of text-based queries by allowing the retrieval of videos that are more similar to an image query.

In this project various technologies were chosen for its development. As a framework, JavaServer Faces was used. For the management of the database, Mongo DB was used. In addition, the OpenCV computer vision library was used. Taking into account the development environment, NetBeans was chosen, since it offers an optimal environment to take full advantage of these technologies.

Introducción

Se propuso un proyecto en el Instituto Tecnológico de Orizaba en el año 2000 que se enfocó en la recuperación de una amplia colección de datos multimedia. A medida que el proyecto avanzaba, se observó que conforme aumentaba el contenido multimedia, era más desafiante administrar eficientemente estos datos. Por tal motivo, se desarrolló en una tesis anterior una base de datos llamada HITO y un sistema para la gestión de los datos de HITO. El sistema también permite la recuperación de imágenes con base en contenido. Sin embargo, hasta el inicio de este proyecto de tesis, el sistema solo administraba texto e imágenes. Actualmente, se agregaron videos de eventos relevantes del Instituto a HITO, por lo que fue necesario extender el sistema para que facilite la gestión de los videos en HITO. Adicionalmente, se desarrolló un módulo para la recuperación de videos con base en contenido que representa un aumento en la efectividad de la búsqueda basada en texto.

Este documento se estructura en cinco capítulos. El primero engloba el marco teórico, donde se incluye el planteamiento del problema, los objetivos generales y específicos, además de la justificación del estudio. El segundo contiene el estado de la práctica y muestra un análisis comparativo; el capítulo tres presenta el análisis y selección de los trabajos más relacionados con el módulo de CBVR (*Content-Based Video Retrieval*, Recuperación de Video Basada en Contenido) desarrollado en esta tesis, también está conformado por el diseño del sistema y el proceso CBVR; el capítulo cuatro abarca el resultado del análisis, el sistema de recuperación de videos con base en contenido de la base de datos multimedia HITO y, por último, el capítulo cinco comprende las conclusiones y recomendaciones.

Capítulo 1. Antecedentes

Este capítulo incluye aspectos fundamentales de la investigación, tales como el marco teórico, el problema abordado, los objetivos específicos y generales, y la justificación de lo que se llevará a cabo.

1.1 Marco teórico

Enseguida, se establecen algunas definiciones de términos importantes para el desarrollo de la investigación.

1.1.1 Bases de datos multimedia

En la actualidad, es común que los datos multimedia, como imágenes, audio y video, se guarden en sistemas de archivos en lugar de bases de datos. Esta práctica no suele ser un problema cuando se trata de un número reducido de objetos multimedia, puesto que las características de las bases de datos no suelen ser necesarias. No obstante, a medida que aumenta la cantidad de objetos multimedia almacenados, se vuelven relevantes las características propias de las bases de datos, como las actualizaciones transaccionales, la facilidad de recuperación e indexación. Los datos multimedia generalmente cuentan con atributos descriptivos, p. ej., información sobre su fecha de creación, su autor y la categoría a la que pertenece. Una estrategia para crear una base de datos para este tipo de objetos multimedia implica utilizar sistemas almacenamiento de información para guardar características descriptivas y llevar un registro de los archivos que contienen los objetos multimedia [1].

1.1.2 Recuperación de Video con Base en Contenido

La recuperación de videos con base en contenido es el método donde los videos más relevantes para la consulta se obtienen desde conjuntos de datos basados en características (contenido del video) extraídas del video y la consulta [2]. En los sistemas CBVR, normalmente hay dos fases principales: extracción de características y recuperación. Ambas consumen recursos del CPU (*Central Processing Unit*, Unidad Central de Procesamiento) y requieren mucho tiempo causado por las operaciones altamente complejas y repetitivas. En la mayoría de los casos, las aplicaciones de visión artificial utilizan una pequeña cantidad de videos por la dificultad para adquirir recursos computacionales y almacenamiento para grandes conjuntos de datos. A

medida que aumenta el progreso de los datos no estructurados, los sistemas analíticos son capaces de asimilar e interpretar imágenes y videos, así como interpretar datos estructurados como el texto. Además, el análisis en tiempo real de conjunto de datos a gran escala es cada vez más importante en muchos campos, como la medicina, la video vigilancia, la publicidad, entre otros [1].

1.1.2.1 Consultas en CBVR

Con el rápido crecimiento en la cantidad de datos de video, los métodos eficientes de indexación y recuperación de video se convirtieron en uno de los desafíos más críticos en la gestión multimedia. Con este propósito las consultas de sistemas de recuperación de video basada en contenido hoy en día son un área activa de investigación. A continuación, se presentan los tipos de consulta más comunes.

Consulta por imagen

Este tipo de consulta consiste en recuperar los videos más relevantes para una gran base de datos de videos utilizando una imagen como consulta; es decir, dado un conjunto de video y una imagen de consulta, el objetivo es recuperar los k videos más relevantes, cada uno de los cuales contiene fotogramas que son similares a la imagen de consulta. Por ejemplo, los usuarios logran encontrar el video de una película por medio de una foto sin ingresar el nombre en Internet. Otra aplicación es que un usuario es capaz de buscar video de una conferencia en línea por medio de una diapositiva.

La comunidad de investigación invierte una considerable cantidad de dinero en el problema de recuperación de imagen a video. La solución más sencilla a este problema es considerar cada fotograma del video como una imagen individual. Por lo tanto, el problema de determinar si una imagen está contenida en un video se convierte en el problema de descubrir si una imagen es similar a otra imagen. Aunque esta solución logra una alta precisión, se llevan a cabo comparaciones de imagen a imagen, lo que consume mucho tiempo y es prohibitivo para las bases de datos de video a gran escala [3].

Consulta por video

Tradicionalmente, los videos requeridos se extraen mediante consultas basadas en texto, de modo que los videos asociados a las consultas se recuperan para el usuario. Por ejemplo, se considera el título, la descripción o los comentarios del video, cuando se realiza la consulta de texto. Sin embargo, la recuperación de video basada en consultas de texto no consigue extraer videos relevantes de un grupo de estos, ya que la recuperación de video depende completamente del texto de la consulta y del texto que se utiliza para describir o comentar videos. En cambio, las consultas basadas en contenido logran extraer videos similares a un videoclip proporcionado como consulta [4].

1.1.2.2 Descriptores

El descriptor es el término utilizado para indexar un conjunto de datos de acuerdo con criterios compartidos entre ellos, con el objetivo de facilitar su búsqueda y análisis. A continuación, se presentan algunos descriptores más comunes acompañados de su descripción.

Histogramas de Gradientes Orientados

Para obtener los HOG (*Histogram of Oriented Gradients*, Histogramas de Gradientes Orientados) se calcula el gradiente de borde de toda la imagen, después, se determina la orientación de cada píxel y, por último, se generan los histogramas. Estos histogramas relevantes extraídos se utilizan para recuperar el fotograma de video pertinente a partir de la base de datos de secuencias de video a través de una medida de similitud, por ejemplo, IHBM (*Integrated Histogram Bin Matching*, Coincidencia de grupo de Histograma Integrado) [5].

Bolsa de Palabras Visuales

El modelo BoVW (*Bag of Visual Words*, Bolsa de Palabras Visuales) tiene como objetivo construir un diccionario visual para la base de datos de video de destino basado en la salida del extractor de características. Su objetivo es extraer las características visuales de los videos de una base de datos y luego construir un diccionario visual mediante el método de agrupación en clústeres [3].

Características Robustas Aceleradas

Speeded-Up Robust Features (SURF) [2] es un algoritmo de visión por computadora que tiene la capacidad de obtener una representación gráfica de una imagen y extraer información detallada y específica de su contenido. Estos datos se procesan para llevar a cabo diversas operaciones, como la detección y reconocimiento de objetos específicos, personas o rostros, la creación de escenas en 3D, realizar el rastro de objetos y extraer puntos relevantes de interés. Estos procesos forman parte de la inteligencia artificial, donde un sistema se entrena para interpretar imágenes y determinar su contenido. El algoritmo SURF fue presentado por Herbert Bay en una conferencia internacional de visión por computadora celebrada en Austria en mayo del 2006.

Puntos Clave Binarios Robustos Invariantes Escalables

Binary Robust Invariant Scalable Keypoints (BRISK) [2] se basa en un patrón de muestreo circular fácilmente configurable a partir del cual se calcula la comparación de brillo para formar una cadena de descriptor binario. Las propiedades únicas de BRISK son útiles para un amplio espectro de aplicaciones, en particular para tareas con duras restricciones en tiempo real o potencia de cálculo limitada, ofrece la calidad de características de alta gama en aplicaciones que requieren mucho tiempo.

Características Harris

Detector de Esquinas Harris [2] es un método utilizado en sistemas de visión por computadora para identificar y extraer características específicas de una imagen. La detección de esquinas es ampliamente utilizada en diversas aplicaciones, como la detección de movimiento, análisis de imágenes, seguimiento de videos, modelado 3D y reconocimiento de objetos, entre otras. Esta técnica está relacionada con un concepto más amplio conocido como la detección de puntos de interés.

1.1.2.3 Redes Neuronales Convolucionales

La CNN (*Convolutional Neural Network*) se trata de una clase de red neuronal artificial en la cual las neuronas artificiales responden a los campos receptivos que funcionan de forma similar a las neuronas encontradas en la corteza visual primaria del cerebro

humano. Este tipo de red es una variante de perceptrón multicapa y su aplicación se centra en matrices bidimensionales. Se ha comprobado que estas redes son altamente efectivas en tareas de visión artificial, como la clasificación y segmentación de imágenes, así como en otras aplicaciones relacionadas [3].

1.1.3 Base de datos HITO (Historia del Instituto Tecnológico de Orizaba)

Con el fin de preservar la memoria histórica del Instituto Tecnológico de Orizaba, se emprendió un proyecto para recopilar, digitalizar y sistematizar el material histórico de la institución. Como resultado de este proyecto, se logró recolectar una amplia cantidad de datos multimedia. Por tal motivo, se desarrolló una base de datos llamada HITO y un sistema para gestionar adecuadamente los datos de HITO, específicamente, texto e imágenes. El sistema permite recuperar imágenes con base en contenido y utiliza un método de fragmentación horizontal para optimizar las consultas [6].

1.2 Situación tecnológica, económica y operativa de la empresa

El Instituto Tecnológico de Orizaba (ITO) es una institución educativa perteneciente al Tecnológico Nacional de México y se encuentra ubicado en la ciudad de Orizaba, Veracruz, específicamente, en oriente 9, colonia Emiliano Zapata. En el ITO se ofrecen programas de licenciatura, maestría y doctorado. Entre las maestrías disponibles se encuentran la 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

Se realizó un proyecto en el Instituto Tecnológico de Orizaba para preservar la memoria histórica de la institución. Este proyecto consistió en recopilar, digitalizar y organizar el material histórico del Instituto, lo cual resultó en la obtención de una gran colección de datos multimedia. No obstante, al concluir el proyecto, se constató que la gestión de estos datos se volvía más complicada a medida que aumentaba su volumen. Además, se creó un sistema que habilita la realización de consultas basadas en contenido de una base de datos multimedia. Se descubrió que uno de los

problemas con estas consultas es que se vuelven más lentas a medida que se agregan más objetos multimedia a la base de datos. Por tal motivo, en una tesis anterior se creó un sistema con la finalidad de facilitar la gestión de los datos históricos del Instituto. No obstante, los datos que se administraron fueron solo texto e imágenes. También, en el sistema es posible realizar recuperación de imágenes basada en contenido. Hasta el inicio de este proyecto de tesis, se contaba con una gran cantidad de videos de eventos relevantes del Instituto en formato VHS (*Video Home System*, sistema de video doméstico), los cuales se digitalizaron y se agregaron a la base de datos HITO. Por lo tanto, en este trabajo, se desarrolló un módulo que permite la gestión, así como la recuperación de video con base en contenido de la base de datos multimedia HITO.

1.4 Objetivos generales y específicos

Enseguida, se detallan el objetivo general y los objetivos específicos.

1.4.1 Objetivo general

Desarrollar un módulo que permita recuperar video con base en contenido de la base de datos multimedia HITO.

1.4.2 Objetivos específicos

1. Estudiar los conceptos principales de bases de datos multimedia y consultas basadas en contenido para entender el alcance de la investigación.
2. Analizar los métodos para recuperación de video con base en contenido para seleccionar el más adecuado.
3. Identificar las técnicas y tecnologías que se utilizarán para el desarrollo del módulo por medio de pruebas de concepto.
4. Diseñar el módulo de acuerdo con la metodología para desarrollo de software seleccionada.
5. Implementar el módulo del sistema para recuperar videos HITO con base en contenido.
6. Validar el módulo por medio de un caso de estudio.

1.5 Justificación

El módulo que se desarrolló facilita la gestión de los videos en la base de datos multimedia HITO, esto redujo la complejidad de trabajar directamente con el gestor para usuarios que no son expertos en la utilización de sistemas gestores de bases de datos. Además, es posible recuperar videos similares de HITO no solo con base en texto, sino también proporcionando una imagen. Los principales beneficiarios del sistema desarrollado son los estudiantes y profesores del Instituto Tecnológico de Orizaba, quienes podrán acceder de manera eficiente a la información histórica de la institución.

Capítulo 2. Estado de la práctica

En la actualidad, los sistemas de recuperación basada en contenido son esenciales para aprovechar al máximo las bases de datos multimedia. También es común encontrar investigaciones que incluyen diferentes métodos de recuperación de video. Por lo tanto, antes de presentar un enfoque relacionado con dichas tecnologías, es fundamental examinar investigaciones previas similares. En este capítulo se exponen diversas investigaciones que han abordado el tema, seleccionadas cuidadosamente que se enfocaron en el uso de sistemas de recuperación basada en contenido con algoritmos para optimizar y mejorar su rendimiento. También se exploraron métodos de recuperación de videos para bases de datos multimedia. Se llevó a cabo una síntesis de estos trabajos mediante una tabla y al final se realizó el análisis de la propuesta establecida en este estudio.

2.1 Trabajos relacionados

En [6] el aumento del contenido multimedia dificultaba la gestión óptima y recopilación de datos históricos del Instituto Tecnológico de Orizaba. Por lo tanto, se propuso la creación de un sistema para administrar los datos multimedia que usa una técnica de partición horizontal para mejorar la recuperación basada en contenido. Este sistema gestionará el almacenamiento y la administración de los datos históricos. La fragmentación adquiere una gran importancia debido a su capacidad para minimizar el acceso a datos intrascendentes, lo que reduce el tiempo de respuesta y los costos de ejecución de las consultas. Se logró desarrollar un sistema que genera un impacto económico positivo, ya que el uso de la fragmentación resultó en ventajas, como una disminución en los costos de ejecución o reducción de tiempos de recuperación.

Araujo y Girod [7] afirmaron que recuperar videos de repositorios grandes mediante consulta de imágenes es importante para muchas aplicaciones, como el monitoreo de marcas o la vinculación de contenido, y que hacer coincidir una imagen con un video es un problema inherentemente asimétrico. Los autores propusieron una técnica de comparación asimétrica para los vectores de Fisher y exploración sistemática de los elementos de consulta, así como también nuevos descriptores de video que se comparan directamente con los descriptores de imagen. Además, construyeron

vectores *Fisher* para segmentos de video y exploraron diferentes técnicas de agregación. Los resultados del primer experimento demostraron una mejora de hasta el 25% para dos tipos de asimetría. Estas técnicas se consiguen ver como incrustaciones de alta dimensión donde se comparan imágenes y videos.

A causa de la necesidad de sistemas eficientes de indexación y recuperación es crucial contar con sistemas automáticos de análisis de contenido de video capaces de reproducir, moderar, indexar, recuperar, navegar o buscar información almacenada en grandes bases de datos multimedia. Estas técnicas se agrupan en un solo concepto de sistemas CBVR. En [1] se desarrolló un sistema CBVR que proporciona videos similares de un gran conjunto de datos multimedia con base en un video de búsqueda. El método propuesto se implementó tanto en una máquina como en un clúster distribuido en tiempo real para evaluar el aspecto de rendimiento en tiempo real, especialmente cuando el número y el tamaño de los videos son grandes. Los resultados experimentales mostraron que el método mejoró significativamente la precisión de la recuperación sobre los métodos de última generación. También demostraron que los fotogramas clave seleccionados son significativos y capturan el contenido del video de manera integral. Los resultados de la implementación distribuida en tiempo real revelaron que las topologías propuestas mejoraron considerablemente el tiempo de procesamiento para aproximarse al rendimiento en tiempo real.

La mayoría de los sistemas CBVR existentes no siempre dan resultados precisos para todo tipo de bases de datos de video con diferentes características de color, forma y textura. A veces, las imágenes o videos que se ven similares no son semánticamente similares. Entonces, Ram *et al.* [5] sugirieron una metodología CBVR centrada en la selección de descriptores de características de gradiente de borde conocidos como HOG, que calculan el gradiente de borde de toda la imagen, determinan la orientación de cada pixel y generan los histogramas. El resultado experimental del enfoque propuesto determinó que el número de muestras de datos de video recuperadas relevantes es mayor en comparación con el trabajo existente de CBVR basado en

IHBM. El rendimiento del enfoque propuesto es mejor cuando se combina con otros enfoques existentes.

En [8], se desarrolló un marco de recuperación de video basada en contenido. Una de las características esenciales del proceso de recuperación de video y CBVR es un valor de color. DCT (*Discrete Cosine Transform*, Transformada de Coseno Discreta) se utiliza para extraer las características del video de búsqueda para compararlas con las características de los videos almacenadas en la base de datos. El video también incluye todo el resto del contenido digital, como fotos, voces y textos. Se distingue además por su consistencia temporal. La recuperación de video tiene dos enfoques: estructura basada en texto y marcos basados en contenidos. El descriptor de texto se utiliza en la estructura basada en texto para anotar video manualmente. La arquitectura basada en contenido permite a la aplicación acceder a un clip de video desde una lista de elementos visuales. El resultado permitió al sistema adquirir un video de una lista de videos con base en contenido visual que se extrae de forma completamente automática, como color, texturas, formas, pero no con base en atributos que no están relacionados con el contenido.

El archivo alemán de radiodifusión mantiene el patrimonio cultural de las emisiones de radio y televisión de la antigua República Democrática Alemana (RDA), por tal motivo, Mülhing *et al.* [9] implementaron un sistema para el análisis y la recuperación automática de contenido de video para facilitar la búsqueda en colecciones históricas de grabaciones de televisión de la RDA. El sistema CBVR se basó en una arquitectura distribuida y orientada a servicios con algoritmos de análisis de video para la detección de límites de disparo, clasificación de conceptos, reconocimiento de personas, reconocimiento de texto y búsqueda de similitudes. Como resultado el sistema de recuperación de video basada en contenido complementa las anotaciones humanas y ayuda a los archivistas y usuarios a encontrar tomas de video relevantes.

En [3] resolvieron el problema de la recuperación de video a gran escala mediante una imagen de búsqueda. Definieron el problema de recuperación de k imágenes más similares (*top-k*) para consultas de video, combinaron los méritos de las redes

neuronales convolucionales y un módulo de *Bag of Visual Words* (BoVW) donde diseñaron un modelo para la extracción y representación de información de fotogramas de video. Para cumplir con los requisitos de la recuperación de video a gran escala, propusieron un VWI (*Visual Weighted Inverted Index*, índice invertido ponderado visualmente) y un algoritmo relacionado para la mejora de la eficiencia y la precisión del proceso de recuperación. Los experimentos mostraron que las técnicas presentadas lograron mejoras sustanciales hasta un orden de magnitud de aceleración.

Las aplicaciones de CBVR están aumentando porque es un enfoque para recuperar los videos más relevantes de una base de datos de videos. Potluri *et al.* [2] desarrollaron un sistema que identifica en primer lugar los fotogramas clave del video utilizando el método del histograma de color. El componente de color se utiliza para identificar el fotograma clave. Tres características se derivan de todos los videos en la base de datos. Las tres características también se calculan para la imagen de búsqueda mediante el uso de técnicas de coincidencia de semejanza, todas las características se utilizan conjuntamente para asignar clasificaciones a los videos en la base de datos con base en las características de la imagen de búsqueda. Los videos que tienen rangos por debajo del umbral se logran recuperar como videos más relevantes para la imagen de consulta dada. También se utilizó SURF, BRISK y el detector de esquinas HARRIS para recuperar los videos similares.

La tarea de recuperación de video basada en contenido es un desafío a causa de varias distorsiones y variaciones de contenido a las que los videos de consulta están sujetos. En [10] se presentó una arquitectura de sistema CBVR que es a la vez robusta y escalable. Se utilizó un método modificado de generación de huellas dactilares de marco rHash, el cual es extremadamente robusto a las distorsiones y rápido de calcular. También se usó la biblioteca Faiis, desarrollada por Facebook Research, para indexar vectores binarios de Fingerprint. Esto se hace para acelerar las consultas de la base de datos. El conjunto de datos VCDB (*Video Copy DataBase*) se utilizó para la evaluación comparativa.

CBVR es tan popular, por la mayor utilización de sistemas analíticos basados en video. El análisis basado en video es bastante efectivo, ya que el video captura una serie de acciones. Esto termina con una mejor capacidad de toma de decisiones. El sistema CBVR juega un papel importante en el impulso de la interacción humano-computadora. En [4] se presentó un CBVR multimodal que tiene en cuenta tanto la información visual como la de audio para recuperar videos relevantes para el usuario. Se emplearon dos módulos para tratar con datos de videos y audio. Los datos de video se procesan para detectar el fotograma significativo de las tomas, esto se logra mediante LOA (*Lion Optimization Algorithm*, Algoritmo de Optimización de León). Las características se extraen de los datos visuales y con respecto a los datos de audio, las características extraídas están agrupadas por el algoritmo KFCM (*Kernel-based Fuzzy C-Means*, C-Medias Difuso basado en *Kernel*). En conclusión, se propuso un sistema CBVR multimodal, que se basa en datos visuales y de audio. Los datos visuales en combinación con los datos de audio funcionan mejor que cuando los modos se emplean individualmente. Los datos visuales y de audio se segregan para su posterior procesamiento. Los datos visuales se procesan para detectar el fotograma significativo de las tomas con la ayuda de LOA y las características se extraen de él.

En [11] se mostró un sistema de recuperación de video basada en contenido que contiene diferentes pasos, como video de búsqueda, extracción de características, ventanas en número de planos, clasificación de valores RGB y búsqueda de los valores medios, técnica TSnbTC (*Thepade's Sorted n-ary Block Truncation Coding*, Método de Codificación de Truncamiento de Bloque Ordenado de Thepade) aplicada, resultado del video y recuperación del mejor video. La técnica mostró un buen resultado en el caso del espacio de color RGB (rojo, verde y azul) y YCbCr (utilizado en sistemas de fotografía y video digital, que define el color en función de un componente de luminancia y dos de crominancia), la comparación se realiza con base en las versiones TSnbTC. A partir de la comparación, se identificó que el espacio de color YCbCr devolvió una mejor precisión en cada versión TSnbTC en comparación con el espacio de color RGB.

Chen *et al.* [12] construyeron un marco de trabajo de extremo a extremo llamado DSVH (*Deep Supervised Video Hashing*, *Hashing de Video Supervisado Profundo*), que emplea una red neuronal convolucional 3D para obtener características espacio-temporales de los videos, también entrenaron un conjunto de funciones *hash* mediante *hash* supervisado para que transfiera las funciones de video al espacio binario y obtenga los códigos binarios compactos de los videos. Finalmente, utilizaron la pérdida de tripletes para el entrenamiento de la red. Realizaron experimentos en tres conjuntos de datos de video público: UCF-101, JHMDB (*Joint-annotated Human Motion Data Base*, Base de Datos de Movimiento Humano con anotaciones Conjuntas) y HMDB-51. Los resultados mostraron que el método propuesto tiene ventajas sobre muchos métodos de recuperación de video de última generación. En comparación con el método DVH, el valor *mAP* (*Mean Average Precision*, Media de la Precisión Promedio) del conjunto de datos UCF-101 se mejora en un 9.3%, y la mejora mínima en el conjunto de datos JHMDB también se incrementa en un 0.3%. Al mismo tiempo, también se demostró la estabilidad del algoritmo en el conjunto de datos HMDB-51.

La generación y transmisión de contenido multimedia tiene un crecimiento explosivo, lo que plantea la necesidad de una gestión eficiente del almacenamiento de video, así como también de un sistema de recuperación de video. Por lo tanto, en [13] se propuso un algoritmo de recuperación de video basado en contenido utilizando la fusión de fotogramas y HOG. Los fotogramas representativos de los videos de la base de datos se procesan previamente mediante la fusión de fotogramas para obtener un perfil representativo de alta resolución de *frames* y HOG. Los resultados experimentales demostraron que el rendimiento del método propuesto funcionó mejor que otros métodos de última generación.

En [14], se afirmó que la mayoría de los hospitales a menudo registran y almacenan videos quirúrgicos. El uso de estas grabaciones incluye muchas aplicaciones potenciales, como para la educación médica, el análisis de enfermedades del paciente, el análisis de errores quirúrgicos y la evaluación de habilidades. Sin embargo, la búsqueda manual en las bases de datos de videos es excepcionalmente laboriosa y requiere mucho tiempo. Por lo tanto, existe una necesidad crítica de un

sistema de búsqueda de video basada en contenido eficiente y eficaz para videos quirúrgicos. Se propuso un marco para CBSVR (*Content-based Surgical Video Retrieval*, Recuperación de Videos Quirúrgicos Basada en Contenido) a través de *Multi-Deep Features Fusion* (Fusión de Características Multiprofundas) para apoyar la enseñanza y la capacitación quirúrgica, evaluación de la calidad, análisis post operatorio de la enfermedad y otras aplicaciones. Para validar el enfoque se tomaron 160 conjuntos de datos de acciones quirúrgicas y los resultados demostraron la efectividad de la fusión de características.

Por causa del rápido crecimiento del contenido multimedia que se originan de múltiples fuentes, los métodos convencionales de recuperación de video basada en palabras clave no son adecuados para enfrentar los desafíos técnicos. En [15] el modelo considera una combinación de cuatro características: textura, color, árbol binario y patrones binarios locales, para la indexación y recuperación de objetos de video distribuidos. El resultado obtenido ofrece flexibilidad, escalabilidad y fácil implementación en aplicaciones de recuperación de video basada en contenido, como la búsqueda académica y la recuperación de información web. Este trabajo se lograría ampliar aún más mediante la inclusión de un mecanismo de aprendizaje profundo para analizar las imágenes visuales.

2.2 Análisis comparativo

La Tabla 2.1 presenta una comparación de las propuestas asociadas con este trabajo. Uno de los desafíos más significativos en la gestión y dirección de bases de datos es el creciente número de videos generados diariamente, lo que plantea un reto importante en la búsqueda de videos basada en contenido. Por esta razón, se desarrolló un módulo para recuperar videos de la base de datos multimedia HITO utilizando características de contenido. Basándose en los temas previamente analizados, se abordaron diferentes enfoques para abordar los desafíos que surgieron, ya que demostraron efectividad, mejoras en el rendimiento, ventajas sobre métodos de recuperación de video de última generación y estabilidad de los algoritmos en los conjuntos de datos

Tabla 2.1 Análisis comparativo de artículos asociado

Artículo	Problema	Contribución	Tecnologías	Resultados
Rodríguez-Arauz et al. [6]	El aumento del contenido multimedia dificulta la gestión óptima y recopilación de datos históricos del Instituto Tecnológico de Orizaba.	Creación de un sistema de gestión de datos multimedia que utiliza un método de fragmentación horizontal para optimizar las consultas basadas en contenido.	Fragmentación horizontal, Recuperación basada en contenido, base de datos multimedia, MongoDB, JavaServer Faces, PrimeFaces, BoofCV.	El sistema generó un impacto económico, ya que al utilizar un método de fragmentación hay ventajas, como menores costos de ejecución o menores tiempos de recuperación.
Araujo y Girod [7]	Recuperar videos de repositorios grandes mediante consulta de imágenes, hacer coincidir imagen con video es un problema inherentemente asimétrico.	Propuso una técnica de comparación asimétrica para los vectores de Fisher y exploración sistemáticamente de los elementos de consulta.	Vectores Fisher, Comparación asimétrica.	Mejora de hasta el 25% para dos tipos de asimetría.

Tabla 2.1 Análisis comparativo de artículos asociados (continuación)

Artículo	Problema	Contribución	Tecnología	Resultados
Saoudi y Jai-Andaloussi [1]	La necesidad de sistemas eficientes de indexación y recuperación.	CBVR proporciona videos similares de un gran conjunto de datos multimedia con base en un video de búsqueda.	Clúster distribuido, CBVR.	Mejóro significativamente la precisión de la recuperación sobre los métodos de última generación.
Ram et al. [5]	Los sistemas CBVR existentes no siempre dan resultados precisos para todo tipo de bases de datos de video con diferentes características de color, forma y textura.	Sugirió una nueva metodología CBVR centrada en la selección de descriptores de características de borde conocidos como HOG.	CBVR, HOG, IHBM.	El rendimiento del enfoque propuesto es mejor cuando se combina con otros enfoques existentes.

Tabla 2.1 Análisis comparativo de artículos asociados (continuación)

Artículo	Problema	Contribución	Tecnología	Resultados
Hamad <i>et al.</i> [8]	La búsqueda de los elementos multimedia requeridos en video basados en grandes conjuntos de datos de Internet se consideran un desafío importante.	Creación de un marco de recuperación de video basada en contenido.	CBVR, DCT, segmentación de imagen.	Permitió al sistema adquirir un video de una lista de videos con base en contenido visual que se extrae de forma completamente automática, como color, texturas, formas, pero no atributos que no están relacionados con el contenido.

Tabla 2.1 Análisis comparativo de artículos asociados (continuación)

Artículo	Problema	Contribución	Tecnología	Resultados
Mühling <i>et al.</i> [9]	Mantener el patrimonio cultural de las emisiones de radio y televisión de la antigua república democrática alemana.	Sistema para el análisis y la recuperación automática de contenido de video para facilitar la búsqueda en colecciones históricas de grabaciones de televisión de la RDA.	Algoritmos de análisis de video para la detección de límites de disparo, clasificación de conceptos, reconocimiento de personas, reconocimiento de texto y búsqueda de similitudes.	El sistema de CBVR complementó las anotaciones humanas y ayudó a los archivistas y usuarios a encontrar tomas de video relevantes.
Zhang <i>et al.</i> [3]	Recuperar los videos más relevantes para una gran base de datos de videos utilizando una imagen como consulta.	Una solución novedosa para mejorar el rendimiento de recuperación de video mediante una imagen de consulta.	CNN, BoVW, Top-k, WWII.	La evaluación experimental en conjuntos de datos mostró que el enfoque superó métodos de vanguardia.

Tabla 2.1 Análisis comparativo de artículos asociados (continuación)

Artículo	Problema	Contribución	Tecnología	Resultados
Potluri y Gnaneswara [2]	La tarea desafiante fue recuperar los fotogramas clave del video más video utilizando el método del histograma de color. grandes conjuntos de datos.	Sistema que identifica fotogramas clave del video utilizando el método del histograma de color.	<i>Query-by-image</i> , CBVR, SURF, BRISK, HARRIS <i>corner detector</i> .	La combinación de características funcionó mejor en el cambio de iluminación y rotación, la métrica de coincidencia se utilizó para extraer las coincidencias exactas y recuperar los videos sobre la base del rango.

Tabla 2.1 Análisis comparativo de artículos asociados (continuación)

Artículo	Problema	Contribución	Tecnología	Resultados
Tseytlin y Makarov [10]	Buscar en una base de datos de videos de referencia dado un video de consulta para que coincidiera si se trataba de una subsecuencia corta distorsionada de un video de referencia (copia parcial).	Una arquitectura de sistema CBVR que es robusta y escalable.	CBVR, Quadrant rHash, Faiis, Vectores binarios, Fingerprint, VCDB.	Se evaluaron estrategias diferentes de coincidencia de secuencias de video, los resultados mostraron un ligero aumento del rendimiento en comparación con enfoques similares.
Prathiba y Kumari [4]	La recuperación de video basada en contenido visual no lograba extraer videos relevantes de un grupo de videos.	Sistema de recuperación de video que emplea dos módulos para tratar con datos de video y audio.	CBVR, CBVR Multimodal, LOA, KFCM.	El sistema CBVR demostró un mejor rendimiento que las aplicaciones existentes a través de los resultados obtenidos.

Tabla 2.1 Análisis comparativo de artículos asociados (continuación)

Artículo	Problema	Contribución	Tecnología	Resultados
Soniminde y Thepade [11]	Los métodos de recuperación de video basada en la descripción del texto o palabras clave enfrentan dificultades. Es posible utilizar varios métodos de extracción de características para CBVR.	Sistema de recuperación de video basada en contenido con diferentes pasos.	de CBVR, TSnbTC, YCbCR, BTC, RGB.	Devolvió un buen resultado para el espacio de color RGB y YCbCR, la comparación se realizó entre estos basada en las versiones TSnbTC. A partir de la comparación, identificó que el espacio de color YCbCR devuelve una mejor precisión en cada versión TSnbTC en comparación con el espacio de color RGB.

Tabla 2.1 Análisis comparativo de artículos asociados (continuación)

Artículo	Problema	Contribución	Tecnología	Resultados
Chen et al. [12]	La necesidad de un sistema de recuperación de video con velocidad de recuperación rápida y alta precisión.	Un método de tres componentes. El primero es elegir fotogramas representativos, el segundo es usar una CNN 3D y el tercero ajustar la red con la capa <i>hash</i> .	<i>Hash</i> , DSVH, CNN 3D, UCF-101, JHMDB, HMDB-51.	El método tiene ventajas sobre muchos métodos de recuperación de video de última generación. Demostró estabilidad del algoritmo en los datos HMDB-51.
Mounika y Khare [13]	La necesidad de una gestión del almacenamiento de video, un sistema de recuperación de videos capaz de soportar el crecimiento de contenido multimedia.	Algoritmo CBVR que utiliza la fusión de fotogramas y el histograma de gradientes orientados.	HOG, CBVR, índice Jaccard.	Los resultados experimentales demostraron que el rendimiento del método propuesto funcionó mejor que otros métodos de última generación.

Tabla 2.1 Análisis comparativo de artículos asociados (continuación)

Artículo	Problema	Contribución	Tecnología	Resultados
Kumar et al. [14]	En la mayoría de los hospitales se registran y almacenan videos quirúrgicos. Sin embargo, la búsqueda manual en la base de datos de videos es laboriosa y requiere mucho tiempo.	Sistema de búsqueda de video basada en contenido para videos quirúrgicos.	CNN, CBSVR, <i>Multi-Deep Features Fusion</i> .	Para validar los datos se tomaron 160 conjuntos de datos de acciones quirúrgicas, los resultados demostraron la efectividad de la fusión de características.
Kaliaperumal et al. [15]	Los métodos convencionales de recuperación de video basada en palabras clave no son adecuados para enfrentar los desafíos técnicos.	Un marco para almacenar y recuperar grandes volúmenes de objetos de video distribuidos de manera eficiente.	CBVR, BPT, CIE, COMB_FETS, DQEE, HSV, LBP.	El resultado ofreció flexibilidad, escalabilidad y fácil implementación en aplicaciones de recuperación de video basada en contenido.

2.3 Propuesta de la solución

Se eligió MongoDB debido a su rapidez y su facilidad para consultar contenidos en la base de datos. JavaServer Faces se seleccionó debido a su flexibilidad para desarrollar aplicaciones utilizando el lenguaje Java, su facilidad de desarrollo y su disponibilidad gratuita. Se seleccionó NetBeans como entorno de desarrollo integrado (IDE) por su facilidad de uso y su compatibilidad con el marco de trabajo elegido. Se escogió la metodología de desarrollo UWE porque su enfoque se centra en desarrollar aplicaciones web y multimedia. Por último, se optó por OpenCV, ya que es una biblioteca libre de visión artificial y porque tiene aplicaciones en detección de movimiento, reconocimiento de objetos, reconstrucción 3D a partir de imágenes, por mencionar algunas.

Capítulo 3. Aplicación de la Metodología

Con el propósito de aplicar la metodología, este capítulo incluye dos secciones. La primera contiene la explicación del proceso de realizar una búsqueda exhaustiva y un profundo análisis de técnicas de CBVR que utilizan diferentes descriptores con el objetivo de determinar si existe una técnica que cumpla con ciertos criterios, sin embargo, no se encontró ninguna por lo que fue necesario proponer un nuevo método. La segunda parte se enfoca en el diseño de una aplicación web para implementar el enfoque propuesto. Además, el presente capítulo detalla el proceso del método propuesto.

3.1 Análisis

Se llevó a cabo una exhaustiva revisión de los trabajos relacionados con CBVR en las principales bibliotecas digitales de Ciencias de Computación (IEEE Xplore, ACM digital library, SpringerLink, ScienceDirect, entre otras). Posteriormente, los trabajos se analizaron con base en seis criterios: 1) El tipo de recuperación de videos que consideran; 2) El tipo de consulta que realizan: por video, por imagen, entre otras; 3) Las tecnologías que utilizan; 4) El conjunto de datos que emplean; 5) Los descriptores usados, y 6) Las medidas de similitud. Solo se eligieron trabajos publicados en los últimos cinco años (2018-2022) en revistas o memorias de congresos y se excluyeron tesis de maestría o doctorado, libros y artículos escritos en idiomas diferentes al inglés.

Se realizaron todas las etapas de la metodología propuesta en la Figura 3.1. Tal como se señaló anteriormente, se llevó a cabo una búsqueda exhaustiva de artículos en las principales bibliotecas digitales de editoriales científicas, ACM, IEEE, Springer y Elsevier. Los artículos encontrados que no estaban publicados por estas editoriales se clasificaron como “Otras”. La búsqueda se centró en obtener todos los artículos que contenían la palabra clave: CBVR. Una vez recopilados todos los trabajos se aplicó un filtro para descartar las tesis de maestría o doctorado, además de los libros y artículos escritos en idiomas distintos al inglés. Los trabajos obtenidos se categorizaron por editorial y por año. La fase final consistió en examinar cada uno de ellos según seis categorías: CBVR, Tipo de consulta, Tecnologías, Conjunto de datos, Descriptores y Medida de similitud.

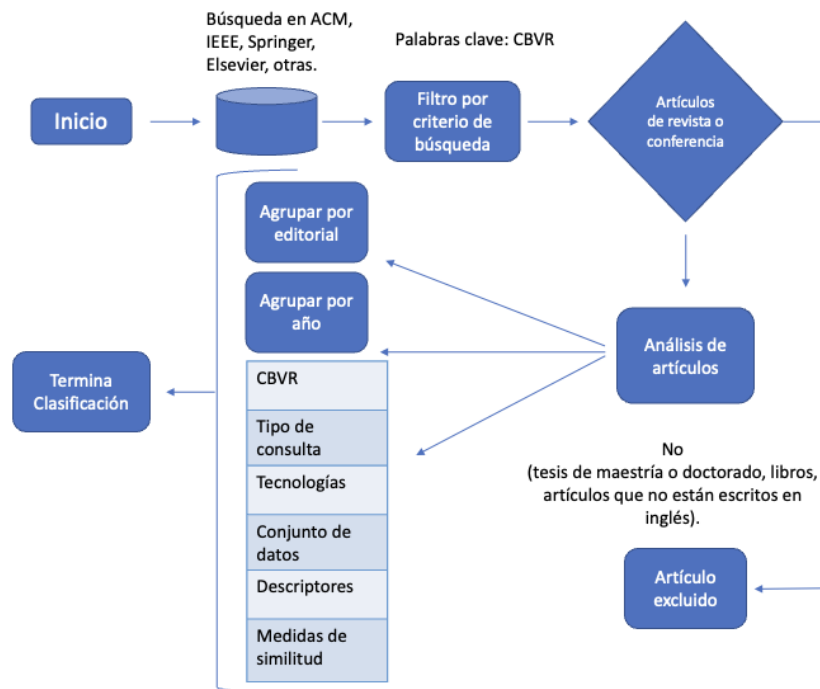


Figura 3.1 Metodología de búsqueda y análisis de artículos relacionados.

Como resultado del proceso de selección, se obtuvieron 32 artículos que se analizaron en relación con las seis características, registrando si cumplían con cada una de ellas. Los criterios considerados fueron: 1) Recuperación de video basada en contenido; 2) Ejecución de consultas por imagen; 3) Utilización del descriptor SURF; 4) Empleo de un SGBD (Sistema Gestor de Bases de Datos); 5) Uso de la distancia Euclidiana, y 6) Consideración de datos reales.

Las Tabla 3.1 se procedió a realizar el registro de cada artículo y se constató que ninguno de ellos cumplía con los seis criterios considerados, la mayoría de los trabajos revisados no consideran un sistema gestor de bases de datos para la administración de datos multimedia y utilizan datos sintéticos, mientras que el módulo propuesto toma una imagen como entrada para recuperar videos con base en contenido de eventos relevantes del Instituto Tecnológico de Orizaba, utiliza el descriptor SURF, emplea la distancia Euclidiana y utiliza el SGBD MongoDB. La Tabla 3.2 proporciona mayor información del análisis realizado en cada uno de los artículos.

Tabla 3.1 Comparación de trabajos

Artículo	1	2	3	4	5	6
Saodi y Jai-Andalousi [1]	Sí	No	No	No	No	No
Potluri y Gnaneswara [2]	Sí	Sí	Sí	No	Sí	No
Zhang <i>et al.</i> [3]	Sí	Sí	No	No	No	Sí
Prathiba y Kumari [4]	Sí	No	No	No	Sí	No
Ram <i>et al.</i> [5]	Sí	Sí	No	No	No	Sí
Araujo y Girod [7]	Sí	Sí	No	No	No	No
Hamad <i>et al.</i> [8]	Sí	No	No	No	Sí	Si
Mühling <i>et al.</i> [9]	Sí	No	No	No	No	Si
Tseytlin y Makarov [10]	Sí	No	No	No	No	Sí
Soniminde y Thepade [11]	Sí	No	No	No	No	No
Chen <i>et al.</i> [12]	Sí	No	No	No	Sí	No
Mounika y Khare [13]	Sí	Sí	No	No	Sí	Sí
Kumar <i>et al.</i> [14]	Sí	No	No	No	No	No
Kaliaperumal <i>et al.</i> [15]	Sí	No	No	No	No	Sí
Jiang <i>et al.</i> [16]	Sí	No	No	No	No	No
Shao <i>et al.</i> [17]	Sí	No	No	No	No	No
Rossetto <i>et al.</i> [18]	Sí	Sí	No	Sí	No	No
Sharma y Reetu [19]	Sí	Sí	No	No	No	No
Jo <i>et al.</i> [20]	Sí	Sí	No	No	No	No
Kordopatis-Zilos <i>et al.</i> [21]	Sí	Sí	No	No	No	No
Zhou <i>et al.</i> [22]	Sí	No	No	No	No	No
Gabeur <i>et al.</i> [23]	No	No	No	No	No	No
He <i>et al.</i> [24]	Sí	No	No	No	No	No
Yasin <i>et al.</i> [25]	Sí	No	No	No	Sí	Sí
Shao <i>et al.</i> [26]	Sí	Sí	No	No	No	No
Zhang <i>et al.</i> [27]	No	No	No	No	Sí	Sí
Amato <i>et al.</i> [28]	Sí	Sí	No	No	No	No
Wu <i>et al.</i> [29]	No	No	No	No	No	No
Liang <i>et al.</i> [30]	Sí	No	No	No	No	No
Kordopatis-Zilos <i>et al.</i> [31]	Sí	No	No	No	No	No
Bansal y Chakraborty [32]	Sí	No	No	No	Sí	Sí
Tan <i>et al.</i> [33]	Sí	Sí	No	No	No	Sí
Este trabajo	Sí	Sí	Sí	Sí	Sí	Sí

1) CBVR, 2) Consultas por imagen, 3) SURF, 4) SGBD, 5) Distancia Euclidiana, 6) Datos reales

Tabla 3.2 Análisis de artículos

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	de Conjunto de datos	de Mapas de color
Saodi y Jai-Andalousi [1]	Sí	Por video	Firmas basadas en movimiento, matrices RGB	k-means, PCA (<i>Principal Component Analysis</i> , Análisis de Componentes Principales), Apache Storm, FFMPEG, Ubuntu 16.04 server	Similitud BCS (<i>Bounded Principal Component</i> , Componente Principal Acotado), Similitud ZNCC (<i>Zero-mean Normalized Cross Correlation</i> , Correlación Cruzada Normalizada Cero-promedio)	HOLLYWOOD2, HMDB51, UCF50 (Conjunto de datos de reconocimiento de acciones de la Universidad Central de Florida) y deportes olímpicos	RGB
Potluri y Gnaneswara Rao [2]	Sí	Por imagen	SURF, BRISK, Características del detector de esquinas Harris, Histograma de color	No se mencionan	Distancia Euclidiana, Distancia Hamming	UCF 11	RGB, HSV

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	Conjunto de datos	Mapas de color
Zhang et al. [3]	Sí	Por imagen	Características CNN (Alexnet)	CNN, Bag of Visual Words (BoVW), Visual Word Inverted Index, Ubuntu, Java, Python	Similitud visual basada en TF-IDF (Term Frequency-Inverse Document Frequency, Frecuencia de Término- Frecuencia Inversa de Documentos)	YouTube-8M, Sports-1M	NE
Prathiba y Kumari [4]	Sí	Por video y audio	Descriptor GIST, Mean Hilbert Envelope Coefficients (MHEC), Linear Prediction Cepstral Coefficients (LPCC), Momentos de color, Momentos Zernike.	MATLAB, Kernelized Fuzzy C Mean, Double Density Discrete Wavelet Transform (DD-DWT), Lion Optimization Algorithm, Bionic Wavelets (BW)	Distancia Euclidiana	Sports 1-M	HSV

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	de Conjunto de datos	de Mapas de color
Ram et al. [5]	Sí	Por imagen	HOG	MATLAB	IHBM	Propio	RGB
Araujo y Girod [7]	Sí	Por imagen	SIFT (<i>Scale Invariant Feature Transform</i> , Transformación de Características Invariantes de Escala) y Vectores Fisher	PCA, Filtros Bloom	Coseno	Búsqueda Visual Móvil de Stanford, Holidays, MIRFLICKR-1M, SI2V (<i>Stanford Image to Video</i>), VB (Video Bookmarking) y ClassX (videos de clases con consultas de imágenes de diapositivas)	NE
Hamad et al. [8]	Sí	Por video	Características de color DCT	DCT	Euclidiana y error cuadrático medio	Propio	RGB

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	Conjunto de datos	Mapas de color
Mühling et al. [9]	Sí	Por video	Características CNN (VGG-16), Hash semántico	CNN GoogleNet, Deep learning framework Caffe, BoVW, Video Optical Character Recognition (OCR), FFMPEG, GPU Nvidia Geforce GTX 770	Distancia Hamming	RDA, German Broadcasting Archive, Conjunto de datos de escenas NUS-WIDE, ImageNet e ILSVRC (ImageNet Large Scale Visual Recognition Challenge) para entrenar GoogleNet	NE
Tseytlin y Makarov [10]	Sí	Por video	rHash, Quadrant rHash	Biblioteca Faiss, Temporally informative representative images (TIRI), Python, Ubuntu 18	Distancia Hamming	Conjunto de datos VCDB	NE

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	de Conjunto de datos	Mapas de color
Soniminde y Thepade [11]	Sí	Por video	Vectores de características de espacio de color	MATLAB, TSnBTC, BTC (Block Truncation Coding, Codificación de truncamiento en bloque) multinivel	Mean Squared Error (MSE)	Propio	YCbCr, RGB, LUV, YIQ, YUV, CIEXYZ, CIEXYZrec, CIEXYZccir, CMY, CMYK
Chen et al. [12]	Sí	Por video	DSVH, Características CNN (modelo 3D ResNet)	CNN 3D NVIDIA Tesla K40c GPU	Distancia Hamming, Distancia Euclidiana	UCF-101 JHMDB HMDB-51 Kinetics	NE
Mounika y Khare [13]	Sí	Por imagen	HOG	Mosaico de imágenes, Super resolución	Distancia Euclidiana	Propio	NE

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	de Conjunto de datos	de Mapas de color
Kumar et al. [14]	Sí, <i>Content based surgical video retrieval</i>	Por video	<i>Multi-Deep Features Fusion</i> (Alexnet, GoogleNet, VggNet, ResNet)	CNN, MATLAB, Tesla k40c GPU	Distancia coseno	Surgical Actions 160	NE
Kaliaperumal et al. [15]	Sí	Por video	Histogramas de color, BPT (<i>Binary Parititon Tree</i> , Árbol de partición binario), coeficientes wavelet, LBP (<i>Local Binary Patterns</i> , Patrones Binarios Locales)	Python	Distancia chi cuadrada	Propio (Videos de escenarios naturales con una duración de 7 a 10 minutos y tamaño de 15.7 a 112.7 MB).	HSV CIELUV

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	de Conjunto de datos	Mapas de color
Jiang et al. [16]	Sí, S-CBVR (<i>Segment-level CBVR</i> , CBVR a nivel de segmento)	Por video	Características CNN (R-MAC, ViSiL)	SSAN (<i>Segment Similarity and Alignment Network</i> , Red de similitud y alineación de segmentos)	Patrón de similitud	de VCDB, CC_WEB, FIVR-200k, VCDB_plus, SVD (<i>Short Video Dataset</i> , Conjunto de Datos de Videos Cortos), ImageNet	No especifica (NE)
Shao et al. [17]	Sí	Por video	<i>Context encoding for video retrieval with contrastive learning</i> (CECL)	PCA, <i>Fine-grained Incident Video Retrieval</i> (FIVR), <i>Event Video Retrieval</i> (EVR), PyTorch, Horovod	<i>Deep metric learning</i> (DML), mAP	VCDB, CC_WEB_VIDEO, Event VidEo (EVVE), FIVR-200K y FIVR-5K	NE

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	de Conjunto de datos	de Mapas de color
Rossetto et al. [18]	Sí	Por texto, por conceptos, bocetos, imagen, audio, video y datos 3D	Características CNN (EAST, CRNN, 3DconvNets, LSTM, DeepLab)	CNN, motor de recuperación Cineast, ADAMpro, TensorFlow Java API, recuperación multimedia de código (vitivr)	Distancia semántica	MJSynth	NE
Sharma y Retu [19]	Sí	Por modelo, Por boceto	MST-CSS (<i>Multi-Spectro-Temporal Curvature Scale Space</i>), <i>Trajectory-Pooled Deep-Convolutional Descriptors</i> (TDDs)	FaceVACS toolbox	No se menciona	Propio (8 hrs. de películas y 8 hrs. de programas de TV)	RGB

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	Conjunto de datos	Mapas de color
Jo et al. [20]	Sí	Por video	Histogramas de color, SIFT Comprimido, Scalable Compressed Fisher Vector (SCFV), Características CNN	Adaptive Binary Arithmetical Coding (ABAC), NVIDIA V100 GPUs	Similitud NIP, Similitud SCFV	FIVR-5K, CC_WEB_VIDEO, Kinetics-400	NE
Kordopatis-Zilos et al. [21]	Sí (FIVR)	Por video	Características CNN (Deep Metric Learning Network, Red de Aprendizaje Métrico Profundo)	CNN, Scikit-learn, 3D CNN, YouTube API, BoVW, NDVR (Near Duplicate Visual Retrieval, Recuperación Visual Casi Duplicada), NLTK toolkit, Part-of-Speech (PoS) tagging, lematizador NLTK basado en WordNet	Similitud coseno entre representaciones TF-IDF de sus palabras visuales	FIVR-200K, CC_WEB_VIDEO, VCBD	NE

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	de Conjunto de datos	de Mapas de color
Zhou et al. [22]	<i>Video Moment Retrieval</i> (VMR)	Por video	Características CNN (C3D, LSTM), <i>Multi-Graph Feature Fusion Module</i>	PCA, MultiLayer Perceptron (MLP), PyTorch	Similitud semántica nivel de fotograma	ActivityNet v1.2, Thumos14	No se especifica
Gabeur et al. [23]	Recuperación de video multimodal	Por texto	Características RGB de S3D en el conjunto de datos de acciones Kinectics, Características de audio del modelo VGGish en el conjunto de datos YouTube8M, Transcripciones ASR codificadas con BERT-base-cased.	Codificador de video multimodal, V-100 GPU	NE	HowTo100M, How2R, YouCook2, MSR-VTT, <i>Condensed Movie Dataset</i> (CMD), ActivityNet Captions	RGB

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	Conjunto de datos	Mapas de color
He et al. [24]	Sí, <i>Duplicate Scene Video Retrieval</i> (DSVR), <i>Complementary Scene Video Retrieval</i> (CSV), <i>Incident Scene Video Retrieval</i> (ISVR)	Por video	Características CNN (codificador de características ResNet50), IsoHash	No se menciona	Distancia Hamming	FIVR-200K, SVD, Propio	NE
Yasin et al. [25]	Sí	Por texto	Características CNN (FaceNet, codificador ResNet-101, decodificador basado en LSTM)	Arquitectura Inception ResNet v1, Word2Vec (CNNs), <i>Long-short term memory networks</i> (LSTMs)	Distancia Euclidiana, similitud coseno	Base de datos VGGFace2, conjunto de datos MS COCO (<i>Common Objects in Context</i> , Objetos Comunes en Contexto), Propio	NE

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	Conjunto de datos	Mapas de color
Shao et al. [26]	Sí (NDVR, FIVR y EVR)	Por imagen, por video	Características CNN (iMAC, L ₃ -iMAC, NETVLAD y NeXtVLAD, LSTM, GRU, ResNet-50)	PCA, PyTorch, Harovod, 4 Tesla-V100-SXM2-32GB GPUs	Similitud coseno, similitud chamfer	CC WEB VIDEO, FIVR-200K, VCDB, EVVE	RGB
Zhang et al. [27]	<i>Video content-based advertising</i>	Por video	Características CNN (SSD300 con ResNet 101 para detección del cuerpo humano, GoogLeNet para selección de poses, SSD300 con VGG-16 para detección de ropa), LBP, <i>Local Texture Pattern</i> (LTP)	Biblioteca CNN, Cafee, 4-ways Geforce GTX 1080 GPU, Motor de reconocimiento de caras de código abierto SeetaFace, <i>Density Peaks Clustering Algorithm</i> (DPCA), Binary Hash Code, MATLAB, Python	Similitud coseno, distancia Hamming, distancia Euclidiana, Normalized Discounted cumulative gain (NDCG)	PASCAL VOC, Street-Shop, Propio,	NE

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	de Descriptores	Tecnologías	Medida de similitud	de Conjunto de datos	de Mapas de color	
Amato et al. [28]	Sí	Por texto, ubicación objeto, ubicación color, imagen	por el profundo del MAC, por histogramas de colores espaciales	Descriptores visuales R- MAC, histogramas de colores espaciales	Detector de objetos YOLO, Apache Lucene, Cafee Model Zoo, ResNet-101, Cuantificación Escalar, VISIONE	Earth Mover's Distance, BM25, TFIDF, Producto punto (TF ranker), NormTF	YFCC100M, V3C1	CIELAB
Wu et al. [29]	Large-scale video retrieval	Por video	Características CNN (Temporal Segment Networks)	PCA, Canonical Correlation (CCA), MATLAB, Servidor Ubuntu, TITAN 1080i GPUs, OpenCV, Cafee	Distancia Hamming	FCVID, YFCC, ActivityNet, ImageNet, UCF101	RGB	

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	de Conjunto de datos	de Mapas de color
Liang et al. [30]	Sí (DSVR, CSV, ISVR)	Por video	Características CNN (VGG-16, ResNet-50)	PCA, BoW, LSH (<i>Locality-Sensitive Hashing</i> , Hashing Sensible a la Localidad), 8 Tesla-V100 32GB GPUs, FFMPEG para extracción de cuadros clave, CNN, Faiss, Unsupervised teacher-student model (UTS Net)	Similitud coseno	FIVR	NE
Kordopatis-Zilos et al. [31]	Sí (NDVR, FIVR, EVR, <i>Action Video Retrieval</i> (AVR))	Por video	Características CNN (ResNet-50, I3D)	PCA CNN	<i>Tensor Dot</i> (TD) <i>Product</i> , Similitud Chamfer	VCDB, CC_WEB-VIDEO, FIVR-200K, FIVR-5K, EVVE, ActivityNet	NE

Tabla 3.2 Análisis de artículos (continuación)

Artículo	CBVR	Tipo de consulta	Descriptores	Tecnologías	Medida de similitud	de Conjunto de datos	de Mapas de color
Bansal y Chakraborty [32]	Sí	Video, lenguaje natural	Características CNN (VGG16, modelo codificador-decodificador de Red Neuronal Recurrente con celdas LSTM), incrustación de oraciones	OpenCV, TensorFlow	Distancia Euclidiana	Microsoft Video Description Dataset (MSVD), Movie Description Dataset (MPII-MD), Montreal Video Annotation Dataset (MVAD)	NE
Tan et al. [33]	Sí, PVCD (<i>Partial Video Copy Detection</i> , Detección de Copia de Video Parcial)	Por video	Características CNN (Resnet29-RMAC y VGG-16 pool5)	KNN CNN Nvidia GTX 1080TI GPU Ubuntu 16.04 LTS OS	Similitud coseno	VCDB y un conjunto de datos PVCD propio	NE

3.2 Selección

Después de analizar los 32 artículos, se constató que ninguno de los artículos cumplió con los seis criterios establecidos, por lo que se considera que los trabajos más relacionados con el módulo CBVR desarrollado en este trabajo son dos artículos ([2], y [13]), ya que el módulo CBVR permite realizar consultas de video con base en contenido ingresando una imagen como entrada, utiliza descriptores SURF, usa la medida euclidiana para comparar la imagen consulta con las imágenes de los cuadros del video y la base de datos multimedia empleada incluye datos históricos del Instituto Tecnológico de Orizaba, por lo que [2] y [13] también consideran consultas por imagen y para la medida de similitud usaron la distancia Euclidiana; sin embargo, solo [2] implementó el descriptor SURF y únicamente [13] utilizó datos reales; además ninguno de los dos utilizó un SGBD para almacenar los datos multimedia.

3.3 Diseño del sistema

En este apartado se expone el diseño de la aplicación empleando la metodología de desarrollo UWE (UML- *based web engineering*, UML-basado en ingeniería Web). Se siguen todos los pasos de UWE y se elaboran los diagramas correspondientes de esta metodología para desarrollar el módulo.

La arquitectura del sistema se basó en el patrón arquitectónico MVC (Modelo Vista Controlador), ya que permite distribuir los componentes del sistema de manera que su mantenimiento sea más fácil. En la Figura 3.2, se muestra de forma abstracta esta arquitectura. La vista consiste en las páginas que interactúan con el usuario, donde se solicita el contenido almacenado en la base de datos HITO y proporcionan acceso a esta. En el controlador se encuentran los *beans* administrados de JSF (*JavaServer Faces*) encargados de manejar el flujo de los datos. Por otro lado, las reglas del negocio se establecen en el modelo, las cuales abarcan la administración de las colecciones históricas del Instituto Tecnológico de Orizaba, incluyendo recuperación de video basada en contenido.

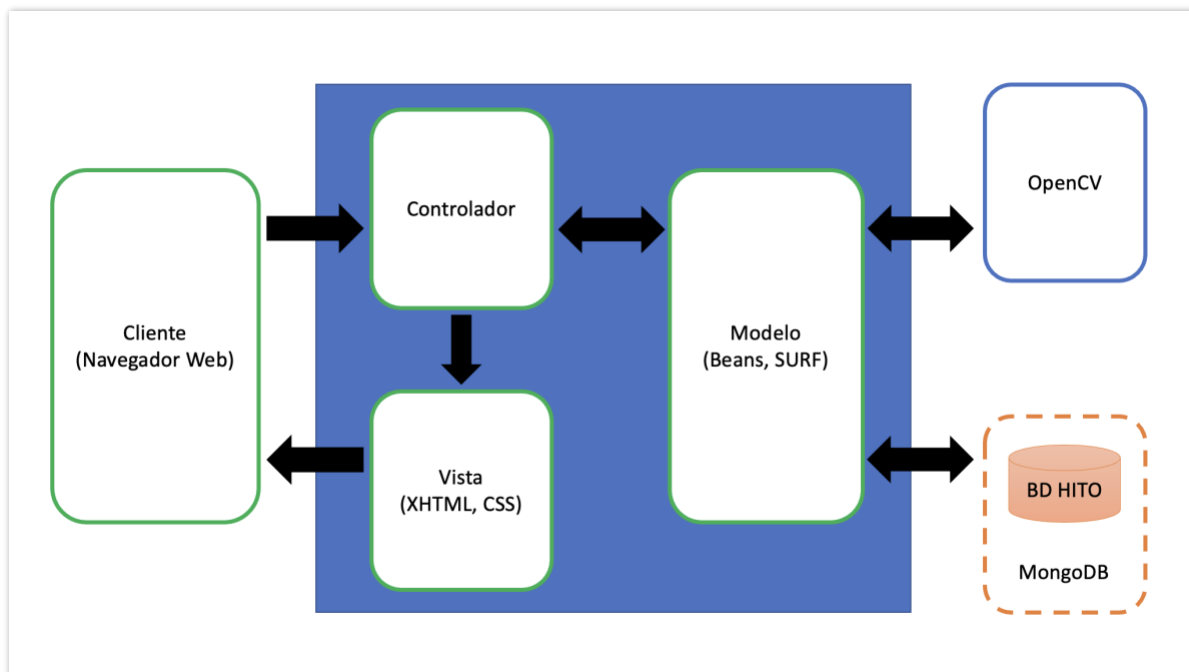


Figura 3.2 Arquitectura de la aplicación Web

3.3.1 Análisis de requisitos

El análisis de requisitos tiene como objetivo principal definir los requisitos funcionales de la aplicación web, los cuales se representan a través de diagramas, siendo el diagrama de casos de uso uno de ellos. Este diagrama muestra la relación entre los actores y las actividades, y permite visualizar las actividades con que cada actor está relacionado. Los actores de la aplicación se establecen en la Tabla 3.3 y se proporciona una explicación concisa de la función desempeñada por cada uno.

Tabla 3.3 Actores de la aplicación

Actor	Descripción
Administrador	Posee pleno control sobre la base de datos y es la única persona autorizada para eliminar contenido.
Colaborador	Tiene permisos para capturar información y acceder a la estructura de HITO, aunque de forma restringida. No tiene

Actor	Descripción
	acceso a la gestión de usuarios ni a las opciones de eliminación de datos.
Docente	Tiene la capacidad de visualizar la información general del Instituto y el contenido multimedia. Sin embargo, su acceso se limita únicamente a las consultas basadas en texto y en contenido, y también se le permite proponer contenido.
Usuario General	Únicamente tiene permisos para ver la información general del Instituto y tiene la capacidad de proponer contenido.
Usuario Anónimo	Tiene la capacidad de acceder y visualizar únicamente los datos generales del Instituto.

Ya que se definieron los actores en la Tabla 3.3, la Figura 3.3 presenta el diagrama de casos de uso, este establece la relación entre los actores y las actividades que realizan. Las Figuras 3.4 – 3.17 muestran los diagramas de actividades correspondientes a los casos de uso de la Figura 3.3.

El diagrama de actividad de la Figura 3.4 presenta el caso de uso *Iniciar Sesión*. La Como se observa en la Figura 3.4 el proceso comienza con la visualización de un formulario en pantalla, a través del cual el usuario ingresa sus credenciales de acceso. Luego, el sistema verifica y autentica las credenciales proporcionadas, iniciando así la sesión correspondiente y finalizando el caso de uso.

La Figura 3.5 presenta el diagrama de actividad correspondiente al caso de uso *Cerrar Sesión*. El diagrama ilustra que el caso de uso empieza con la visualización de un formulario en pantalla, donde el usuario elige la alternativa de cerrar sesión. A continuación, la aplicación procede a cerrar la sesión actual, finalizando de esta manera el caso de uso.

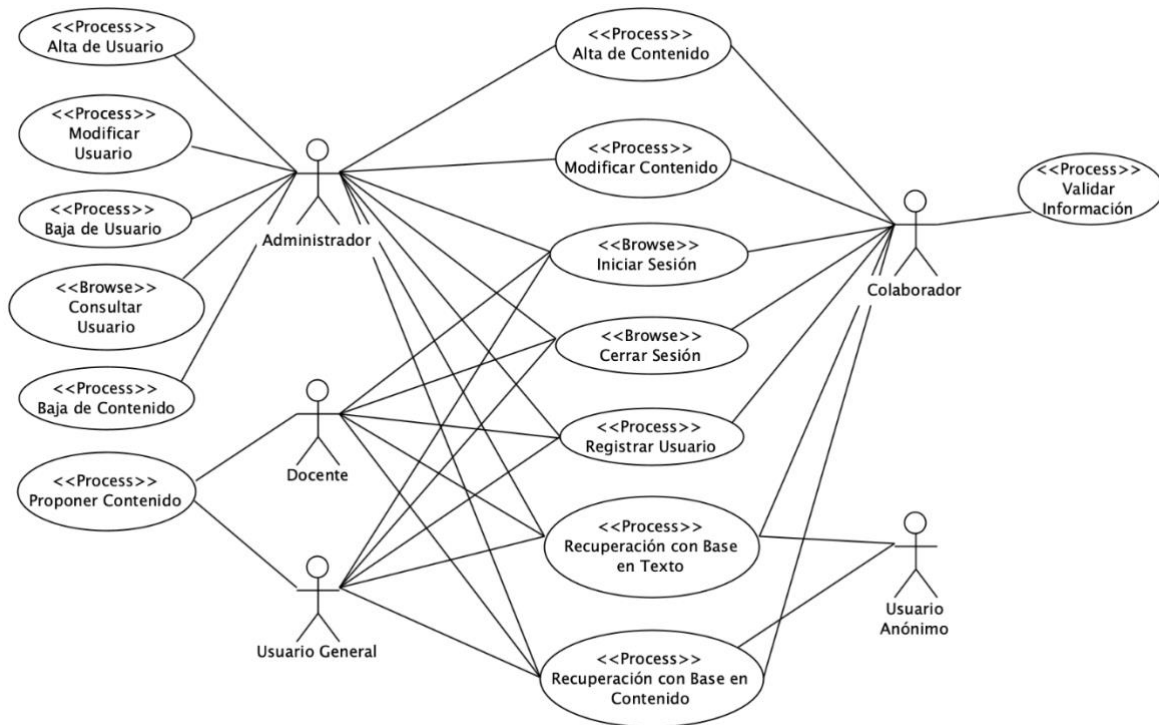


Figura 3.3 Diagrama de casos de uso

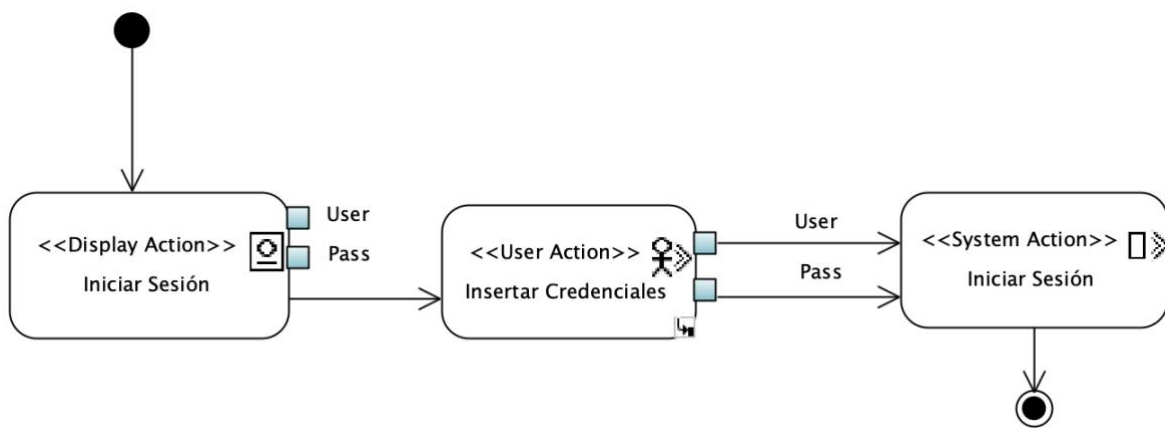


Figura 3.4 Diagrama del caso de uso Iniciar Sesión

La Figura 3.6 muestra el diagrama de actividad correspondiente al caso de uso *Registrar Usuario*. El diagrama representa que el caso de uso comienza presentando en pantalla un formulario donde el usuario autorizado ingresa los datos requeridos. A continuación, el sistema guarda la información del nuevo usuario y finaliza el caso de uso.

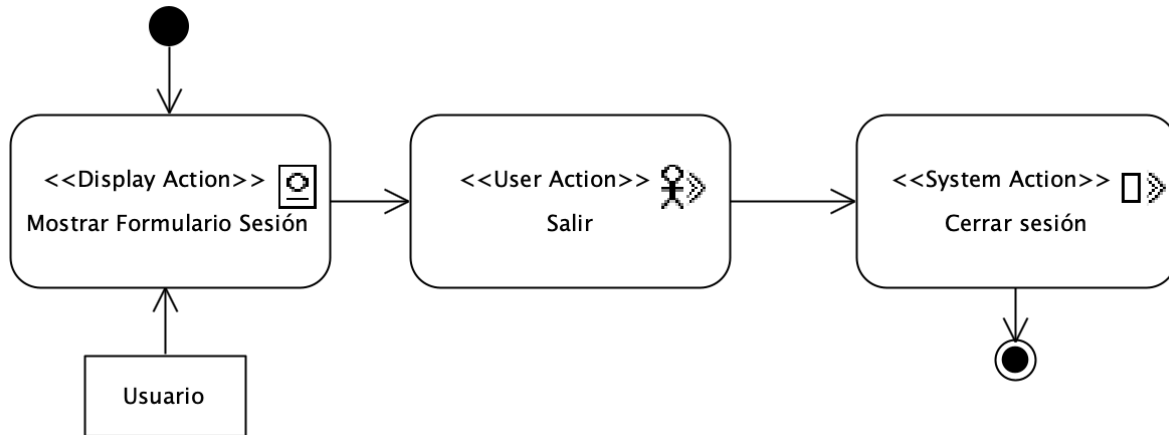


Figura 3.5 Diagrama del caso de uso Cerrar Sesión

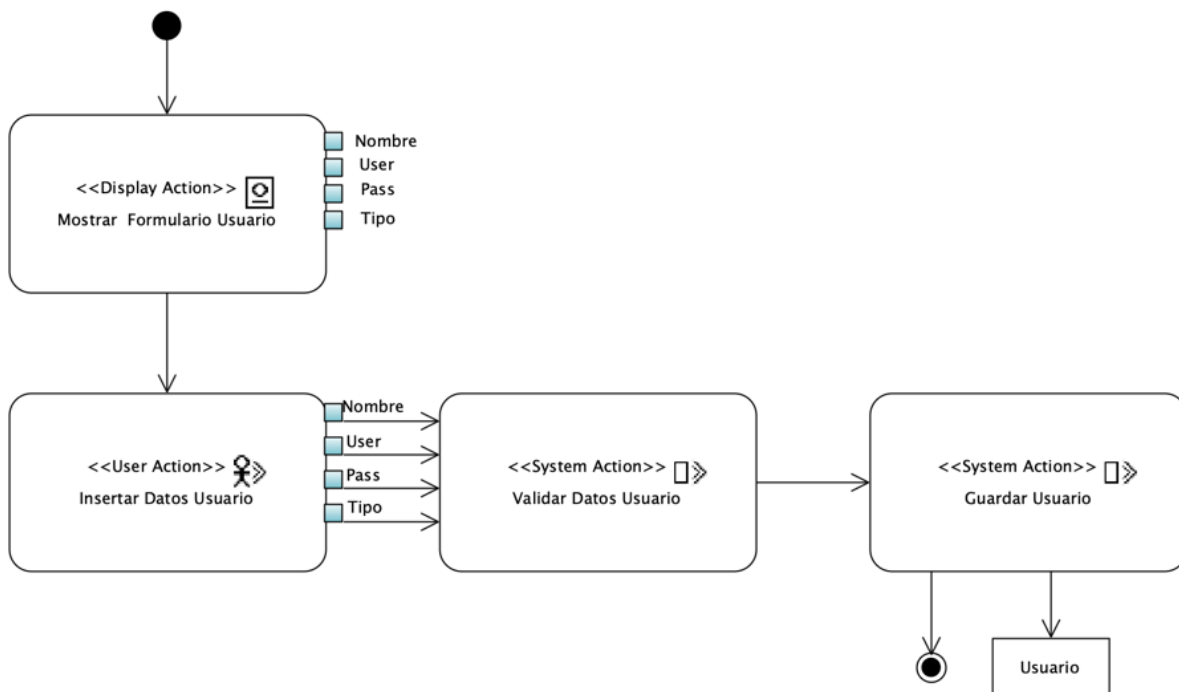


Figura 3.6 Diagrama del caso de uso Registrar Usuario

La Figura 3.7 muestra el diagrama de actividad correspondiente al caso de uso *Validar Información*. El caso de uso inicia presentando en pantalla una lista de contenidos propuestos en el sistema, junto con un formulario donde el usuario con autorización proporciona el identificador del contenido que desea validar. Luego, la aplicación regresa el contenido seleccionado y el usuario indica su intención de validar los datos. A continuación, la aplicación valida el contenido y concluye el caso de uso.

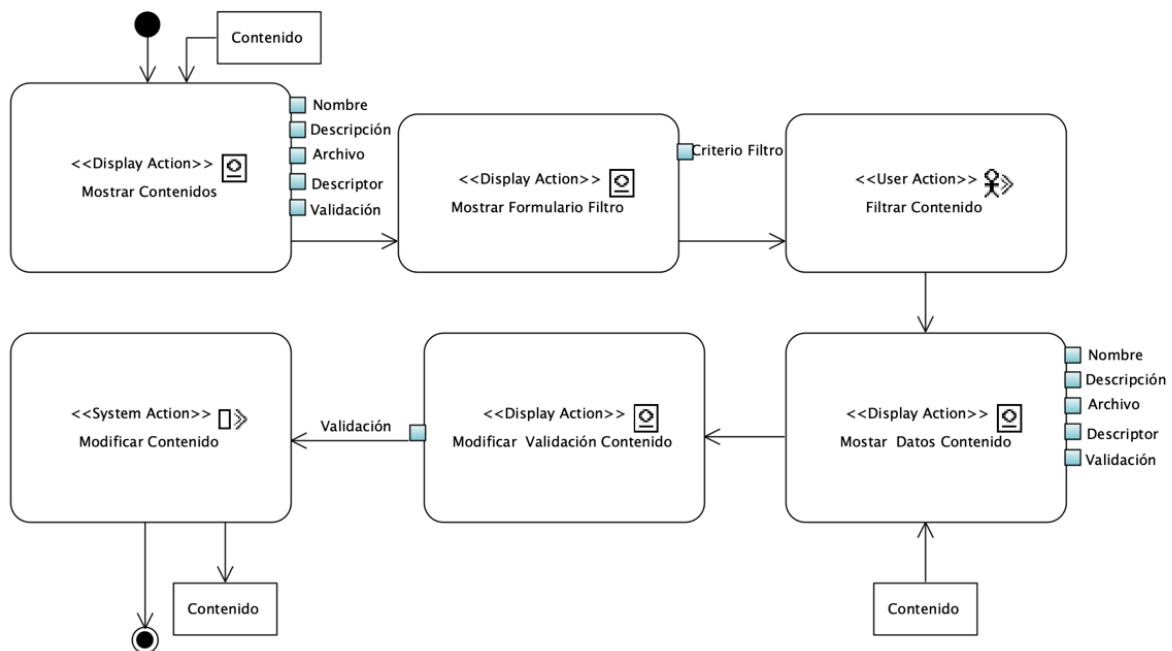


Figura 3.7 Diagrama del caso de uso Validar Información

La Figura 3.8 muestra el diagrama de actividad correspondiente al caso de uso *Proponer Contenido*. El caso de uso inicia con la visualización de un formulario en pantalla, a través del cual el usuario introduce el contenido y proporciona una descripción del mismo. Luego, el sistema valida y guarda la nueva propuesta de contenido. Posteriormente, la propuesta se valida por algún usuario colaborador o administrador. Una vez que el contenido se registra exitosamente en la aplicación, el caso de uso finaliza.

El diagrama de actividad de la Figura 3.9 corresponde al caso de uso *Alta de Contenido*. Como se observa en el diagrama el caso de uso comienza con la visualización de un formulario en pantalla, donde el usuario autorizado suministra los datos requeridos para el nuevo contenido. A continuación, el sistema guarda el nuevo contenido y finaliza el caso de uso.

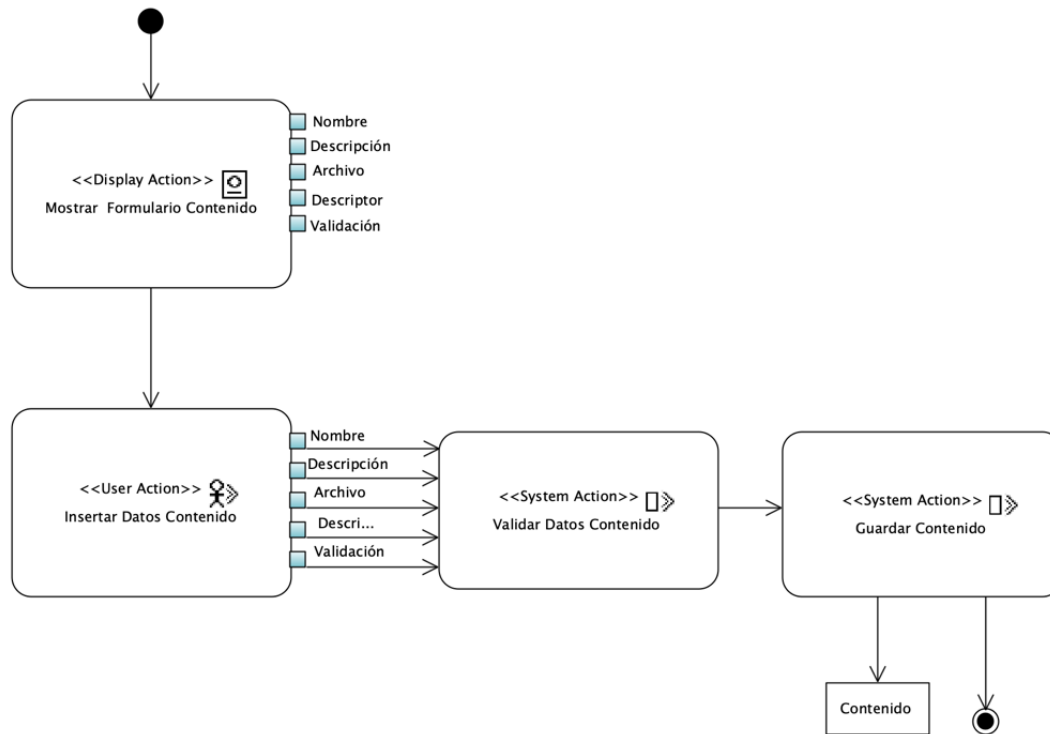


Figura 3.8 Diagrama del caso de uso Proponer Contenido

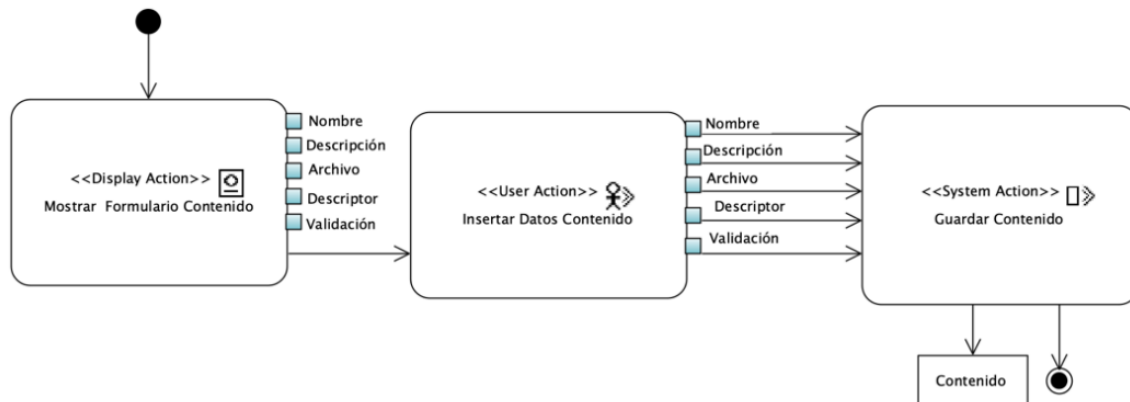


Figura 3.9 Diagrama del caso de uso Alta de Contenido

La Figura 3.10 muestra el diagrama de actividad correspondiente al caso de uso *Modificar Contenido*. El caso de uso inicia por mostrar en la pantalla una lista de los contenidos en el sistema, junto con un formulario donde el usuario con autorización introduce el identificador del contenido que quiere cambiar. Luego, el sistema muestra los datos del contenido seleccionado y el usuario realiza las modificaciones deseadas

en el contenido. El caso de uso concluye una vez que se ha realizado la edición del contenido.

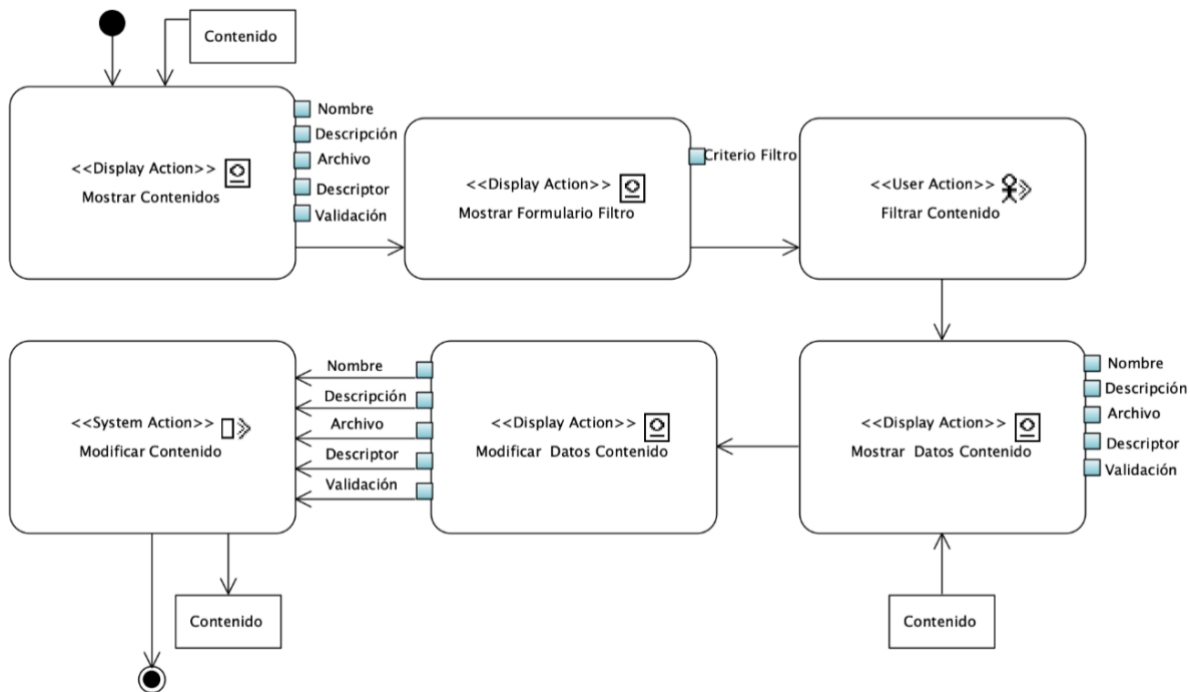


Figura 3.10 Diagrama del caso de uso Modificar Contenido

La Figura 3.11 presenta el diagrama de actividad correspondiente al caso de uso *Baja de Contenido*. El caso de uso empieza por desplegar en la pantalla una lista de los contenidos registrados en la aplicación, al mismo tiempo que presenta un formulario donde el usuario con autorización introduce el identificador del contenido que espera borrar. Luego, la aplicación proporciona el contenido seleccionado y el usuario elige la alternativa de eliminarlo. A continuación, la aplicación da de baja el contenido. El caso de uso concluye una vez que se ha eliminado exitosamente el contenido.

La Figura 3.12 presenta el diagrama de actividad correspondiente al caso de uso *Alta de Usuario*. Este diagrama muestra que el caso de uso comienza con la visualización de un formulario en pantalla, mediante este proceso, el usuario autorizado suministra la información necesaria. A continuación, el sistema guarda el nuevo usuario y finaliza el caso de uso.

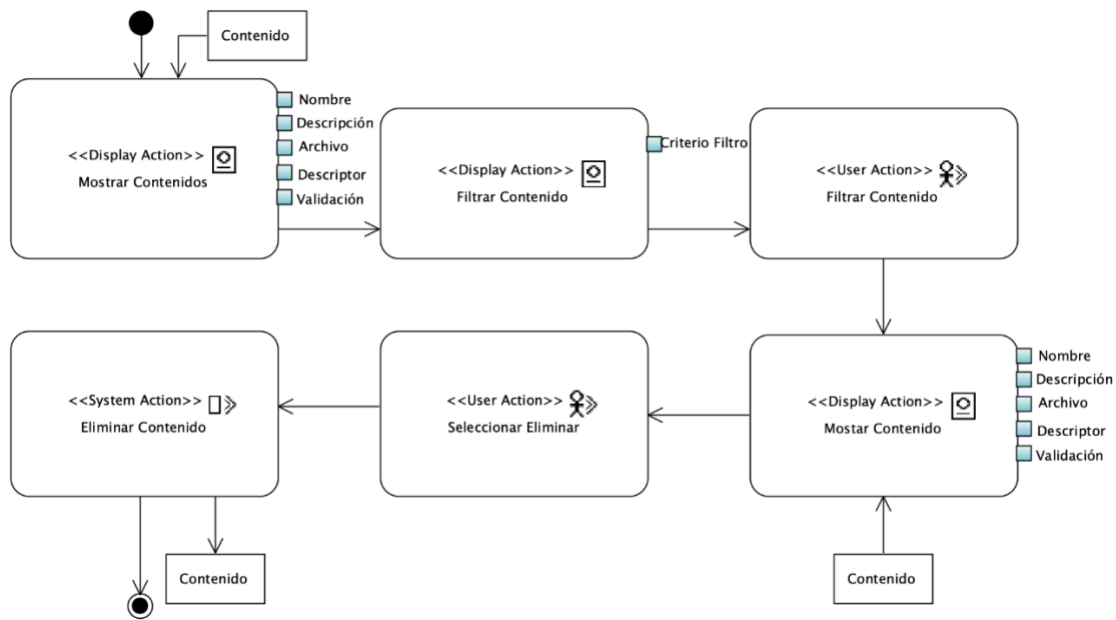


Figura 3.11 Diagrama del caso de uso Baja de Contenido

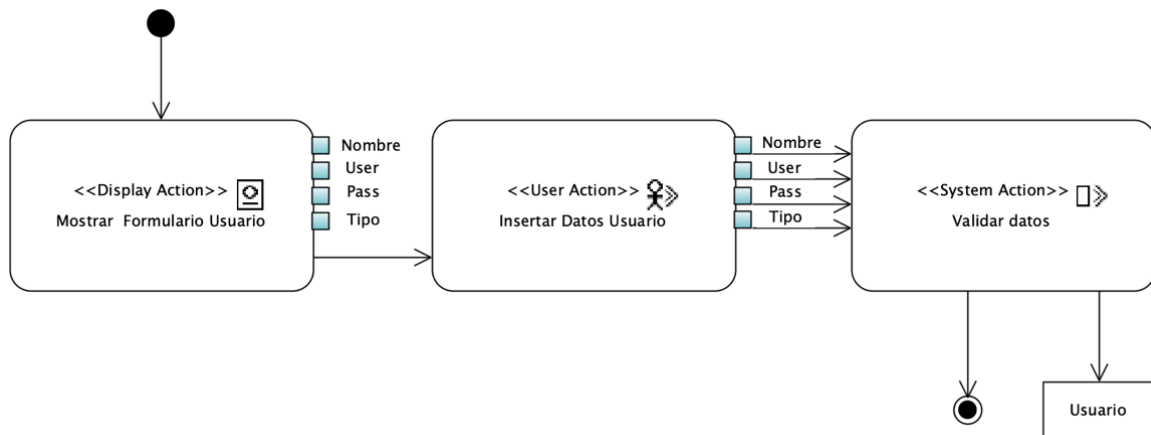


Figura 3.12 Diagrama del caso de uso Alta de Usuario

La Figura 3.13 presenta el diagrama de actividad correspondiente al caso de uso *Baja de Usuario*. Este caso de uso comienza al mostrar en la pantalla una lista de usuarios existentes del sistema, se proporciona un formulario al usuario autorizado para ingresar el identificador del usuario que desea dar de baja. Luego, el sistema proporciona los datos del usuario seleccionado y el usuario indica su intención de eliminar al usuario. A continuación, el sistema lleva a cabo la operación de baja de usuario. El caso de uso concluye una vez que se ha eliminado exitosamente al usuario.

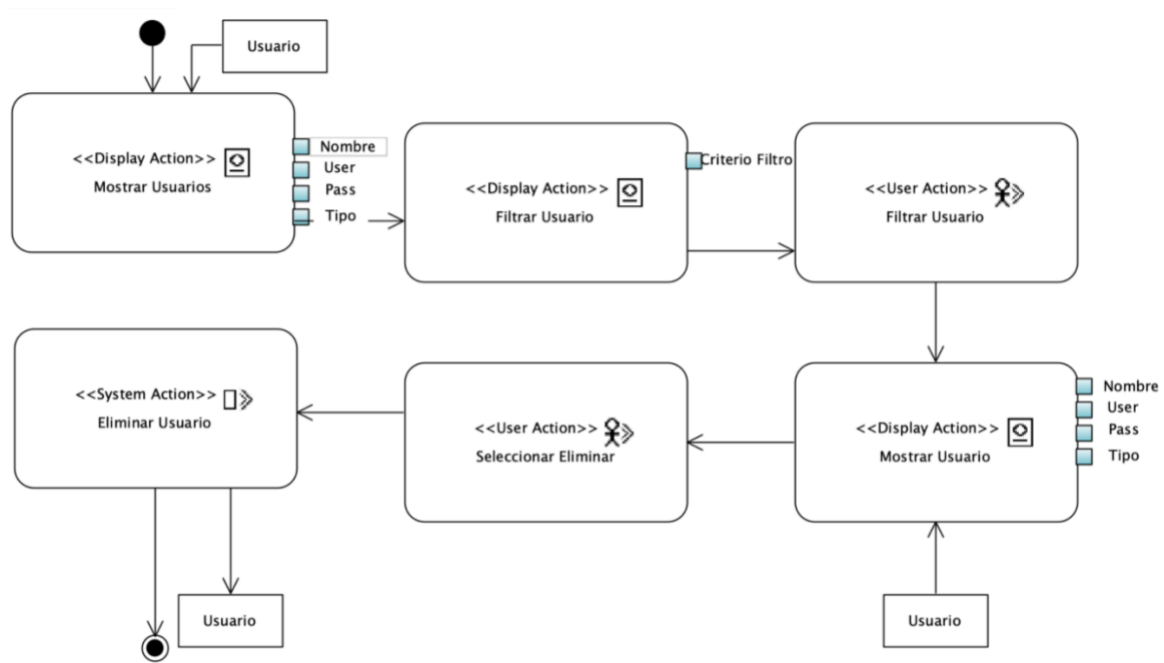


Figura 3.13 Diagrama del caso de uso Baja de Usuario

La Figura 3.14 muestra el diagrama de actividad correspondiente al caso de uso *Modificar Usuario*. Este caso de uso inicia presentando en la pantalla una lista de usuarios existentes del sistema, junto con un formulario donde el usuario autorizado ingresa el identificador del usuario que desea modificar. Luego, el sistema muestra los datos del usuario buscado. A continuación, el usuario realiza las modificaciones deseadas en los datos del usuario. El caso de uso concluye una vez que se ha realizado la actualización requerida.

El caso de uso *Consultar Usuario* que representa el diagrama de actividad ilustrado en la Figura 3.15, comienza al desplegar en la pantalla un listado de los usuarios existentes del sistema, junto con un formulario donde el usuario con autorización filtra los resultados. Para realizar el filtrado, el usuario ingresa el identificador del usuario buscado. Posteriormente, la aplicación responde mostrando la información correspondiente del usuario requerido. El caso de uso concluye una vez que se ha proporcionado la respuesta al usuario y se han mostrado los datos del usuario consultado.

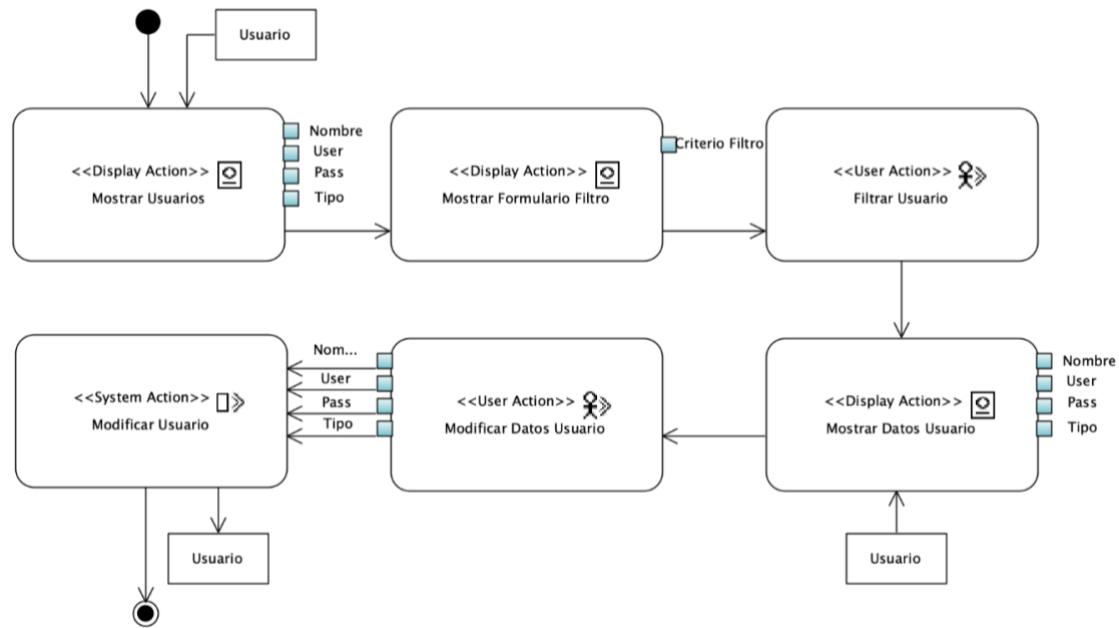


Figura 3.14 Diagrama del caso de uso Modificar Usuario

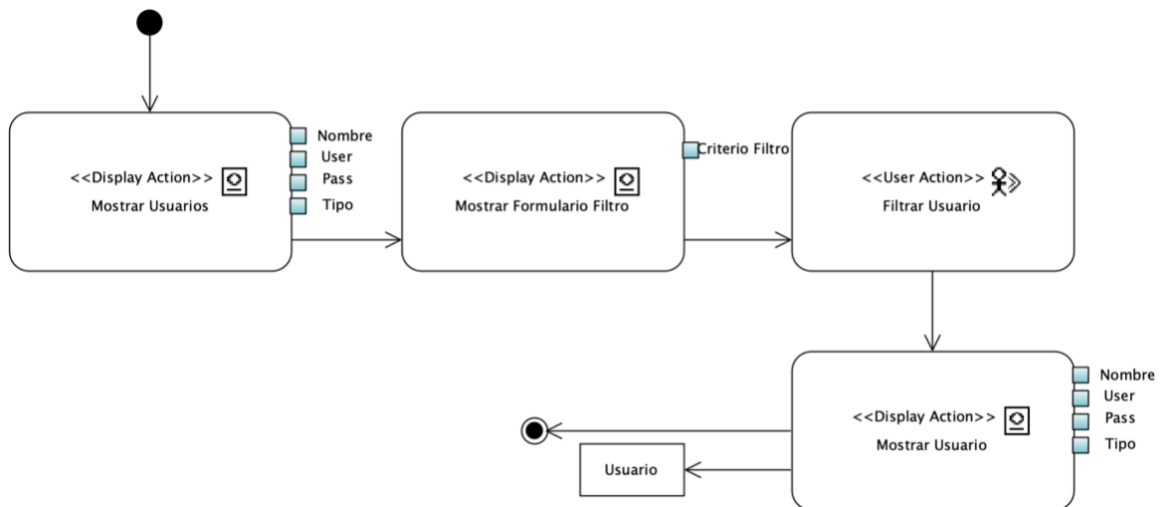


Figura 3.15 Diagrama del caso de uso Consultar Usuario

El caso de uso *Recuperación con Base en Texto* que ilustra el diagrama de actividad de la Figura 3.16, comienza al desplegar en pantalla un formulario donde el usuario con autorización ingresa los datos requeridos para realizar la búsqueda de contenidos. Luego, el sistema utiliza los criterios de búsqueda ingresados para buscar los contenidos correspondientes. Los contenidos encontrados se mostrarán como resultado de la búsqueda. Una vez que se presentan los contenidos obtenidos, el caso de uso concluye.

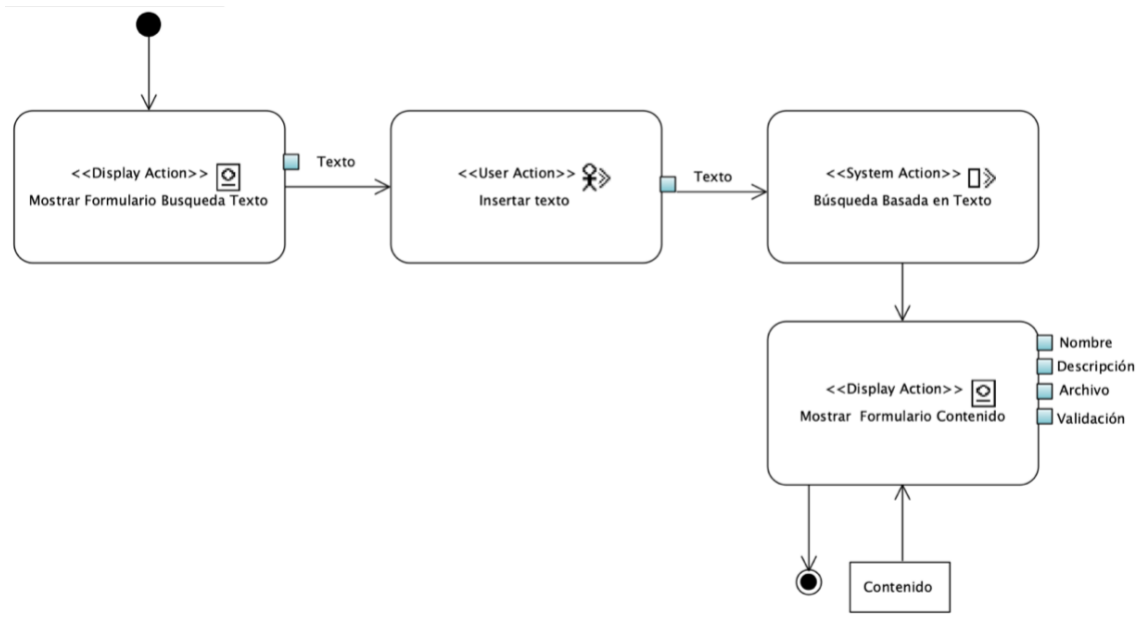


Figura 3.16 Diagrama del caso de uso Recuperación con Base en Texto

El caso de uso *Recuperación con Base en Contenido*, como se ilustra en el diagrama de actividad de la Figura 3.17, comienza desplegando en la pantalla un formulario donde el usuario con autorización ingresa los datos multimedia necesarios para realizar la consulta basada en contenido. Posteriormente, el sistema utiliza estos datos multimedia para llevar a cabo la consulta y encontrar los contenidos relevantes. Los contenidos encontrados se mostrarán como resultado de la consulta. Una vez que se visualizan los contenidos encontrados el caso de uso concluye.

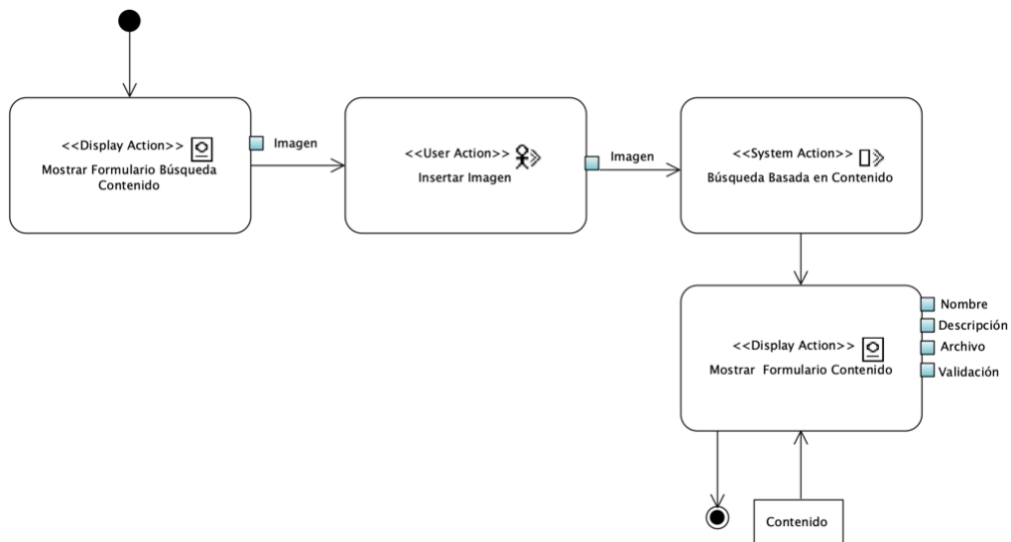


Figura 3.17 Diagrama del caso de uso Recuperación con Base en Contenido

3.3.2 Modelo Conceptual

El modelo conceptual, que también se conoce como modelo de dominio, proporciona una representación visual de cómo se relacionan los requisitos de la aplicación y muestra el comportamiento del sistema. En este trabajo, el modelo conceptual se representa mediante tres diagramas principales: el diagrama conceptual, el diagrama lógico y el diagrama físico de la base de datos. Estos diagramas se muestran en las Figuras 3.18, 3.19 y 3.20.

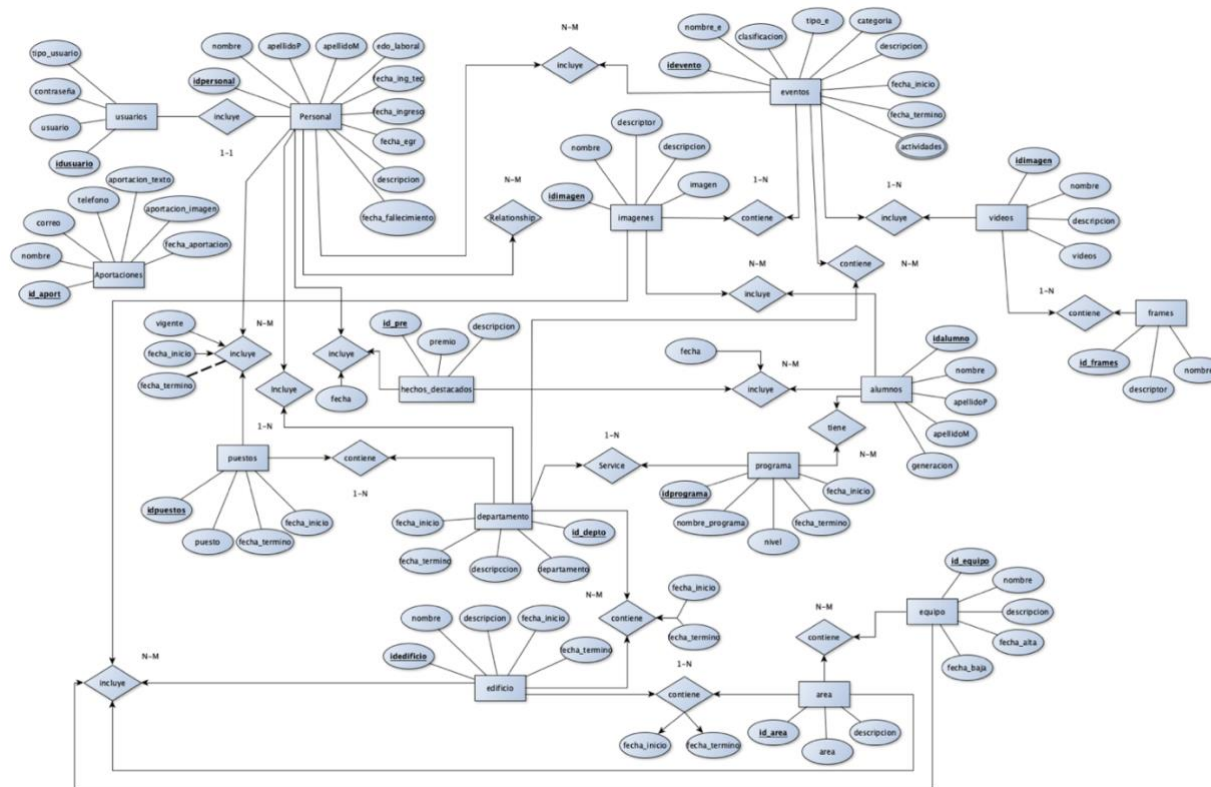


Figura 3.18 Diagrama conceptual

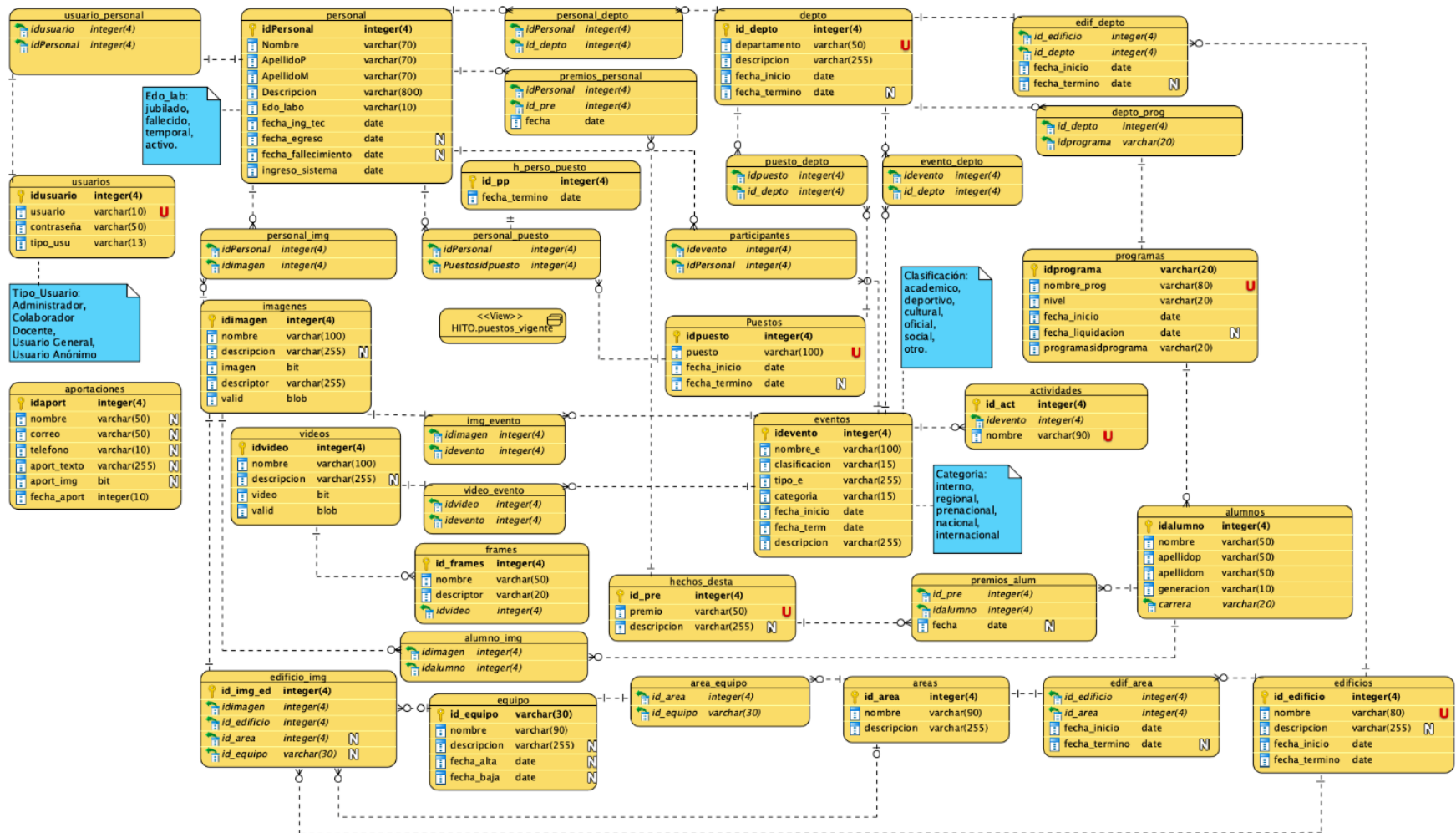


Figura 3.19 Diagrama Lógico

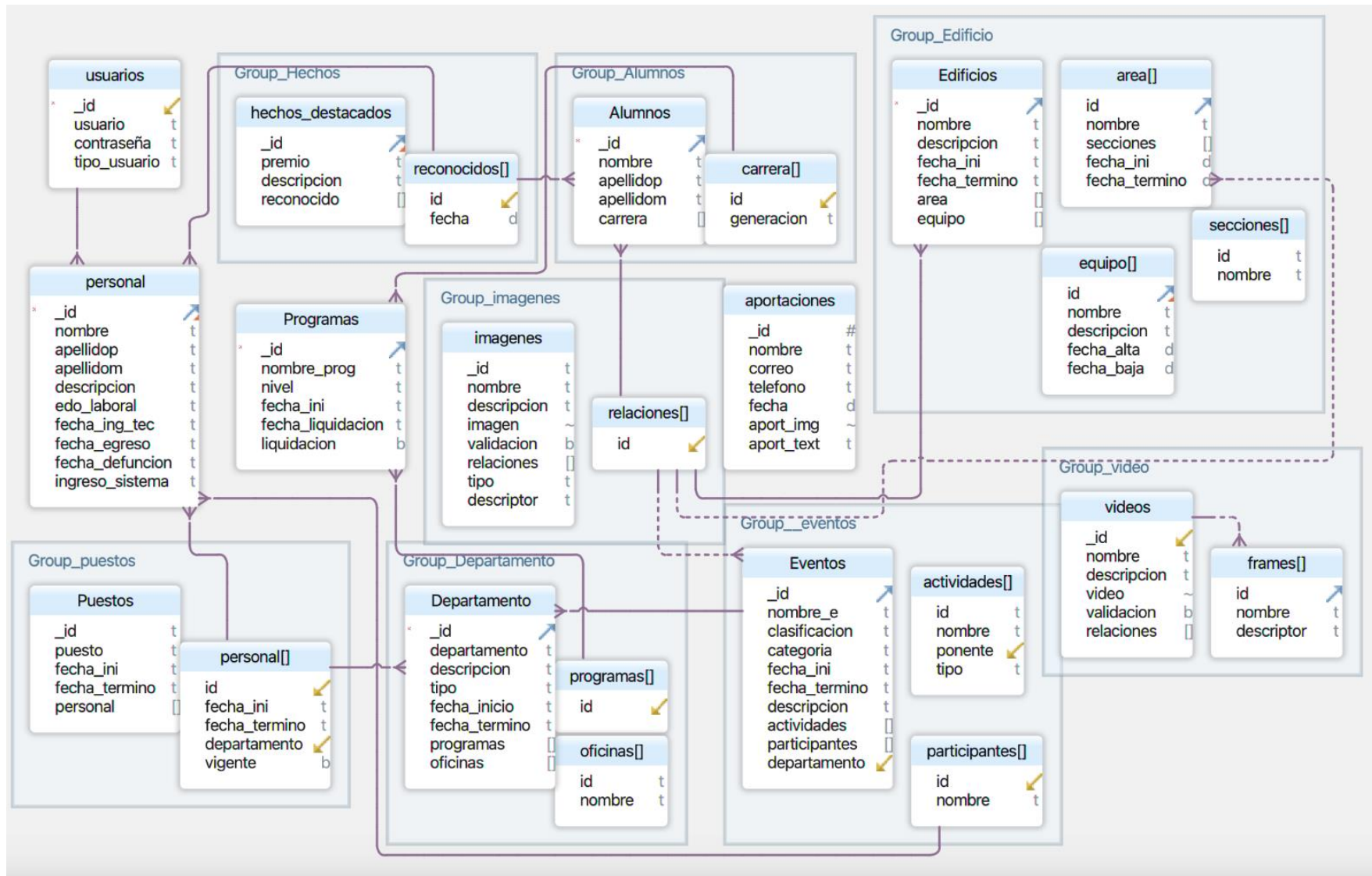


Figura 3.20 Diagrama Físico

Tablas de la base de datos

Usuarios: la tabla de usuarios almacena los datos de los usuarios registrados en el sistema, lo cual es necesario para su identificación en el sistema.

Personal: la tabla de personal almacena la información del personal del Instituto, permitiendo conocer fechas de ingreso y egreso para tener información actualizada sobre el personal.

Grupo Puestos

Puestos, Personal[]: establece la relación entre los puestos y la información del personal del Instituto, permitiendo su clasificación y gestión.

Grupo Hechos

Hechos Destacados, Reconocidos[]: guarda los datos relacionados con los hechos relevantes dentro del Instituto.

Programas: registra los nombres de los programas educativos impartidos en el Instituto.

Grupo departamento

Departamento, Programas, Oficinas: almacena la información de los departamentos y oficinas del Instituto, estableciendo relaciones entre ellos.

Grupo Alumnos

Alumno, Carrera: almacena los datos de los alumnos del Instituto, así como su relación con las carreras a las que pertenecen.

Grupo Imágenes

Imágenes, Relaciones: contiene información multimedia histórica del Instituto Tecnológico de Orizaba, específicamente imágenes. Se aplicó un método de fragmentación horizontal en esta tabla para optimizar las consultas basadas en contenido.

Aportaciones: guarda los datos correspondientes a las contribuciones realizadas al Instituto Tecnológico de Orizaba.

Grupo Eventos:

Eventos, Actividades[], Participantes[]: contiene la información de los eventos realizados por el Instituto, así como su relación con los participantes y actividades asociadas.

Grupo Edificios

Edificios, Área, Secciones, Equipo: almacena los datos de los edificios del Instituto, como también su relación con las áreas, secciones y el equipo presente en ellos.

Grupo videos

Videos, Frames: contiene información sobre los videos del instituto, así como su relación con los *frames* y sus descriptores.

3.3.3 Modelo de Navegación

El modelo de navegación describe todos los posibles caminos de navegación dentro de la aplicación para los usuarios. Este modelo utiliza diagramas de clase estereotipados para crear mapas de navegación, que representan todas las rutas disponibles para un usuario dentro de un sitio o aplicación. En las Figuras 3.21, 3.22, 3.23 y 3.24 se presentan los diferentes modelos de navegación de la aplicación para los usuarios.

Desde la página de inicio, los usuarios tienen la opción de registrarse como usuarios o iniciar sesión desde el menú principal. Si eligen la segunda opción, existen hasta cinco posibles caminos a seguir, dependiendo del tipo de usuario. La primera opción permite a cualquier usuario registrado navegar hacia la página de consulta de contenido y realizar búsquedas en el sistema. La segunda alternativa está disponible solo para usuarios autorizados, quienes son capaces de proponer contenido al llenar un formulario con los datos necesarios. La tercera opción proporciona al usuario la funcionalidad para cerrar sesión en la aplicación. La cuarta alternativa brinda a los

usuarios autorizados el acceso a la página de usuarios, donde son capaces de agregar nuevos usuarios, seleccionar uno específico y realizar diversas acciones relacionadas con los usuarios, como ver sus datos, eliminarlos o modificarlos.

Por último, la quinta opción otorga al usuario autorizado la capacidad de navegar hacia la página de contenidos, donde tiene la elección de agregar nuevo contenido, seleccionar uno específico y realizar acciones como ver los datos del contenido deseado, eliminarlo o modificarlo.

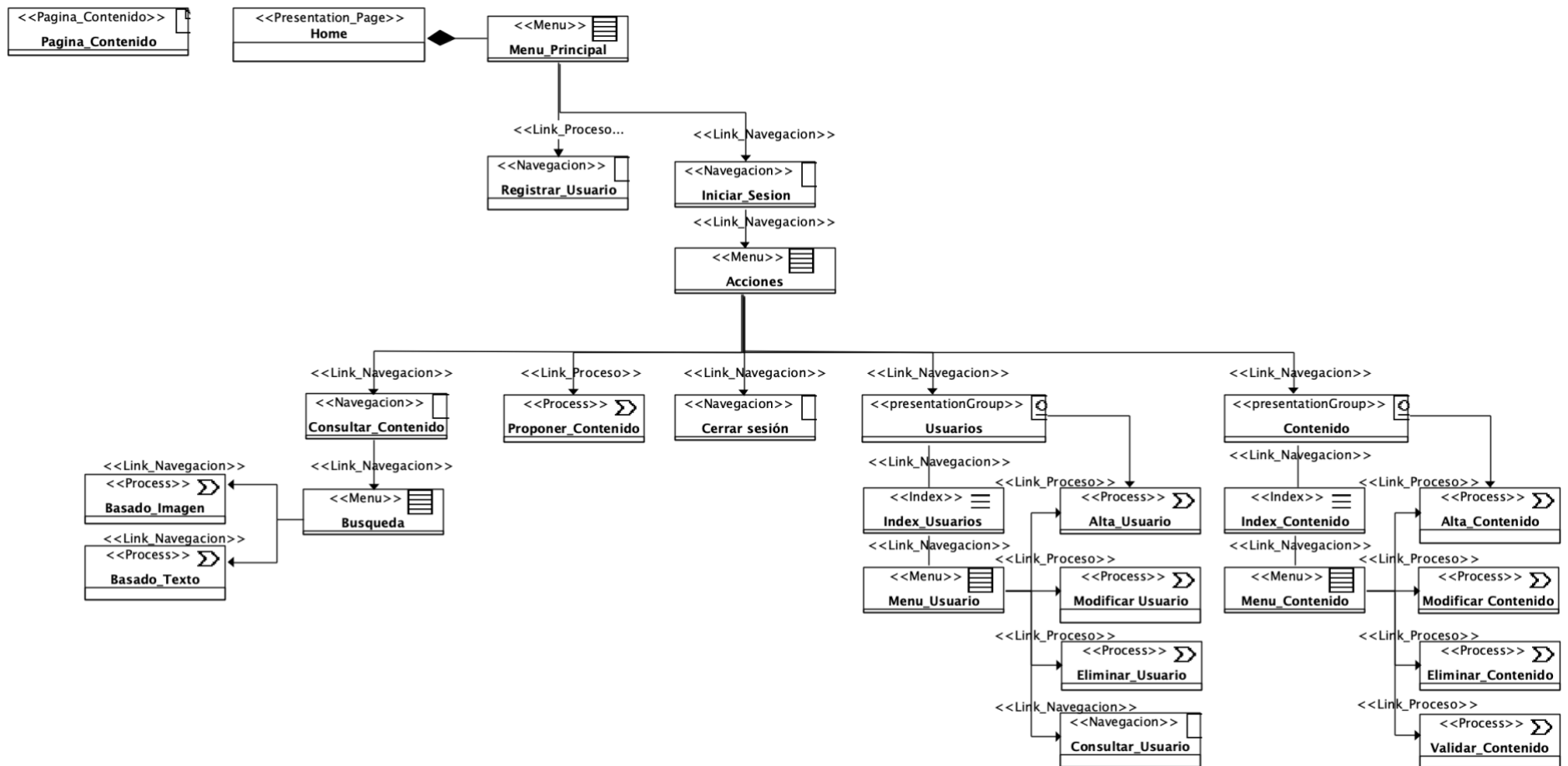


Figura 3.21 Diagrama de navegación del Administrador

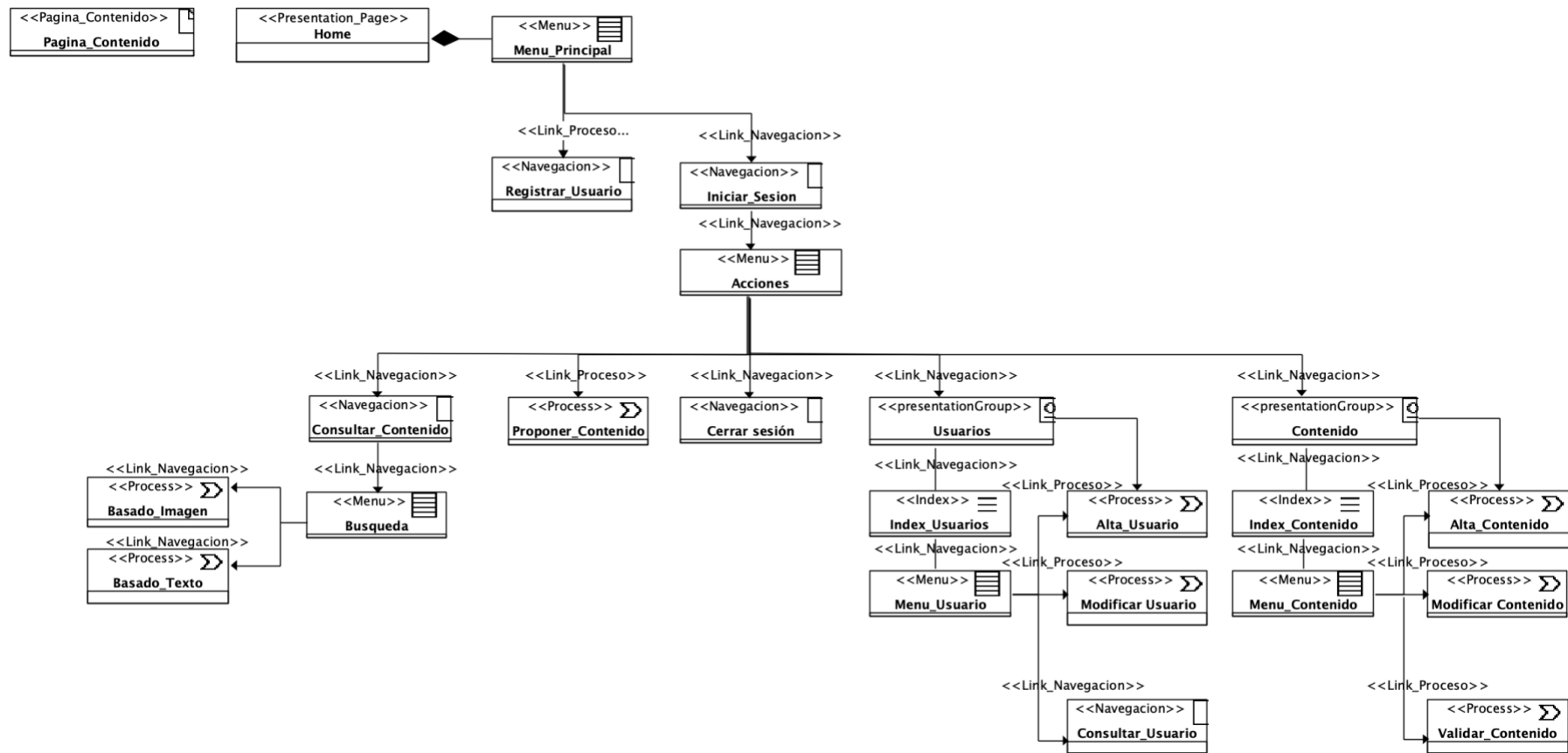


Figura 3.22 Diagrama de navegación del Colaborador

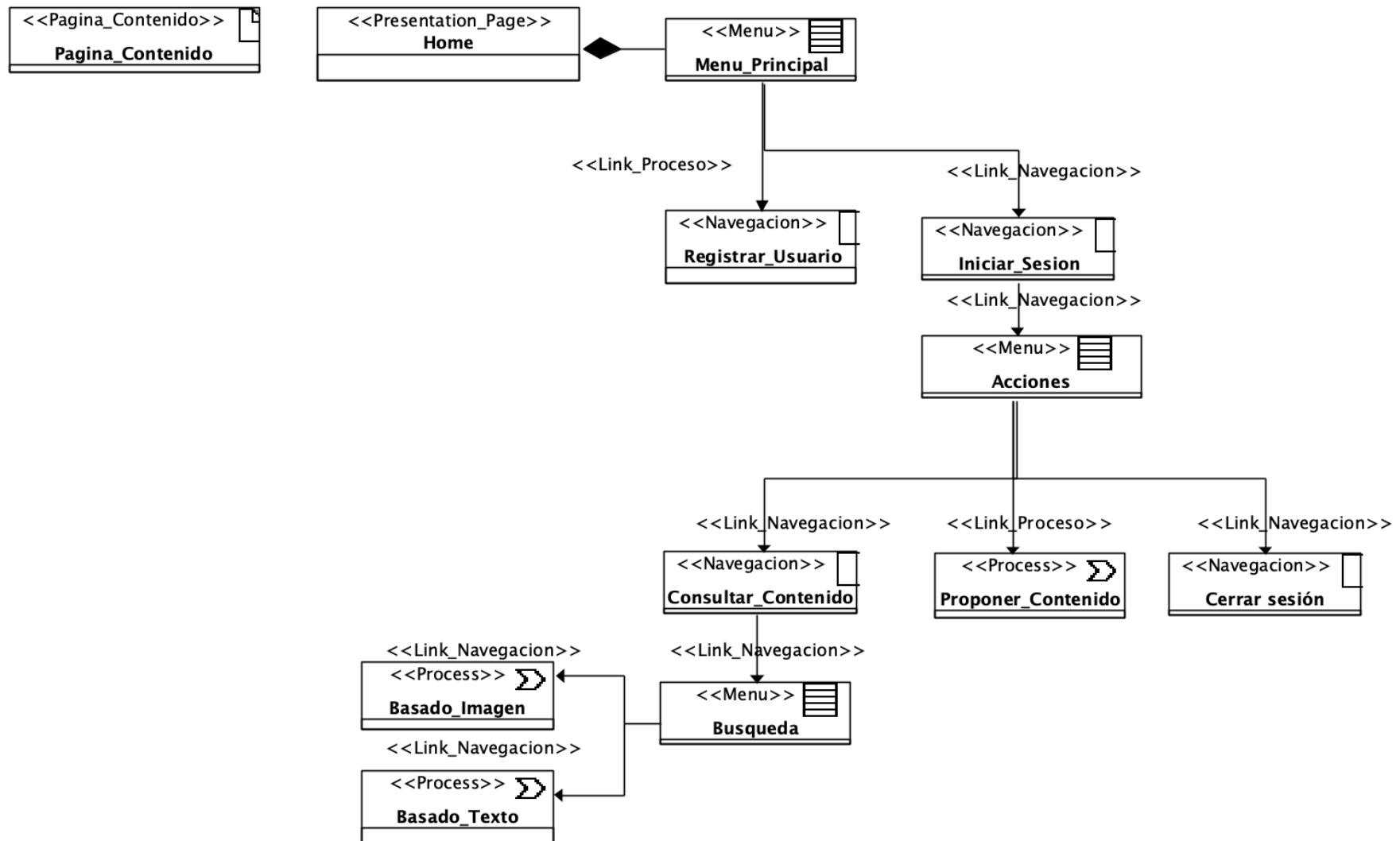


Figura 3.23 Diagrama de navegación del Docente/ Usuario General

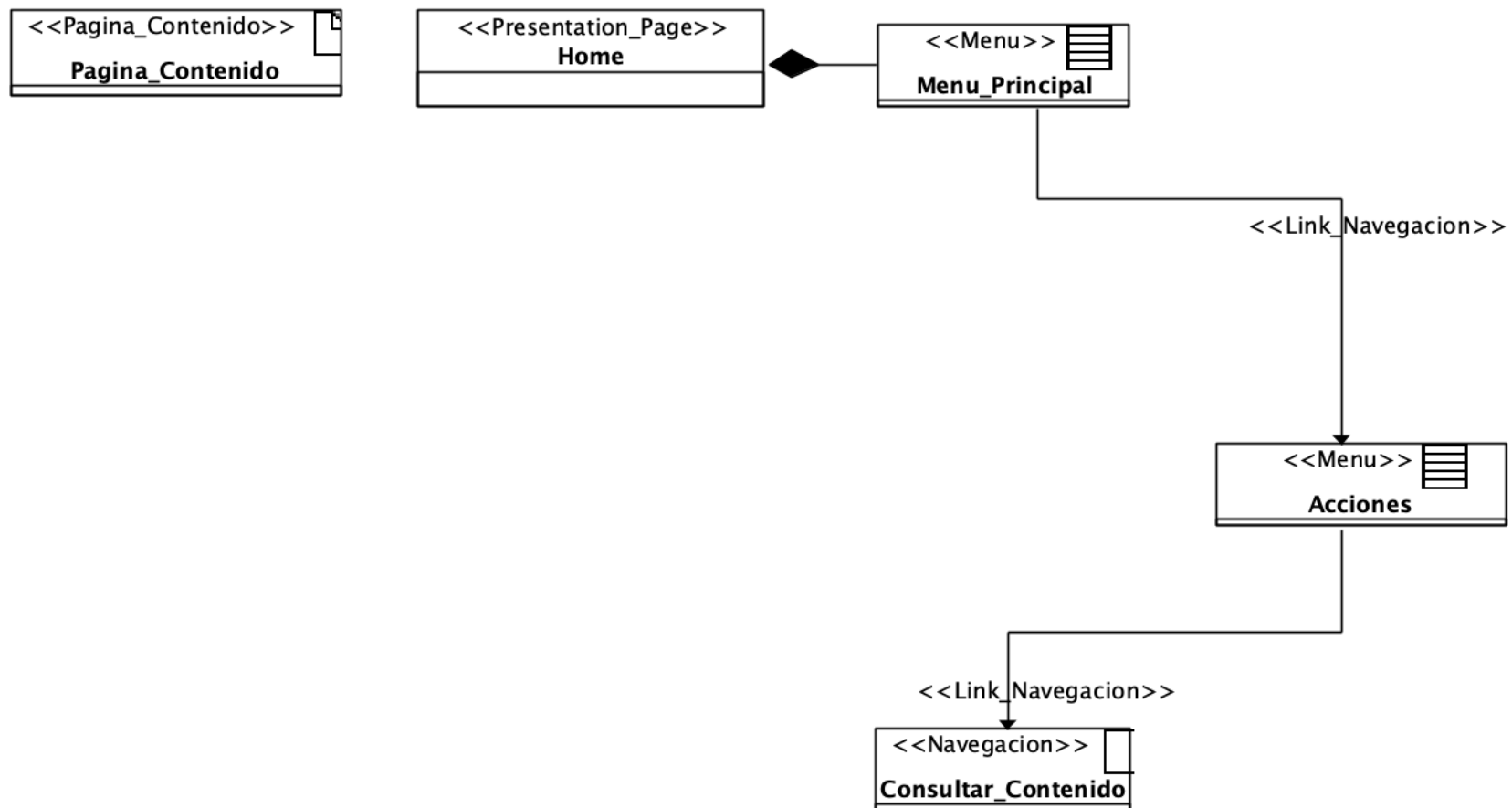


Figura 3.24 Diagrama de navegación del usuario Anónimo

3.3.4 Modelo de presentación

El modelo de presentación establece las páginas del sistema, su estructura y sus componentes asociados. El sistema consta de seis páginas, las cuales se describen en seguida. La Figura 3.25 muestra la página *Inicio*, la cual da la bienvenida a los usuarios que ingresan a la aplicación, esta página muestra información sobre el Tecnológico Nacional de México, Campus Orizaba. Además, se proporciona un enlace para iniciar sesión en la aplicación y acceder al contenido disponible.

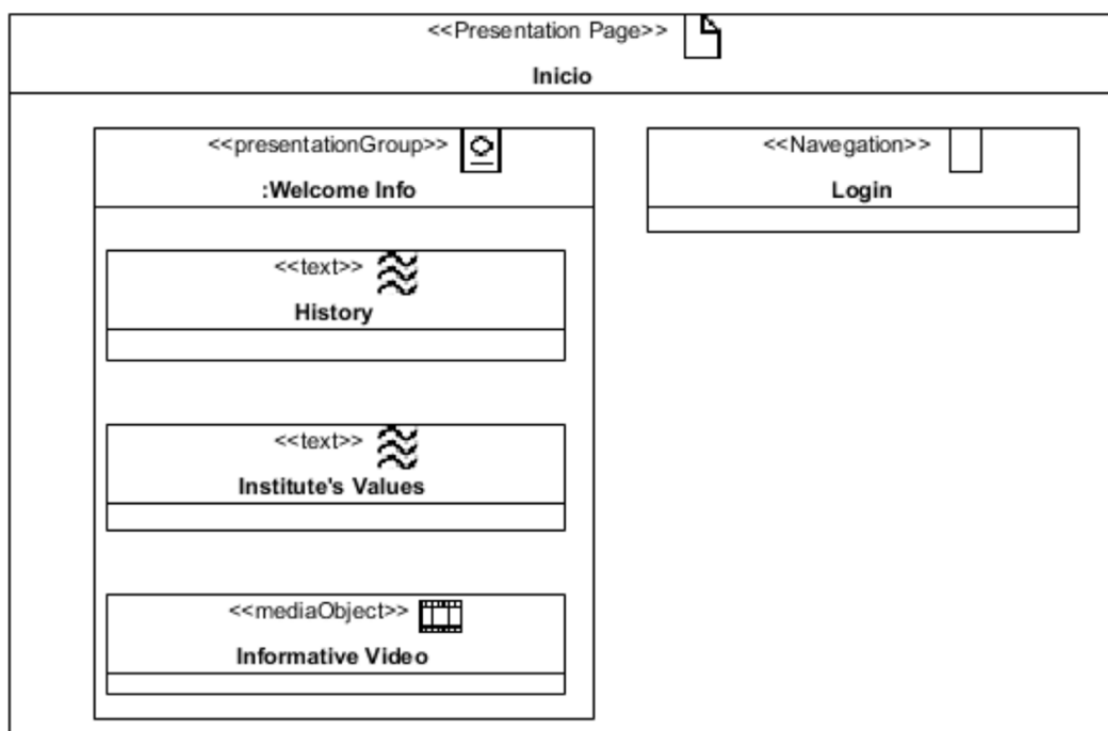


Figura 3.25 Página Inicio del modelo de presentación

La página *Iniciar Sesión*, como se muestra en la Figura 3.26, presenta un formulario donde se deben ingresar las credenciales requeridas para acceder al sistema, se encuentra disponible la opción de crear una cuenta en la parte inferior del formulario, la cual actualiza dicho formulario para permitir el registro de los datos de un nuevo usuario.

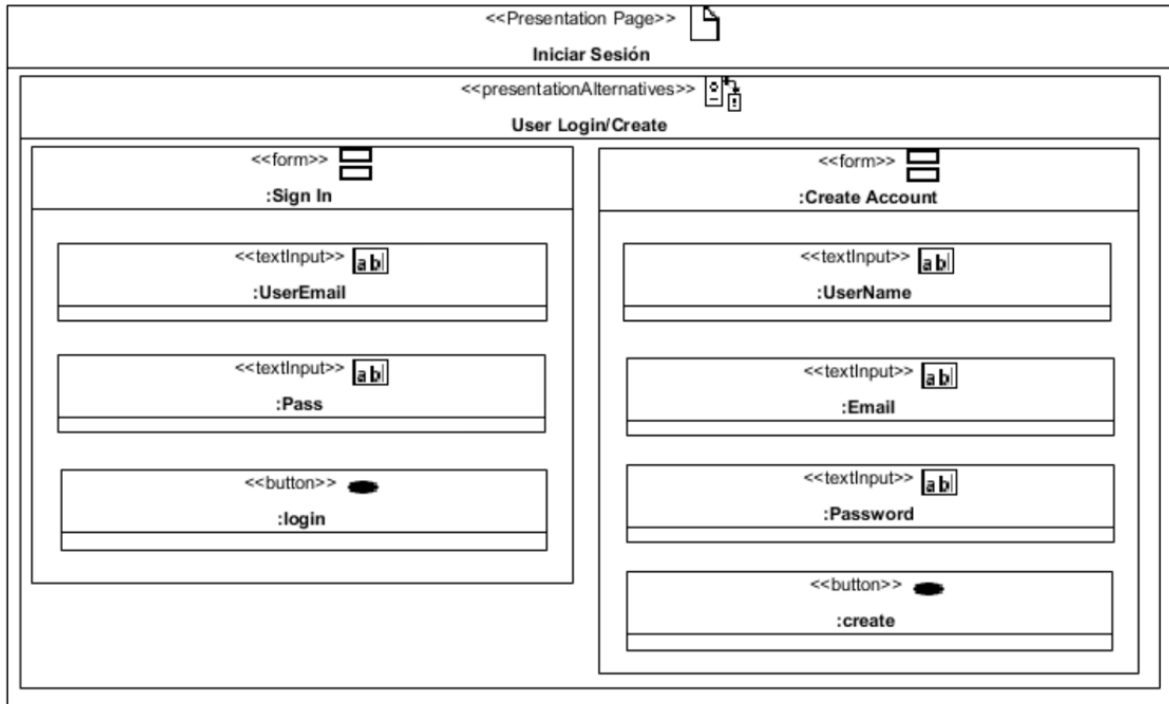


Figura 3.26 Página Iniciar Sesión del modelo de presentación

En la página de *Contenidos*, la primera pestaña de izquierda a derecha, como se observa en la Figura 3.27, proporciona datos acerca de la historia del Instituto Tecnológico de Orizaba. En las pestañas siguientes, se presenta información detallada sobre distintos tipos de contenido, como edificios, eventos, equipamiento, personal y videos. En la parte superior de la página se muestra un menú que varía según el tipo de usuario y contiene opciones para cerrar sesión, gestionar usuarios o gestionar un tipo específico de contenido. Además, se encuentra disponible la opción de proponer nuevo contenido a través de un botón que redirige al usuario a la página *Proponer Contenido*, como se muestra en la Figura 3.28, en esta página se presenta un formulario donde se capturan los datos necesarios para proponer un nuevo contenido. También se encuentra disponible la opción de consultar con base en contenido o texto a través de un botón que redirige al usuario como se muestra en la Figura 3.29, en esta página se presenta un formulario donde se captura la información requerida para llevar a cabo el tipo de consulta especificado.

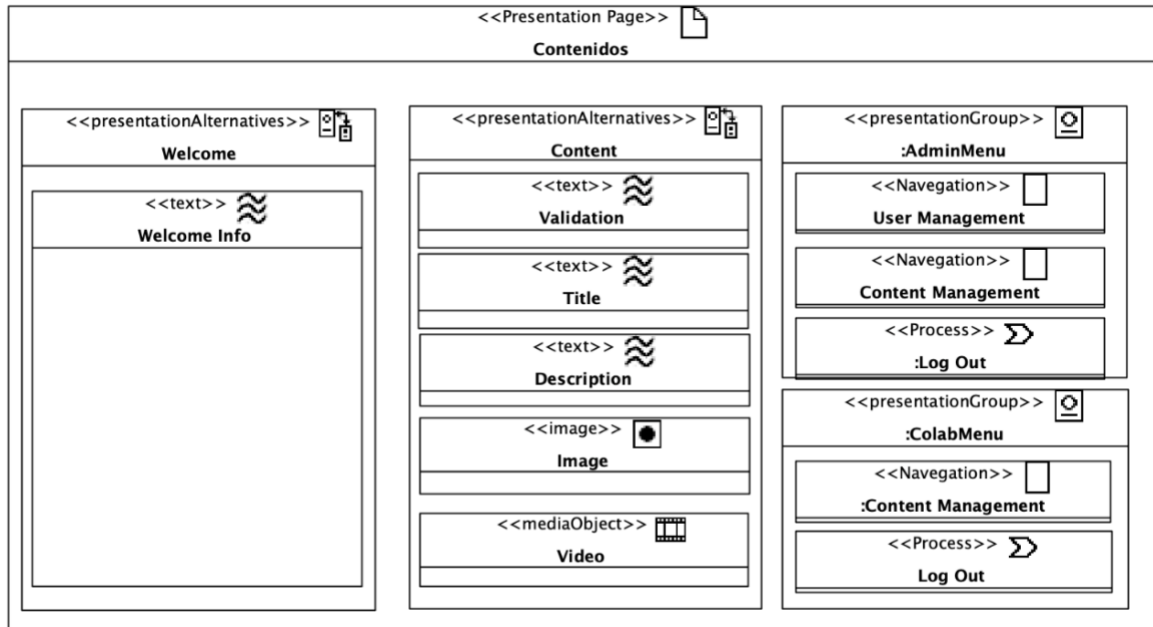


Figura 3.27 Página de Contenidos del modelo de presentación

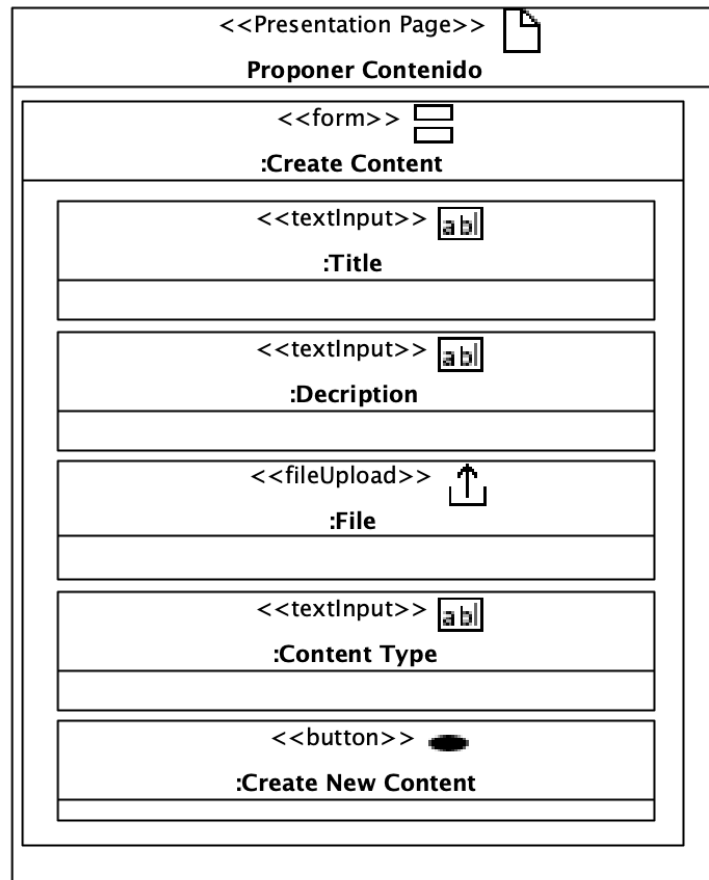


Figura 3.28 Página de Proponer Contenido del modelo de presentación

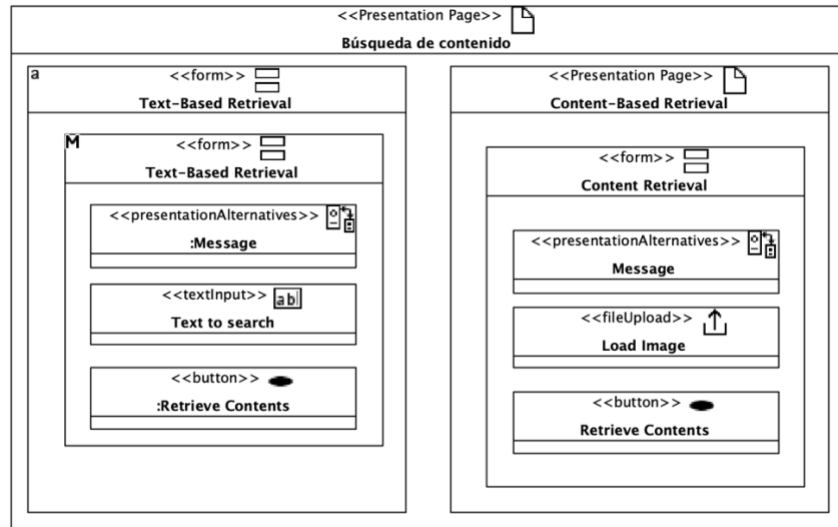


Figura 3.29 Página de Búsqueda de Contenido del modelo de presentación

La Figura 3.30 muestra la página de *Gestionar Usuarios* que permite al administrador gestionar las cuentas de los usuarios dados de alta en la aplicación. En esta página se presenta una tabla que contiene la información de los usuarios, y se proporcionan las opciones para modificar y dar de baja asociadas a cada uno de ellos. Además, se cuenta con un filtro de búsqueda que permite buscar un usuario específico usando su nombre de usuario o dirección de correo electrónico. La opción de crear un nuevo usuario también está disponible a través de un botón, el cual, al ser seleccionado, muestra un formulario emergente donde se ingresan los datos necesarios para realizar el registro de un nuevo usuario.

La Figura 3.31 presenta la página de *Gestionar Contenido*, donde se administra la gestión del contenido de diferentes tipos de imágenes, como edificios, equipamiento, eventos, personal y videos. En esta página se observa una tabla con los datos de las imágenes o videos almacenados, junto con las opciones de actualizar y eliminar asociadas a cada uno de ellos. Además, se proporciona un filtro de búsqueda que permite buscar un contenido específico utilizando su nombre. La opción de crear un nuevo contenido también está disponible a través de un botón, el cual, al ser seleccionado, muestra un formulario emergente para ingresar la información necesaria para realizar el registro de un nuevo contenido.

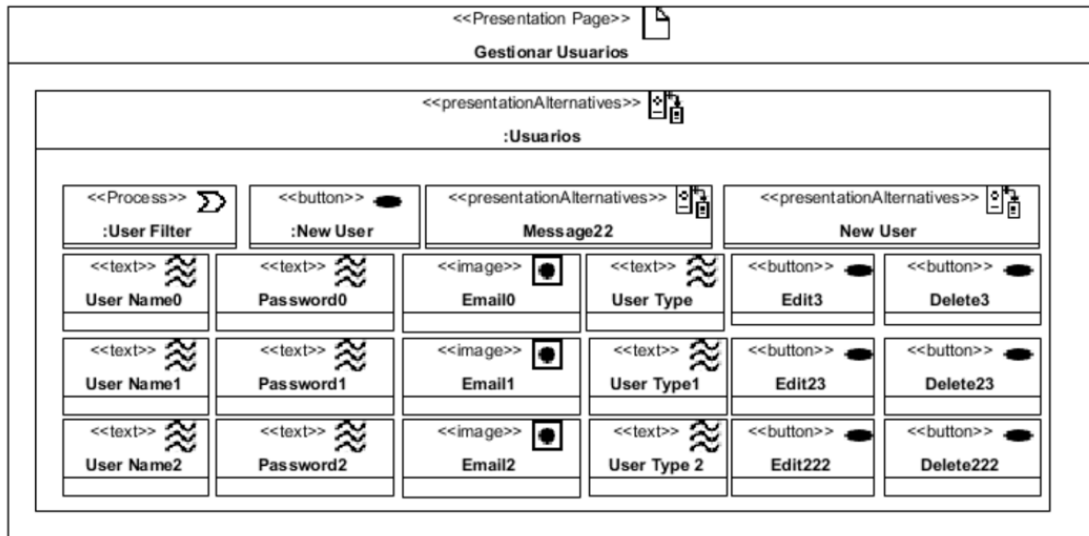


Figura 3.30 Página de Gestionar Usuarios del modelo de presentación

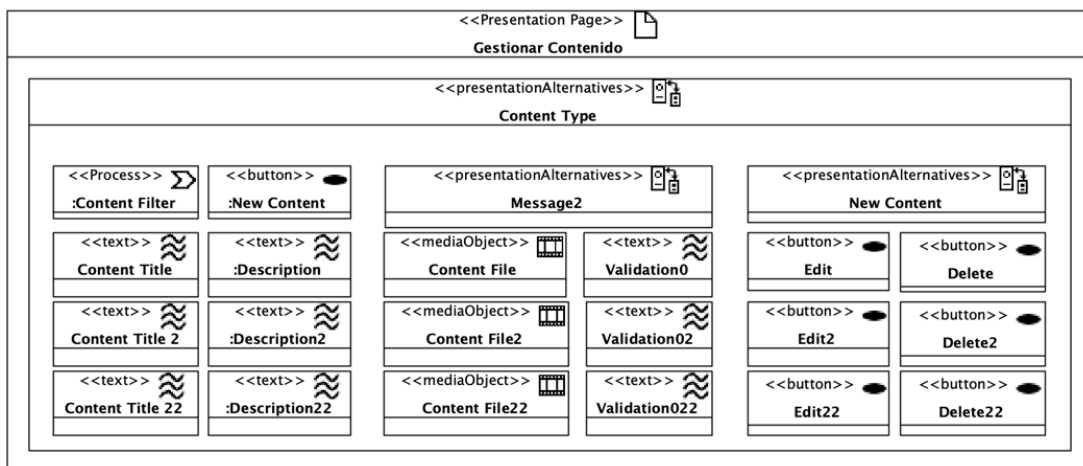


Figura 3.31 Página de Gestionar Contenido del modelo de presentación

3.3.5 Modelo de procesos

El diagrama de procesos se fundamenta en el diagrama de actividades y agrega varios controles. La Figura 3.32 muestra el diagrama de proceso del caso de uso *Iniciar Sesión* que incluye varios componentes como nodos de decisión y la representación de objetos dentro del flujo de actividades. Dicho proceso comienza presentando en pantalla un formulario donde el usuario otorga sus datos de acceso. Luego, la aplicación verifica las credenciales y dependiendo de su validez, inicia la sesión correspondiente o muestra un mensaje de error. Esas acciones concluyen el caso de uso de *Iniciar Sesión*.

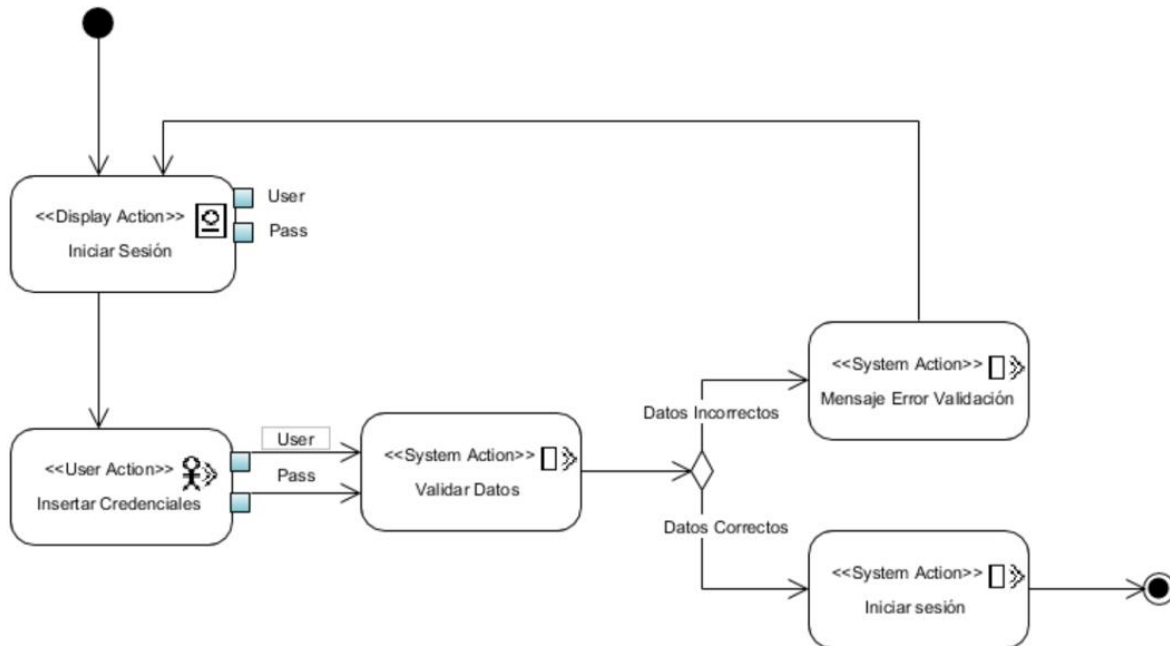


Figura 3.32 Diagrama de proceso del caso de uso *Iniciar Sesión*

La Figura 3.33 presenta el diagrama de proceso del caso de uso *Cerrar Sesión*. Este diagrama muestra que el proceso comienza exhibiendo en pantalla un formulario donde el usuario elige la alternativa para finalizar sesión. Una vez seleccionada, se presenta un breve mensaje emergente para confirmar la decisión. Luego, la aplicación procede a cerrar la sesión actual y finaliza el caso de uso *Cerrar Sesión*.

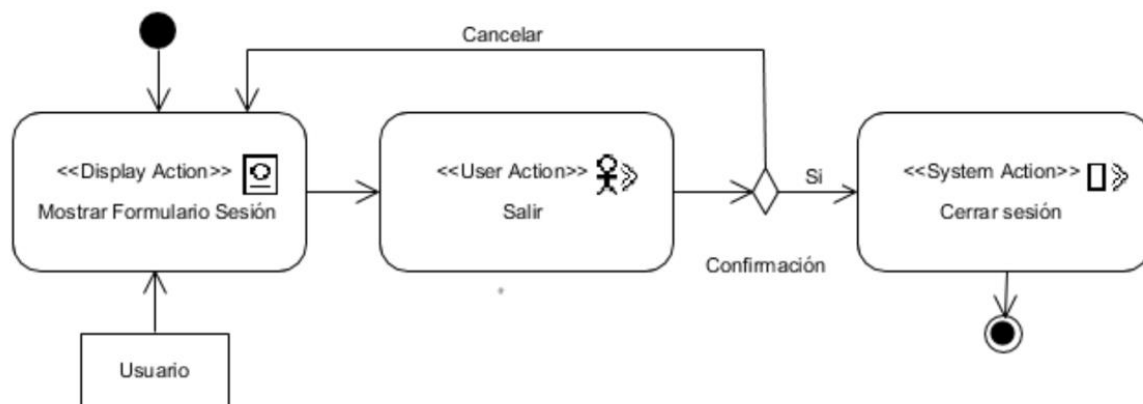


Figura 3.33 Diagrama de proceso del caso de uso *Cerrar Sesión*

La Figura 3.34 representa el diagrama de proceso para el caso de uso *Registrar Usuario*. Este diagrama muestra que el proceso comienza desplegando en pantalla un

formulario donde el usuario con autorización ingresa los datos requeridos. A continuación, la aplicación realiza una validación de los datos introducidos. Si los datos son correctos, la aplicación registra el nuevo usuario y finaliza el caso de uso, registrando exitosamente al nuevo usuario en el sistema.

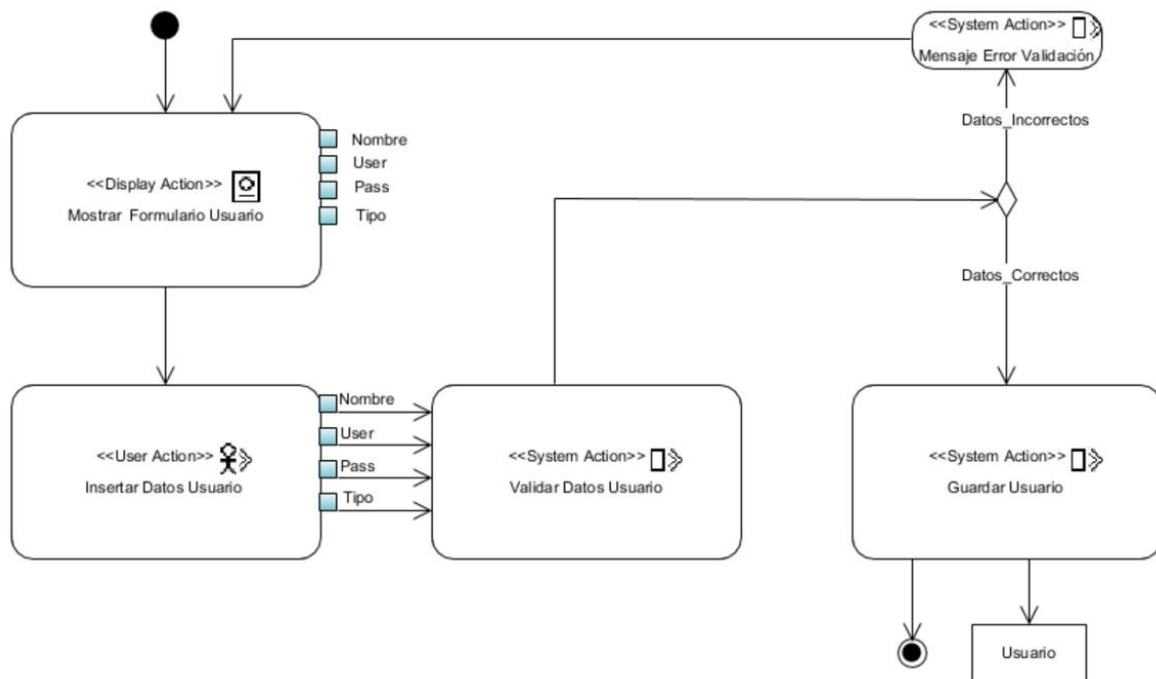


Figura 3.34 Diagrama de proceso del caso de uso Registrar Usuario

La Figura 3.35 muestra el diagrama de procesos para el caso de uso *Validar Información*. El proceso inicia al desplegar en la pantalla una lista de los contenidos de la base de datos HITO, al mismo tiempo que se presenta un formulario donde el usuario con autorización captura el identificador del contenido que quiere validar. Entonces, la aplicación proporciona los datos del contenido correspondiente. El usuario solicita la validación de la información y el sistema ejecuta la operación correspondiente de validación. El caso de uso concluye con la validación exitosa del contenido.

La Figura 3.36 presenta el diagrama de procesos para el caso de uso *Proponer Contenido*. el proceso comienza desplegando en pantalla un formulario donde el usuario ingresa el contenido y otorga una descripción del mismo. A continuación, el

sistema valida y guarda la nueva propuesta de contenido. Luego, se requiere la validación por parte de un usuario colaborador o administrador. Una vez que la aplicación guarda el contenido, el proceso se completa.

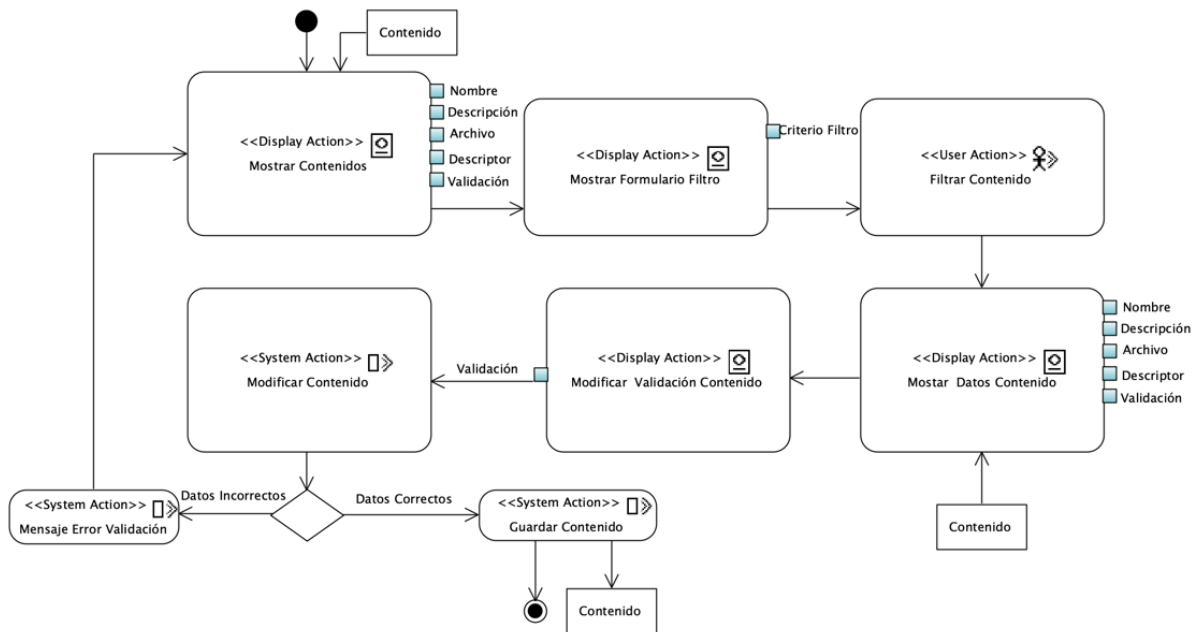


Figura 3.35 Diagrama de proceso del caso de uso Validar Información

La Figura 3.37 muestra el diagrama de procesos para el caso de uso *Alta de Contenido*. Este proceso comienza desplegando en pantalla un formulario donde el usuario con autorización ingresa los datos requeridos. Posteriormente, la aplicación guarda el nuevo contenido después de validar los datos proporcionados.

La Figura 3.38 presenta el diagrama de procesos para el caso de uso *Modificar Contenido*. El proceso inicia cuando en la pantalla se muestra tanto la lista de contenidos propuestos en el sistema como un formulario asociado donde el usuario autorizado introduce el identificador del contenido que quiere actualizar. La aplicación proporciona la información del contenido y el usuario realiza las modificaciones necesarias. El proceso finaliza cuando se completa la modificación del contenido.

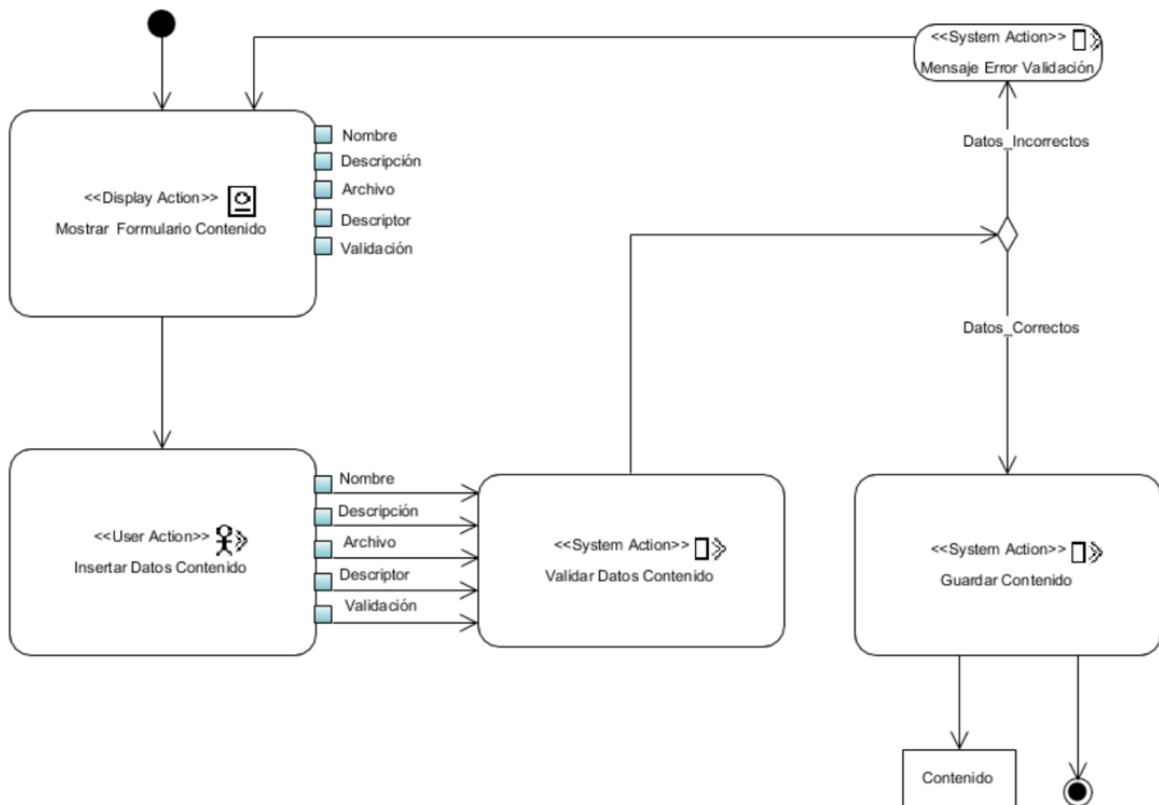


Figura 3.36 Diagrama de proceso del caso de uso Proponer Contenido

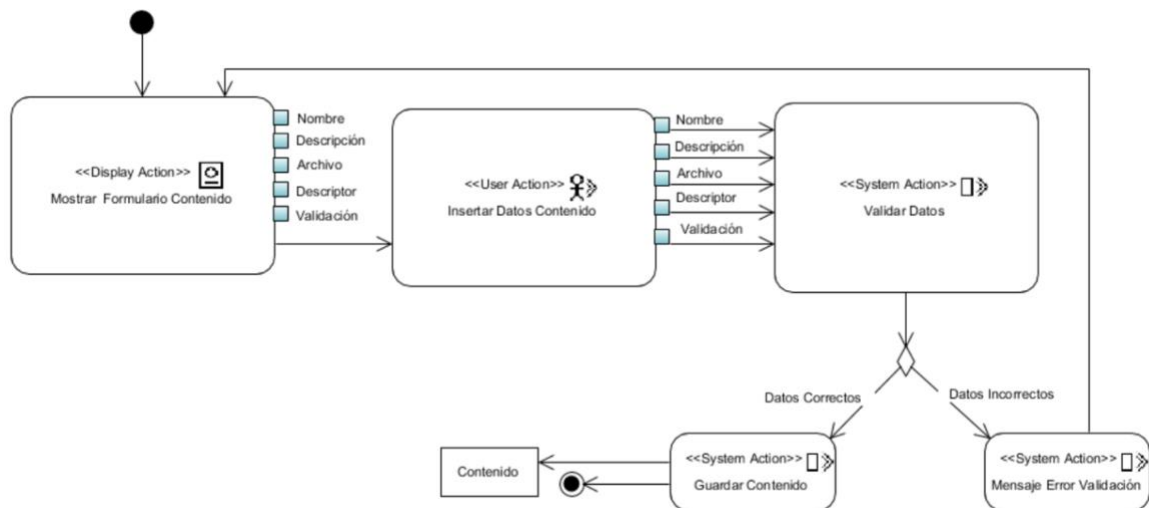


Figura 3.37 Diagrama de proceso del caso de uso Alta de Contenido

La Figura 3.39 muestra el diagrama de procesos para el caso de uso *Baja de Contenido*. Como se observa, en la pantalla se muestra tanto la lista de contenidos propuestos en el sistema como un formulario asociado donde el usuario con autorización proporciona el identificador del contenido que quiere borrar. La aplicación

proporciona los datos del contenido y el usuario indica su intención de eliminar el contenido. Luego, la aplicación procede a eliminar el contenido y el proceso se completa.

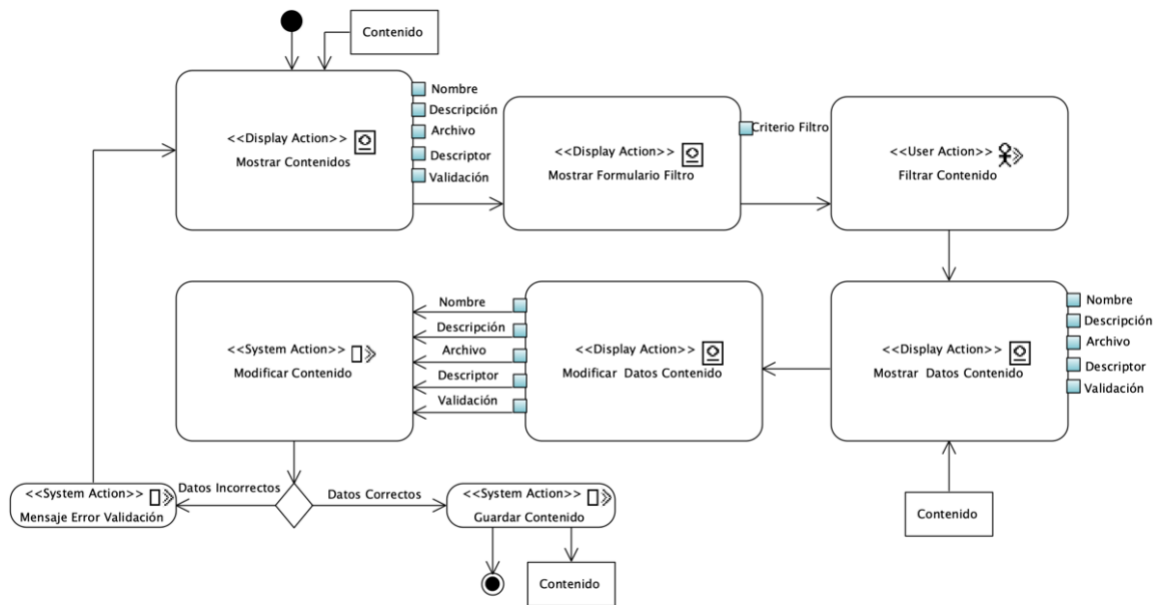


Figura 3.38 Diagrama de proceso del caso de uso Modificar Contenido

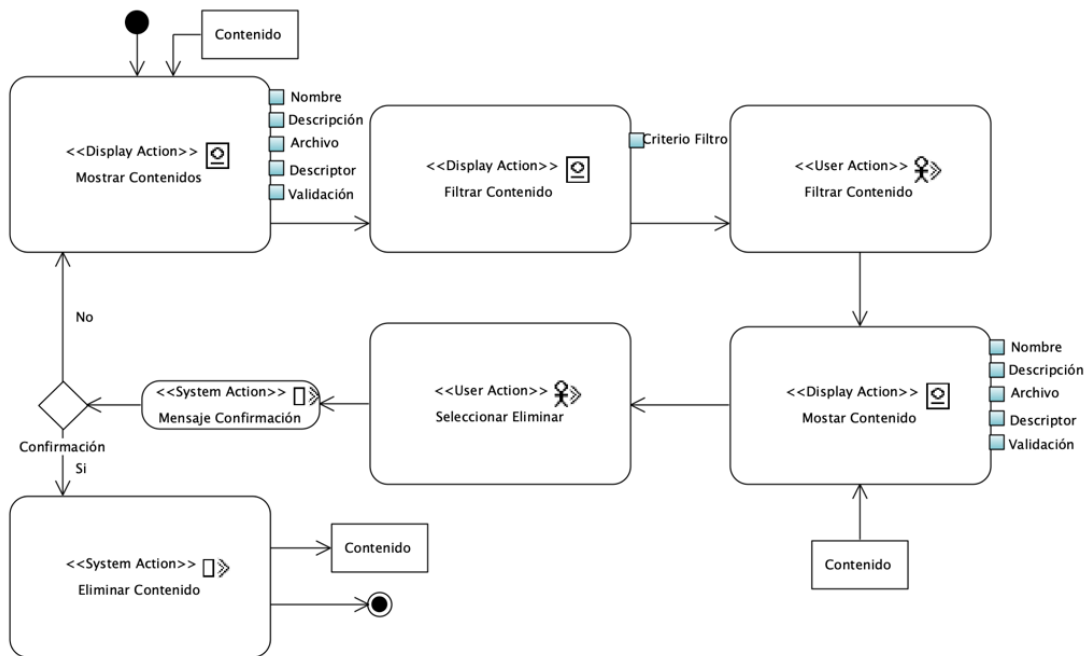


Figura 3.39 Diagrama de proceso del caso de uso Baja de Contenido

La Figura 3.40 muestra el diagrama de procesos para el caso de uso *Alta de Usuario*. Este proceso comienza al mostrar en pantalla un formulario donde el usuario con autorización ingresa los datos requeridos. Luego, el sistema guarda el nuevo usuario y finaliza el proceso correspondiente.

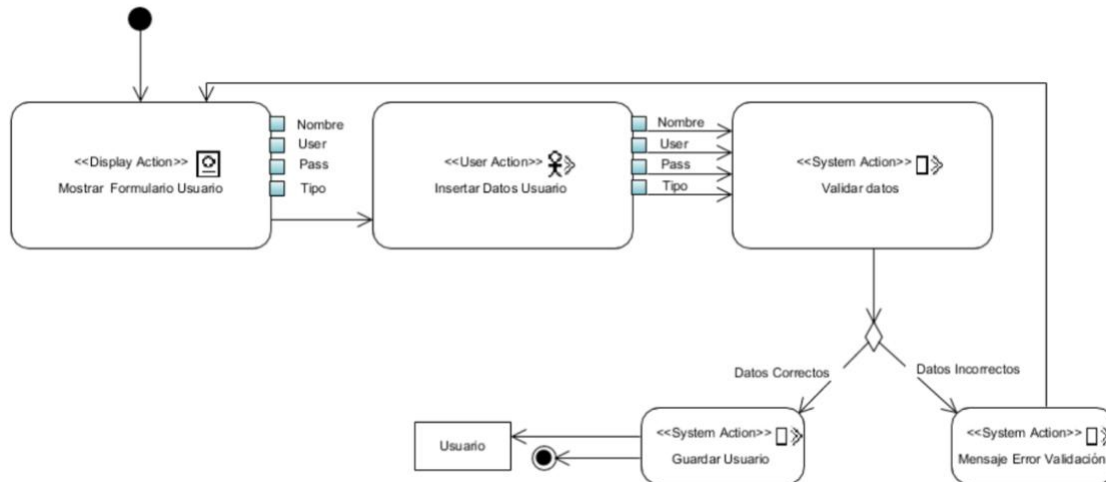


Figura 3.40 Diagrama de proceso del caso de uso *Alta de Usuario*

La Figura 3.41 presenta el diagrama de procesos para el caso de uso *Baja de Usuario*. El proceso inicia al mostrar en pantalla la lista de usuarios del sistema y un formulario donde el usuario autorizado ingresa el identificador del usuario que quiere borrar. Luego, la aplicación proporciona los datos del usuario seleccionado y el usuario autorizado indica su deseo de eliminar al usuario, para que la aplicación realice la operación correspondiente. El proceso finaliza cuando se completa la baja del usuario.

La Figura 3.42 ilustra el diagrama de proceso para el caso de uso *Modificar Usuario*. En la pantalla se muestra tanto la lista de usuarios del sistema como un formulario asociado donde el usuario autorizado ingresa el identificador del usuario que desea modificar. Después de proporcionar los datos del usuario seleccionado, el usuario autorizado realiza las modificaciones necesarias. El proceso finaliza cuando se completa la modificación de los datos del usuario.

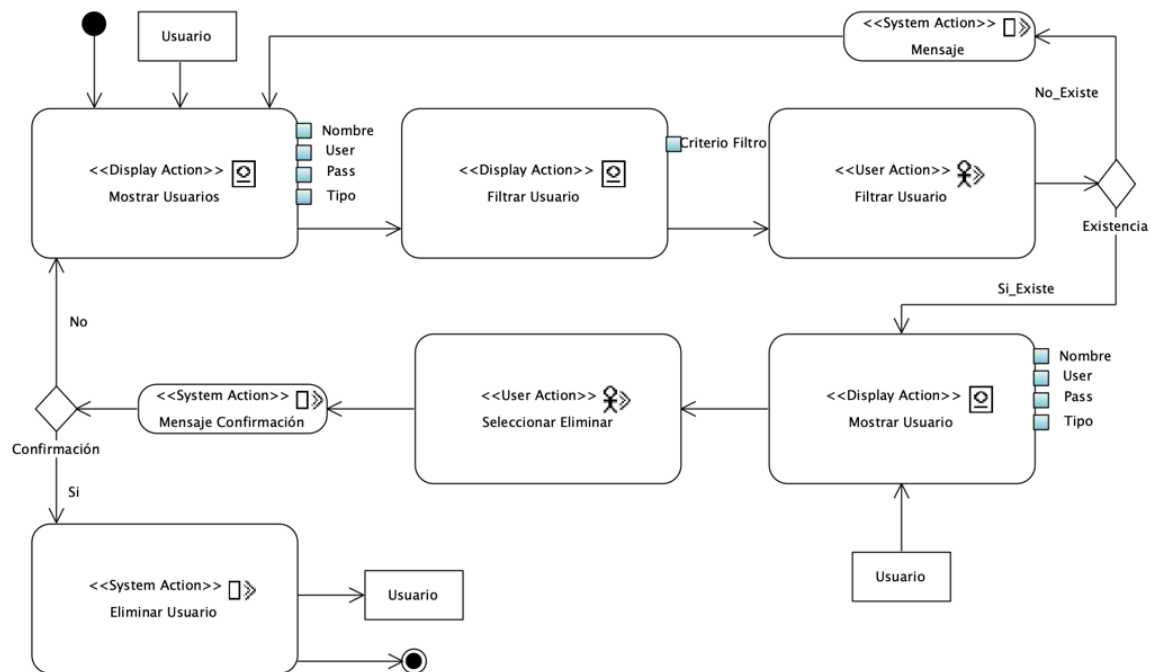


Figura 3.41 Diagrama de proceso del caso de uso Baja de Usuario

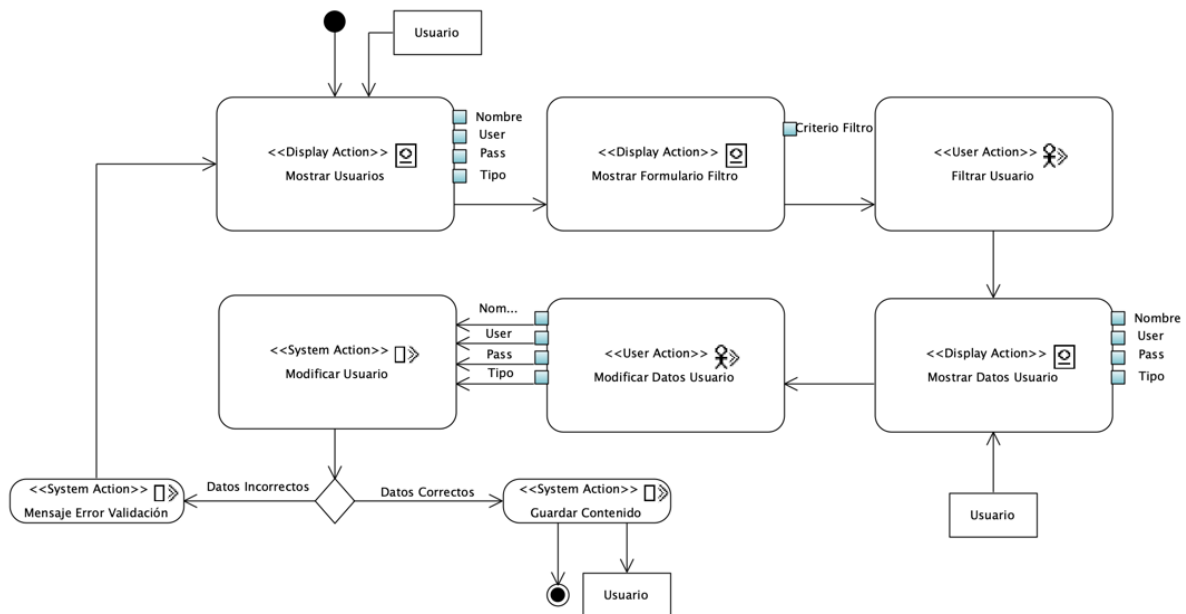


Figura 3.42 Diagrama de proceso del caso de uso Modificar Usuario

El caso de uso *Consultar Usuario*, como se muestra en el diagrama de proceso en la Figura 3.43, inicia cuando en la pantalla se muestra tanto el listado de usuarios registrados en aplicación como un formulario asociado donde el usuario autorizado es capaz de filtrar los resultados. El usuario autorizado ingresa el identificador del usuario

que desea consultar y el sistema responde proporcionando la información correspondiente al usuario consultado. El proceso finaliza una vez que se obtiene la respuesta del sistema

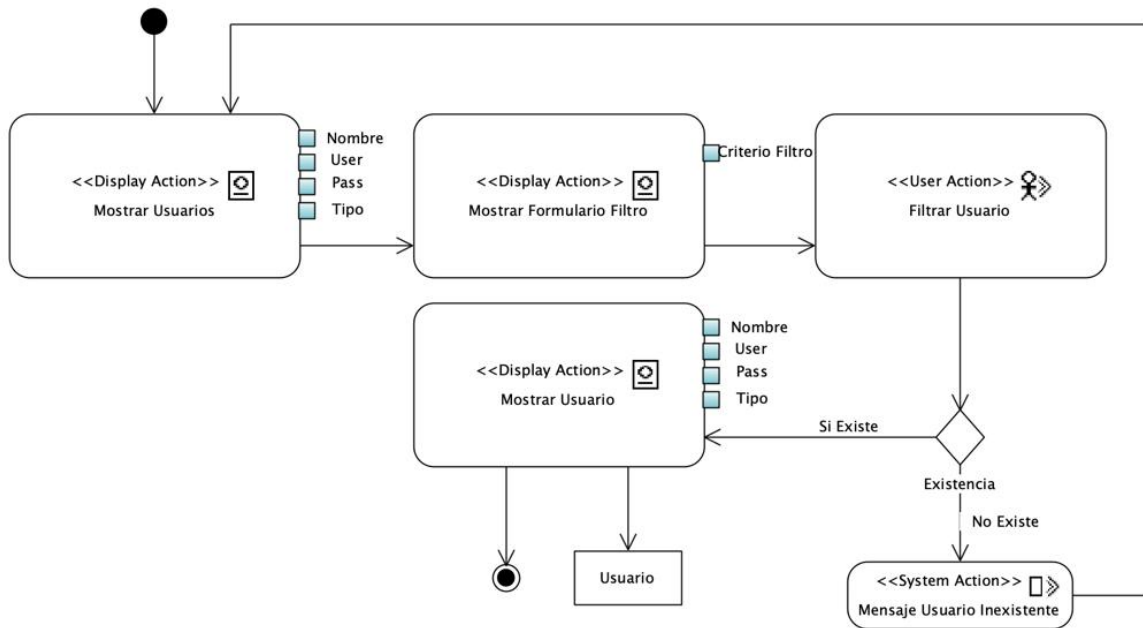


Figura 3.43 Diagrama de proceso del caso de uso Consultar Usuario

El caso de uso *Recuperación con Base en Texto*, representado en el diagrama de proceso de la Figura 3.44, comienza desplegando en pantalla un formulario en el cual el usuario autorizado ingresa la información requerida para realizar la búsqueda de contenido. A continuación, el sistema realiza la búsqueda utilizando los criterios proporcionados por el usuario. Los contenidos encontrados se muestran en pantalla y el proceso finaliza en este punto.

El caso de uso *Recuperación con Base en Contenido*, representado en el diagrama de proceso de la Figura 3.45, comienza desplegando en pantalla un formulario en el cual el usuario con autorización ingresa los datos multimedia necesarios. Luego, la aplicación recupera los datos multimedia con base en contenido utilizando los datos proporcionados por el usuario. Los contenidos encontrados se muestran en pantalla y el proceso se termina en este punto.

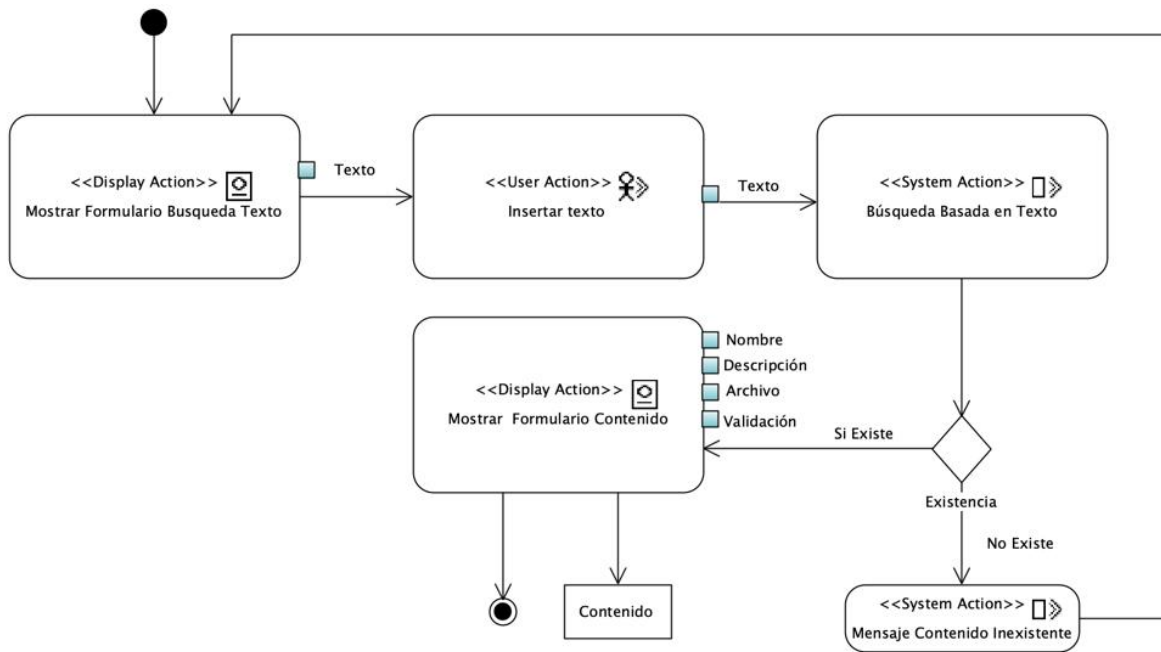


Figura 3.44 Diagrama de proceso del caso de uso Recuperación con Base en Texto

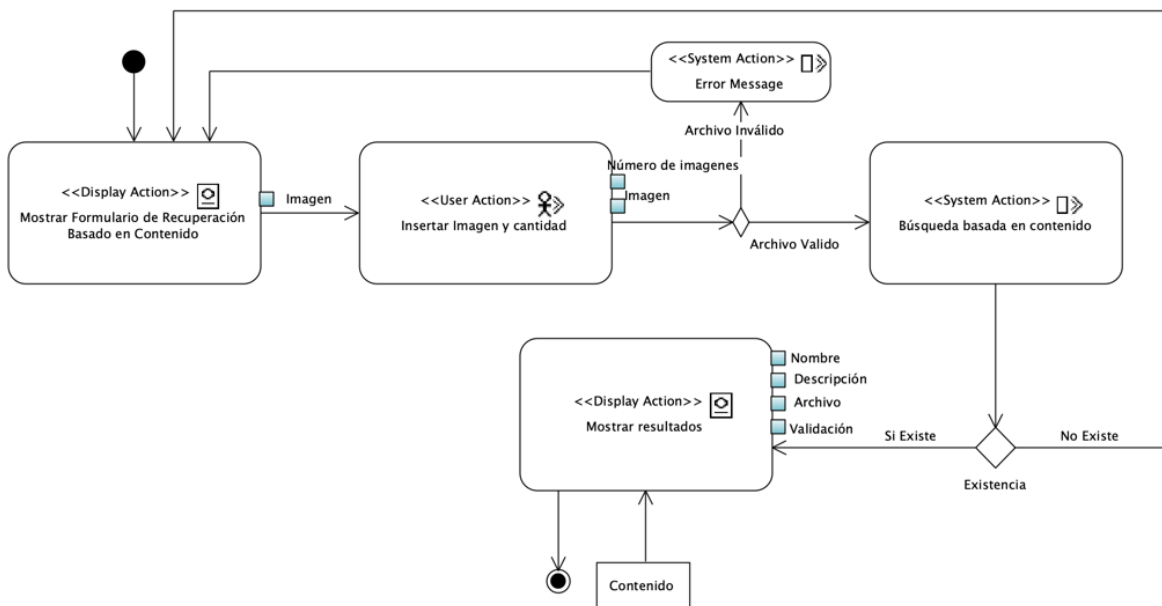


Figura 3.45 Diagrama de proceso del caso de uso Recuperación con Base en Contenido

3.4 Proceso CBVR

En esta parte del documento se describe el funcionamiento del módulo CBVR desarrollado, abarcando desde el flujo de trabajo, la forma en cómo al proponer un video se extraen los cuadros clave y tanto el video como los cuadros clave se guardan en la base de datos, cómo se obtiene los resultados de búsqueda de contenido y, por último, el algoritmo de recuperación de video con base en contenido propuesto en este trabajo.

3.4.1 Flujo de trabajo

En la Figura 3.46, se muestra el flujo de trabajo del módulo CBVR en el cual se observan dos fases, el preprocesamiento y la ejecución de consulta. En el preprocesamiento se selecciona el video, luego se obtienen los cuadros clave y sus descriptores SURF y se almacenan los valores en la base de datos MongoDB.

En la ejecución de consulta se ingresa una imagen de consulta, se obtiene los descriptores SURF de esa imagen, se comparan con los descriptores de los cuadros guardados en la base de datos y se recuperan los videos relevantes.

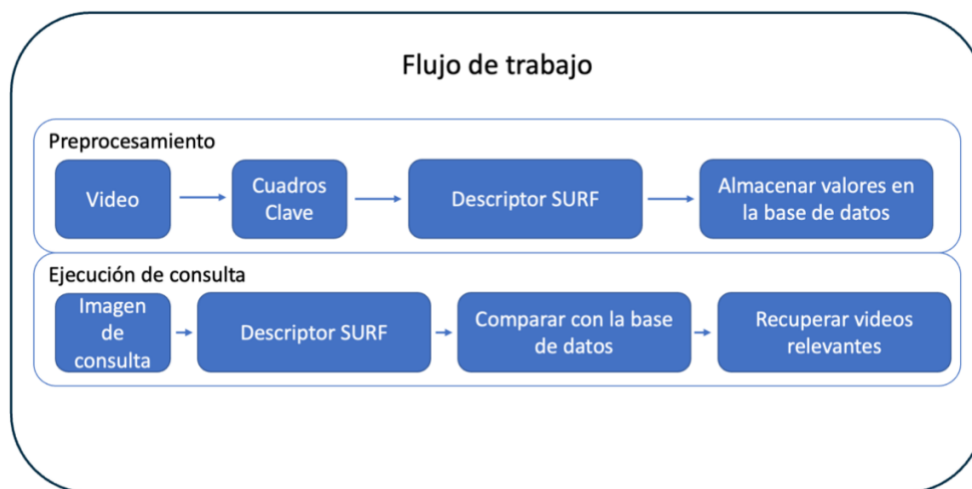


Figura 3.46 Flujo de trabajo

3.4.2 Propuesta de video

Para proponer un video se ingresa el título del video, su descripción y se selecciona un archivo de video, por último, se propone como nuevo video tal como se observa en la Figura 3.47.

Después de proponer un nuevo video, en el Código 3.1, se muestra el código del método *insertarvideo* que realiza diversas operaciones relacionadas con la inserción de un video en la base de datos de MongoDB usando la especificación GridFS como se observa en las líneas 10 a 20.

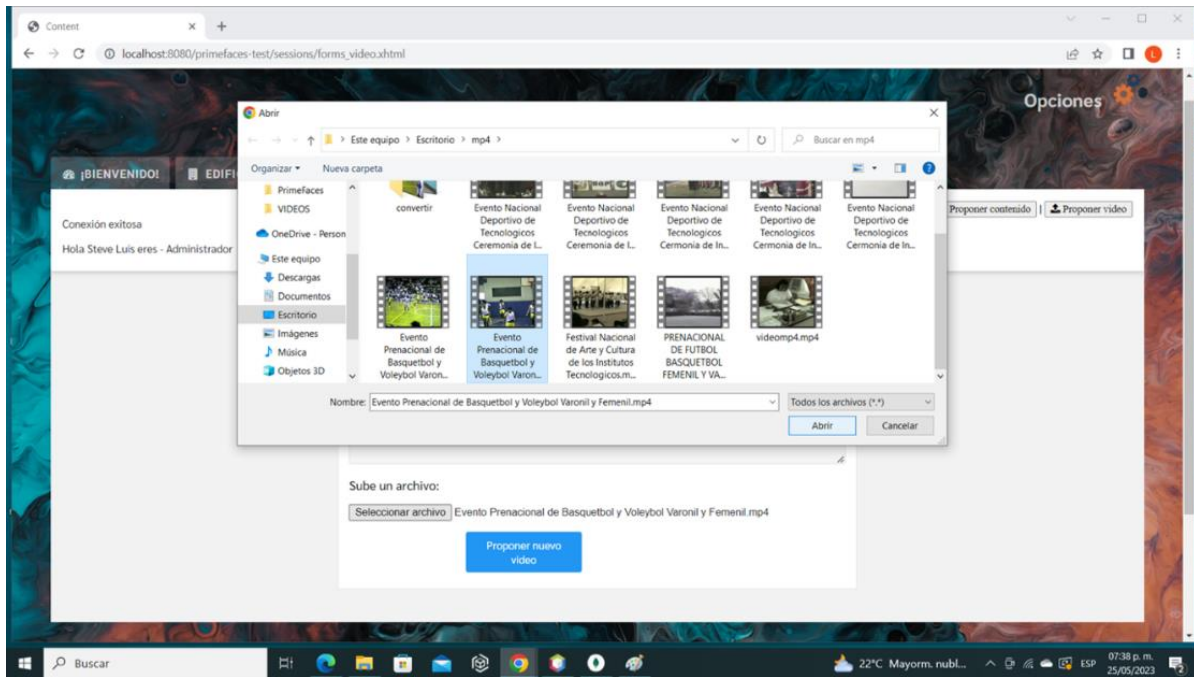


Figura 3.47 Proponer video

```

1. //video
2. public String insertarvideo() throws IOException, Exception{
3.     MongoClient mongoClient = new MongoClient();
4.     String sMsg = "";
5.     PropuestaVideo oCon = null;
6.     ObjectId fileId = null;
7.     String shortId="";
8.     FacesMessage msg;
9.     MongoDB database = mongoClient.getDatabase("HITO");
10.    GridFSBucket B=GridFSBuckets.create(database, "videos");
11.    FileInputStream streamToUploadFrom=null;
12.    try {
13.        Document doc=new Document();
14.        doc.append("Título", nombre);
15.        doc.append("Descripción",descripcion);
16.        doc.append("Validado","no");
17.        sMsg = nombre;
18.        shortId = RandomStringUtils.random(8,"0123456789abcdef");
19.        GridFSUploadOptions
            options = new GridFSUploadOptions().chunkSizeBytes(7
                16800).metadata(new Document("Título",

```

```

        nombre).append("Descripción",descripcion).append("Validado","no").append("Id", shortId));
20.    fileId = gridFSFilesBucket.uploadFromStream(file.getFile
        ileName(), file.getInputStream(), options);

```

Código 3.1 Código insertar video 1

Como se muestra en las Figura 3.48 y el Código 3.2 (línea 17), para la obtención de cuadros clave del video, se dividió el video en 25 partes, teniendo en cuenta que los cuadros que están juntos son similares, se usaron los cuadros que están separados por ser más significativos. Así se evita usar un método basado en histograma porque no es necesario comparar cada cuadro considerando un umbral de distancia, se utilizó OpenCV para la extracción de los cuadros clave como se observa en la línea 2 del Código 3.2.



Figura 3.48 Diagrama obtención de cuadros clave

```

1.  String input="C:\\videos2\\"+file.getFileName();
2.  OpenCV.loadShared();
3.  try {
4.      File f=new File(input);
5.      FileUtils.copyInputStreamToFile(file.getInputStream(),
        f);
6.  } catch (FileNotFoundException exception) {
7.      exception.printStackTrace();
8.  } catch (IOException ioException) {
9.      ioException.printStackTrace();
10. }
11. VideoCapture cap = new VideoCapture();
12. String output = "C:\\Frames";
13. System.out.println(input);
14. cap.open(input);
15. int video_length = (int) cap.get(Videoio.CAP_PROP_FRAME_COUNT
    );
16. int frames_per_second = (int) cap.get(Videoio.CAP_PROP_FPS);
17. int step=video_length/25;

```

```

18. Mat frame = new Mat();
19. int frames_remaining=0;
20. int frame_number=0;
21. if (cap.isOpened()){
22.     while(cap.read(frame)){
23.         if(frames_remaining%step==0){
24.             Imgcodecs.imwrite(output + "/" + nombre + frame_
                number + ".jpg", frame);

```

Código 3.2 Código insertar video 2

En el Código 3.3, se observa que se realizó la extracción de descriptores de los cuadros clave utilizando la biblioteca *BoofCV* en las líneas 2 a 13. Posteriormente, en las líneas 15 a 21 se insertan los datos del video en la base de datos. En la Figura 3.49, se observan los cuadros claves que se extrajeron del video propuesto en una carpeta local, esto con fines de prueba y demostración.

```

1. //Descriptores
2. List<Double> descriptor;
3. Class imageType = GrayF32.class;
4. DetectDescribePoint
   detDesc = FactoryDetectDescribe.surfStable(new ConfigFastHess
       ian(1, 2, 200, 1, 9, 4, 4), null, null, imageType);
5. ScoreAssociation
   scorer = FactoryAssociation.defaultScore(detDesc.getDescriptionType());
6. AssociateDescription
   associate = FactoryAssociation.greedy(new ConfigAssociateGreedy(true), scorer);
7. BoofCV surf = new BoofCV(detDesc, associate, imageType);
8. InputStream oImagen = new BufferedInputStream(new FileInputStream(output + "/" + nombre + frame_number + ".jpg"));
9. byte[] bytes = IOUtils.toByteArray(oImagen);
10. String encoded = Base64.getEncoder().encodeToString(bytes);
11. // Se extraen los descriptores de la imagen
12. descriptor = surf.extract(encoded);
13. System.out.println(surf);
14. try {
15.     oCon = new PropuestaVideo();
16.     oCon.setNombre(this.nombre + frame_number);
17.     oCon.setIdVideo(shortId);
18.     oCon.setImagen(encoded);
19.     oCon.setDescriptor(descriptor);
20.     sMsg = oCon.getNombre();
21.     oCon.insertar();
22. } catch (Exception e) {
23.     e.printStackTrace();
24.     sMsg = "Error al actualizar bd";
25. }

```

```

26.         frame_number++;
27.     }
28.         frames_remaining++;
29.     }
30.         cap.release();
31.     }
32.     else{
33.         System.out.println("Fail");
34.     }
35.     msg = new FacesMessage("Video propuesto insertado:",
        nombre);
36.     FacesContext.getCurrentInstance().addMessage(null, msg);
37.     return "/sessions/forms_video";
38. }

```

Código 3.3 Código insertar video 3

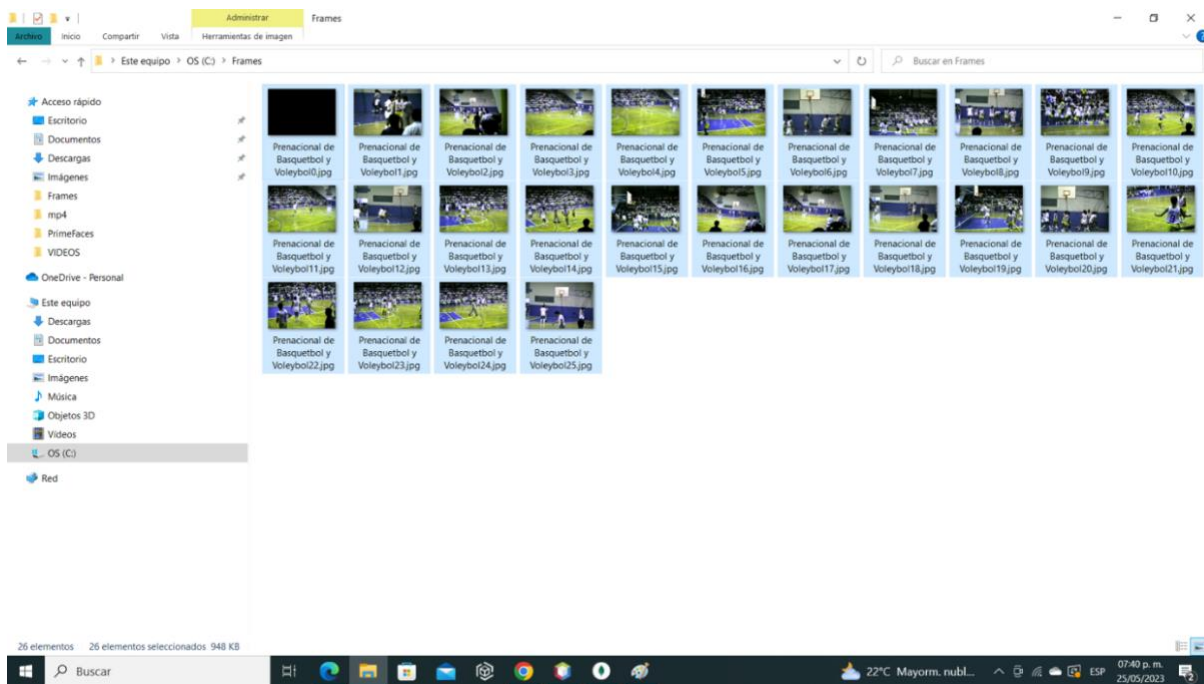


Figura 3.49 Cuadros clave

3.4.3 Base de datos

En la Figura 3.50, se observa la colección `HITO.videos.files` en la base de datos HITO donde se guardaron todos los videos propuestos con sus metadatos: *Título*, *Descripción*, *Validado* e *Id*. En la Figura 3.51, se muestra la colección `HITO.frames` donde se almacenaron todos los cuadros clave de cada video y sus descriptores SURF.

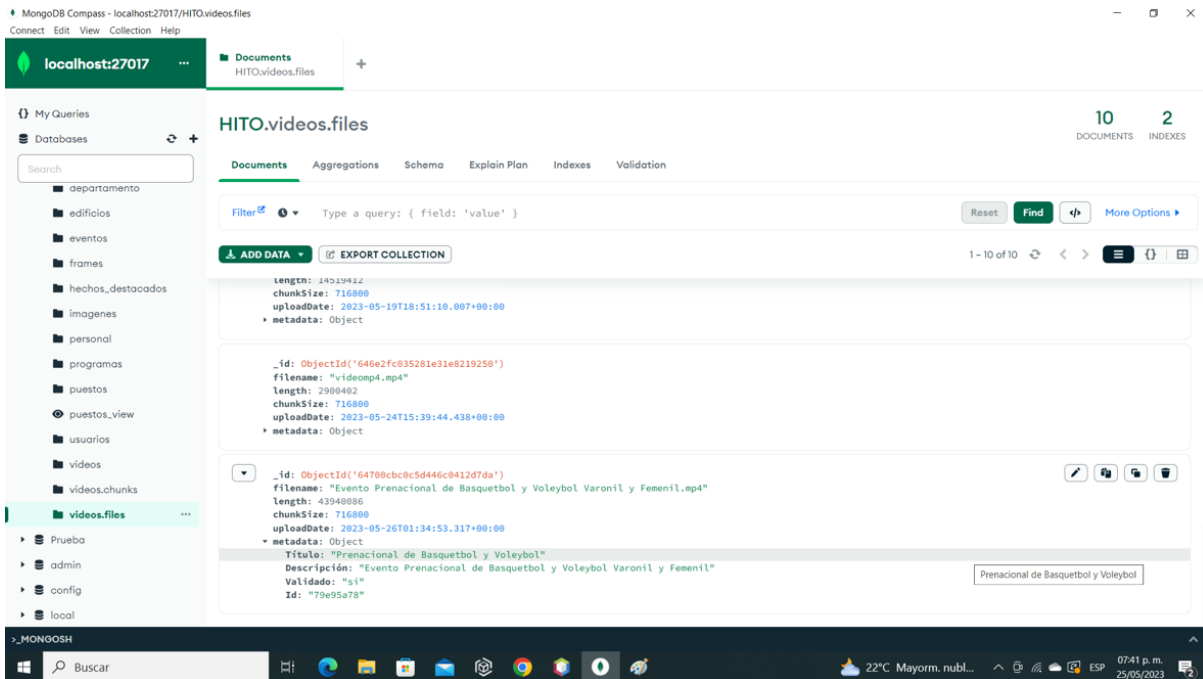


Figura 3.50 Colección videos MongoDB.

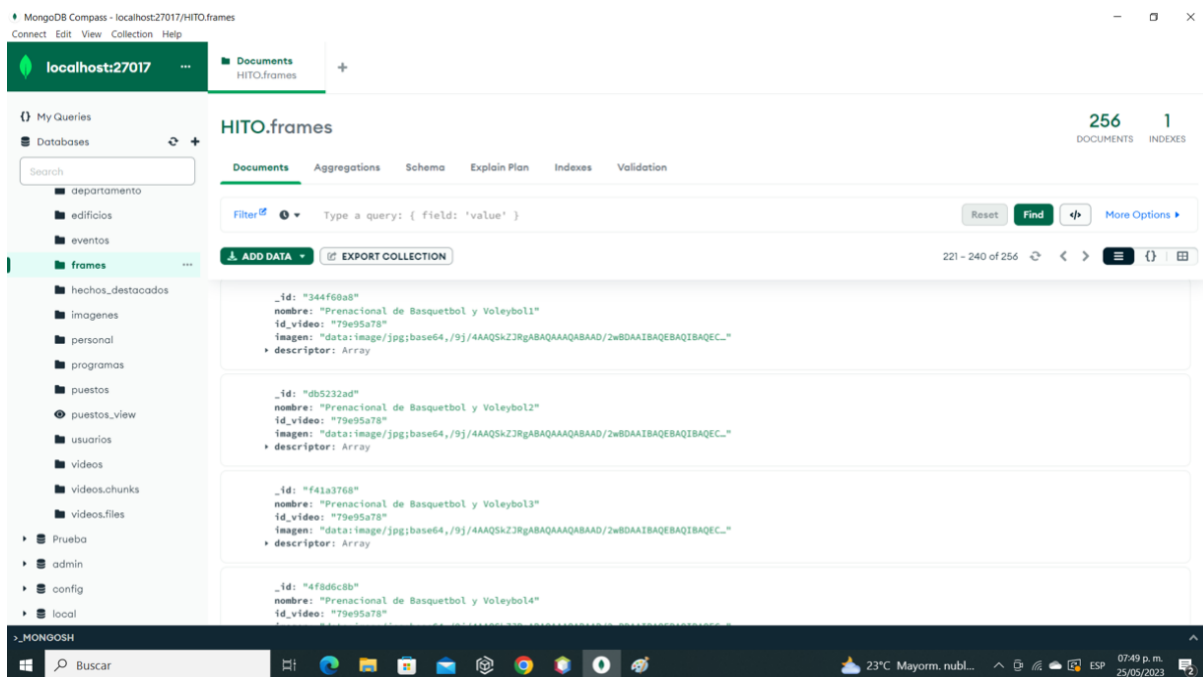


Figura 3.51 Colección frames MongoDB.

3.4.4 Algoritmo

El Algoritmo 1 muestra el proceso de comparación de cuadros clave para encontrar los k mejores videos con base en una medida de distancia Euclidiana.

Algoritmo 1. Recuperación de videos basada en contenido

Entrada: Imagen de búsqueda i , Base de datos multimedia, k

Salida: k mejores videos

```
1      foreach video  $v$ 
2           $mFoto=1$ 
3          foreach fotograma  $f$ 
4              obtener  $d(i,f)$ 
5              if  $d(i,f)<mFoto$  then
6                   $mFoto=d(i,f)$ 
7              end if
8          end foreach
9           $d(i,v)=mFoto$ 
10     end foreach
10     ordenar distancias
11     Recuperar los primeros  $k$  videos
```

En resumen, este algoritmo itera sobre cada video y posteriormente sobre cada fotograma dentro de cada video para encontrar la distancia mínima entre la imagen proporcionada por el usuario para realizar la búsqueda del video y los fotogramas de la base de datos. Luego, ordena las distancias y selecciona los k mejores videos asociados a las distancias más pequeñas entre la imagen buscada y los fotogramas de los videos de la base de datos.

Capítulo 4. Resultados

En este capítulo se presentan los resultados encontrados en esta tesis de maestría, abarcando los resultados del análisis llevado a cabo, se explica el funcionamiento de la aplicación Web y finalmente, se presentan los resultados obtenidos a partir del caso de estudio.

4.1 Resultados del análisis

En la presente sección, se presentan los resultados utilizando gráficas y tablas, lo cual proporciona una perspectiva más clara y visual de los datos mostrados en la Tabla 3.1. La Figura 4.1 compara la cantidad de trabajos que consideraron el tipo de recuperación de video basada en contenido y el número de trabajos que no tomaron en cuenta CBVR. En la Figura 4.1, se observa que CBVR fue un tema muy común entre los 32 artículos. Además, hay trabajos que incluyeron otro tipo de recuperación de contenido, como recuperación de video multimodal [23], publicidad basada en contenido de video [27] y recuperación de video a gran escala [29].

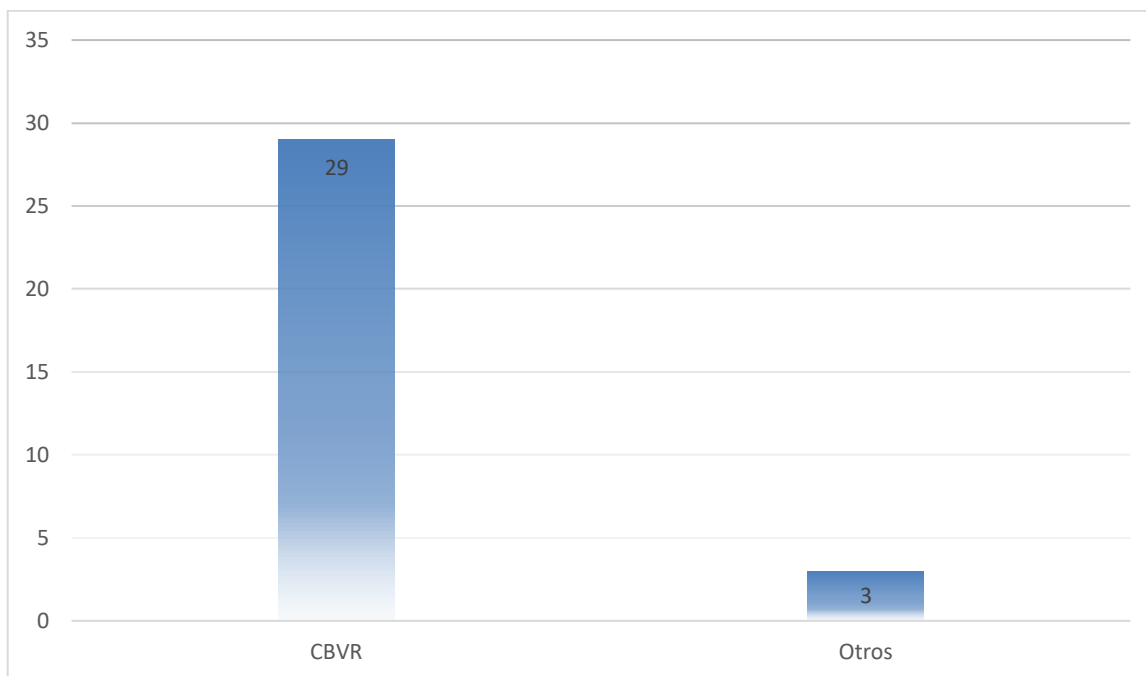


Figura 4.1 Tipo de recuperación

En la Figura 4.2 se exhibe un gráfico de barras que establece una relación entre el número de artículos y la editorial correspondiente; de los doce artículos de IEEE ocho utilizaron como tipo de consulta por video ([11], [14], [20], [21], [29], [30], [31] y [33]), dos por texto ([23] y [25]), uno por imagen ([7]) y uno multimodal de imagen con video ([26]).

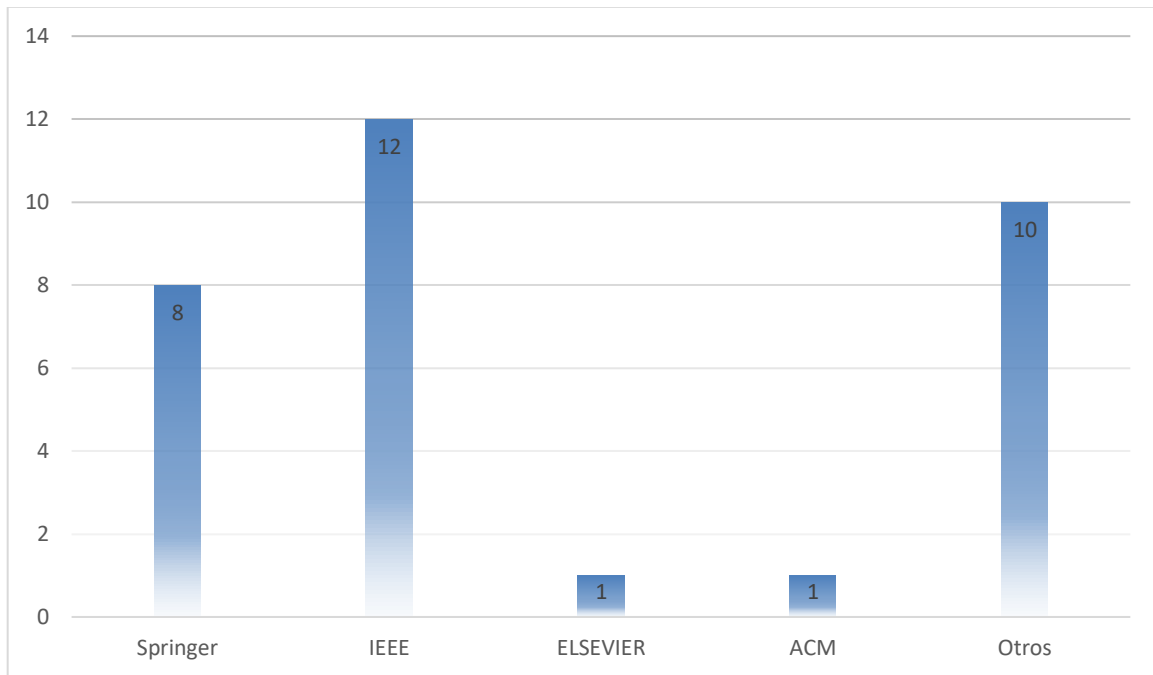


Figura 4.2 Número de artículos por editorial

Se visualiza la distribución de los artículos analizados según el año en que fueron publicados en la Figura 4.3. Se observa que los artículos son de 2018 a 2022. En 2019 y 2021 se publicaron más artículos y se observó que de los ocho artículos del 2019 cinco usaron como descriptor características CNN ([18], [21], [30]-[32]), otros trabajos utilizaron HOG [5] y SURF [2], y en el 2021 también cinco usaron características CNN ([12], [14], [16], [24], [28]).

La Figura 4.4 presenta un gráfico de barras que establece una relación entre el número de artículos y su tipo correspondiente de consulta. El número de artículos que tomaron en cuenta el tipo de consulta por video son 19 ([1], [8]-[12], [14]-[17], [20]-[22], [24], [27], [29]-[31], [33]), cinco consideran consulta por imagen ([2], [3], [5], [7], [13]) y dos

consultas por texto ([23], [25]). En seis artículos los autores realizaron consulta multimodal ([4], [18], [19], [26], [28], [32]).

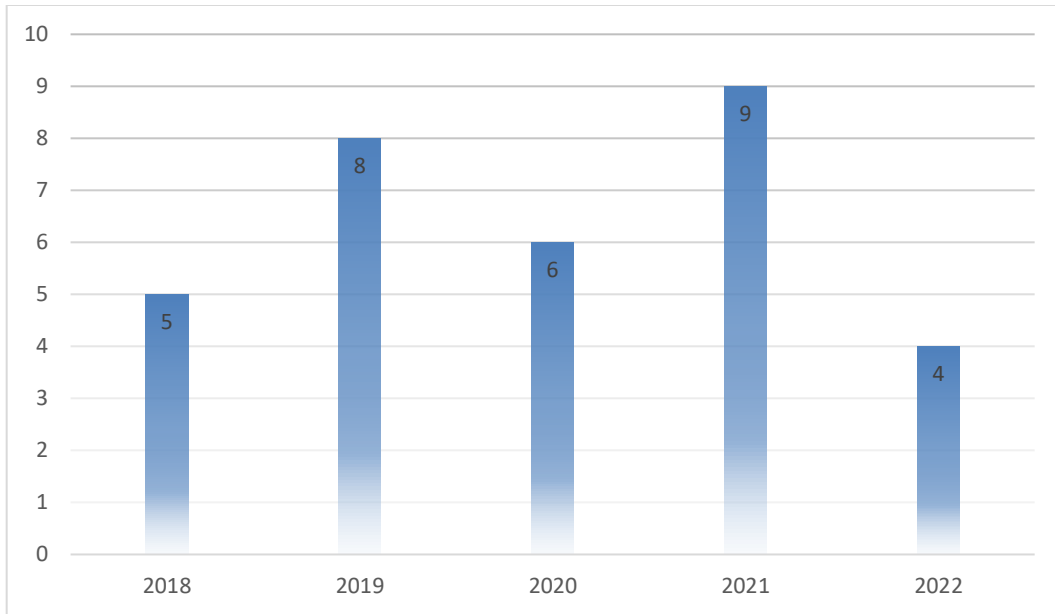


Figura 4.3 Cantidad de artículos por año

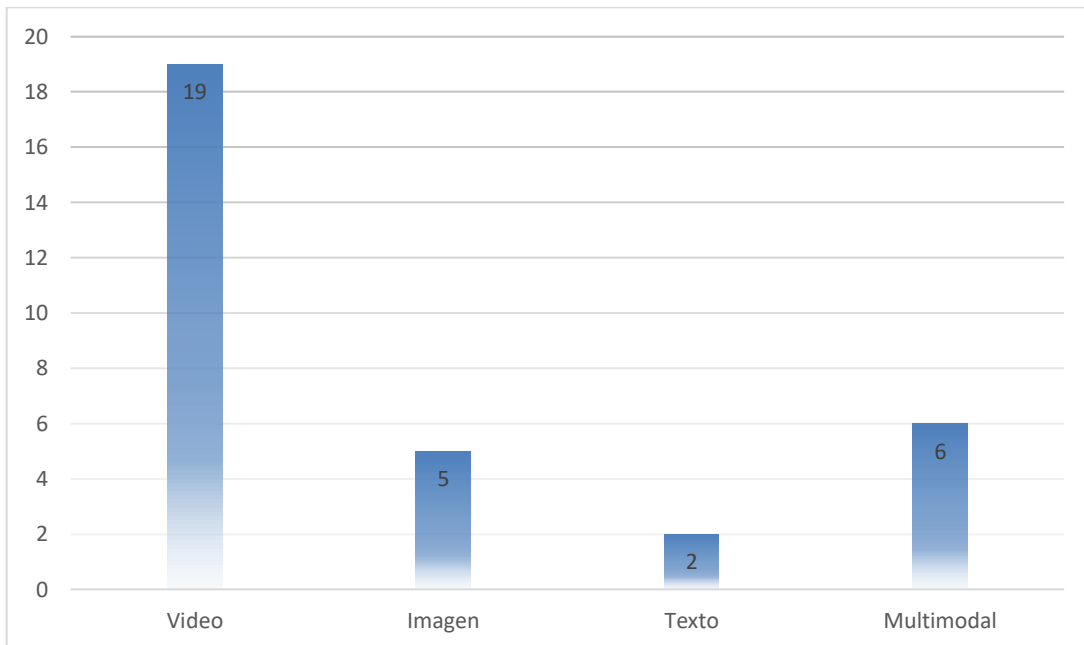


Figura 4.4 Tipo de consulta

En la Figura 4.5 se comparó la cantidad de tecnologías que se consideraron en los trabajos, donde las tecnologías más usadas fueron CNN en 19 trabajos ([3], [9], [12], [14], [16], [18], [20]-[22], [24]-[33]), PCA ([1], [7], [17], [22], [26], [29]-[31]) y Python ([3], [10], [15], [17], [21], [22], [27]) en ocho y MATLAB en seis ([4], [5], [11], [14], [27], [29]). En la gráfica solo se muestran las tecnologías que se emplearon en más de dos trabajos.

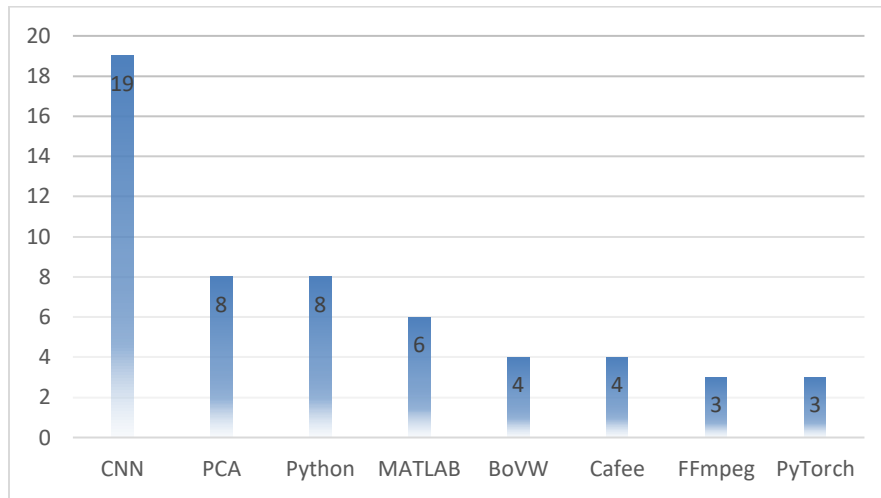


Figura 4.5 Tecnologías más utilizadas

La Figura 4.6 presenta un gráfico de barras que establece una relación con el número de artículos que tuvieron un conjunto de datos propio; 11 de 32 trabajos utilizaron un conjunto de datos propio ([5], [8], [9], [11], [13], [15], [19], [24], [25], [27] y [33]). Por lo tanto, 21 artículos optaron por usar datos sintéticos de diferentes *benchmark* como VCDB ([10], [16], [17], [21], [26], [31], [33]), FIVR-200K ([16], [17], [21], [24], [26], [30], [31]), CC_WEB_VIDEO ([16], [17], [20], [21], [26], [31]), FIVR-5K ([17], [20], [26], [31]), EVVE ([17], [26], [31]), ActivityNet ([22], [29], [31]) e ImageNet ([9], [16], [29]), por mencionar los más empleados.

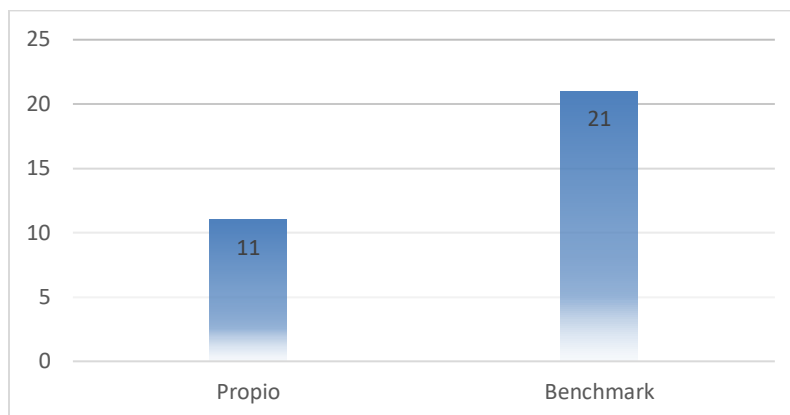


Figura 4.6 Conjunto de datos

4.2 Sistema de Recuperación de Videos con Base en contenido de la Base de Datos Multimedia HITO

Se requiere realizar el inicio de sesión para acceder al sistema, ya sea como administrador, colaborador, profesor o usuario general. En la Figura 4.7 se visualiza la página de propuesta de contenido, la cual está disponible para cualquier tipo de usuario existente del sistema. En dicha página se muestra un formulario que debe completarse para registrar la propuesta de contenido, la cual se almacena en la base de datos junto con el atributo correspondiente “validado” con un valor negativo, ya que antes de que este contenido se muestre en las páginas correspondientes, se requiere la aprobación previa por parte de un colaborador o administrador, también se extraen los fotogramas clave del video y se guardan en la base de datos HITO. En la Figura 4.8, se muestra la colección de videos y se observa que ya aparece el nuevo video propuesto.

Los usuarios con rol de administrador tienen la capacidad de ingresar a todas las características de gestión de datos dentro del sistema. Por otro lado, los usuarios con rol de colaborador solo tienen acceso a las opciones de inserción, modificación y validación de información de contenido como edificios, eventos, equipo, personal y videos. Mientras que los usuarios de tipo profesor y usuario general, no tienen acceso a ninguna de estas opciones. Esto se representa en la Figura 4.9.

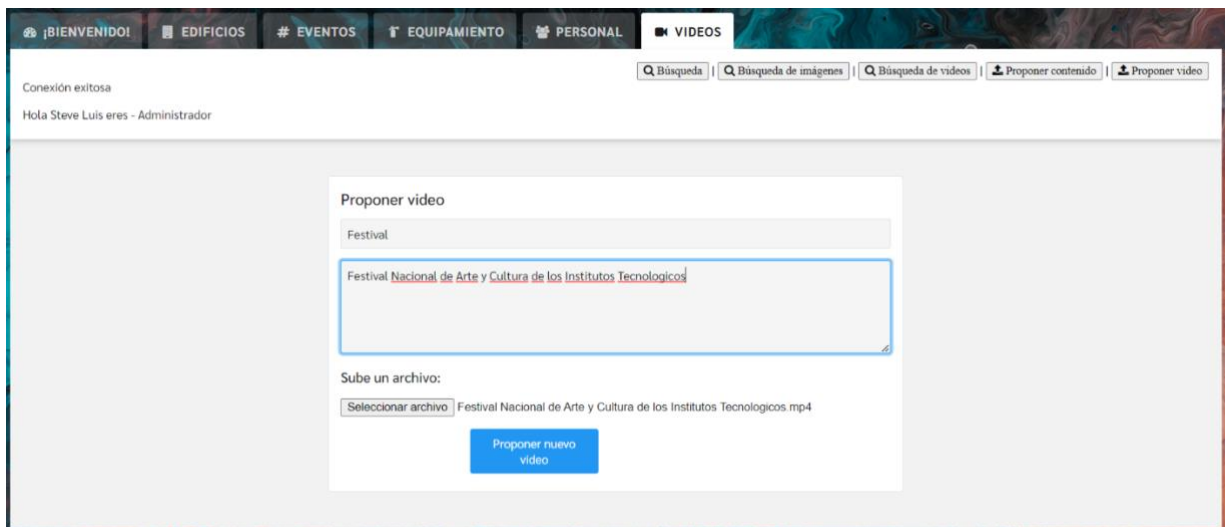


Figura 4.7 Proponer video

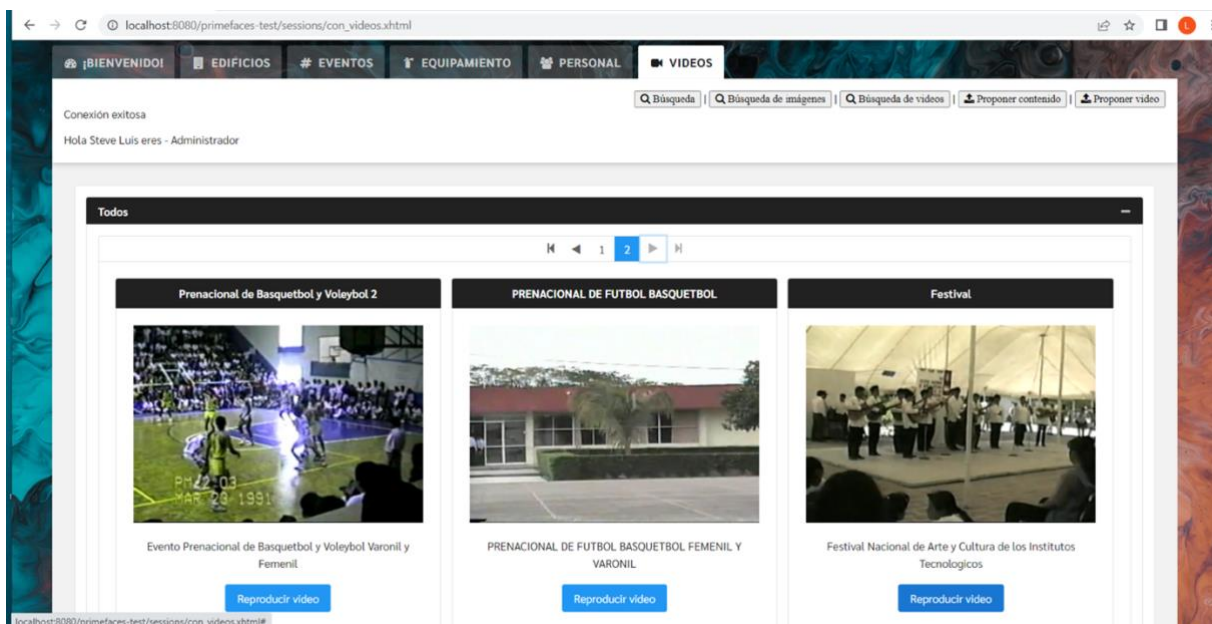


Figura 4.8 Colección de videos

En la Figura 4.10, se observa que cuando se accede al botón ubicado a la derecha del título de la lista de videos, se muestra la interfaz que incluye el formulario requerido para registrar un nuevo video con su título y descripción, al hacer clic en *Registrar Nuevo Video* se obtienen sus cuadros clave y descriptores y se almacenan en la base de datos para su posterior uso en el módulo de búsqueda de videos.

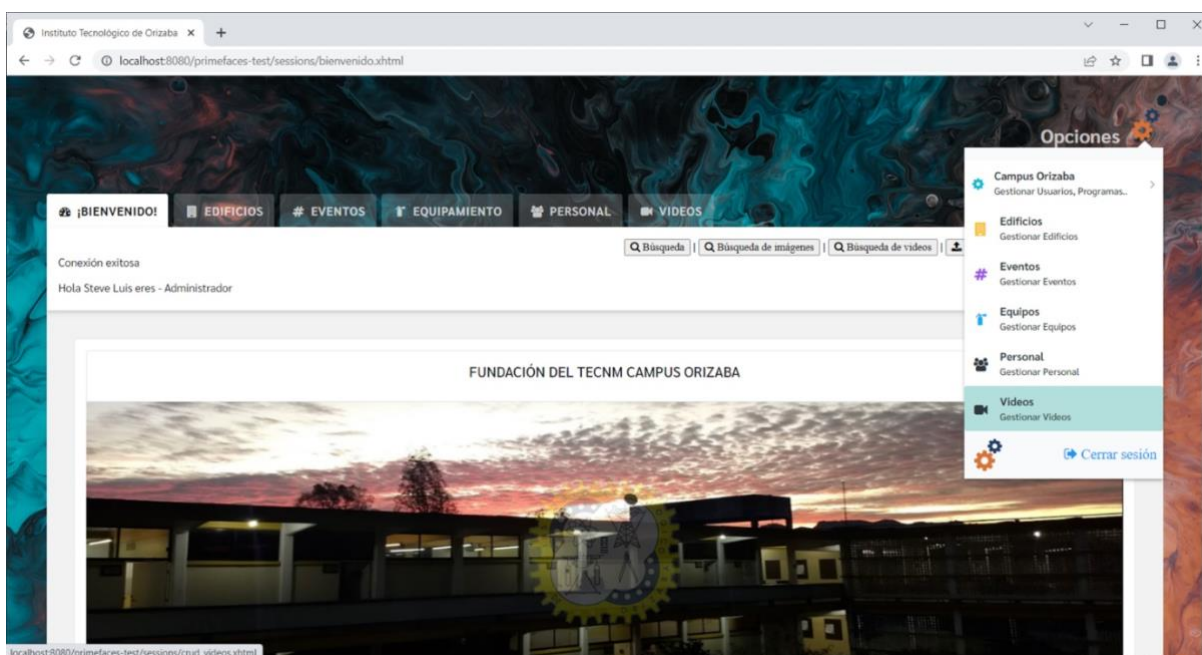


Figura 4.9 Opciones para gestionar contenido

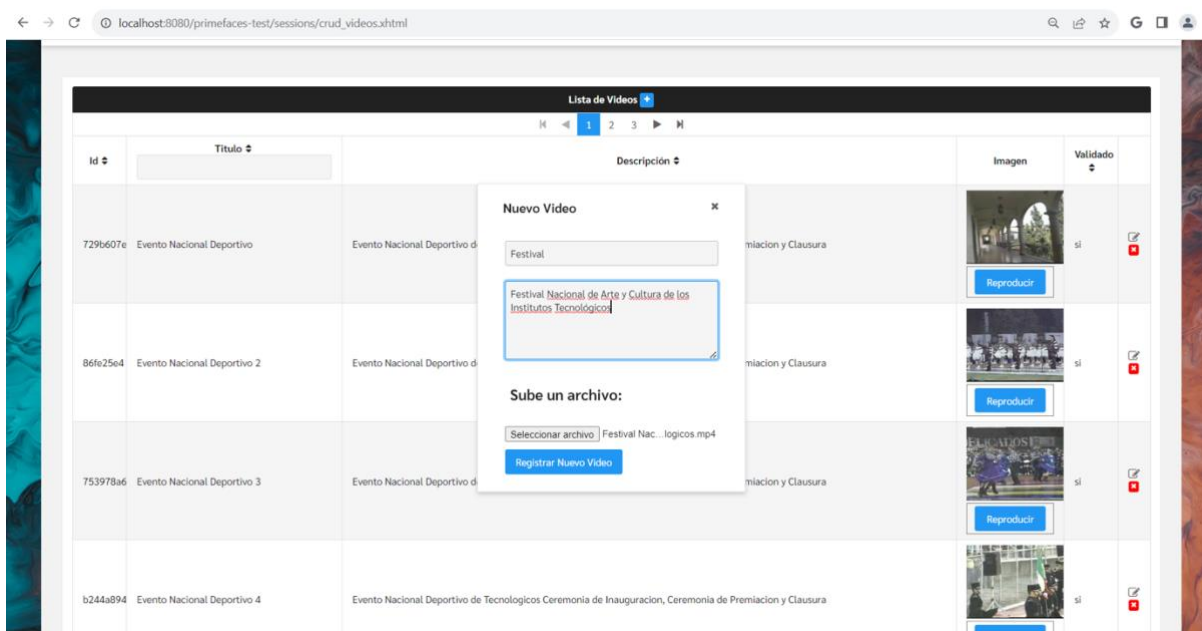


Figura 4.10 Formulario para inserción de video desde la página de gestión

En la Figura 4.11, se presenta la página para la gestión de información de los videos, la cual posee una tabla de datos donde se muestra un fotograma previamente almacenado con el video correspondiente. Si se elige el botón de edición de una fila

específica, esa misma fila se modifica y se muestran los campos editables para realizar modificaciones posteriores. En caso de que se haya seleccionado alguna fila, pero ningún cambio se realiza o se cancela la edición, se procede a cambiar el modo de edición de la fila a un modo de solo lectura. Si una fila se modifica correctamente, el sistema emite una notificación indicando que la fila seleccionada fue modificada exitosamente.

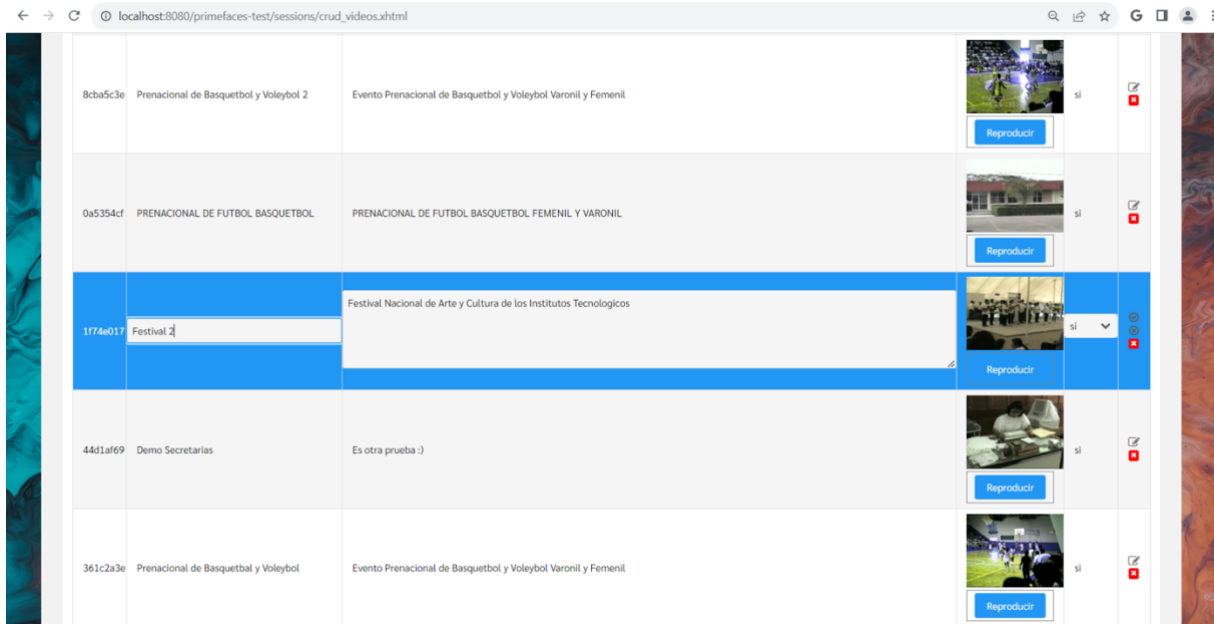


Figura 4.11 Editar video

En la Figura 4.12, se observa que cuando se selecciona el botón para eliminar una fila, se muestra un cuadro de diálogo que solicita la confirmación del administrador y si se confirma, elimina el video, en caso contrario, se cierra el diálogo sin cambios en la tabla de datos. La Figura 4.13 presenta la notificación del sistema de que el video se eliminó exitosamente.

En la Figura 4.14, se presenta la página que contiene los videos, los cuales se visualizan al hacer clic en el ícono + a la derecha de *Todos*. La Figura 4.15 muestra la página que contiene los datos multimedia validados encontrados en la colección de `videos` de la base de datos.

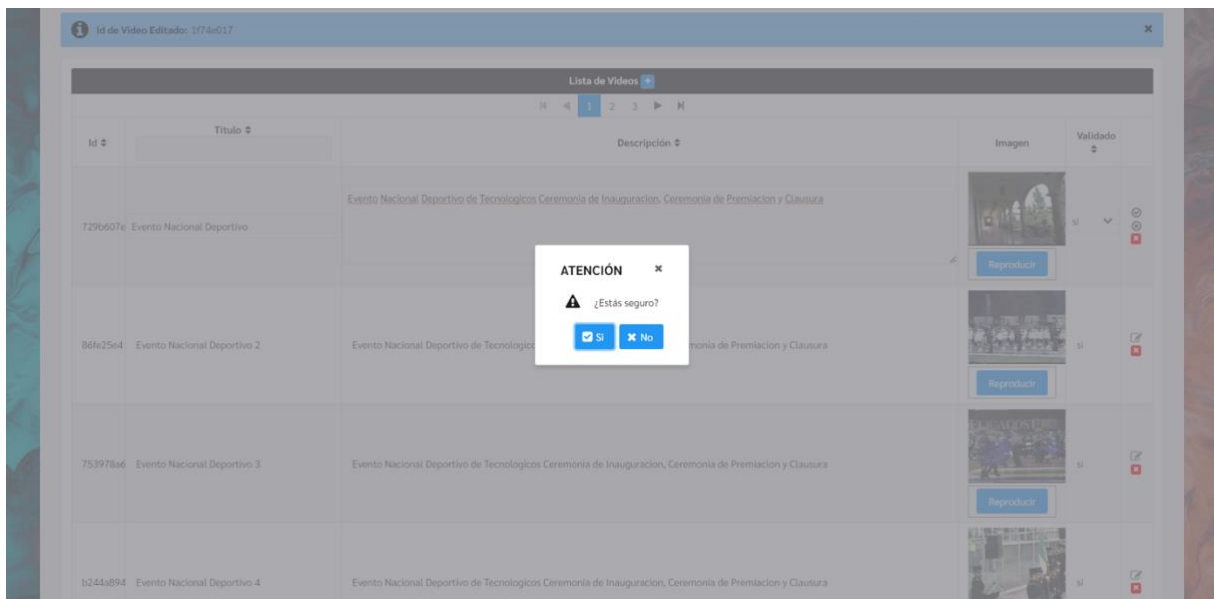


Figura 4.12 Eliminar video

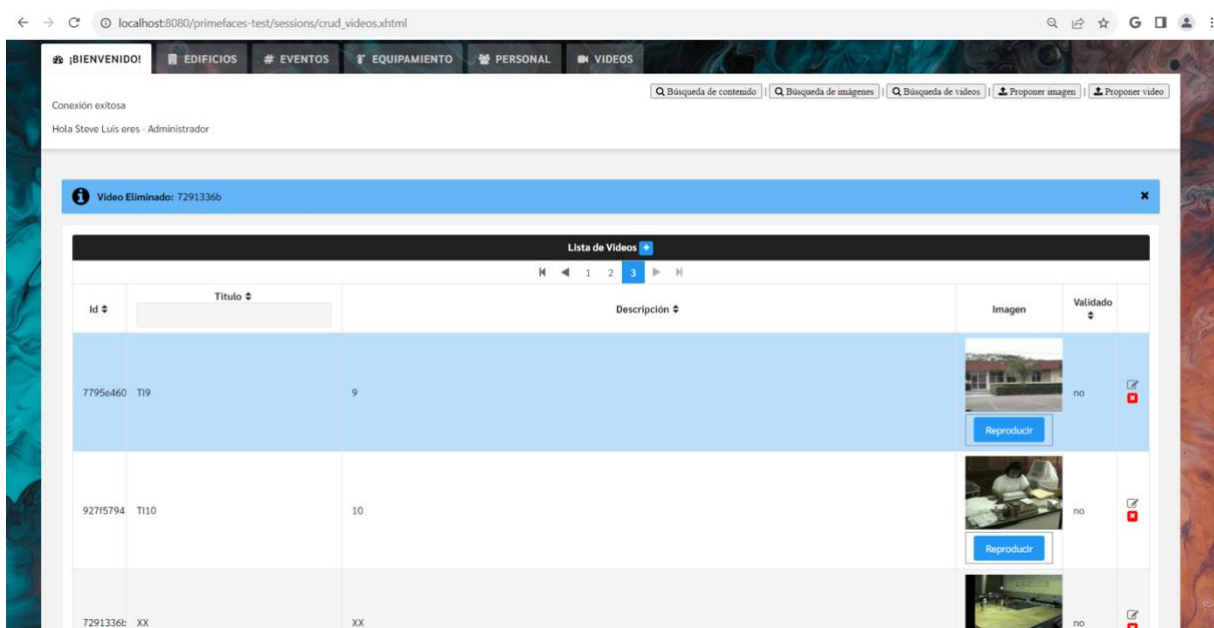


Figura 4.13 Video eliminado

Como se observa en la Figura 4.15, cada una de las imágenes que se muestran en la pantalla de la pestaña de videos tienen debajo un botón para reproducir el video correspondiente a la imagen. La Figura 4.16 muestra la reproducción de un video seleccionado.

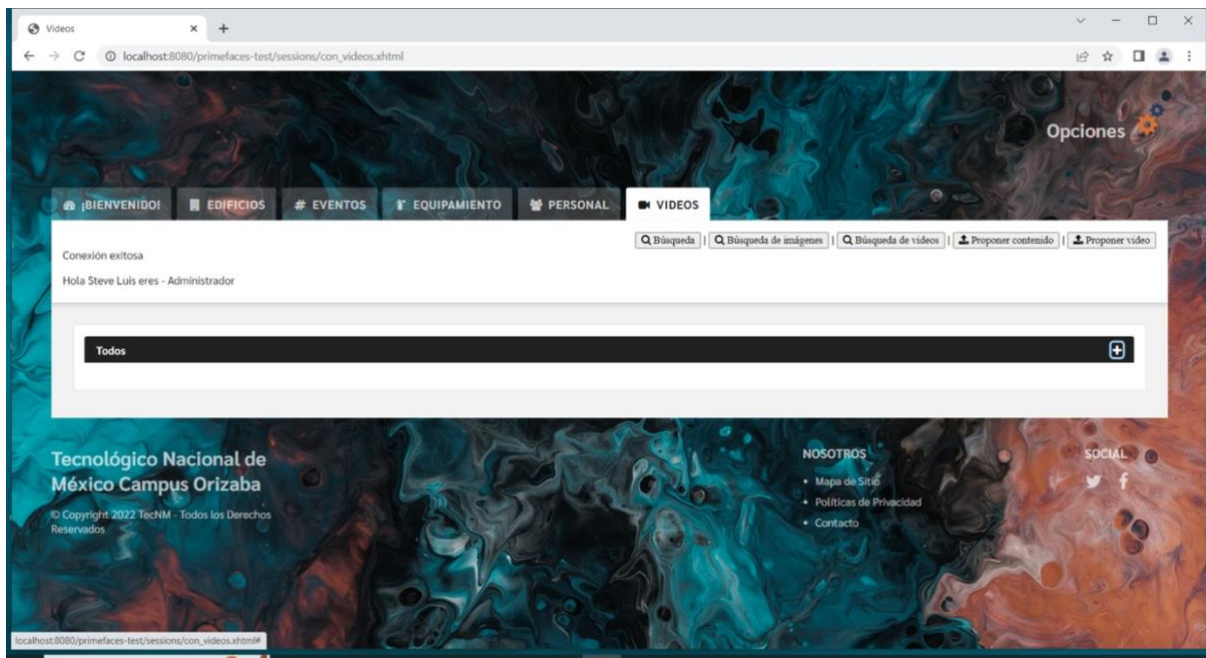


Figura 4.14 Pestaña videos

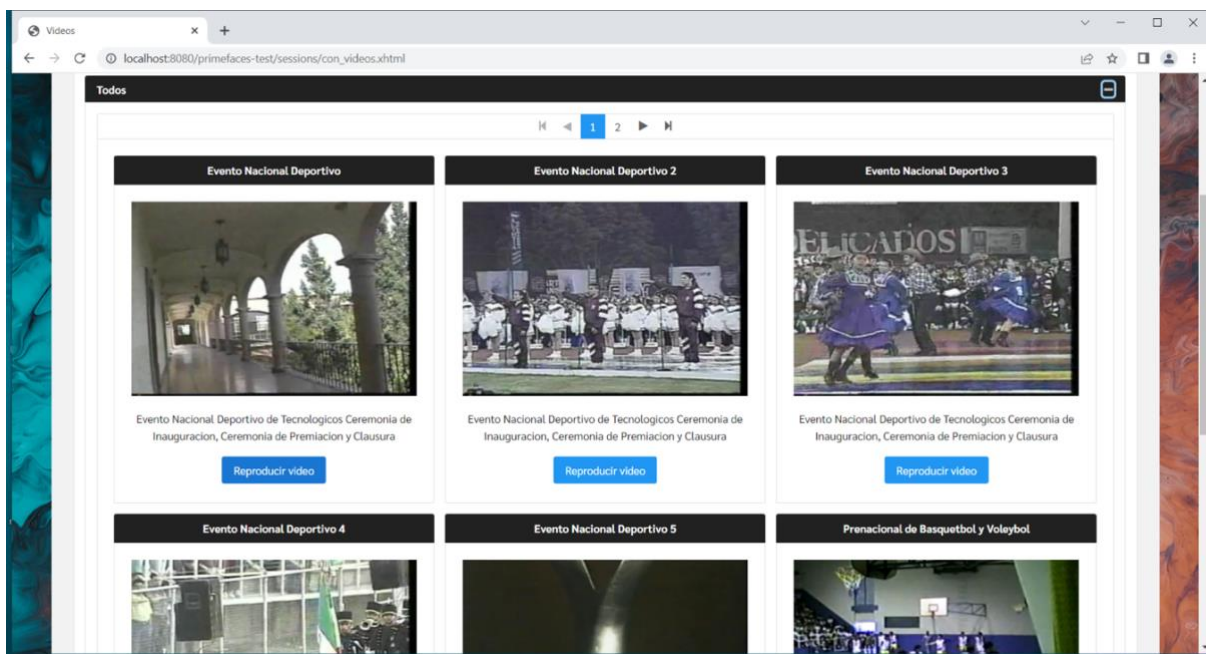


Figura 4.15 Todos los videos

Se muestra la pestaña de búsqueda basada en contenido en la Figura 4.17, donde se selecciona cuántos contenidos se desean recuperar y se ingresa el archivo para la búsqueda.

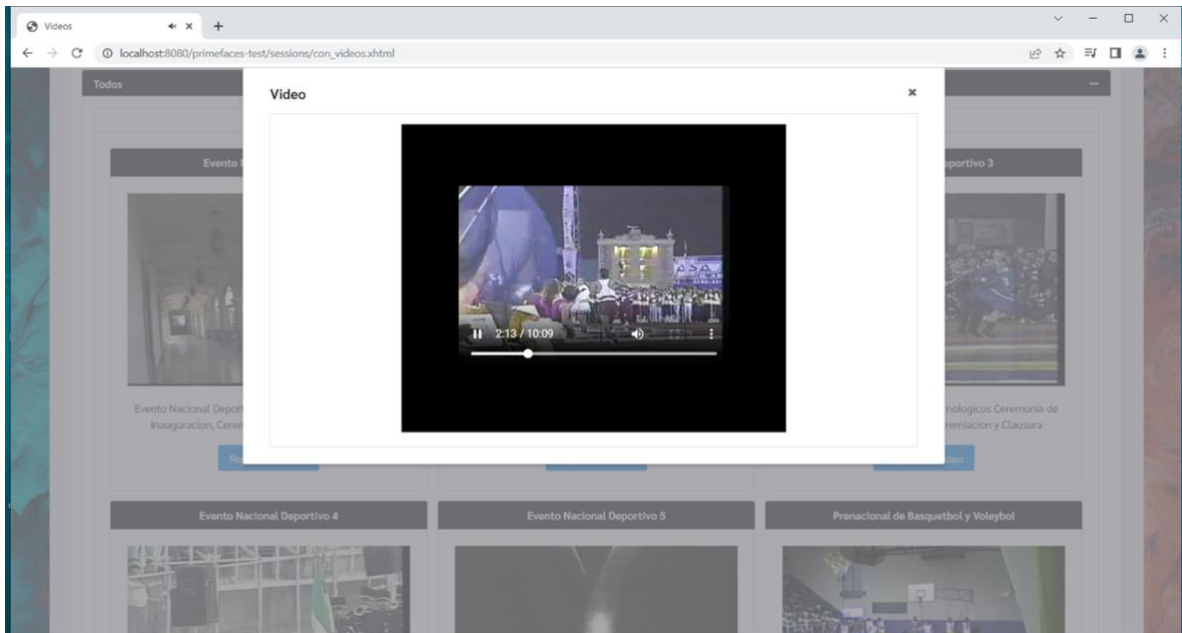


Figura 4.16 Reproductor de video

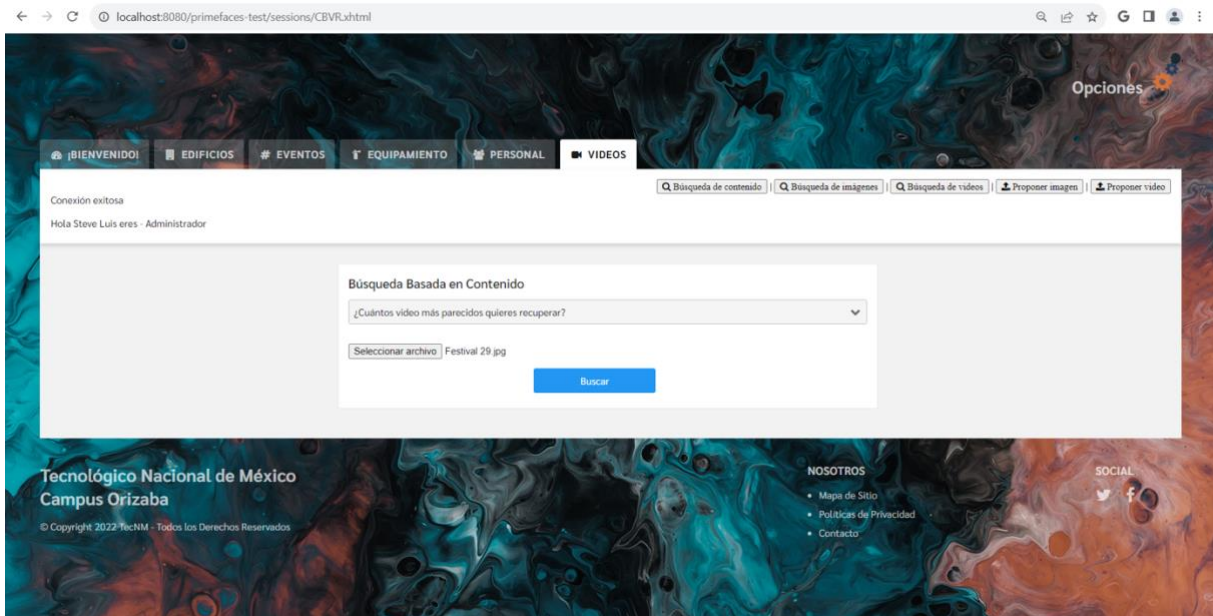


Figura 4.17 Búsqueda basada en contenido

En la Figura 4.18, se observa la imagen de búsqueda y los resultados con sus porcentajes de similitud, con la opción de reproducir el video del fotograma clave de los resultados. Posteriormente en la Figura 4.19, se muestran los cinco resultados de

la búsqueda realizada. El reproductor de video de los resultados obtenidos se visualiza en la Figura 4.20.

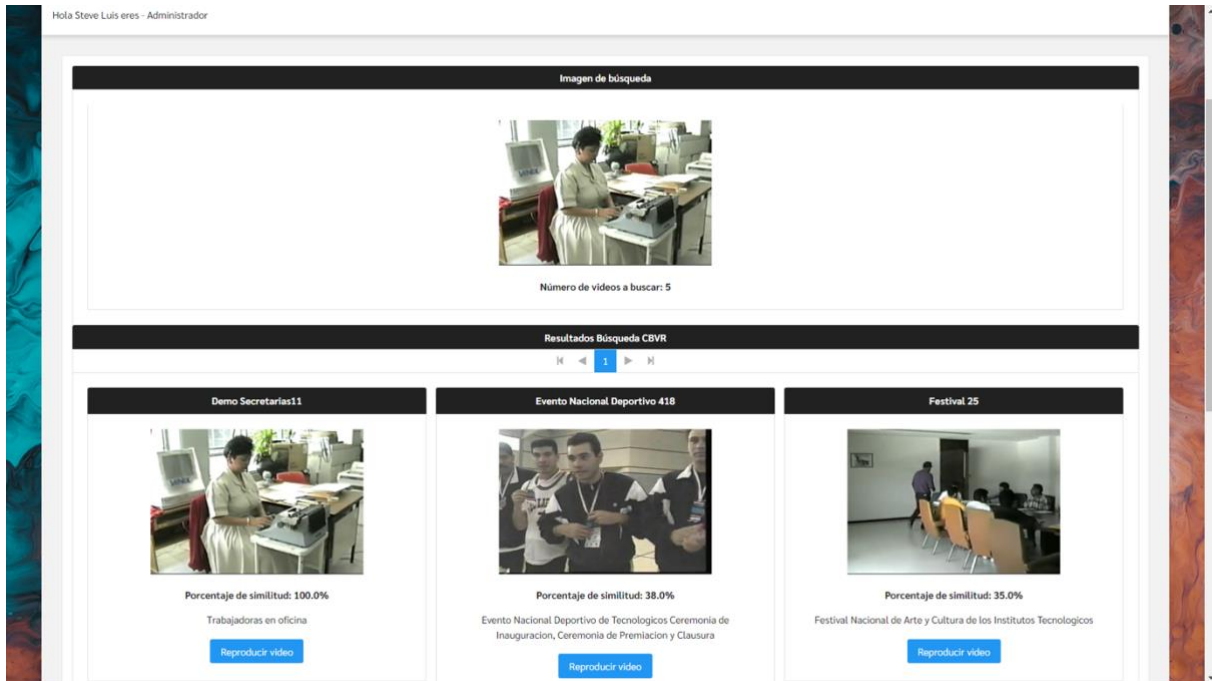


Figura 4.18 Resultado de búsqueda basada en contenido 1

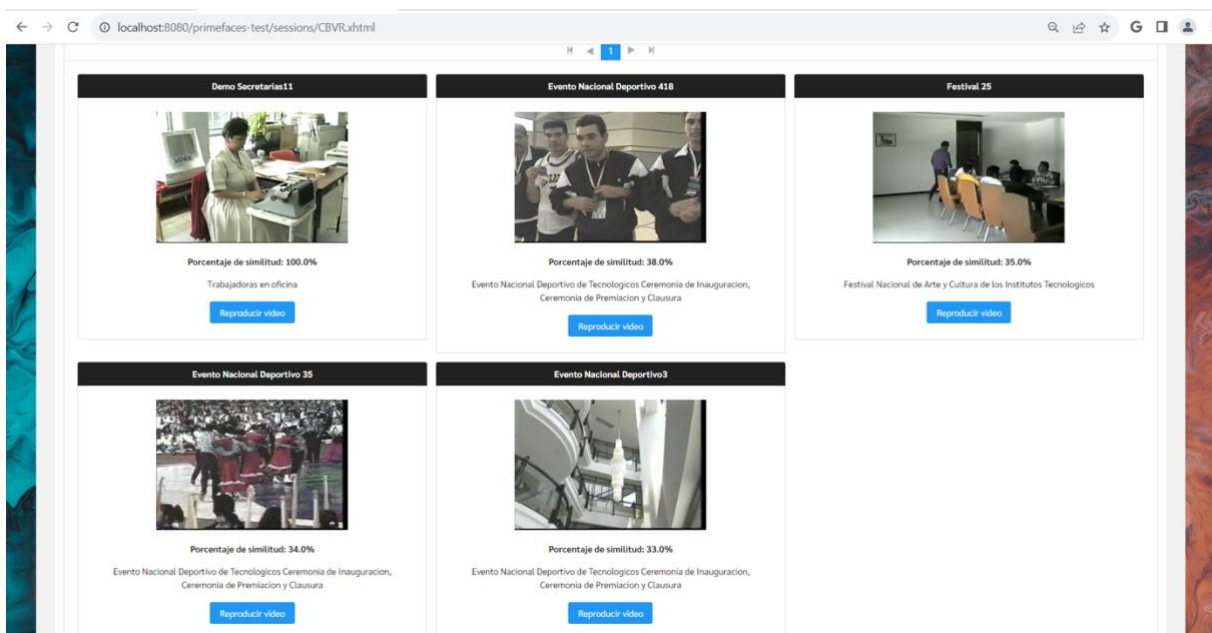


Figura 4.19 Resultado de búsqueda basada en contenido 2

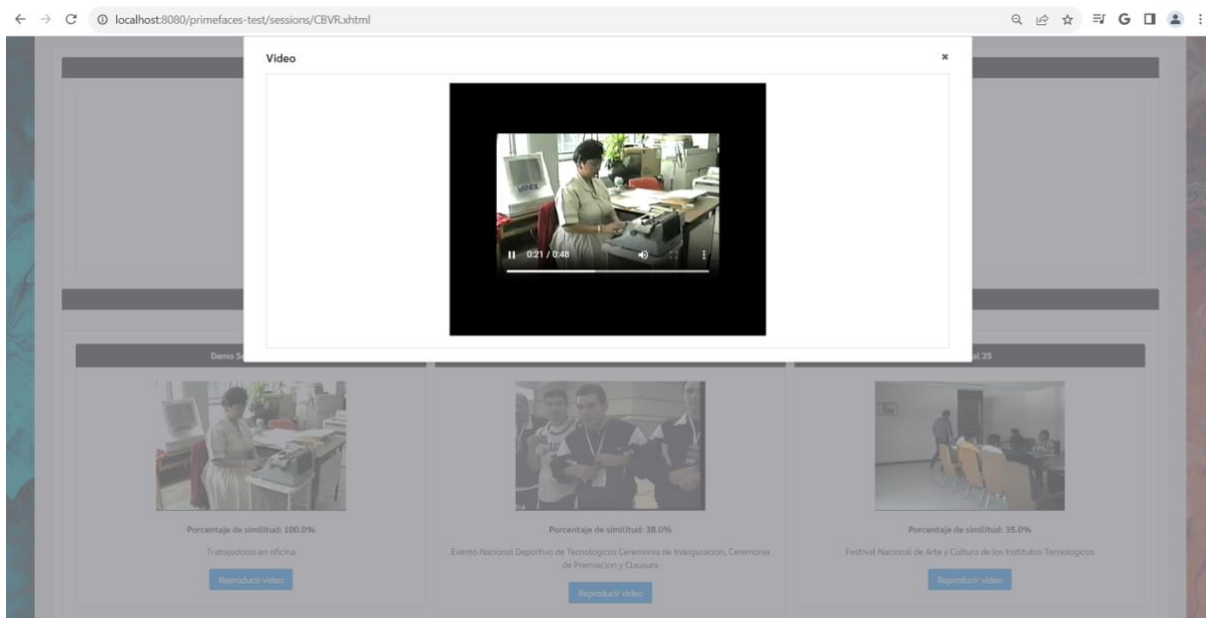


Figura 4.20 Reproductor de video en resultado

Se muestra el formulario de recuperación de contenido basada en texto en la Figura 4.21, donde se escribe el texto de consulta y se selecciona el tipo de contenido que se desea buscar, ya sea imágenes o videos tal como se presenta en las Figuras 4.22 y 4.23.

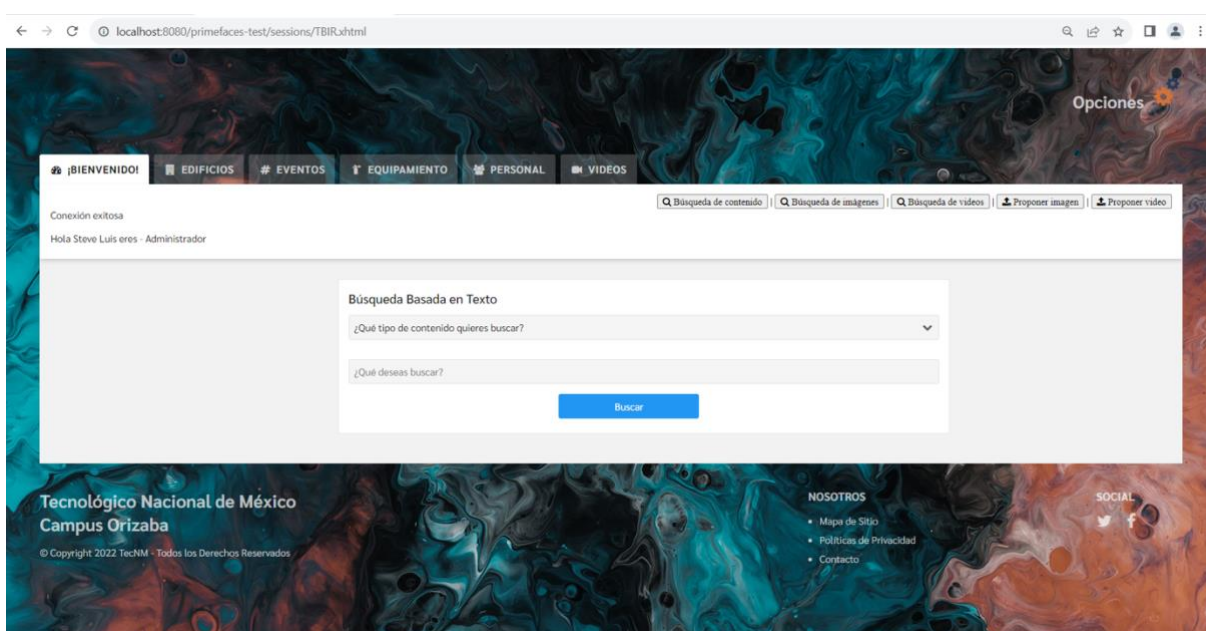


Figura 4.21 Página búsqueda basada en texto

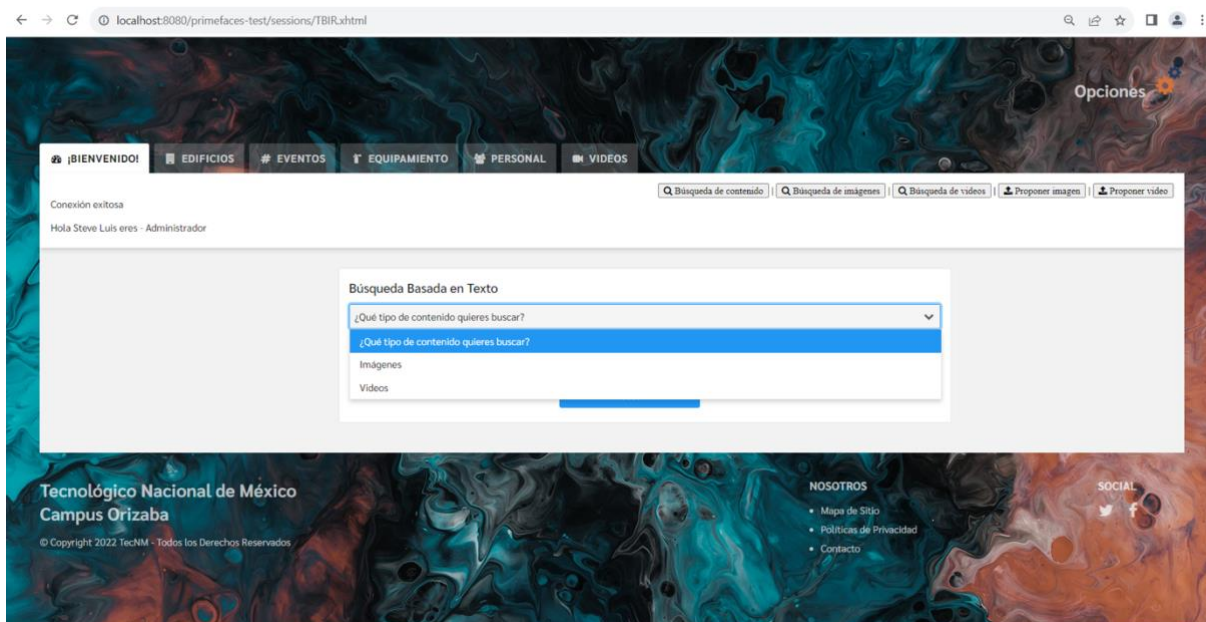


Figura 4.22 Selección búsqueda basada en texto

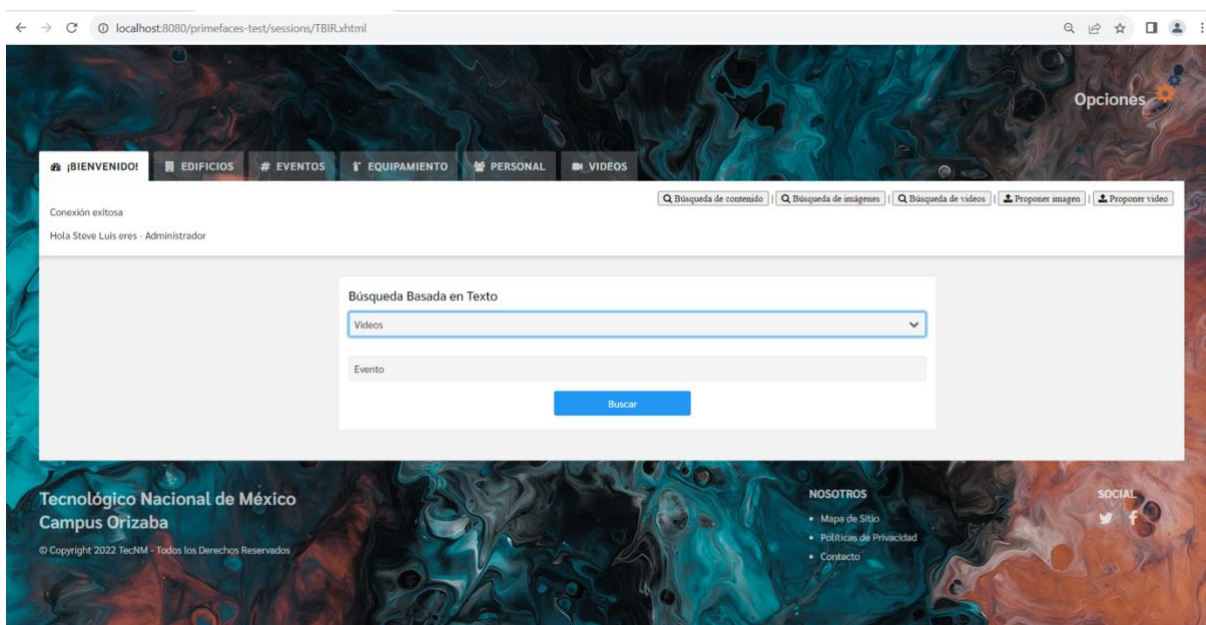


Figura 4.23 Búsqueda basada en texto

En la Figura 4.24, se presentan los resultados de la recuperación de contenido basada en texto y posteriormente en la Figura 4.25, se observa la reproducción de los resultados mostrados de la búsqueda basada en texto.

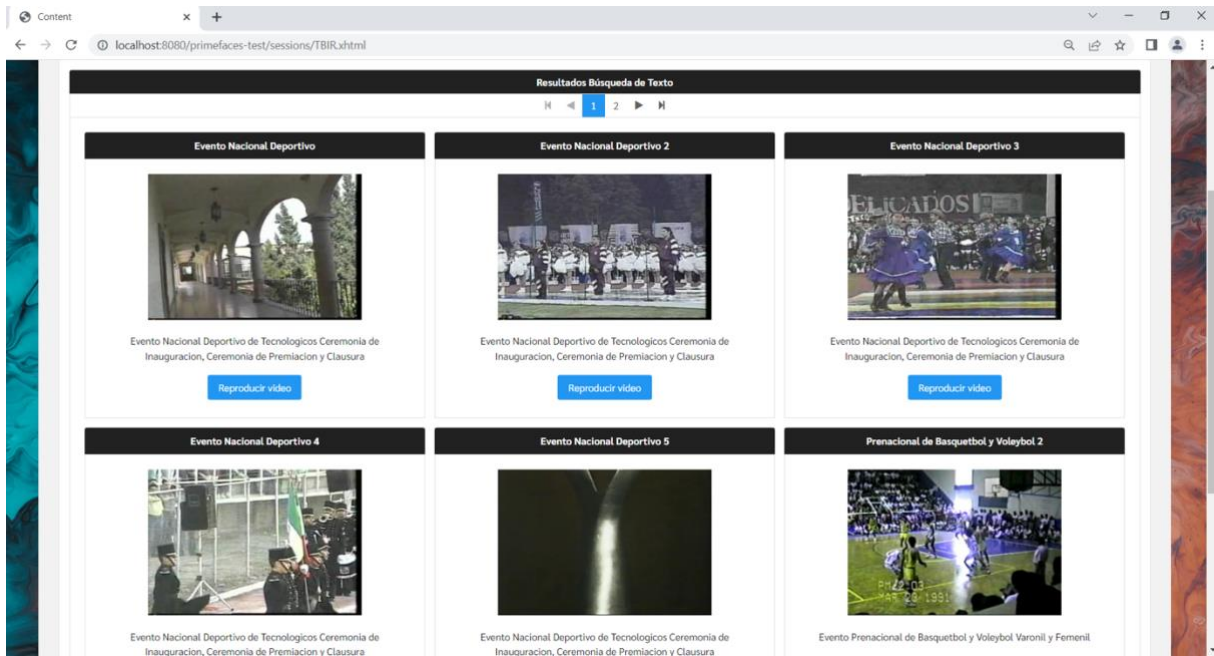


Figura 4.24 Resultado búsqueda basada en texto

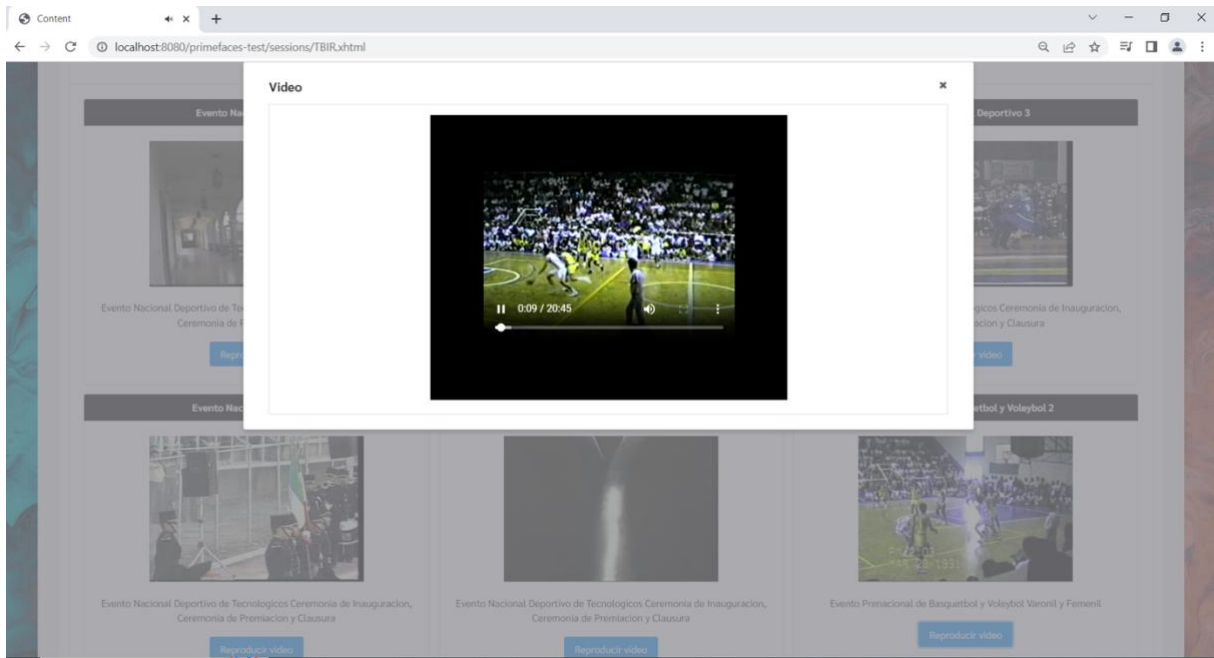


Figura 4.25 Reproducción resultado basada en texto

Capítulo 5. Conclusiones y recomendaciones

En este capítulo se describe las conclusiones después de haber realizado el módulo de la recuperación de videos con base en contenido y las recomendaciones para futuros trabajos.

5.1 Conclusiones

Este trabajo tiene un impacto social porque permite aumentar el conocimiento sobre la historia del Instituto Tecnológico de Orizaba. Además, esta iniciativa también genera un impacto económico, ya que, al utilizar un método de CBVR se obtienen mejores resultados en cuanto a consultas de videos, permitiendo el acceso al contenido multimedia solicitado por los usuarios (profesores, estudiantes, entre otros) de una manera más efectiva y fácil.

La recuperación de videos con base en contenido (CBVR) es relevante en las bases de datos multimedia, ya que permite obtener videos similares proporcionando una imagen o video como entrada, esto logra resultados más precisos comparado con las consultas basadas en texto. Por tal motivo, en este proyecto se agregó un módulo para CBVR en el sistema de gestión de datos multimedia de HITO.

Uno de los principales problemas en el desarrollo de un método de CBVR es la determinación de los fotogramas clave de los videos, ya que es necesario intercambiar precisión por eficiencia, puesto que el aumento del número de fotogramas obtenidos permitirá mejorar la precisión de las consultas, pero la recuperación de los videos será más lenta. Por tal motivo, en este trabajo se tomaron 25 fotogramas de cada video distribuidos a través de la duración del video. De esta forma, el número de fotogramas clave no depende del tamaño del video y se obtienen fotogramas muy diferentes entre sí. Además, se propuso un algoritmo para CBVR, en el que para cada video se encuentra el fotograma más similar con la imagen de búsqueda y se recuperan los k videos correspondientes a los k mejores fotogramas.

El módulo CBVR se desarrolló siguiendo la metodología UWE, utilizando el Sistema Gestor de Bases de Datos MongoDB, el lenguaje de programación Java, el marco de trabajo JSF y las bibliotecas de visión artificial OpenCV y BoofCV.

5.2 Recomendaciones

El módulo CBVR tiene mucho potencial para hacer varias mejoras a futuro, una de ellas es extender el módulo para varios formatos de video, ya que por el momento el sistema funciona adecuadamente con videos en formato MP4.

También como trabajo a futuro se propone implementar el reconocimiento facial, esto debido a que se tiene una amplia colección de fotografías del personal del Instituto Tecnológico de Orizaba. Además, se recomienda analizar otros descriptores para integrarlos en el módulo desarrollado como HOG, SIFT o CNN para mejorar la efectividad de la recuperación de videos. Además, se recomienda implementar estrategias que permitan garantizar la seguridad de los videos y su privacidad evitando que sea posible su descarga.

Finalmente, para una óptima gestión se recomienda que el tamaño de los videos sea menor a 500 MB y como trabajo futuro sería conveniente implementar en el sistema que al proponer un video se divida en partes si el tamaño del video es mayor a 500 MB. El módulo funciona hasta con videos con una duración mayor a 3 horas y con un tamaño mayor a 900 MB.

Productos Académicos



Gustavo Steve Luis Peña Riquer, Lisbeth Rodríguez Mazahua, Mario Leoncio Arrioja Rodríguez, Ulises Juárez Martínez, María Antonieta Abud Figueroa.

Revisión de Métodos de Recuperación de Video con base en Contenido

FEGLININ (ISSN 2594-2298), Número 23, Vol. 3.3, pp. 56-63, 2022.

Estado: *Publicado*



Gustavo Steve Luis Peña Riquer, Lisbeth Rodríguez Mazahua, Mario Leoncio Arrioja Rodríguez, Ulises Juárez Martínez, María Antonieta Abud Figueroa.

“MÓDULO DE RECUPERACIÓN DE VIDEOS BASADA EN CONTENIDO DE LA BASE DE DATOS MULTIMEDIA HITO”.

2023.

Estado: *En proceso*

Referencias

- [1] E. M. Saoudi and S. Jai-Andaloussi, "A distributed Content-Based Video Retrieval system for large datasets," *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00479-x.
- [2] T. Potluri and N. Ganeswara Rao, "Content Based Video Retrieval Using SURF, BRISK and HARRIS Features for Query-by-image," in *Communications in Computer and Information Science*, 2019, vol. 1035, pp. 265–276. doi: 10.1007/978-981-13-9181-1_24.
- [3] C. Zhang, Y. Lin, L. Zhu, A. Liu, Z. Zhang, and F. Huang, "CNN-VWII: An Efficient Approach for Large-Scale Video Retrieval by Image Queries," *Pattern Recognition Letters*, vol. 123, pp. 82-88 Oct. 2018, [Online]. Available: <http://arxiv.org/abs/1810.06030>
- [4] T. Prathiba and R. S. S. Kumari, "Content based video retrieval system based on multimodal feature grouping by KFCM clustering algorithm to promote human–computer interaction," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 6, pp. 6215–6229, Jun. 2021, doi: 10.1007/s12652-020-02190-w.
- [5] R. S. Ram, S. A. Prakash, M. Balaanand, and C. B. Sivaparthipan, "Colour and orientation of pixel based video retrieval using IHBM similarity measure," *Multimedia Tools and Applications*, vol. 79, no. 15–16, pp. 10199–10214, Apr. 2020, doi: 10.1007/s11042-019-07805-9.
- [6] M. Joaquín Rodríguez-Arauz, L. Rodríguez-Mazahua, M. Leoncio Arriola-Rodríguez, M. Antonieta Abud-Figueroa, S. Gustavo Peláez-Camarena, and L. del Carmen Martínez-Méndez, "Design of a Multimedia Data Management System that Uses Horizontal Fragmentation to Optimize Content-based Queries", IMMM 2020, The Tenth Internacional Conference on Advances in Information Mining and Management, pp. 15-21.
- [7] A. Araujo and B. Girod, "Large-Scale Video Retrieval Using Image Queries," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 6, pp. 1406–1420, Jun. 2018, doi: 10.1109/TCSVT.2017.2667710.

- [8] S. Hamad, A. S. Farhan, and D. Y. Khudhur, "Content based video retrieval using discrete cosine transform," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, no. 2, pp. 839–845, 2020, doi: 10.11591/ijeecs.v21.i2.pp839-845.
- [9] M. Mühling *et al.*, "Content-based video retrieval in historical collections of the German Broadcasting Archive," *International Journal on Digital Libraries*, vol. 20, no. 2, pp. 167–183, Jun. 2019, doi: 10.1007/s00799-018-0236-z.
- [10] B. Tseytlin and I. Makarov, "Content Based Video Retrieval System for Distorted Video Queries", *Proceedings of MacsPro*, 2020, vol.20. [Online]. Available: <http://ceur-ws.org>
- [11] N. v. Soniminde and S. D. Thepade, "Performance Improvement of Content-Based Video Retrieval using Thepade's Sorted n-ary Block Truncation Coding(TSnBTC) with Various Color Spaces", *IEEE International Conference on Communication information and Computing Technology (ICCICT)* Aug. 2021, pp. 1–5. doi: 10.1109/iccict50803.2021.9509934.
- [12] Chen *et al.*, A supervised video hashing method based on a deep 3d convolutional neural network for large-scale video retrieval. *Sensors*, Oct 2021 (9), 3094.
- [13] R. M. Bommisetty and A. Khare, "Content based video retrieval using histogram of gradients and frame fusion," *SPIE 11433, Twelfth International Conference on Machine Vision (ICMV 2019)* Jan. 2020, p. 65. doi: 10.1117/12.2558559.
- [14] V. Kumar, V. Tripathi, and B. Pant, "Content based Surgical Video Retrieval via Multi-Deep Features Fusion," *IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)* Dec. 2021, pp. 1–5. doi: 10.1109/conecct52877.2021.9622562.
- [15] N. Kaliaperumal, A. Das, and V. Balakrishnan, "A Content-Based Retrieval Model with Combinational Features and Indexing for Distributed Video Objects," *International Journal of Engineering Research and Technology*, volumen 13, number 12 (2020), pp. 5142-5148. [Online]. Available: <http://www.irphouse.com>
- [16] Jiang *et al.*, Learning segment similarity and alignment in large-scale content based video retrieval. In *Proceedings of the 29th ACM International Conference on Multimedia*, Oct 2021 (pp. 1618-1626).

- [17] J. Shao, X. Wen, B. Zhao, C. Wang y X. Xue, "Context Encoding for Video Retrieval with Contrastive Learning". *arXiv preprint arXiv:2008.01334*. 2020
- [18] L. Rossetto, A. Parian, M. Gasser, R. Giangreco y I. Heller, "Deep learning-based concept detection in vitivr", *Springer Int. Publishing, MultiMedia Modeling: 25th International Conference, Thessaloniki, Greece*, p. 616–621, enero de 2019.
- [19] Y. Sharma y M. Reetu, "Designing Content based Video Retrieval System for Real Time Applications", *J. Comput. Inf. Syst.*, p. 118–127, 2019.
- [20] W. Jo, G. Lim, J. Kim, J. Yun y Y. Choi, "Exploring the Temporal Cues to Enhance Video Retrieval on Standardized CDVA", *IEEE*, 2022. 38973-38981.
- [21] G. Kordopatis-Zilos, S. Papadopoulos, I. Patras y I. Kompatsiaris, "FIVR: Fine-grained incident video retrieval", *IEEE, Transactions on Multimedia*, 21(10), 2638-2652, 2019.
- [22] Y. Zhou, M. Wang, R. Wang y S. Huo, "Graph neural network for video relocalization", *arXiv preprint arXiv:2007.09877*, 2020.
- [23] V. Gabeur, A. Nagrani, C. Sun, K. Alahari y C. Schmid, "Masking modalities for cross-modal video retrieval", *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vision*, p. 1766–1775, 2022.
- [24] X. He, Y. Pan y M. Tang, "Self-supervised Video Retrieval Transformer Network", *arXiv preprint arXiv:2104.07993*, 2021.
- [25] D. Yasin, A. Sohail y I. Siddiqi, "Semantic video retrieval using deep learning techniques", *17th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST) IEEE*, p. 338–343, enero de 2020.
- [26] J. Shao, X. Wen, B. Zhao y X. Xue, "Temporal context aggregation for video retrieval with contrastive learning", *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vision*, p. 3268–3278, 2021.
- [27] H. Zhang, Y. Ji, W. Huang y L. Liu, "Sitcom-star-based clothing retrieval for video advertising: a deep learning framework", *Neural comput. applications*, 31(11), vol. 7361-7380, 2019.
- [28] G. Amato *et al.*, "The visione video search system: exploiting off-the-shelf text search engines for large-scale video retrieval", *J. Imaging*, 7(5), 76, 2021.

- [29] G. Wu *et al.*, "Unsupervised deep video hashing via balanced code for large-scale video retrieval", *IEEE Trans. Image Processing*, 28(4), p. 1993–2007, 2018.
- [30] D. Liang, L. Lin, R. Wang, J. Shao, C. Wang y Y. Chen, "Unsupervised teacher-student model for large-scale video retrieval", *Proc. IEEE/CVF Int. Conf. Comput. Vision Workshops*, 2019.
- [31] G. Kordopatis-Zilos, S. Papadopoulos, I. Patras y I. Kompatsiaris, "Visil: Fine-grained spatio-temporal video similarity learning", *Proceedings IEEE/CVF Int. Conf. Computing. Vision*, p. 6351–6360, 2019.
- [32] R. Bansal y S. Chakraborty, "Visual content based video retrieval on natural language queries", *Proceedings. 34th ACM/SIGAPP Symp. Appl. Computing.*, p. 212–219, abril de 2019.
- [33] W. Tan, H. Guo y R. Liu, "A fast partial video copy detection using KNN and global feature database", *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vision*, p. 2191–2199, 2022. *lications of Computer Vision* (pp. 2191-2199).

Anexo

En las siguientes tablas se detallan las pruebas realizadas en el sistema donde se muestra la duración de los videos, su tamaño, así como también el tiempo de inserción y el tiempo de búsqueda. Se realizó tres veces cada uno y el dato mostrado es el promedio.

Para estas pruebas, se utilizó una computadora DELL Latitude 7420 con las siguientes características:

- Procesador 11th Gen Intel Core i7 3.00 GHz 1.80 GHz
- RAM 16 GB
- Disco 500 GB
- Windows 10 Pro
- Sistema operativo de 64 bits

Tabla Inserción

Número de video	Duración de video (min)	Tamaño (MB)	Tiempo de inserción (ms)
1	0.48	2.76	4
2	2.15	11.1	2
3	4.54	13.8	5
4	7.55	23	18
5	9.38	40.1	13
6	5.48	41.9	10
7	7.57	42.4	9
8	8.36	45.1	10
9	10.09	47	10
10	19.56	56.3	18
11	8.19	59.3	10
12	20.45	68	18
13	1.16.02	77.6	33
14	45.39	86.7	49
15	22.44	87.2	18

Número de video	Duración de video (min)	Tamaño (MB)	Tiempo de inserción (ms)
16	48.37	103	105
17	1.02.46	110	58
18	37.15	116	67
19	1.09.26	117	40
20	1.36.21	126	277
21	1.03.24	133	132
22	7.55	177	118
23	1.54.07	182	124
24	1.46.53	184	54
25	1.02.56	184	56
26	1.40.08	186	26
27	1.34.28	186	113
28	1.22.21	197	88
29	56.09	203	118
30	1.29.39	203	215
31	1.39.38	208	182
32	1.29.22	210	215
33	1.14.13	211	167
34	1.29.05	215	109
35	1.23.49	217	195
36	1.24.56	219	104
37	1.29.49	226	86
38	2.03.48	227	39
39	1.31.18	230	90
40	1.27.52	240	101
41	1.58.44	247	336
42	1.29.22	270	122
43	2.02.13	271	350
44	1.39.16	273	105

Número de video	Duración de video (min)	Tamaño (MB)	Tiempo de inserción (ms)
45	2.02.18	278	145
46	1.31.08	279	240
47	1.55.57	283	163
48	1.25.52	291	114
49	1.46.48	291	117
50	1.48.16	301	165
51	1.44.30	310	287
52	1.09.14	320	75
53	1.42.13	321	142
54	1.48.20	321	297
55	2.05.22	331	168
56	2.02.31	348	153
57	2.02.03	349	152
58	2.03.30	357	192
59	1.51.07	357	313
60	2.02.30	360	195
61	1.46.04	367	309
62	2.03.33	372	194
63	2.03.23	379	179
64	1.39.17	382	274
65	2.01.44	393	358
66	1.08.15	400	40
67	1.16.02	409	801
68	1.50.49	559	116
69	1.59.59	667	193
70	3.11.36	909	387

Tabla Tiempo de recuperación

Número de videos en BD	K videos a recuperar	Tiempo de recuperación de contenido (ms)
10	1	20
10	2	19
10	5	22
10	10	22
10	Búsqueda texto	2.38
20	1	18
20	2	21
20	5	20
20	10	19
20	Búsqueda texto	3.04
30	1	24
30	2	26
30	5	26
30	10	25
30	Búsqueda texto	3.66
40	1	26
40	2	27
40	5	26
40	10	26
40	Búsqueda texto	4.01
50	1	27
50	2	26
50	5	26
50	10	27
50	Búsqueda texto	5.10
60	1	25
60	2	26
60	5	26

60	10	26
60	Búsqueda texto	6.06
70	1	26
70	2	27
70	5	27
70	10	27
70	Búsqueda texto	6.86
80	1	26
80	2	26
80	5	27
80	10	27
80	Búsqueda texto	7.26