



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO
FACULTAD DE CIENCIAS DE LA INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS

Proyecto de Investigación previo a la
obtención del título de Ingeniero en
Sistemas.

Título del Proyecto de Investigación:

**“GESTIÓN AUTOMÁTICA DE DOCUMENTOS DIGITALES BASADO EN EL
CONTENIDO TEXTUAL, PARA FACILITAR EL ACCESO A LA
INFORMACIÓN DE LA FACULTAD DE CIENCIAS DE LA INGENIERÍA.”**

Autores:

José Luis Agual Espín
Jonathan Aladino Mendoza Intriago

Director de Proyecto de Investigación:

Ing. Gleiston Cicerón Guerrero Ulloa
Quevedo – Los Ríos - Ecuador.

2016

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

Yo, **José Luis Agual Espín** y **Jonathan Aladino Mendoza Intriago**, declaramos que la investigación aquí descrita es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Universidad Técnica Estatal de Quevedo, puede hacer uso de los derechos correspondientes a este documento, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

f. _____

José Luis Agual Espín

C.C. 120600085-1

f. _____

Jonathan Aladino Mendoza Intriago

C.C. 131119493-8

CERTIFICACIÓN DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN

El suscrito **Ing. Gleiston Cicerón Guerrero Ulloa, Msc.**, ,Docente de la Universidad Técnica Estatal de Quevedo, certifica que los estudiantes **José Luis Agual Espín** y **Jonathan Aladino Mendoza Intriago**, realizaron el Proyecto de Investigación de grado titulado “**Gestión automática de documentos digitales basado en el contenido textual, para facilitar el acceso a la información de la Facultad de ciencias de la Ingeniería.**”, previo a la obtención del título de Ingeniero en Sistemas, bajo mi dirección, habiendo cumplido con las disposiciones reglamentarias establecidas para el efecto.

Ing. Gleiston Cicerón Guerrero Ulloa.

DIRECTOR DEL PROYECTO DE INVESTIGACIÓN

CERTIFICADO DEL REPORTE DE LA HERRAMIENTA DE PREVENCIÓN DE COINCIDENCIA Y/O PLAGIO ACADÉMICO.

INFORME DEL DIRECTOR DE PROYECTO DE INVESTIGACIÓN SOBRE EL SISTEMA URKUND

Ing. Gleiston Cicerón Guerrero Ulloa, Msc., en calidad de director del proyecto de investigación: **“GESTIÓN AUTOMÁTICA DE DOCUMENTOS DIGITALES BASADO EN EL CONTENIDO TEXTUAL, PARA FACILITAR EL ACCESO A LA INFORMACIÓN DE LA FACULTAD DE CIENCIAS DE LA INGENIERÍA.”**, me permito manifestar a usted lo siguiente:

Que, los Sres. AGUAL ESPÍN JOSÉ LUIS Y MENDOZA INTRIAGO JONATHAN ALADINO, egresados de la Facultad de Ciencias de la Ingeniería, han cumplido con las correcciones pertinentes, e ingresado su proyecto de investigación al sistema URKUND, tengo a bien certificar la siguiente información sobre el informe del sistema anti plagio con un porcentaje del 6%, como se observa en la siguiente figura:

URKUND	
Documento	Proyecto de investigacion copia-Final.docx (D20954878)
Presentado	2016-06-27 19:50 (-05:00)
Presentado por	luisagualespin@gmail.com
Recibido	gguerrero.uteq@analysis.urkund.com
Mensaje	Titulacion Mostrar el mensaje completo
	6% de esta aprox. 48 páginas de documentos largos se componen de texto presente en 10 fuentes.

.....
Ing. Gleiston Cicerón Guerrero Ulloa

DIRECTOR DEL PROYECTO DE INVESTIGACIÓN



UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO
FACULTAD DE CIENCIAS DE LA INGENIERÍA
CARRERA DE INGENIERIA EN SISTEMAS

PROYECTO DE INVESTIGACION

Título:

“Gestión automática de documentos digitales basado en el contenido textual, para facilitar el acceso a la información de la Facultad Ciencias de la Ingeniería.”

Presentado a la Consejo Académico de la Facultad como requisito previo a la obtención del título de Ingeniero en Sistemas.

Aprobado por:

Phd. Pavel Novoa Hernández

PRESIDENTE DEL TRIBUNAL

Ing. Eduardo Samaniego Mena

MIEMBRO DEL TRIBUNAL

Ing. Ariosto Vicuña Pino

MIEMBRO DEL TRIBUNAL

QUEVEDO – LOS RIOS – ECUADOR

2016

AGRADECIMIENTO

A Dios y mi familia por el apoyo, dedicación y motivación ofrecida a lo largo de los años transitados, por la universidad, siendo mis soportes principales en los momentos difíciles y compartiendo mi felicidad en los éxitos alcanzado en este camino.

A la Universidad Técnica Estatal de Quevedo por ser parte de mi vida profesional, y abrirme las puertas para obtener este título.

A mi estimado director Ing. Gleiston Cicerón Guerrero Ulloa, docente guía de este proyecto de investigación que con su enseñanza, exigencia y asesoría me ha orientado en este proceso para alcanzar mi objetivo.

Jonathan Aladino Mendoza Intriago

A Dios por permitirme existir.

A mis padres y hermanos por haberme apoyado y empujado siempre en todo lo que inicié.

Al Ing. Gleiston Cicerón Guerrero Ulloa, director de este proyecto de investigación que, con sus sabios conocimientos, la asesoría y paciencia han permitido que este proyecto se haga realidad.

Al PhD. Pavel Novoa Hernández quien nos brindó todo el apoyo necesario, con sus conocimientos, sus consejos, permitió culminar el presente proyecto de investigación.

Y a aquellas personas y amistades que de una u otra manera me ayudaron a seguir adelante.

José Luis Agual Espín

DEDICATORIA

A Dios quien supo guiarme por el buen camino, porque a pesar de las dificultades y momentos difíciles me dió las fuerzas, gana y motivo para salir adelante y no desmayar en los problemas que se presentaban, enseñándome a encarar las adversidades y de esta manera poder cumplir mis objetivos.

A mi madre, por ser la base que estructura mi vida y el ejemplo a seguir por todas las enseñanzas brindadas y hacer de mí una mejor persona a través de sus consejos.

A mi esposa Stefanía Navarrete por el amor, apoyo incondicional los cuales fueron mi pilar principal en este camino de alcanzar mi sueño.

A mis hijos Josenka y Dylan, por ser la razón principal y el motivo más importante de inspiración para romper todos los obstáculos y cumplir con esta meta.

Jonathan Aladino Mendoza Intriago

A Dios por guiarme en todo momento y permitirme cumplir una meta más en la vida.

A mis padres Manuel Agual y María Espín quienes nunca dejaron de creer en que podría lograr todas las metas que me proponga y siempre estuvieron para apoyarme en esta travesía, por haberme dado la educación, por inculcarme valores, por las sonrisas, las broncas, los besos, los abrazos, por la confianza y el esfuerzo brindado hacia mí, por todo ello y por el hecho de haberme dado la vida, nunca me cansaré en decirles gracias, gracias papá y gracias mamá.

A la persona especial que llegó a mi vida, gracias por animarme, apoyarme y creer en mí siempre, por estar a mi lado.

José Luis Agual Espín

RESUMEN

La clasificación de documentos en el área de la Inteligencia Artificial (IA), consiste en organizar los archivos de manera automática mediante un modelo de aprendizaje generado por un algoritmo. En la Secretaría de Facultad de Ciencias de la Ingeniería (FCI), la gestión de los documentos digitales se realiza de manera manual. Esto provoca el consumo de recursos al almacenar los archivos, duplicaciones, eliminación, alteraciones en los documentos y al no contar con un módulo de respaldo de la documentación se dificulta la restauración de los documentos alterados y eliminados, aumentando la carga de trabajo para la persona encargada. La búsqueda de archivos es un proceso que se debe agilizar en la secretaría, por el incremento de los documentos digitales de los trámites que se generan en la FCI. Por tal razón se debe automatizar este proceso a través de una aplicación que facilite las tareas que se realizan en dicha Secretaría.

Se propuso diseñar una herramienta que mejore la gestión de los documentos digitales en la FCI, implementado diferentes técnicas de la IA como; el algoritmo clasificación de documentos Naive Bayes, permitiendo obtener una organización adecuada de los archivos y clasificándolos de manera automática, la biblioteca Apache Lucene para optimizar el proceso de búsqueda de los mismos y la biblioteca OpenNLP que permitió el procesamiento de los diferentes tipos de archivos.

Para el desarrollo de este sistema clasificación de automática de documentos digitales se realizó un análisis de las tareas que realiza la Secretaria de la FCI, para la implementación se elaboró un módulo de clasificación con el algoritmo Naive Bayes y una interfaz de fácil manejo bajo la plataforma de desarrollo Netbeans.

Palabras Clave: lucene, clasificación automática, naive bayes, opennlp, inteligencia artificial.

ABSTRACT

The classification of documents in the area of Artificial Intelligence (AI), consists of organize files automatically through a learning model generated by an algorithm. In the Secretaria of the Faculty of Engineering Sciences (FCI), the management of digital files is done manually. This causes resource consumption by storing files, duplication, deletion, changes in the documents and not having a support module documentation restoration of altered documents is difficult and eliminated, thus increasing the workload for the person in charge thereof. File searching is a process that should speed up in the secretariat, by the increase of digital documents paperwork generated in the FCI. For this reason you should automate this process through an application that facilitates the tasks performed in the Secretariat of the FCI.

It was proposed to design a tool that improves the management of digital documents in the FCI, implemented different AI techniques like; the algorithm document classification Naive Bayes, allowing to obtain an adequate organization of files and sorting them automatically, the Apache Lucene library to optimize the search process them and OpenNLP library that allowed processing different types of files.

For the development of this classification system automatic digital document an analysis of the tasks performed by the Secretariat of the FCI, to implement a module classification was developed with the algorithm Naive Bayes and an interface easy to use underperformed the Netbeans development platform.

Keywords: lucene, classification automatic, naive Bayes, OpenNLP, intelligence artificial.

TABLA DE CONTENIDO

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS	iii
CERTIFICACIÓN DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN.....	iv
CERTIFICADO DEL REPORTE DE LA HERRAMIENTA DE PREVENCIÓN DE COINCIDENCIA Y/O PLAGIO ACADÉMICO.	v
AGRADECIMIENTO	vii
DEDICATORIA.....	viii
RESUMEN	ix
ABSTRACT	x
TABLA DE CONTENIDO	xi
INDICE DE FIGURAS	xv
INDICE DE TABLAS.....	xvii
INDICE DE ECUACIONES	xviii
CODIGO DUBLIN	xix
INTRODUCCIÓN.....	1
CAPITULO I.....	2
1.1. Problema de investigación.....	3
1.1.1. Planteamiento del problema.	3
1.1.2. Formulación del problema.....	5
1.1.3. Sistematización del problema.....	6
1.2. Objetivos.....	6
1.2.1. Objetivo General.....	6
Objetivos Específicos.....	6
1.3. Justificación.	7
CAPÍTULO II.....	8
FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN	8
2.1. Marco conceptual.....	9
2.1.1. Recuperación de Información.....	9
2.1.2. Clasificación automática de textos.	9
2.1.3. Tipos de clasificadores automáticos.....	10
2.1.3.1. Clasificación Supervisada.	10
2.1.3.2. Clasificación no supervisada.....	11
2.1.4. Técnicas de clasificación automática de textos.	11

2.1.4.1.	Modelo de probabilístico de Naive Bayes.....	11
2.1.4.2.	Algoritmo basado en ejemplos k-vecinos más cercanos (K-NN, k-Narest Neighbor). 13	
2.1.4.3.	Árboles de decisión.	14
2.1.4.4.	Máquina de soporte vectorial o Máquina de vectores soporte.	15
2.1.4.5.	Algoritmo K-Means.	15
2.1.5.	Procesamiento del Lenguaje Natural.	16
2.1.6.	Fases de un clasificador automático de documentos.	17
2.1.6.1.	Pre-procesado y representación del documento.	17
2.1.6.2.	Reducción de dimensiones.	17
2.1.6.3.	Asignación de pesos.	18
2.1.6.4.	Entrenamiento.	18
2.1.6.5.	Clasificación.....	18
2.1.7.	Pre - procesado del documento y representación estructura.....	18
2.1.7.1.	Palabras “vacías” (StopWords).	18
2.1.7.2.	Expansión por sinónimos.	19
2.1.7.3.	Lematización (stemming).....	20
2.1.7.4.	Asignación de pesos.	20
2.1.8.	Xtreme Programming (XP).	21
2.1.8.1.	Ciclo de vida de XP.....	22
2.1.9.	Java.	22
2.1.10.	NetBeans.....	23
2.1.11.	Lucene.	23
2.1.12.	Lucene Analyzers.	23
2.1.13.	MySQL.	24
2.1.14.	Tika Version 11.0	24
2.1.15.	Poi.....	24
2.1.16.	Tm – extractors.....	25
2.1.17.	Apache OpenNLP.....	25
2.2.	Marco referencial.....	26
2.2.3.	Clasificación de textos académicos en función de su contenido léxico-semántico* (Academic text classification based on lexical-semantic content).....	26
2.2.4.	Clasificación de textos en lenguaje natural usando la Wikipedia.	26
2.2.5.	Marco de trabajo para Indexación, Clasificación y Recopilación Automática de Documentos Digitales.	26
2.2.6.	Clasificación Automática de Videos.	27

2.2.7.	Clasificación de documentos usando Maquinas de Vectores de Apoyo	27
2.2.8.	Minería de texto para la categorización automática de documentos	28
CAPÍTULO III	29
MÉTODOLOGÍA DE LA INVESTIGACIÓN.	29
3.1.	Localización.....	30
3.2.	Tipo de investigación.....	30
3.2.1.	Investigación de campo	30
3.2.2.	Investigación descriptiva.	31
3.2.3.	Investigación aplicada.	31
3.3.	Métodos de investigación.	32
3.3.1.	Método Analítico.....	32
3.3.2.	Método Deductivo.	32
3.3.3.	Método Inductivo.	33
3.4.	Fuentes de recopilación de información.	33
3.4.1.	Primarias.....	33
3.4.2.	Secundarias.....	33
3.5.	Diseño de investigación.	34
3.6.	Instrumento de investigación.....	34
3.7.	Recursos humanos y materiales	35
3.7.1.	Recursos Humanos	35
3.7.2.	Hardware y Software	35
3.8.	Metodología de desarrollo del Software.....	36
3.8.1.	Planificación de los requisitos del proyecto	36
3.8.1.1.	Representación conceptual de la aplicación.	37
3.8.1.2.	Diagrama de caso de uso	38
3.8.2.	Análisis y diseño del proyecto.....	38
3.8.2.1.	Casos de uso extendido	39
3.8.2.2.	Diagramas de secuencias.....	80
3.8.3.	Codificación del proyecto.....	83
3.8.3.1.	Definición de las categorías	83
3.8.3.2.	Entrenamiento por categoría.	84
3.8.3.3.	Pre – procesamiento en lenguaje natural de los archivos de entrenamiento	84
3.8.3.4.	Eliminar o suprimir los signos de puntuación.....	85
3.8.3.5.	Conversión de las palabras a minúsculas	87

3.8.3.6.	Eliminación de los signos de Acentuación de las palabras.	88
3.8.3.7.	Filtro de Palabras Vacías (StopWords).	88
3.8.3.8.	Stemming – Lematización de términos.	90
3.8.3.9.	Creación de Datasets por categorías.	91
3.8.3.10.	Clasificación de documentos aplicando el Algoritmo Naive Bayes.	92
3.8.3.11.	Indexación de los documentos.	92
3.8.3.12.	Búsqueda por categoría	92
3.8.3.13.	Búsqueda global	92
3.8.3.14.	Diagramas de Clase	93
3.8.4.	Pruebas del Sistema	97
CAPÍTULO IV	98
RESULTADOS Y DISCUSIÓN	98
4.1.	Resultados.	99
4.1.1.	Gestión de documentos digitales en la Secretaría de la FCI.	99
4.1.2.	Gestión de documentos digitales en la FCI, utilizando el aplicativo elaborado. 100	
4.2.	Pruebas y resultados.	104
4.3.	Discusión.	106
CAPITULO V	108
CONCLUSIONES Y RECOMENDACIONES	108
5.1.	Conclusiones.	109
5.2.	Recomendaciones.	110
CAPITULO VI	111
BIBLIOGRAFÍA	111
6.1.	Bibliografía	112
CAPITULO VII	115
ANEXOS	115
	Manual de Usuario del aplicativo “Gestión automática de documentos digitales basado en el contenido textual, para facilitar el acceso a la información de Facultad de Ciencias de la Ingeniería”	116

INDICE DE FIGURAS

Figura.- 1. Esquema de los sistemas de recuperación de información.	9
Figura.- 2. Tipos de clasificadores automáticos.	10
Figura.- 3. Clasificación supervisada.	10
Figura.- 4. Clasificación no – supervisada	11
Figura.- 5. Técnicas de clasificación automática de textos.	11
Figura.- 6. Ejemplo de KNN.	14
Figura.- 7. Árboles de decisión – Clasificación.	14
Figura.- 8. Fases de un clasificador de documentos.	17
Figura.- 9. Localización del proyecto de investigación.	30
Figura.- 10. Representación conceptual del sistema gestión automática de documentos digitales.	37
Figura.- 11. Casos de uso del Sistema Gestión automática de documentos digitales.	38
Figura.- 12. Diagrama de secuencia "Obtener el modelo de aprendizaje "	81
Figura.- 13. Diagrama de secuencia " Clasificar documentos "	81
Figura.- 14. Diagrama de secuencia " Búsqueda de documentos por contenido"	82
Figura.- 15. Diagrama de secuencia " Búsqueda de documentos por metadatos"	82
Figura.- 16. Diagrama de secuencia "Búsqueda Global de documentos"	83
Figura.- 17. Código de supresión de los signos de puntuación.	86
Figura.- 18. Código de conversión de las palabras a minúsculas.	87
Figura.- 19. Código de eliminación de Acentuación.	88
Figura.- 20. Filtrado de Palabras Vacías.	89
Figura.- 21. Lematización de Términos.	91
Figura.- 22. Diagramas de clase Utilidades, DeleteDoc , Searcher , database, StopWords, Indexador.	93
Figura.- 23. Diagramas de clase NaiveBayes, NaiveBayesKnowledgeBase, FeatureStats, FeatureExtraction, TextTokenizer	94
Figura.- 24. Diagrama de Clases ClasificarDoc, naivebayes,categorias,BusquedaDocumentos	95
Figura.- 25. Diagramas de clase ArchiEntren, categ, ExtensionFileFiltrer	96
Figura.- 26. Diagrama de paquetes	96
Figura.- 27. Diagrama de Despliegue	97
Figura.- 28. Situación actual de organizar los documentos en la Secretaría de la FCI.	100
Figura.- 29.- Generación del modelo de aprendizaje.	101
Figura.- 30. Clasificación automática de los archivos.	102
Figura.- 31. Búsqueda de los documentos.	103
Figura.- 32. Organización de los archivos mediante el sistema.	103
Figura.- 33. Funcionamiento del Sistema - Ingreso de la dirección ip de la Base de datos.	116
Figura.- 34. Funcionamiento del sistema - Inicio de sesión.	117
Figura.- 35. Funcionamiento del sistema - Seleccionar el directorio de archivos del servidor.	117
Figura.- 36. Funcionamiento del sistema - Interfaz principal (Usuario Administrador)...	117
Figura.- 37. Funcionamiento del sistema - Interfaz principal (Usuario Invitado).	118
Figura.- 38. Funcionamiento del sistema - Importar Archivo de Entrenamiento.	119

Figura.- 39. Funcionamiento del sistema - Importación de archivos de entrenamiento....	120
Figura.- 40. Funcionamiento del sistema - Manipulación de los archivos	121
Figura.- 41. Funcionamiento del sistema - Generar el modelo de aprendizaje.	121
Figura.- 42. Funcionamiento del sistema - Clasificar documentos.	122
Figura.- 43. Funcionamiento del sistema - Importar documentos a clasificar	123
Figura.- 44. Funcionamiento del sistema - Operaciones con los archivos importados.	124
Figura.- 45. Funcionamiento del sistema - Clasificar los documentos importados.....	124
Figura.- 46. Funcionamiento del sistema - Búsqueda de documentos.	125
Figura.- 47. Funcionamiento del sistema - Búsqueda por categoría.	125
Figura.- 48. Funcionamiento del sistema - Búsqueda por metadatos.....	126
Figura.- 49. Funcionamiento del sistema - Búsqueda Global.	127
Figura.- 50.-Funcionamiento del sistema - Manipulación de archivos.	127
Figura.- 51. Funcionamiento del sistema – Borradores de documentos.....	128
Figura.- 52. Funcionamiento del sistema - Manipular archivos	129
Figura.- 53. Funcionamiento del sistema - Restaurar archivo.....	129
Figura.- 54. Funcionamiento del sistema -Administrador de usuarios.....	130
Figura.- 55. Funcionamiento del sistema - Opciones de Administrador de usuarios.....	130
Figura.- 56. Funcionamiento del sistema - Crear nuevo usuario (invitado.)	131
Figura.- 57. Funcionamiento del sistema - Editar usuario.....	131
Figura.- 58. Funcionamiento del sistema - Eliminar usuario.	132
Figura.- 59. Funcionamiento del sistema - Vista registro de búsquedas.	133
Figura.- 60. Funcionamiento del sistema - Formulario Vista registro de búsquedas.	133
Figura.- 61. Funcionamiento del sistema –Búsqueda por categoría.....	134
Figura.- 62. Funcionamiento del sistema - Búsqueda global por fecha.	135
Figura.- 63. Funcionamiento del sistema – Manipulación de los archivos.	135
Figura.- 64. Funcionamiento del sistema - Respaldo de información.	136
Figura.- 65. Funcionamiento del sistema - Generar respaldo.....	136
Figura.- 66. Funcionamiento del sistema - Lista de respaldos generados.	137
Figura.- 67. Funcionamiento del sistema - Exportar respaldo.....	137
Figura.- 68. Funcionamiento del sistema. Elegir dirección donde exportar el respaldo. ..	138
Figura.- 69. Funcionamiento del sistema - Eliminar respaldo.....	138
Figura.- 70. Situación actual de la Gestión de documentos de la FCI.....	139
Figura.- 71. Organización de los documentos mediante el sistema desarrollado.....	139
Figura.- 72. Solicitud de permiso para proporcionarnos la documentación de la Secretaría de la FCI	140

INDICE DE TABLAS

Tabla 1. Listado de Palabras Vacías.....	19
Tabla 2. Listado de Lucene Analyzers	24
Tabla 3. Investigación descriptiva. Identifiación de causas y consecuencia.	31
Tabla 4. Hardware usado en el desarrollo.	35
Tabla 5. Software usado en el desarrollo.....	35
Tabla 6. Planificación de requisitos.....	36
Tabla 7. Inicio de Sesión	39
Tabla 8. Registrar Nuevo Usuario.	40
Tabla 9. Editar Usuario.....	42
Tabla 10. Eliminar Usuario.	44
Tabla 11. Nueva Categoría.	45
Tabla 12. Importar archivos de entrenamiento Único	47
Tabla 13. Importar varios archivos de entrenamiento.	49
Tabla 14. Importar carpeta de archivos de entrenamiento.....	51
Tabla 15. Entrenar archivos por categoría.....	53
Tabla 16. Importar archivo único.	54
Tabla 17. Importar Varios archivos.....	56
Tabla 18. Importar carpeta de archivos.	58
Tabla 19. Clasificar.....	60
Tabla 20. Búsqueda por categoria.	61
Tabla 21. Búsqueda por Metadatos.	64
Tabla 22. Búsqueda Global.	66
Tabla 23. Restaurar documentos.	68
Tabla 24. Búsqueda por categoría.	70
Tabla 25. Búsqueda global por fecha.	72
Tabla 26. Respalidar Información.	74
Tabla 27. Exportar respaldo.....	76
Tabla 28. Eliminar respaldo.....	78
Tabla 29. Información textual extraída de una resolución académica.	85
Tabla 30. Supresión de los signos de puntuación.....	86
Tabla 31. Conversión de las palabras a minúsculas.	87
Tabla 32. Eliminación de los signos de Acentuación.....	88
Tabla 33. DataSet de palabras en español.	89
Tabla 34. Proceso de filtrado de palabras vacías.....	90
Tabla 35. Proceso de stemming o lematización de términos.....	91
Tabla 36. Dataset generado..	91
Tabla 37. Ventajas y desventajas de los algoritmos estudiados.	105
Tabla 38. Prueba 1 Creación de carpetas.....	106
Tabla 39. Prueba 2 Organización de los archivos.	106
Tabla 40. Prueba 3 Búsqueda de los documentos.	106

INDICE DE ECUACIONES

Ecuación 1. Ecuación (MAP) Maximum a posteriori	12
Ecuación 2. Suma de los logaritmos.....	13
Ecuación 3. Representación de los datos de entrenamiento.	13
Ecuación 4. Formula de la distancia Euclidea.	14
Ecuación 5. Máquina de Soporte de vectores –Riesgo Empírico.....	15
Ecuación 6. Factores de Peso $tf - idf$	20
Ecuación 7. Ecuación del $tf - idf$	21
Ecuación 8. Ecuación del tf	21

CODIGO DUBLIN

Título:	Gestión automática de documentos digitales basados en el contenido textual, para facilitar el acceso a la información de la Facultad de Ciencias de la Ingeniería.			
Autores	José Luis Agual Espín, Jonathan Aladino Mendoza Intriago			
Palabra clave:	Sistema	Clasificación	documentos	Lucene
Fecha de Publicación:				
Editorial:	Quevedo: UTEQ, 2016.			
Resumen:	<p>La clasificación de documentos en el área de la Inteligencia Artificial (IA), consiste en organizar los archivos de manera automática mediante un modelo de aprendizaje generado por un algoritmo. En la Secretaría de Facultad de Ciencias de la Ingeniería (FCI), la gestión de los documentos digitales se realiza de manera manual. Esto provoca el consumo de recursos al almacenar los archivos, duplicaciones, eliminación, alteraciones en los documentos y al no contar con un módulo de respaldo de la documentación se dificulta la restauración de los documentos alterados y eliminados, aumentando la carga de trabajo para la persona encargada. La búsqueda de archivos es un proceso que se debe agilizar en la secretaría, por el incremento de los documentos digitales de los trámites que se generan en la FCI. Por tal razón se debe automatizar este proceso a través de una aplicación que facilite las tareas que se realizan en dicha Secretaría.</p> <p>Se propuso diseñar una herramienta que mejore la gestión de los documentos digitales en la FCI, implementado diferentes técnicas de la IA como; el algoritmo clasificación de documentos Naive Bayes, permitiendo obtener una organización adecuada de los archivos y clasificándolos de manera automática, la biblioteca Apache Lucene para optimizar el proceso de búsqueda de los mismos y la biblioteca OpenNLP que permitió el procesamiento de los diferentes tipos de archivos.</p> <p>Para el desarrollo de este sistema clasificación de automática de</p>			

	<p>documentos digitales se realizó un análisis de las tareas que realiza la Secretaria de la FCI, para la implementación se elaboró un módulo de clasificación con el algoritmo Naive Bayes y una interfaz de fácil manejo bajo la plataforma de desarrollo Netbeans.</p> <p style="text-align: center;">ABSTRACT</p> <p>The classification of documents in the area of Artificial Intelligence (AI), consists of organize files automatically through a learning model generated by an algorithm. In the Secretaria of the Faculty of Engineering Sciences (FCI), the management of digital files is done manually. This causes resource consumption by storing files, duplication, deletion, changes in the documents and not having a support module documentation restoration of altered documents is difficult and eliminated, thus increasing the workload for the person in charge thereof. File searching is a process that should speed up in the secretariat, by the increase of digital documents paperwork generated in the FCI. For this reason you should automate this process through an application that facilitates the tasks performed in the Secretariat of the FCI.</p> <p>It was proposed to design a tool that improves the management of digital documents in the FCI, implemented different AI techniques like; the algorithm document classification Naive Bayes, allowing to obtain an adequate organization of files and sorting them automatically, the Apache Lucene library to optimize the search process them and OpenNLP library that allowed processing different types of files.</p> <p>For the development of this classification system automatic digital document an analysis of the tasks performed by the Secretariat of the FCI, to implement a module classification was developed with the algorithm Naive Bayes and an interface easy to use underperformed the Netbeans development platform.</p>
Descripción:	
URI:	

INTRODUCCIÓN

La generación de información digital ha dejado atrás el uso de documentos físicos por lo que las empresas han recurrido de manera exponencial al uso de documentos en formatos electrónicos; como resoluciones, memorandos, solicitudes, sin embargo, este crecimiento de información ha hecho que se tengan dificultades en el almacenamiento y recuperación de los mismos. Debido a la carencia de un mecanismo semi-automático que gestione los documentos, el encargado de este proceso se enfrenta a varios retos, que imposibilita realizar su trabajo de forma adecuada.

En la Facultad de Ciencias de la Ingeniería (FCI) habitualmente se generan muchos documentos importantes tanto para el personal administrativo como para los estudiantes, la masiva cantidad de archivos ha ocasionado que sufran vulnerabilidades de integridad y seguridad, al no contar con un lugar idóneo de almacenamiento.

La inadecuada forma de clasificar los documentos sin técnica alguna ha provocado inconvenientes al momento de realizar una búsqueda, obteniendo resultados que no superan las expectativas, recurriendo al proceso manual de acceder al contenido de cada documento para determinar su utilidad y veracidad de la información.

Los usuarios para obtener acceso a la información de manera oportuna, proceden a leer varios fragmentos de los mismos en lugar del texto completo. De esta forma, muchos de los documentos son siempre desconocidos por los usuarios; por lo tanto, de nada sirve tener información íntegra en formato digital, si no se tiene una correcta gestión de documentos, con la finalidad que cualquier usuario pueda tener la información de una forma ágil y eficiente.

Por tal motivo se propone la gestión automática de documentos, aplicando el aprendizaje automático identificar a qué categoría pertenece cada documento que son generados en la FCI y a través de la revisión del contenido textual agilizar el proceso de búsqueda. Este proyecto pretende involucrarse en el área de pre-procesamiento, clasificación automática e indexación de documentos mediante las técnicas, algoritmos y las diferentes metodologías para el manejo de la información que nos proporciona la Inteligencia Artificial (IA).

CAPITULO I
CONTEXTUALIZACIÓN DE LA INVESTIGACIÓN

1.1. Problema de investigación.

1.1.1. Planteamiento del problema.

“La gestión documentos se encarga de controlar de manera eficiente y sistemática la producción, organización, conservación, acceso, disposición y destino final de los documentos que han sido producto de las funciones de una organización” [1]. La norma *ISO 15489* es el ente encargado de regular “la gestión de documentos que producen las organizaciones, ya sean públicas o privadas, con fines externos o internos, independiente de su formato y de las tecnologías utilizadas, creados o recibidos por una organización transcurso de sus actividades” [2].

En la Secretaría de Facultad de Ciencias de la Ingeniería (FCI), se manejan diferentes documentos ya sean creados por la misma o recibidos de otros departamentos de la Facultad, el masivo incremento de los archivos ha ocasionado que se presenten algunos problemas en el manejo de la documentación digital que en esta dependencia se gestionan.

La metodología de clasificación de los documentos que posee la secretaría de la Facultad, es una tarea que se está llevando de manera inadecuada, al tratar de clasificar los archivos con realizar una breve revisión del mismo y decidir a qué trámite pertenece, lo cual no cumple con lo dispuesto en la fase de clasificación de la norma *ISO 15489*, el mismo que menciona que “se ha identificar la categoría a la que pertenece un documento, identificando la operación o actividad que el documento atestigua y ubicar la operación o actividad en el sistema de clasificación”.

La inadecuada conservación de los documentos en la FCI, ha ocasionado problemas significativos como la duplicación de archivos, pérdidas de documentos por permanecer almacenados en el computador donde labora la secretaria, utilizando de manera excesiva los recursos de disco duro del computador, y a su vez no cumple con la norma *ISO 15489*, en la fase de almacenamiento nos manifiesta que “los documentos se deben conservar en un entorno seguro, garantizando su autenticidad, integridad, fiabilidad de los mismos ”.

Cuando un usuario se acerca a solicitar un trámite, el encargado de la secretaría procede a la exploración de los archivos, la recuperación o búsqueda de los archivos se realiza mediante un filtro con las palabras claves referente a la información que se necesita, en algunos casos

obteniendo resultados positivos en otro no, proceso que implica utilizar cierto tiempo para la naturaleza de lo requerido y resulta tediosa para quien posee esta responsabilidad, incumpliendo con lo establecido en la *ISO 15489*, que manifiesta a “la indización como la técnica de recuperación de documentos en las clasificaciones, categorías y soportes ,el cual permite localizar los documentos mediante la búsqueda en su contenido”.

Diagnóstico.

Al tratar de clasificar los archivos, revisando de manera breve el documento, produce que los archivos sean clasificados en categorías que no van de acuerdo a la naturaleza del documento, al no realizarse un análisis del contenido para identificar el formato y asignar de manera adecuada a la categoría pertenece.

La conservación de los documentos es de vital importancia en la FCI, pero un manejo erróneo ocasiona problemas como la duplicación de archivos, pérdida, vulnerabilidad en la integridad de los mismos, por almacenar los documentos en el computador donde labora la secretaria, y al no contar con un entorno seguro de almacenamiento, incrementa el trabajo a la secretaria al generar los documentos que han sufrido daño alguno.

La recuperación o búsqueda de los archivos en la Secretaría de la FCI, ha provocado inconvenientes a los encargados de este departamento, al tener que recurrir abrir cada documento y revisar si el requerimiento del usuario está en el mismo y al no obtener una respuesta positiva sobre su requerimiento, el interesado se siente molesto y se ve en la necesidad de acercarse a la Institución por repetidas ocasiones.

Pronóstico.

De persistir el problema en la gestión de documentos digitales en la Facultad de Ciencias de la Ingeniería (FCI), la gran cantidad de archivos importantes que en ella disponen podrían sufrir daños de integridad, disponibilidad, además del consumo excesivo de recursos de disco duro al almacenar los documentos en el mismo computador donde labora la secretaria. La búsqueda de los trámites que requieran los usuarios, se ven retrasados por la inadecuada clasificación de los documentos, lo que conlleva a que existan reclamos a las autoridades competentes ocasionando así un inconveniente en el proceso de acreditación de la Universidad.

1.1.2. Formulación del problema.

¿Cómo favorecer la gestión de los documentos digitales de la Facultad de Ciencias de la Ingeniería en la UTEQ?

1.1.3. Sistematización del problema.

- ¿De qué manera se puede clasificar los documentos digitales reduciendo el consumo de recursos de la FCI?
- ¿Cómo mejorar la conservación de los documentos digitales en la FCI?
- ¿Cómo mejorar el proceso de búsqueda de los documentos digitales en la Secretaría de la FCI?

1.2. Objetivos.

1.2.1. Objetivo General.

Informatizar el proceso de gestionar los documentos digitales en la Secretaría de la Facultad de Ciencias de la Ingeniería, incluyendo tecnologías de recuperación de información.

Objetivos Específicos.

- Desarrollar un software que permita búsqueda y clasificación de los documentos digitales en la FCI, analizando el contenido textual para identificar a que categoría pertenece.
- Disminuir el uso de los recursos clasificando los documentos digitales de las actividades realizadas en la FCI.
- Garantizar la conservación íntegra de los documentos digitales en la Secretaría de la FCI, almacenados en lugar seguro.
- Agilitar el proceso de búsqueda de documentos a través del contenido textual, disminuyendo el tiempo.

1.3. Justificación.

En la secretaría de la Facultad de Ciencias de la Ingeniería, se gestiona toda la información documental que se genera de diferentes trámites, esto ha ocasionado dificultades en la organización y el proceso de búsqueda de los mismos. La implementación de una herramienta que ayude en la gestión de documentos digitales, se convierte en una alternativa para facilitar la organización y búsqueda de información, optimizando así tiempo y recursos.

La elaboración del presente proyecto estará orientada dentro de la línea de Investigación de la UTEQ “Ambientes informáticos para las aplicaciones multimedia e inteligencia artificial”, en el estudio de algoritmos para la gestión automática de documentos basado en su contenido textual, implementado para reducir el consumo de recursos.

Por esta razón se desarrolló una herramienta, cuya función es automatizar los procesos empleados en la gestión de documentos digitales de la FCI, creando módulos que permita el almacenamiento de la información, clasificación automática e indexación por categorías, agilizando el proceso de búsqueda.

La Gestión automática de documentos digitales basado en el contenido textual, para facilitar el acceso a la información de la Facultad de Ciencias de la Ingeniería, beneficiará en la búsqueda de información y documentos relevantes que satisfagan las necesidades de todos los usuarios; disminuyendo el tiempo que se recurre al revisar la documentación que se almacena en los computadores, obteniendo una aplicación que va de acorde a la nuevos de mecanismos automáticos.

Por esta razón se justifica la implementación de una herramienta que gestione el manejo de los documentos digitales en función de su contenido textual, de tal manera que los directivos de la FCI, docentes, estudiantes y todo el personal interesado tengan a disposición toda la documentación referente a procesos y actividades que realizan en la institución de forma organizado.

CAPÍTULO II

FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN

2.1. Marco conceptual.

2.1.1. Recuperación de Información.

La recuperación de información (RI, Information Retrieval) se define como el problema a solucionar de la selección de documentos dentro de un corpus documental en respuesta a consultas proporcionadas por parte de un usuario. Los sistemas de RI utilizan bases de datos documentales y procesan minuciosamente las consultas de los usuarios permitiéndoles acceder a la información relevante en un intervalo de tiempo apropiado. El objetivo general, es crear mecanismos para encontrar una determinada información en grandes colecciones de documentos en formato electrónico [3].

Figura.- 1. Esquema de los sistemas de recuperación de información.



FUENTE: CLASIFICACIÓN AUTOMÁTICA DE VIDEOS.

AUTOR: DAVID APARICIO ESCRIBANO.

2.1.2. Clasificación automática de textos.

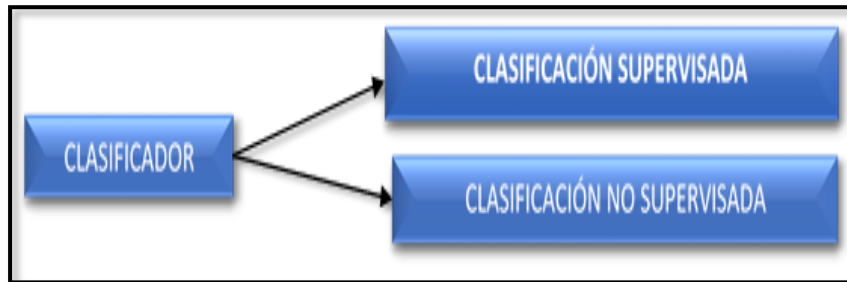
La clasificación automática de textos emergió a partir de la recuperación de información y que en la actualidad se ha convertido en un método indispensable por el incremento de información en formato electrónico. Su función principal es categorizar documentos dentro de un número determinado de categorías, en el cual un documento puede pertenecer a una o varias, todas o ninguna de las categorías. Las categorizaciones de texto cumplen tres etapas: pre-procesamiento de los datos, construcción del clasificador y categorización de los nuevos documentos [4].

Al utilizar el aprendizaje automático de los algoritmos, la finalidad es obtener un modelo de aprendizaje para clasificar a partir de ejemplos de entrenamiento que nos permiten la asignación a las categorías de manera automática [4].

2.1.3. Tipos de clasificadores automáticos.

Los clasificadores automáticos de documentos se clasifican dependiendo de sus características y de criterios, entre los cuales se tiene a la clasificación supervisada y no supervisada.

Figura.- 2. Tipos de clasificadores automáticos.



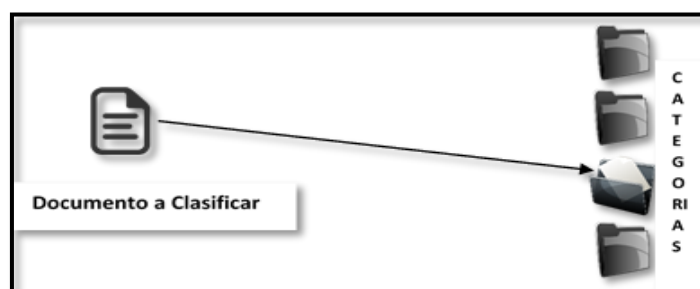
FUENTE: CLASIFICACIÓN AUTOMÁTICA DE VIDEOS

AUTORES: AGUAL; MENDOZA,2016

2.1.3.1. Clasificación Supervisada.

Parte de una serie de clases o categorías pre- establecidas. Su clasificación requiere de un corpus de entrenamiento previamente elaborado manualmente. Este tipo de clasificadores procuran elaborar un patrón representativo para cada una de las categorías entrenadas, posteriormente aplicar una función que encuentre la similitud entre el documento y cada uno de los patrones elaborados, para así de esta manera colocar cada documento en la categoría correspondiente [5].

Figura.- 3. Clasificación supervisada.



FUENTE: CLASIFICACIÓN AUTOMÁTICA DE DOCUMENTOS SUPERVISADOS.

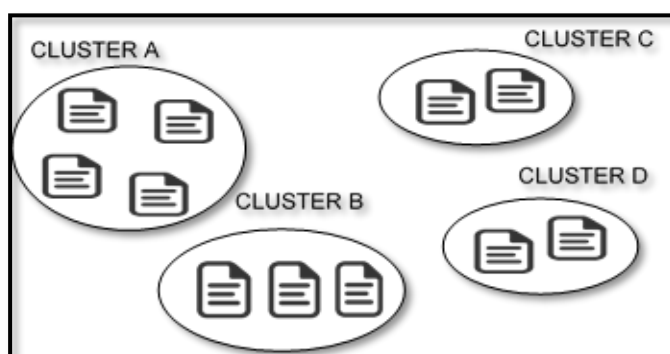
AUTORES: AGUAL; MENDOZA,2016.

Es decir, la clasificación supervisada parte de una serie de categorías conceptuales ya definidas a priori, y que la labor del clasificador (sea manual o automático) es asignar cada documento a la clase o categoría en las cuales hay que colocar a cada uno de los documentos, también conocida como categorización [5].

2.1.3.2. Clasificación no supervisada.

En este tipo de clasificadores no existe una base de datos documental de entrenamiento previo, no hay categorías previamente elaboradas, los documentos se clasifican en función de su contenido de manera automática, se puede decir que se auto-organizan. La clasificación no supervisada se la conoce como clustering, por que la organización de los documentos efectúa de forma automática, sin ninguna supervisión alguna [5].

Figura.- 4. Clasificación no – supervisada



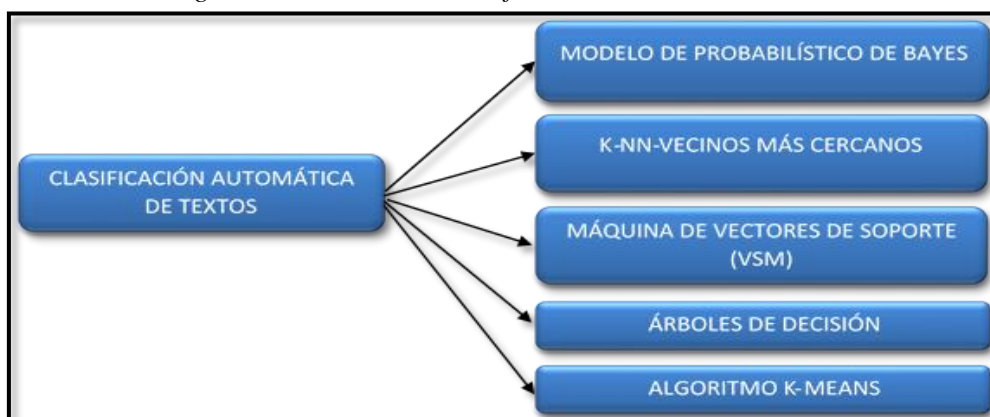
FUENTE: CLASIFICACIÓN AUTOMÁTICA DE DOCUMENTOS NO - SUPERVISADOS.

AUTORES: AGUAL; MENDOZA,2016

2.1.4. Técnicas de clasificación automática de textos.

Entre las técnicas de clasificación automática de texto, entre las más utilizadas podemos mencionar a las siguientes:

Figura.- 5. Técnicas de clasificación automática de textos.



FUENTE: CLASIFICACIÓN AUTOMÁTICA DE VIDEOS.

AUTORES: AGUAL; MENDOZA,2016.

2.1.4.1. Modelo de probabilístico de Naive Bayes.

Su nombre proviene del filósofo ingles Thomas Bayes (1702-1761) en el cual su teoría fundamenta en la probabilística de un suceso a partir de la probabilidad de que ocurra otro

suceso, el teorema puede verse en encontrar una hipótesis significativa dentro de un corpus de entrenamiento proporcionado y una base de conocimiento a priori de la probabilidad de cada hipótesis [6]. Esta técnica de clasificación de textos se ha utilizado en diversas aplicaciones tales como la detección de correos spam, la clasificación de correo electrónico personal, categorización de documentos, detección de contenido sexual explícito, detección de idioma y detección de emociones. Tiene las siguientes ventajas:

- Se desempeña bien en muchos problemas complejos del mundo real.
- Muy eficiente consumo mínimo tanto en la CPU y la memoria.
- Requiere una pequeña cantidad de datos de entrenamiento.
- El tiempo de entrenamiento es significativamente menor.

Se puede decir, que las estimaciones de probabilidad de Bayes son de baja calidad, pero la toma de decisiones del algoritmo es correcta y por lo tanto el modelo es exacto. En la clasificación de texto, se procede a utilizar las palabras (*tokens*) del documento con la finalidad de clasificar en la clase adecuada. Mediante la regla de decisión “máxima a posteriori (MAP)” [7].

Ecuación 1. Ecuación (MAP) Maximum a posteriori

$$C_{map} = \frac{\arg \max_{c \in C} (P(c|d))}{ceC} = \frac{\arg \max}{ceC} (P(c) \prod_{l \leq k \leq n_d} P(t_k|c))$$

Dónde:

- tk son los tokens (términos / palabras) del documento.
- C es el conjunto de clases que se utiliza en la clasificación
- $P(c|d)$ la probabilidad condicional de clase c que figura el documento d
- $P(c)$ la probabilidad a priori de la clase C
- $P(t_k|c)$ la probabilidad condicional de tk contador dada la clase c

Con el objetivo de encontrar a qué clase pertenece un nuevo documento a clasificar, se procede a calcular el producto de la probabilidad de cada palabra del documento dado una clase particular (probabilidad), multiplicado por la probabilidad de la clase particular (*priori*). Después de calcular el anterior para todas las clases de conjunto C , se selecciona el que tiene la mayor probabilidad [7].

El hecho que las computadoras pueden manejar números con una precisión en punto decimal, una vez calculado el producto de las probabilidades nos dará un punto de desbordamiento, lo que significa que vamos a terminar con un número tan pequeño, que no va a ser capaz de encajar en la memoria y por lo tanto será redondeado a cero, lo que hace el proceso de análisis inútil. Para evitar esto en vez de maximizar el producto de las probabilidades vamos a maximizar la suma de sus logaritmos [7].

Ecuación 2. Suma de los logaritmos.

$$c_{map} = \frac{\arg \max}{ceC} (\log P(c) + \sum_{l \leq k \leq n_d} \log P(t_k|c))$$

El teorema de Bayes, estima las probabilidades de que un documento corresponda a una categoría. La pertenencia a la categoría depende de las características conociendo la probabilidad de que aparezcan en los documentos. Las características son los términos de los documentos y su probabilidad de aparición en una categoría concreta, pueden observarse a partir de los datos de entrenamiento [8].

2.1.4.2. Algoritmo basado en ejemplos k-vecinos más cercanos (K-NN, k-Narest Neighbor).

Es un método supervisado, en el cual se incluye el conjunto de datos de entrenamiento, las propiedades multidimensionales utilizadas para el reconocimiento y los clasificadores para deducir la clase de los datos de entrada [9]. Se basa en que las propiedades del dato x de entrada tienen semejanza a los datos de su vecindad, entonces el dato pertenece a la misma clase que la clase más frecuente de sus k vecinos más cercanos. “El algoritmo k-nn asume que todos los ejemplos corresponden a puntos en un espacio p -dimensional \mathfrak{R}^p , los cuales tienen establecida una clase C “. Los datos de entrenamiento son de la forma presentada en [10]

Ecuación 3. Representación de los datos de entrenamiento.

$$(x_i, c_i) = (x_{i1}, x_{i2}, x_{i3}, x_{i4}, \dots, x_{ip}, c_i)$$

Para clasificar un documento, se hace la comparación del documento con los ejemplos de entrenamiento y calcula la distancia Euclídea, con la que determina todas las distancias entre el punto a clasificar y todos los puntos del conjunto de entrenamiento. Una vez encontradas

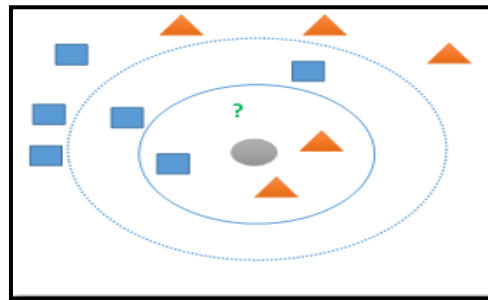
las distancias se determinan los K vecinos más cercanos y, a través del tipo de la clase para determinar, se asigna el punto a una de ellas [9]:

Ecuación 4. Formula de la distancia Euclidea.

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{|c|} (x_{jk} - x_{ik})^2}$$

En la *figura 6* se puede apreciar el proceso del algoritmo. En el entrenamiento el sistema ha adquirido los valores de sus atributos (distribución en el plano) y también los valores de sus clases (forma y color). El entonces el algoritmo pretende, clasificar y resolver qué ejemplos está aproximado al documento (círculo verde) [3].

Figura.- 6. Ejemplo de KNN.



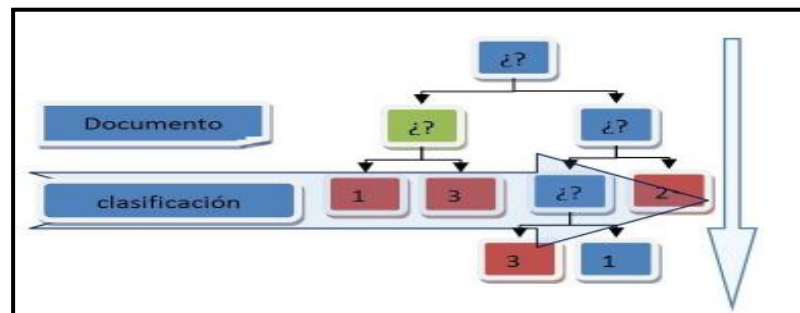
FUENTE: CLASIFICACIÓN AUTOMÁTICA DE VIDEOS.

AUTORES: AGUAL; MENDOZA,2016.

2.1.4.3. Árboles de decisión.

Es un método de clasificación de predicción se basa en el entrenamiento consiste en la elaboración de un árbol de decisión de múltiples, es un árbol con nodos internos etiquetados como términos, las ramas contienen los pesos que tienen en el documento de texto y en las hojas se encuentran las categorías. Cada nodo en el árbol se asocia con un conjunto de casos. Las decisiones de los árboles son binarias [11].

Figura.- 7. Arboles de decisión – Clasificación.



FUENTE: CLASIFICACIÓN AUTOMÁTICA DE VIDEOS.

AUTORES: AGUAL; MENDOZA,2016.

La clasificación se ejecuta en base a un conjunto de preguntas a los valores de sus atributos, iniciando por el nodo principal (raíz), consecuente en relación a la respuesta del nodo principal sigue el camino realizando las preguntas a los nodos internos, continuando así hasta alcanzar a un nodo hoja, que contiene la clasificación [3].

2.1.4.4. Máquina de soporte vectorial o Máquina de vectores soporte.

El método está relacionado con la clasificación y la regresión, útil cuando se quiere elaborar un clasificador utilizando conjunto de entrenamiento, se orienta en el principio de inducción de Minimización del Riesgo Estructural [(SRM, Structural Risk Minimization) y la Dimensión de Vapnik-Chervonenkis] como proceso de inferencia. Se fundamenta en aprender a discriminar los miembros positivos y negativos de una clase de vectores de n-dimensional proporcionada [12].

Según David Aparicio Escribano en su investigación menciona que “la función hallar una hipótesis h para la cual, a partir de la obtención de una cota sobre el riesgo esperado $R(h)$ (tasa de error medio sobre el conjunto de test) se concluye que, para asegurar su minimización, fijado el conjunto de entrenamiento, es necesario minimizar conjuntamente el riesgo empírico $R_{emp}(h)$ (tasa de error media sobre el conjunto de entrenamiento) y la VC (Vapnik Chervonenkis) dimensión del espacio de hipótesis” [3]. El riesgo empírico viene dado por la expresión:

Ecuación 5. Máquina de Soporte de vectores –Riesgo Empírico.

$$R_{emp}(h) = \frac{1}{2n} \sum_{i=1}^n |y_i - f(x_i, h)|$$

2.1.4.5. Algoritmo K-Means.

Es un proceso iterativo de clustering en el que cada objeto es asociado a un único cluster se basa en la idea de que un punto central puede representar un cluster. El algoritmo particiona un repertorio de vectores $\{x_1, x_2, \dots, x_n\}$ en un conjunto de clusters $\{C_1, C_2, \dots, C_k\}$, para ello se necesita k centroides para iniciar el algoritmo, los mismos pueden ser externo o elegidos aleatoriamente entre los vectores [13].

El algoritmo k-means sigue la siguiente metodología [14]:

- Inicialización

- Elegir un punto k al azar.
- Se seleccionan k documentos de manera aleatoria como los centroides iniciales.
- Los documentos restantes son asignados a los clusters que tengan su centroide más cercano, formándose así un k grupos.
- Iteración
 - Se recalcula los nuevo centroides de los grupos k basado en los actuales miembros del cluster.
 - Cada documento es reasignado al cluster que tenga su centroide más cercano o sea más similar.
- Detención:
 - En la detención se pueden usar diferentes condiciones, hasta que los centroides no cambien o que se haya superado las iteraciones previstas [14].

En la clasificación de documentos, los procesos comienzan todos agrupados y posteriormente son distribuidos en grupos más pequeños de documentos de igual semejanza, realizándose varias iteraciones de este proceso hasta alcanzar un criterio detención. Este algoritmo ha sido utilizado ampliamente con documentos de textos, presentando un grado eficiencia aceptable en la clasificación [13].

2.1.5. Procesamiento del Lenguaje Natural.

Es una disciplina de la Inteligencia Artificial (IA) su función es construir sistemas y mecanismos que permitan la comunicación entre personas y máquinas por medio de lenguajes naturales. Es un conjunto de instrucciones que recibe un sistema recibe en lenguaje de programación, que permita una comunicación con el ser humano en su propio lenguaje, este procesamiento presenta diversas aplicaciones tales [15].

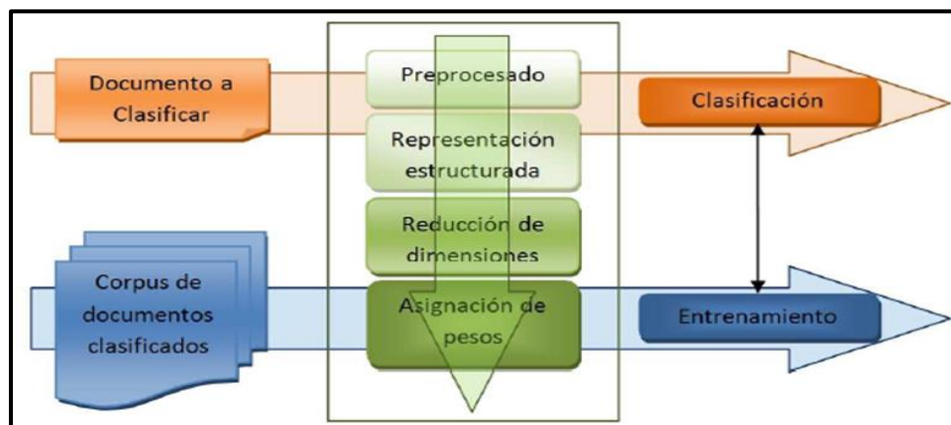
- Corrección de textos.
- Traducción automática.
- Recuperación de la Información.
- Extracción de información y resúmenes.

- Búsqueda de documentos.
- Traducción automática.
- Reconstrucción de discurso.
- Reconocimiento del habla.
- Síntesis de voz [15].

2.1.6. Fases de un clasificador automático de documentos.

La clasificación lleva ciertas fases para el procesamiento de documentos, tal como se muestran en la *fig. 8*. Las mismas que deben realizarse tanto para el documento que va a formar parte del conjunto de entrenamiento como para el documento que va a ser clasificado.

Figura.- 8. Fases de un clasificador de documentos.



FUENTE: CLASIFICACIÓN AUTOMÁTICA DE VIDEOS

AUTOR: DAVID APARICIO ESCRIBANO

2.1.6.1. Pre-procesado y representación del documento.

Esta fase consiste en eliminar elementos textuales que no contienen información relevante, es decir transformar los documentos en una representación más apropiada. El tener información no relevante ocasionaría que costos de procesamiento del algoritmo sea superior, consecuentemente al ejecutar los procesos de clasificación los resultados no serían muy adecuados.

2.1.6.2. Reducción de dimensiones.

El modelo que representa cada documento debe contener únicamente aquellos que se consideran más importantes, despreciando así aquellos que no son relevantes en el documento. Existen diferentes métodos con los cuales se puede hacer una depuración del

conjunto de atributos quedándonos con los más informativos y relevantes. Así esta manera se disminuye la cantidad de atributos con los clasificadores procederían hacer su proceso, reduciendo tanto en tiempo que conlleva la clasificación. Al reducir la dimensionalidad en la fase de pre-proceso, se reduce la complejidad por lo tanto coste del proceso de clasificación va a hacer mínimo [16].

2.1.6.3. Asignación de pesos.

El proceso de asignación de pesos es proveer a cada uno de los términos que tiene los documentos después haber pasa el filtro una ponderación dependiendo la importancia que tienen en el documento.

2.1.6.4. Entrenamiento.

Esta etapa tiene por objetivo elaborar el modelo de lenguaje para la clasificación de documentos usando campos de entrenamiento textos previamente categorizados manualmente por expertos en el área, a los cuales se les nombra conjunto de datos de entrenamiento [17]

2.1.6.5. Clasificación.

La fase de clasificación consiste en designar a categorías nuevos documentos, en base al aprendizaje que fue adquirido fase de entrenamiento de cada una de las categorías.

2.1.7. Pre - procesado del documento y representación estructura.

El punto de partida para la clasificación de documentos es la obtención del conjunto de documentos también denominado corpus documental. Debido que los documentos no pueden ser interpretados por el clasificador, obtener la representación estructurada del documento es necesario tanto para el entrenamiento como para el proceso de clasificación.

El Procesamiento del lenguaje natural nos proporciona técnicas que pretenden entender y procesar el lenguaje. Las técnicas más usadas son las siguientes:

2.1.7.1. Palabras “vacías” (StopWords).

La eliminación de los términos conocido como StopWords es el proceso suprimir las palabras de escasa utilidad que en un documento se presenta con frecuencia y ocasiona errores en el proceso de recuperación de información. El objetivo del *Stopwords* es permitir que en la clasificación de documentos solamente se tomen en cuenta las palabras que aportan mayor información semántica [18].

En la clasificación de textos se hace el uso de listas de palabras conocidas denominadas listas de parada (*stopword list*). Elaborar una lista de palabras vacías para el idioma inglés es simple, pero para idiomas como el español este proceso se complica debido a la inexistencia de corpus, también se puede elaborar una lista de palabras propias, de acuerdo a la naturaleza del proceso a realizar [18].

En la *Tabla 1* se muestra un conjunto palabras vacías que podrían usar para la eliminación:

Tabla 1. Listado de Palabras Vacías.

“la”, “los”, “les”, “las”, “de”, “del”, “a”, “ante”, “con”, “en”, “para”, “por”, “y”, “o”, “u”, “tu”, “te”, “ti”, “le”, “que”, “al”, “ha”, “un”, “han”, “lo”, “su”, “una”, “estas”, “esto”, “este”, “es”, “tras”, “suya”, “a”, “acá”, “ahí”, “ajena”, “ajenas”, “ajeno”, “ajenos”, “al”, “algo”, “algún”, “alguna”, “algunas”, “alguno”, “algunos”, “allá”, “alli”, “allí”, “ambos”, “empleamos”, “ante”, “antes”, “aquel”, “aquella”, “aquellas”, “aquello”, “aquellos”, “aquí”, “aquí”, “arriba”, “asi”, “atrás”, “aun”, “aunque”, “bajo”, “bastante”, “bien”, “cabe”, “cada”, “casi”, “cierta”, “ciertas”, “cierto”, “ciertos”, “como”, “cómo”, “con”, “conmigo”, “conseguimos”, “conseguir”, “consigo”, “consigue”, “consiguen”, “consigues”, “contigo”, “contra”, “cual”, “cuales”, “cualquier”, “cualquiera”, “cualquieras”, “cuan”, “cuán”, “cuando”, “cuanta”, “cuánta”, “cuantas”, “cuántas”, “cuanto”, “cuánto”, “cuantos”, “cuántos”, “de”, “dejar”, “del”, “demás”, “demasiada”, “demasiadas”, “demasiado”, “demasiados”, “dentro”, “desde”, “donde”, “dos”, “el”, “él”, “ella”, “ellas”, “ello”, “ellos”, “empleáis”, “emplean”, “emplear”, “empleas”, “empleo”, “en”, “encima”, “entonces”, “entre”, “era”, “éramos”

AUTORES: AGUAL; MENDOZA, 2016.

2.1.7.2. Expansión por sinónimos.

Un clasificador de textos no siempre va a encontrar que un documento contenga la palabra con la que se tiene especificada en el corpus de entrenamiento. Algunos sistemas para entender mejor la pregunta que se le ha realizado para su búsqueda utilizan la expansión de sinónimos los mismos que se ejecuta sobre las palabras de mayor relevancia de información, incorporando los sinónimos de estas palabras a la consulta [3]. El no incorporar la búsqueda por sinónimos en un sistema de recuperación podría suceder que existan documentos con información relevante por recuperar, pero al momento de ingresar las palabras claves no aparecen, por consiguiente el documento no es recuperado, por tal motivo es casi obligatorio

de un buscador semántico, incorporar la búsqueda sinónimos, con la finalidad de encontrar más documentos [19].

2.1.7.3. Lematización (stemming).

Lematizar es reducir las diferentes formas de escritura de una palabra a la forma canónica o su lema, consiste en reagrupar los distintos tonos de un verbo en el infinitivo, de esta manera ayudar a la correcta indización y recuperación. Existen diferentes teorías y algoritmos, entre las técnicas tenemos la lematización flexiva que procede a eliminar [20];

- Plurales.
- Géneros.
- Terminaciones verbales.

Otra técnica es la lematización derivativa el cual elimina lo presentado por la lematización flexiva además incluye;

- Supresión de sufijos derivativos.

El algoritmo efectivo más conocido para el *stemming* es el de *Porter*, el mismo que extrae los sufijos y prefijos de las palabras comunes pero que tienen una raíz en común que pueden ser consideradas como un sólo término, se puede implementar cualquier programa que se necesiten de la lematización de textos, obteniendo resultados aceptables [20].

Por ejemplo, se tiene las siguientes palabras **hablar, hablará, hablando, hablo**, entre otros., aplicando el *stemming* su raíz en radica **habl**.

2.1.7.4. Asignación de pesos.

En el procesamiento de lenguaje natural (PNL), los textos deben ser convertidos a una representación vectorial, mediante técnicas que proviene la recuperación de información con el que se logra un balance entre el número de ocurrencias de un término por documento con el número de documentos en que dicho término aparece, beneficiando los términos que mejor discriminan entre los tipos de documentos. Entre los factores de peso tenemos *el tf – idf*, denominada **frecuencia inversa** de documento la cual se define como [21]:

Ecuación 6. Factores de Peso tf – idf.

$$idf_t = \log \frac{N}{df_t}$$

Donde:

- N es el número de documentos
- df_t es el número de documentos en la colección que contienen el término t .
 - Asigna un valor alto a los términos raros.
 - Asigna valor bajo a los términos muy comunes.

El peso $tf - idf$ se define como:

Ecuación 7. Ecuación del $tf - idf$.

$$tf - idf_{t,d} = tf_{td} * idf_t$$

Donde tf_{td} es el número de ocurrencias del término t en el documento d .

El peso asignado por $tf - idf$ es:

- muy alto cuando el término t , ocurre muchas veces en un pequeño número de documentos,
- bajo cuando t ocurre pocas veces en un documento u ocurre en muchos documentos,
- muy bajo cuando t ocurre prácticamente en todos los documentos [21].

Otro de los factores de peso es la **Frecuencia de palabra** (TF, Term Frequency)

En donde a cada palabra se le asigna una importancia proporcional al número de veces que aparece en el documento, su expresión es la siguiente:

Ecuación 8. Ecuación del tf .

$$wk_j = tf(t_k, d_j)$$

Donde $tf(t, d)$ es la frecuencia de la palabra t en el documento d [3].

2.1.8. Xtreme Programming (XP).

La metodología de desarrollo XP definida por Kent Beck, Ron Jeffries, Ward Cunningham, es un marco de trabajo que se compone de un conjunto de valores y prácticas importantes los mismos que forman un método para el desarrollo de software, fue diseñado para

desarrollar proyectos pequeños y medianos, entre sus características tenemos [22]

- Los requerimientos del usuario se las puede cambiar de manera rápida.
- La comunicación entre el cliente y los desarrolladores es de vital importancia durante el proceso de desarrollo.
- Lleva todo al extremo para el desarrollo de aplicaciones en corto tiempo [22].

2.1.8.1. Ciclo de vida de XP.

El ciclo de vida de la metodología Xtreme Programming se subdivide en 4 fases cuales se describen a continuación [23]:

- **Fase de Planeación.** - En esta fase se plantean escuchando las historias de clientes, para entender el contexto del negocio, definir las principales características y la funcionalidad.
- **Fase del Diseño.** – En esta etapa se evalúan las historias del cliente y se divide el sistema en tareas dentro del equipo de proyecto, cada tarea representa una característica del sistema a desarrollar.
- **Fase de Desarrollo.** – En esta fase se procede a la programación, la unidad de pruebas y la integración del código, en esta etapa el cliente es de vital importancia para que esté pueda despejar cualquier duda que se presente en el trabajo.
- **Fase de Pruebas.** – En esta fase se realiza una prueba de unidad por cada una de las historias de usuario designado al equipo de trabajo, cada clase representa una característica distinta del sistema, existen pruebas unitarias las cuales son diseñadas para probar cada uno de los métodos y clases, dichas pruebas son realizadas por los programadores [23].

2.1.9. Java.

Lenguaje de programación Orientado a Objetos de código abierto, es una tecnología que se orienta al desarrollo de diferentes programas informáticos, cuáles pueden ser ejecutados en los sistemas operativos existentes. Se puede desarrollar muchas aplicaciones tanto para entorno empresarial como también para la tecnología móvil. Java se compone básicamente por 2 elementos [24]:

- El lenguaje Java.

- Máquina virtual de Java (Java Virtual Machine) para ejecución de las aplicaciones desarrolladas.

Ventajas de utilizar Java:

- Orientado a objetos.
- Muy flexible.
- Funciona en cualquier plataforma.
- Código abierto.
- Lenguaje expandible [24].

2.1.10. NetBeans.

NetBeans IDE es un entorno de desarrollo integrado (IDE), modular, de base estándar escrito en el lenguaje de programación Java de código abierto y una plataforma de aplicación, las cuales pueden ser usadas como una estructura de soporte general (framework) para compilar cualquier tipo de aplicación, en diferentes sistemas operativos existentes [25].

2.1.11. Apache Lucene.

Apache Lucene es un proyecto de código abierto, es una biblioteca de búsqueda de texto desarrollado en Java, es un sistema de recuperación de información basado en texto, funciona como un motor. Es una tecnología adecuada para casi cualquier aplicación que requiera la búsqueda de texto completo, ya sea en una aplicación o página web, especialmente multiplataforma [26].

Su funcionamiento es el siguiente:

- Crea un índice de texto completo
- Realiza consultas en este índice.
- Genera resultados clasificados de acuerdo a la importancia de la consulta.

El contenido a procesar puede ser de diversas fuentes tales como una base de datos, un sistema de archivos e incluso de los sitios web [27].

2.1.12. Lucene Analyzers.

Es una clase abstracta que se usa para proporcionar una interfaz que permitirá que un documento pueda ser indexado. Esta clase sirve como filtro cuando se indexa el texto, en la

eliminación de las palabras vacías que no sirve para la realización de una búsqueda, omisión de caracteres especiales, conversiones de minúscula a mayúscula, entre otros. Existen diferentes listas de *analyzers* [28]:

Tabla 2. Listado de Lucene Analyzers

• <i>BrazilianAnalyzer</i> ,	• <i>GreekAnalyzer</i>
• <i>ChineseAnalyzer</i> ,	• <i>KeywordAnalyzer</i>
• <i>WhitespaceAnalyzer</i>	• <i>PatternAnalyzer</i>
• <i>CJKAnalyzer</i>	• <i>PerFieldAnalyzerWrapper</i>
• <i>CzechAnalyzer</i>	• <i>QueryAutoStopWordAnalyzer</i>
• <i>DutchAnalyzer</i>	• <i>RussianAnalyzer</i>
• <i>FrenchAnalyzer</i>	• <i>ShingleAnalyzerWrapper</i>
• <i>GermanAnalyzer</i>	• <i>SimpleAnalyzer</i>
• <i>SnowballAnalyzer</i>	• <i>StopAnalyzer</i>
• <i>StandardAnalyzer</i>	• <i>SpanishAnalyzer</i>
• <i>ThaiAnalyzer</i>	

AUTORES: AGUAL; MENDOZA,2016.

2.1.13. MySql.

MySQL es un servidor de base de datos SQL (Structured Query Language) veloz, multi-hilo, multiusuario y robusto. MySQL es una base de datos de código abierto más reconocida en el mundo, permite la entrega rentable de aplicaciones de bases de datos basadas en Web y embebidos escalables. Desde enero del 2008 fue una subsidiaria de Sun Microsystems y a partir de abril del 2009 Oracle Corporation. MySQL AB fue creado por David Axmark, Allan Larsson y Michael Widenius [29].

2.1.14. Tika Version 11.0

Es un conjunto de herramientas de análisis de contenido, detecta y extrae los metadatos y el texto de diferentes tipos de archivos tales como PPT, XLS y PDF, etc. Utilizada para la indexación de los motores de búsqueda, análisis de contenido, traducción, ente otros. Tika es un proyecto de la Apache Software Foundation, y anteriormente fue un subproyecto de Apache Lucene [30].

2.1.15. Poi.

Apache Poi es una API que permite mediante código de programación crear, modificar y visualizar los archivos de MS Office utilizando Leguaje de programación Java. Es una biblioteca de código abierto desarrollado y distribuido por la Apache Software. Contiene clases y métodos para decodificar los datos de entrada de usuario o un archivo en documentos de MS Office [31].

2.1.16. Tm – extractors.

Es una API de Java para la minería de textos, que hace posible la extracción del texto de documentos de MS Office, PDF, texto plano.

2.1.17. Apache OpenNLP.

Es un conjunto de herramientas para el procesamiento de texto en lenguaje natural basado en el aprendizaje de máquinas, se utiliza para la tokenización, la segmentación de frases, etiquetado, la extracción de entidades, fragmentación, el parsing y la resolución de la correferencia. Estas tareas suelen ser necesarios para construir servicios de procesamiento de texto más avanzados [32].

2.2. Marco referencial.

En esta sección se presentan las referencias acerca de diferentes trabajos publicados, tesis elaboradas, artículos de revistas, entre otros., los cuales están relacionados con este trabajo de investigación proporcionándonos formulas, procesos y técnicas que ayuden en la elaboración del proyecto.

2.2.3. Clasificación de textos académicos en función de su contenido

léxico-semántico* (Academic text classification based on lexical-semantic content).

El investigador René Venegas (2007), en su proyecto propone realizar un estudio descriptivo-comparativo de orden lingüístico-textual a partir de los textos que son leídos en el ámbito académico y en el profesional. El objetivo de esta investigación es clasificar, utilizando y comparando dos métodos de categorización automática, cuales son Bayes Ingenuo y Máquina de Soporte de Vectores aplicado a los textos académicos incluidos en el Corpus PUCV-2006 perteneciente al trabajo realizado en el proyecto Fondecyt 1060440. Estos métodos están basados en los lexemas de contenido semántico compartidos en el corpus de textos académicos usados en cuatro carreras profesionales de la Pontificia Universidad Católica de Valparaíso, Chile [33].

2.2.4. Clasificación de textos en lenguaje natural usando la Wikipedia.

En el artículo presentado por Jose María Quinteiro, Ernestina Martel, Pablo Hernández, Juan A. Ligerio, Aaron López (2011), proponen un clasificador que toma como base la Wikipedia para la creación del corpus que define una categoría, mediante técnicas de Procesado de Lenguaje Natural (PLN) los mismos que analizan sintácticamente los textos a clasificar. El resultado final del sistema propuesto presenta un alto porcentaje de acierto, incluso cuando se compara con los resultados obtenidos con técnicas alternativas de aprendizaje automático. El presente clasificador, muestra una forma más de salvar la visión que deja el Modelo de Espacio Vectorial de los documentos como meras bolsas de palabras, al valorar la función sintáctica que tiene una palabra dentro del texto a la hora de computar su peso [34].

2.2.5. Marco de trabajo para Indexación, Clasificación y Recopilación Automática de Documentos Digitales.

El objetivo en este trabajo es analizar diferentes enfoques de solución a los problemas de

exploración, clasificación e indexación de documentos digitales en la web, e incorporarlos en una sola herramienta, creando un marco de trabajo para la búsqueda de documentos digitales en la web. En este proyecto los autores Javier Caicedo y Gonzalo Parra (2009), presentan la integración de varias herramientas desarrolladas en el área de inteligencia artificial implementando el algoritmo BestFirst, en complemento con el algoritmo Naive Bayes, clasificar automáticamente documentos digitales, de acuerdo a áreas de temáticas definidas, exploración en la Web y encontrar publicaciones relacionadas con los documentos, las pruebas se realizaron usando tópicos tomados del OpenDirectory [35].

2.2.6. Clasificación Automática de Videos.

La clasificación automática de contenido audiovisual, mediante el proceso de convertir el audio en documentos de texto, para posteriormente ser procesados con las técnicas de procesamiento del lenguaje natural y métodos de clasificación automática es la propuesta elaborado por David Aparicio Escribano (2009) consta dos fases primordiales, el reconocimiento automático del habla y la clasificación automática de texto.

En la fase del reconocimiento automático del habla, permite realizar la transcripción a texto del contenido audiovisual para posteriormente ser clasificado como un documento de texto. Se utilizó las herramientas de software libre Media Mining Indexer (MMI) es un software elaborado Sail Labs de reconocimiento de voz desarrollado, que permite procesar voz de múltiples fuentes en varios formatos y las transcribe en texto en tiempo real, el algoritmo K-NN para su clasificación y la biblioteca de Apache Lucene como herramienta recuperación de información en la aplicación [3].

2.2.7. Clasificación de documentos usando Maquinas de Vectores de Apoyo

La clasificación automática de documentos utilizando el algoritmo supervisado SVM (Maquina de Vectores de Apoyo) en su estudio presentado por Martha Varguez Moo, Víctor Uc Cetina y Carlos Brito Loeza (2014), aplicando aun conjunto de datos de *Reuters*, utilizando para tal efecto las bibliotecas para SVM más usadas como Weka y SVMLight para sus pruebas respectivas. Los autores presentan la solución a las tareas de clasificación de documentos para conseguir el objetivo los autores describen la descripción del problema a solucionar y las funcionalidades del algoritmo implementado [36].

2.2.8. Minería de texto para la categorización automática de documentos

Los autores M. Alicia Pérez Abelleira y Carolina A. Cardoso (2010), en su estudio proponen implementar un buscador semántico que aproveche el resultado del algoritmo SMO (máquinas de vectores soporte) para la clasificación de documentos. Los documentos a clasificar provienen de un corpus de información que contiene más de 8000 archivos en distintos formatos (Microsoft Word, texto plano, Pdf, entre otros.). El buscador implementado utilizando las Apis de software libre SemanticSearch y Apache Lucene se basa la búsqueda de los atributos extraídos de los documentos tales como: personas, unidades académicas, carreras, instituciones, fechas, número, año de las resoluciones, títulos, DNIs. Cada una de las entidades es extraída por un anotador [37].

CAPÍTULO III
MÉTODOLOGÍA DE LA INVESTIGACIÓN.

3.1. Localización.

El proyecto se ha desarrollado en las instalaciones de la Universidad Técnica Estatal de Quevedo la misma que está situada en la Provincia de Los Ríos, Cantón Quevedo, Av. Quito km 1½ vía a Santo. Domingo de los Tsáchilas. Definiendo como eje de estudio la Secretaría de la Facultad de Ciencias de la Ingeniería, lugar donde se ha obtenido la información necesaria para llevar la ejecución este proyecto investigativo.

Figura.- 9. Localización del proyecto de investigación.



FUENTE: GOOGLE MAPS.

AUTOR: GOOGLE.

3.2. Tipo de investigación.

Los tipos de investigación empleados en este proyecto fueron la investigación campo, la investigación descriptiva y la investigación aplicada. Por cuanto se identificaron los inconvenientes que se presentan en la Secretaría de la FCI en el manejo de la información digital, obteniendo todos los aspectos que considerables para el desarrollo del aplicativo informático.

3.2.1. Investigación de campo

Se empleó la investigación de campo aplicando la técnica de la observación directa para evidenciar la realidad de cómo se está llevando la gestión de documentos en la Secretaría de la FCI y por medio de la técnica mencionada se logró obtener la siguiente información pertinente del problema:

- Organización no adecuada de los archivos.
- Duplicación de los archivos.

- Pérdidas de los archivos.
- Alto consumo de tiempo en buscar los documentos digitales almacenados en el computador.

Identificado los problemas principales se procedió a realizar los análisis competentes, para generar una solución que se acople a la necesidad de gestionar los documentos digitales de la Secretaría de FCI.

3.2.2. Investigación descriptiva.

Es una investigación descriptiva, porque se determinó las causas y consecuencias del problema a investigar cómo se puede observar en la *Tabla 3*, posteriormente se procedió a analizar minuciosamente los resultados obtenidos de la investigación, con el objetivo de encontrar una solución al inconveniente presentado.

Tabla 3. Investigación descriptiva. Identificación de causas y consecuencia.

CAUSAS	CONSECUENCIAS
<ul style="list-style-type: none"> • Organización no adecuada de los archivos. • Duplicación de los archivos. 	<ul style="list-style-type: none"> • Alto consumo de recursos electrónicos.
<ul style="list-style-type: none"> • Pérdida de archivos al no contar un lugar adecuado para el almacenamiento. 	<ul style="list-style-type: none"> • Volver a generar los archivos perdidos, aumentando trabajo para el encargado de Secretaría.
<ul style="list-style-type: none"> • Demoras en las búsquedas de los documentos digitales, que requieren los usuarios. 	<ul style="list-style-type: none"> • Molestias por parte de los usuarios, al no obtener respuesta inmediata. • Quejas a las autoridades competentes de la FCI.

AUTORES: AGUAL; MENDOZA,2016.

3.2.3. Investigación aplicada.

Este proyecto se localiza dentro de la investigación aplicada, porque una vez realizado el análisis correspondiente de toda la información obtenida, se plantea una mejora el cual es:

- Disminuir el consumo de recursos.
- Organización adecuada de los archivos.
- Evitar duplicaciones y pérdidas de documentos.

- Clasificación automática de los documentos según el tipo de trámite.
- Búsqueda de archivos en menor tiempo posible.

Al elaborar un aplicativo informático que optimice la gestión de los documentos digitales en la Secretaría de la FCI.

3.3. Métodos de investigación.

Los métodos que han sido utilizados en el presente proyecto de investigación son el método analítico, el método deductivo y el método inductivo siendo estos los más utilizados en la elaboración del presente trabajo.

3.3.1. Método Analítico.

Durante la elaboración de la investigación se utilizó la entrevista y la observación directa, obteniendo a través de las técnicas mencionadas las causas y consecuencias de la metodología actual de la gestión de documentos, consecutivamente haciendo uso de este método se analizó la información obtenida, definiendo las características principales de la gestión de documentos:

- Consumo mínimo de recursos.
- Organización adecuada de los archivos.
- Identificación y agrupación de los archivos según el trámite.
- Evitar duplicación y la pérdida de los archivos.
- Reducir el tiempo de búsqueda de los archivos.

3.3.2. Método Deductivo.

El presente método se utilizó para identificar los problemas existentes en la gestión de documentos de la FCI, método necesario para determinar la forma más adecuada de automatizar sus procesos, los cuales son:

- Organizar automáticamente los archivos digitales.
- Clasificar e indexar los archivos según el trámite en categorías, mediante un algoritmo.
- Reducir el tiempo de búsqueda de archivos.
- Respalidar los archivos digitales para evitar pérdidas de los mismos.

3.3.3. Método Inductivo.

A través del análisis de los resultados obtenidos de la aplicación gestión automática de documentos digitales y el proceso que realiza la Secretaría actualmente en cuanto a la gestión documental, se podrá inferir la eficiencia del sistema para reducir el tiempo y el consumo de recursos en la gestión de documentos digitales.

3.4. Fuentes de recopilación de información.

En el desarrollo del sistema gestión automática de documentos digitales se utilizaron datos que nos fueron proporcionados, mediante las siguientes fuentes de información:

3.4.1. Primarias.

Entrevistas realizadas a:

- Lcda. Virginia Zambrano Menéndez Secretaria de la Facultad de Ciencias de la Ingeniería (FCI).
- Lcda. Hilda Álvarez Álvarez Secretaria saliente por motivos de jubilación.
- Lcdo. Carlos Andrade Secretario del Decanato de la FCI.

Quienes nos proporcionaron de la manera más comedida toda la información que tienen a su cargo, para el desarrollo del proyecto investigativo.

3.4.2. Secundarias.

- Observación directa del proceso de gestión documental digital que se realiza en la Secretaría de la Facultad de Ciencias de la Ingeniería.
- Documentos digitales proporcionados por la Secretaría de la FCI tales como, resoluciones del concejo académico, solicitudes de estancias, solicitudes de permisos, entre otros., diferentes trámites que se realizan en la Facultad de Ciencias de la Ingeniería, que nos sirvieron como archivos de pruebas para el correcto funcionamiento del sistema.
- Materiales bibliográficos:
 - Libros,
 - Revistas científicas (Artículos),
 - Páginas web,

- Bibliotecas virtuales,
- Otras fuentes de información, que contribuyeron a la realización teórica y la obtención del conocimiento necesario para el proyecto de investigación.

3.5. Diseño de investigación.

En la elaboración del presente proyecto se utilizó el diseño no experimental, la investigación se basa como eje primordial en la observación de fenómeno tal, es decir presenciar la realidad de cómo se lleva el proceso de gestión de documentos en la Secretaría de la FCI, para analizarlos minuciosamente con posterioridad. En este tipo de investigación no hay condiciones ni estímulos a los cuales se expongan los sujetos del estudio. Los sujetos son observados en su ambiente natural.

3.6. Instrumento de investigación.

La presente investigación que tiene como objetivo desarrollar un aplicativo que gestione los documentos digitales en la Secretaria de la FCI, para efecto se utilizaron los siguientes instrumentos de investigación.

Entrevista: Se realizaron varias entrevistas al personal a cargo del proceso de gestión de documentos entre los cuales se tiene a:

- La Lcda. Virginia Zambrano Menéndez, se realizó las entrevistas los días 9 y el 18 de febrero del año 2016, obteniendo como resultado los nombres de los diferentes trámites que se generan en la FCI y la situación actual de cómo se gestiona los archivos digitales.
- La Lcda. Hilda Álvarez Álvarez, la entrevista se la realizó el 9 de febrero del año 2016, obteniendo una descripción de las tareas que se debe realizar en caso que un documento se pierda y los inconvenientes que se tiene al momento de buscar un documento al tener que revisar todas las carpetas en los que están almacenados los archivos.
- El Lcdo. Carlos Andrade, la entrevista se la realizó 18 de febrero del año 2016, dándonos a conocer los trámites que se generan en su dependencia y como maneja toda la documentación tiene a su responsabilidad.

Las entrevistas fueron realizadas con el objetivo de recabar la información relevante de la gestión documental digital y así determinar a fondo la actual problemática en la Secretaría

de la FCI en la gestión de documentos digitales.

Observación Directa: Observar el proceso diario que realiza la Secretaria de la FCI al almacenar todos los documentos digitales en carpetas en el computador, de esta manera categorizándolos u organizándolos por un nombre que vaya de acorde al asunto realizado, para su posterior búsqueda.

3.7. Recursos humanos y materiales

3.7.1. Recursos Humanos

- Investigadores y desarrolladores Agual Espín José Luis y Mendoza Intriago Jonathan Aladino.
- Ing. Gleiston Guerrero Ulloa Msc. - Director del Proyecto de Investigación.
- Ing. PhD. Pavel Novoa Hernández– Docente de apoyo en el área de inteligencia artificial.

3.7.2. Hardware y Software

El equipo informático en el cual se desarrolló el sistema, presentaba con las siguientes especificaciones de hardware y software:

Tabla 4. Hardware usado en el desarrollo.

CANTIDAD	EQUIPO	DESCRIPCIÓN
1	Computador (Laptop)	<ul style="list-style-type: none">• Procesador Intel Core i3, 2.2 Ghz• Disco Duro de 500 GB• Memoria RAM 4 GB
1		<ul style="list-style-type: none">• Procesador Intel Core i5, 3.1 Ghz• Disco Duro de 600 GB• Memoria RAM 6 GB

AUTORES: AGUAL; MENDOZA,2016.

Tabla 5. Software usado en el desarrollo

TIPO	DESCRIPCIÓN
Sistema Operativo	<ul style="list-style-type: none">• Windows 10
Software	<ul style="list-style-type: none">• Netbeans (Java)

	<ul style="list-style-type: none"> • Office Profesional 2013 • Bibliotecas Jar <ul style="list-style-type: none"> • Apache Lucene • Lucene Analyzers • Tika • Poi • Tm-extractors • Apache OpenNLP
--	---

AUTORES: AGUAL; MENDOZA,2016

3.8. Metodología de desarrollo del Software.

El desarrollo de este proyecto se está utilizó la metodología Xtreme Programming (XP), conocida como programación al extremo, para entregar un software de calidad al menor tiempo posible (de uno a tres meses), para lo cual se realizaron las siguientes etapas:

- Planificación de los requisitos del proyecto.
- Diseño del proyecto.
- Codificación del proyecto
- Pruebas del proyecto.

3.8.1. Planificación de los requisitos del proyecto.

Para el desarrollo de la aplicación se recopilamos los requerimientos necesarios para poder automatizar los procesos que se evidenciaron mediante la observación directa y las entrevistas a las personas encargadas de la gestión documentos de la FCI. En esta etapa se definieron los alcances y requerimientos que son necesarios para realizar la gestión adecuada de los documentos digitales. Entre los principales procesos tenemos:

Tabla 6. Planificación de requisitos.

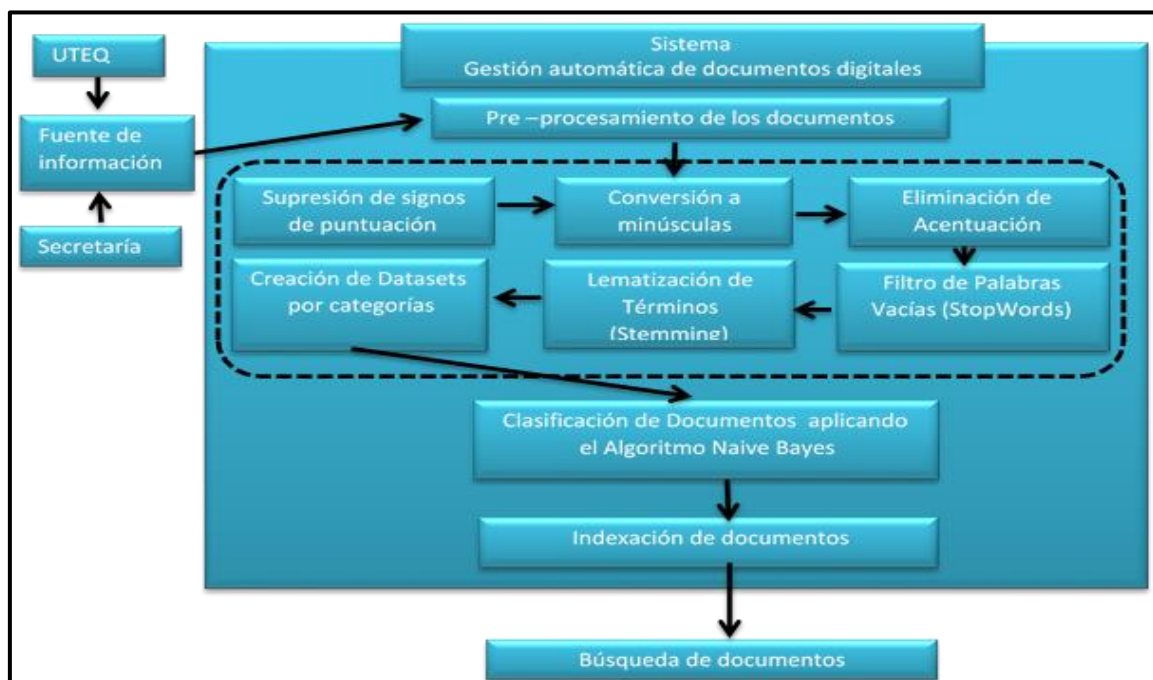
REF.	REQUERIMIENTOS.
R.001	Iniciar Sesión en la Aplicación.
R.002	Registrar Usuarios.
R.003	Importar archivos de entrenamiento por cada categoría.
R.004	Generar modelo de aprendizaje.
R.005	Crear nueva categoría.

R.006	Importar archivos para la clasificación.
R.007	Clasificación de los documentos.
R.008	Restaurar documentos.
R.009	Búsqueda de documentos clasificados por categorías.
R.010	Búsqueda de documentos clasificados global (búsqueda por todas las categorías).
R.011	Búsqueda de documentos clasificados por metadatos.
R.012	Búsqueda de documentos por el historial de búsquedas por categorías.
R.013	Búsqueda de documentos por el historial de búsquedas global (búsqueda por todas las búsquedas realizadas).
R.014	Generar respaldo de información.
R.015	Exportar respaldo de información.
R.016	Eliminar respaldo de información.

AUTORES: AGUAL; MENDOZA,2016.

3.8.1.1. Representación conceptual de la aplicación.

Figura.- 10. Representación conceptual del sistema gestión automática de documentos digitales.



AUTORES: AGUAL; MENDOZA, 2016.

Por medio de la representación conceptual se pudo mostrar la interacción de los procesos más importantes, mostrar el flujo de los procesos internos del sistema gestión automática de documentos, a continuación, se muestra el diagrama de los procesos de la aplicación.

3.8.1.2. Diagrama de caso de uso

Figura.- 11. Casos de uso del Sistema Gestión automática de documentos digitales.



AUTORES: AGUAL; MENDOZA, 2016.

3.8.2. Análisis y diseño del proyecto

A continuación, se detallada los casos de uso extendido que nos explica la interacción que tiene el usuario con el aplicativo.

3.8.2.1. Casos de uso extendido

Tabla 7. Inicio de Sesión

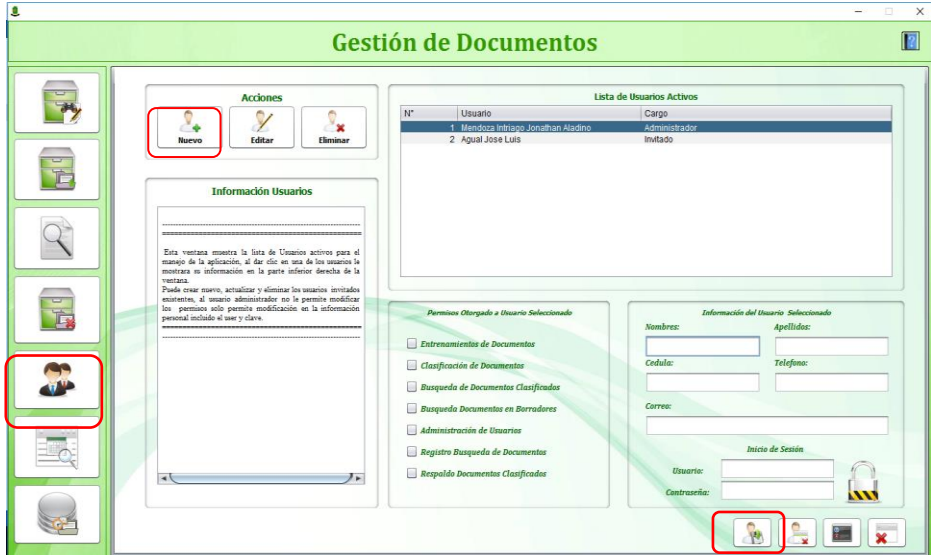
Caso de uso	Inicio de Sesión	
Código	Cu001	
Actores	Usuario (Administrador/Invitado)	
Propósito	Ingresar a todas funciones del sistema.	
Resumen	Para ingresar a la aplicación el usuario debe facilitar al sistema, el nombre de usuario designado, la clave para acceder a los diferentes permisos según sean sus competencias.	
Prioridad	Esencial, Primario.	
Tipo	Primario.	
Pre-condición	Debe estar la aplicación iniciada.	
Flujo normal de eventos	Acción del actor 1.-El caso de uso inicia cuando el usuario inicia la aplicación. 3.-El usuario ingresa los datos. 4.-El usuario da clic en iniciar.	Respuesta del sistema 2.-El sistema muestra el formulario de iniciar sesión. 5.-El sistema verifica los datos ingresados sean correctos. 6.-El sistema muestra el formulario principal según el perfil del usuario. 7.-El caso de uso termina cuando el usuario se encuentra en la interfaz según su perfil.
Flujos alternos	Línea 3.- El usuario ingresa a la aplicación y escribe mal el usuario y clave.	

	Línea 5.- La aplicación no permite el ingreso, mostrando un mensaje con el error sucedido.
Post - condición	El sistema muestra la interfaz de acuerdo al perfil del usuario.
Formulario	

AUTORES: AGUAL; MENDOZA,2016.

Tabla 8. Registrar Nuevo Usuario.

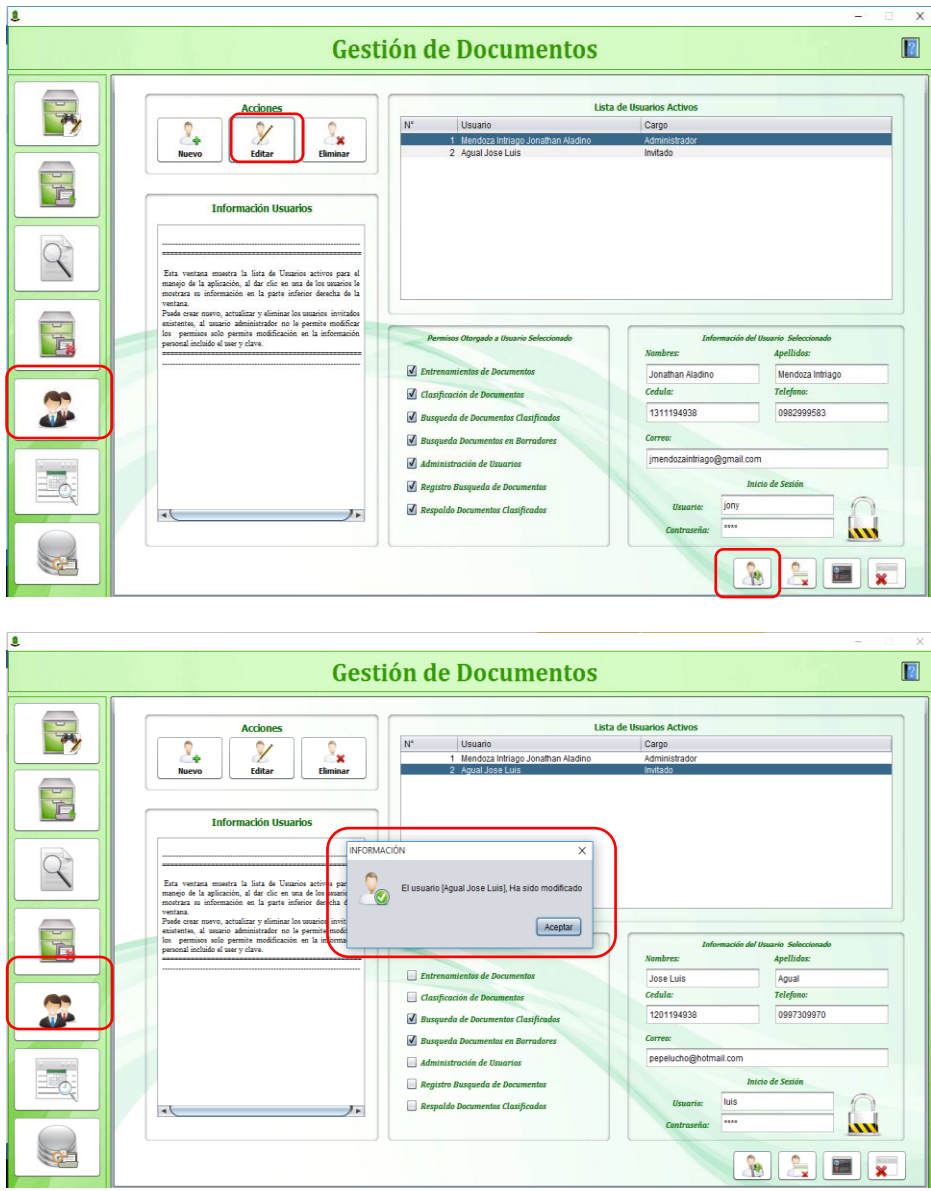
Caso de uso	Registrar Nuevo Usuario.	
Código	Cu002	
Actores	Usuario (Administrador)	
Propósito	Registrar un nuevo usuario para el sistema (usuario invitado)	
Resumen	El usuario (administrador) registra datos de un nuevo usuario, para la manipulación del sistema.	
Prioridad	Esencial.	
Tipo	Primario.	
Pre-condición	La aplicación debe estar iniciada y haber iniciado sesión el usuario.	
Flujo normal de eventos	Acción del actor 1.-El caso de uso inicia cuando el usuario selecciona el modulo “Administrador de Usuarios”. 3.-El usuario selecciona la opción “Nuevo”.	Respuesta del sistema 2.- El sistema muestra interfaz de “Administrador de Usuarios”.

	<p>5.-El usuario ingresa los datos del usuario y selecciona los permisos que se le va a otorgar.</p> <p>6.-El usuario da clic en “Guardar” para almacenar al usuario ingresado.</p>	<p>4.-Se carga el formulario con todos campos para registrar un nuevo usuario.</p> <p>7.-El sistema guarda el nuevo usuario y muestra mensaje de confirmación.</p> <p>8.-El caso de uso termina cuando el/los usuarios para utilizar el sistema han sido registrados correctamente.</p>
Flujos alternos	Línea 7.- Si los campos requeridos para el usuario no son ingresados correctamente, el sistema le mostrara un mensaje de advertencia.	
Post - condición	Registro de usuario, clave y permisos para acceder a las diferentes opciones del sistema.	
Formulario		

AUTORES: AGUAL, MENDOZA,2016.

Tabla 9. Editar Usuario.

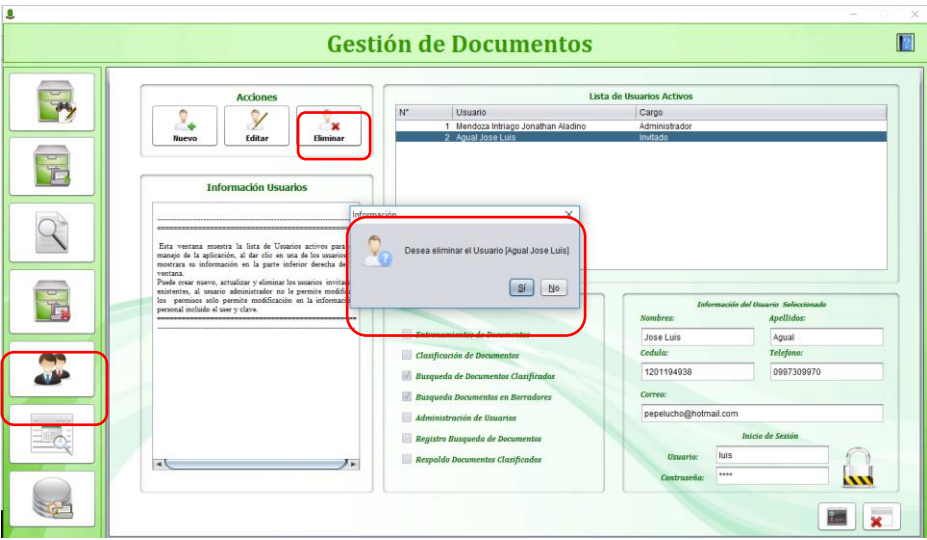
Caso de uso	Editar Usuario.	
Código	Cu003	
Actores	Usuario (Administrador)	
Propósito	Editar los usuarios registrado en el sistema.	
Resumen	El usuario (administrador) edita los datos registrado de los usuarios, para la manipulación del sistema. Los permisos del usuario (administrador) no pueden ser modificados	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	La aplicación debe estar iniciada y haber iniciado sesión el usuario.	
Flujo normal de eventos	<p>Acción del actor</p> <p>1.-El caso de uso inicia cuando el usuario selecciona el módulo “Administrador de Usuarios”.</p> <p>3.-El usuario selecciona un usuario.</p> <p>4.-El usuario selecciona la opción editar.</p> <p>6.-El usuario modifica los datos.</p> <p>7.-El usuario da clic en “Guardar”.</p>	<p>Respuesta del sistema</p> <p>2.-El sistema muestra interfaz de “Administrador de Usuarios” con los usuarios registrados.</p> <p>5.-El sistema permite la edición en los campos del usuario seleccionado.</p> <p>8.-El sistema guarda los datos modificados y muestra mensaje de confirmación.</p>

		9.-El caso de uso termina cuando los datos de un usuario se han modificado correctamente.
Flujos alternos	<p>Línea 7.- Si no desea guardar los cambios da clic en cancelar.</p> <p>Línea 8.- Si los campos requeridos para el usuario no son ingresados, el sistema le mostrara un mensaje de advertencia.</p>	
Post - condición	Datos de un usuario editados	
Formulario	 <p>The image displays two screenshots of a web application titled "Gestión de Documentos".</p> <p>Top Screenshot:</p> <ul style="list-style-type: none"> Acciones: A menu with "Nuevo", "Editar" (highlighted with a red box), and "Eliminar". Lista de Usuarios Activos: A table with columns "N°", "Usuario", and "Cargo". It lists two users: "1 Mendoza Intriago Jonathan Aladino" (Administrador) and "2 Aguil Jose Luis" (Invitado). Información Usuarios: A section with a description of the user management interface and a list of permissions for the selected user. Permisos Otorgados a Usuario Seleccionado: A list of permissions with checkboxes, including "Entrenamiento de Documentos", "Clasificación de Documentos", "Búsqueda de Documentos Clasificados", "Búsqueda Documentos en Servidores", "Administración de Usuarios", "Registro Búsqueda de Documentos", and "Respaldo Documentos Clasificados". Información del Usuario Seleccionado: A form with fields for "Nombre", "Apellido", "Cédula", "Teléfono", and "Correo". It also includes a "Inicio de Sesión" section with "Usuario" and "Contraseña" fields. <p>Bottom Screenshot:</p> <ul style="list-style-type: none"> The same interface as the top screenshot, but with a red box highlighting a message box that says "El usuario (Aguil Jose Luis), Ha sido modificado". 	

AUTORES: AGUAL; MENDOZA,2016.

Tabla 10. Eliminar Usuario.

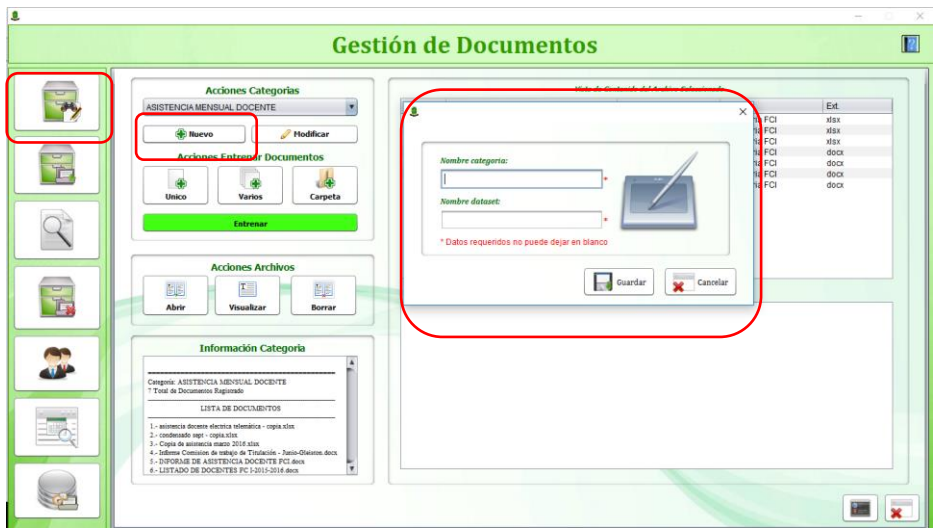
Caso de uso	Eliminar Usuario.	
Código	Cu004	
Actores	Usuario (Administrador)	
Propósito	Elimina los usuarios registrados en el sistema.	
Resumen	El usuario (administrador) elimina a los usuarios (invitados) registrados en el sistema.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	La aplicación debe estar iniciada y haber iniciado sesión el usuario.	
Flujo normal de eventos	<p>Acción del actor</p> <p>1.-El caso de uso inicia cuando el usuario selecciona el módulo “Administrador de Usuarios”.</p> <p>4.-El usuario selecciona un elemento de la lista.</p> <p>5.-El usuario selecciona la opción eliminar.</p>	<p>Respuesta del sistema</p> <p>2.-El sistema muestra la interfaz de “Administrador de Usuarios.</p> <p>3.-El sistema carga el formulario con los usuarios registrados.</p> <p>6.-El sistema muestra un mensaje de confirmación de eliminación.</p> <p>7.- El caso de uso termina cuando un usuario se ha eliminado correctamente.</p>

Flujos alternos	Línea 5.- Si no desea eliminar al usuario seleccionado da clic en cancelar.
Post - condición	Usuario eliminado del sistema.
Formulario	

AUTORES: AGUAL, MENDOZA, 2016.

Tabla 11. Nueva Categoría.

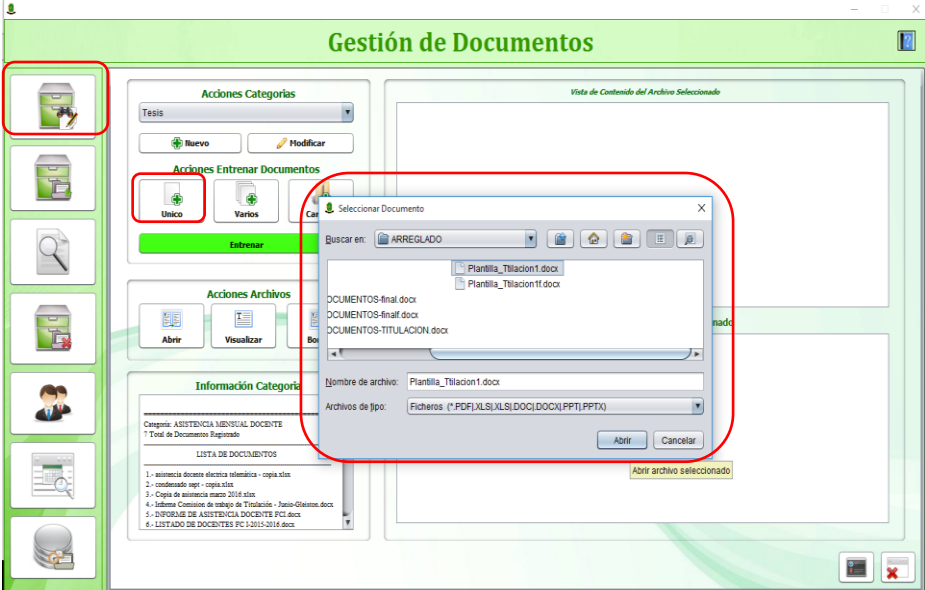
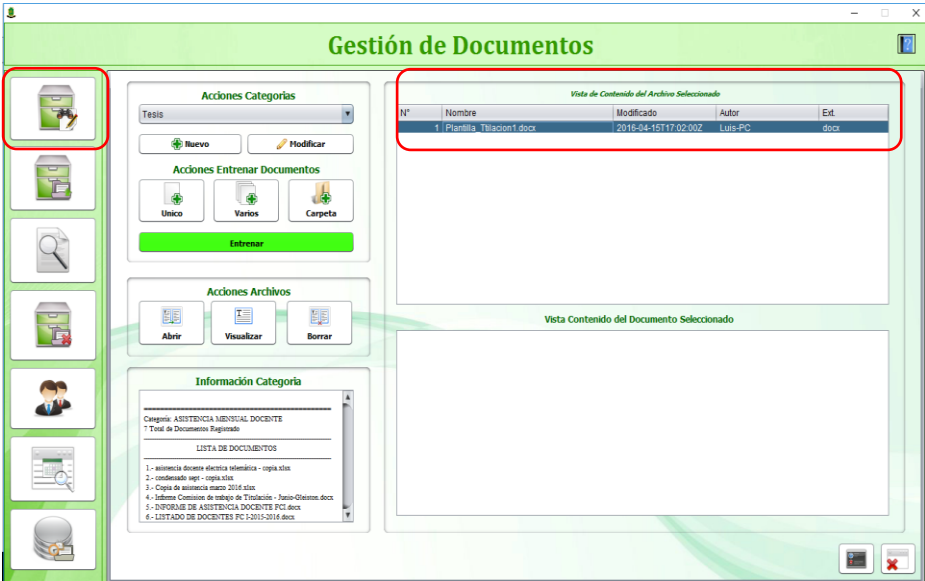
Caso de uso	Nueva Categoría.	
Código	Cu005	
Actores	Usuario (Administrador)	
Propósito	Registrar una nueva categoría al sistema.	
Resumen	El usuario registra una nueva categoría para la gestión de documentos.	
Prioridad	Esencial.	
Tipo	Primario	
Pre- condición	La aplicación debe estar iniciada y haber iniciado sesión el usuario.	
Flujo normal de eventos	Acción del actor 1.-El caso de uso inicia cuando el usuario selecciona el módulo “Importar Archivo de Entrenamiento”.	Respuesta del sistema

	<p>3.-El usuario selecciona la opción “Nueva”.</p> <p>5.-El usuario ingresa los datos requerido en el formulario.</p>	<p>2.-El sistema muestra la interfaz de “Importar Archivo de Entrenamiento”.</p> <p>4.-El sistema muestra el formulario de registrar una nueva categoría, con los campos requeridos.</p> <p>6.-Registra la nueva categoría, mostrando un mensaje de confirmación.</p> <p>7.-El caso de uso termina cuando la nueva categoría es creada correctamente.</p>
Flujos alternos	<p>Línea 6.- Si no desea registrar una nueva categoría, dar clic en cancelar.</p> <p>Línea 6.- Si los campos requeridos para el usuario no son ingresados, el sistema le mostrara un mensaje de advertencia.</p>	
Post - condición	Una nueva categoría registrada en el sistema.	
Formulario		

AUTORES: AGUAL; MENDOZA, 2016.

Tabla 12. Importar archivos de entrenamiento Único

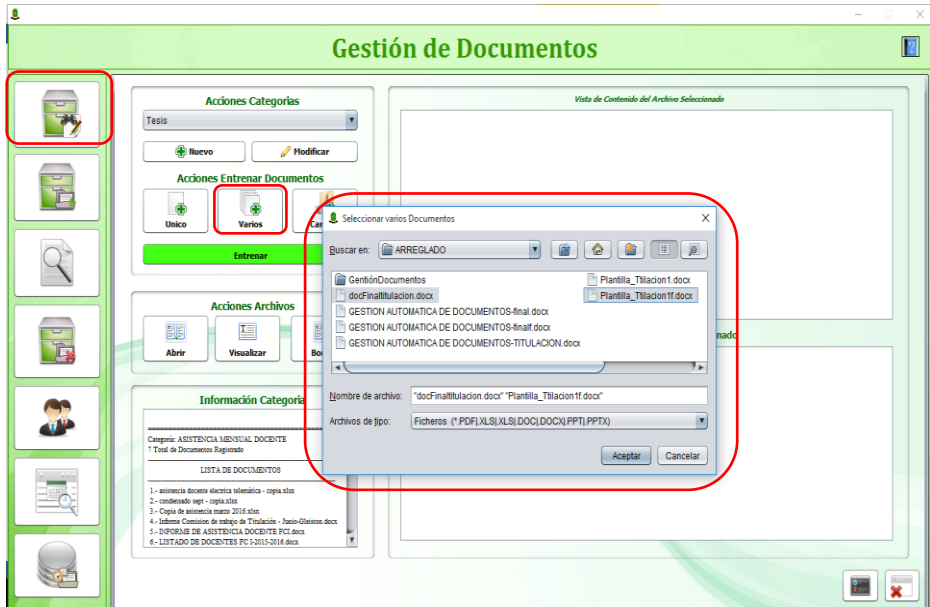
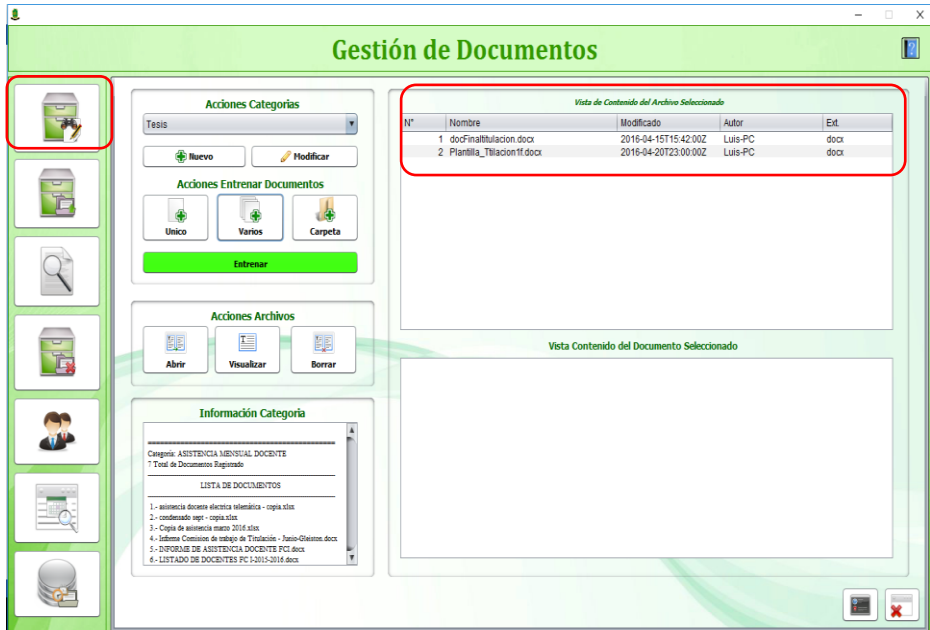
Caso de uso	Importar archivos de entrenamiento único.	
Código	Cu006	
Actores	Usuario (Administrador/Invitado)	
Propósito	Importar un único archivo para el entrenamiento.	
Resumen	El usuario importa un archivo por cada categoría para el entrenamiento de los documentos.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	La aplicación debe estar iniciada y haber iniciado sesión el usuario.	
Flujo normal de eventos	<p>Acción del actor</p> <p>1.-El caso de usos inicia cuando usuario selecciona el módulo “Importar Archivo de Entrenamiento”.</p> <p>3.-El usuario elije una categoría y selecciona la opción “Único”.</p> <p>5.-El usuario selecciona un archivo.</p>	<p>Respuesta del sistema</p> <p>2.-El sistema muestra la interfaz de “Importar Archivo de Entrenamiento”.</p> <p>4.-El sistema muestra el formulario para seleccionar el único archivo para la categoría seleccionada.</p> <p>6.-El sistema carga el archivo seleccionado mostrándolo en una tabla.</p> <p>7.-El sistema muestra el contenido del documento en la parte inferior.</p>

		8.-El caso de uso termina cuando el archivo seleccionado ha sido importado al sistema.
Flujos alternos	Ninguno	
Post - condición	Archivo único para el entrenamiento por categorías importados al sistema.	
Formulario	<div></div> <div></div>	

AUTORES: AGUAL; MENDOZA,2016.

Tabla 13. Importar varios archivos de entrenamiento.

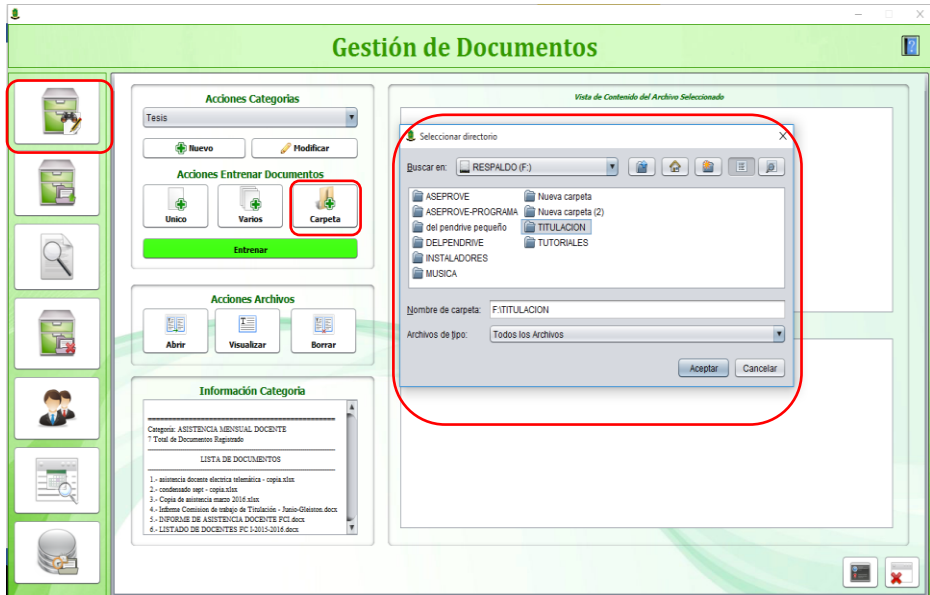
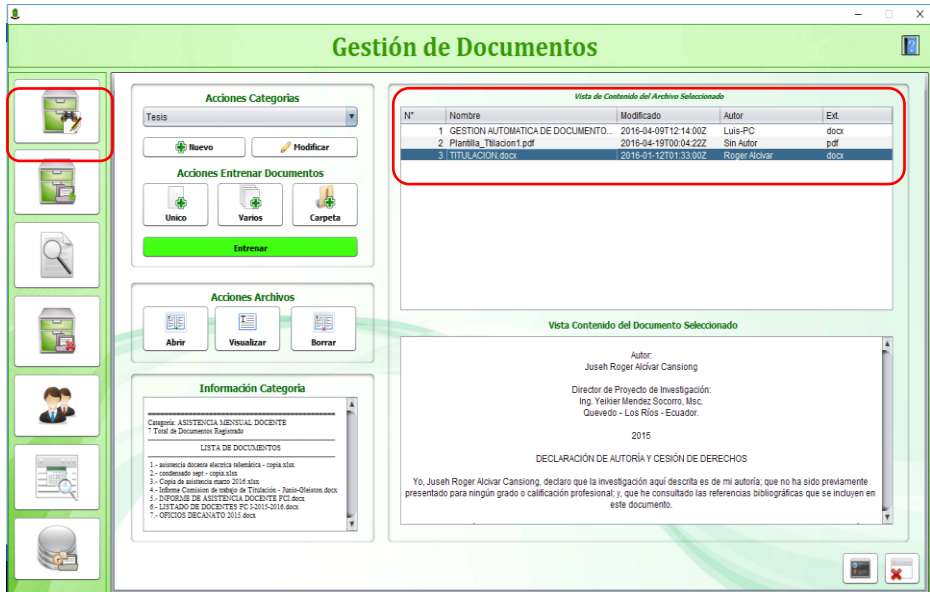
Caso de uso	Importar varios archivos de entrenamiento	
Código	Cu007	
Actores	Usuario (Administrador/Invitado)	
Propósito	Importar varios archivos para el entrenamiento.	
Resumen	El usuario importa varios archivos por cada categoría para el entrenamiento de los documentos.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	La aplicación debe estar iniciada y haber iniciado sesión el usuario.	
Flujo normal de eventos	<p>Acción del actor</p> <p>1.-El caso de uso inicia cuando el usuario selecciona el módulo “Importar Archivo de Entrenamiento”.</p> <p>3.-El usuario elije una categoría y selecciona la opción “Varios”.</p> <p>5.-El usuario selecciona los archivos.</p>	<p>Respuesta del sistema</p> <p>2.-El sistema muestra la interfaz de “Importar Archivo de Entrenamiento”.</p> <p>4.-El sistema muestra el formulario para seleccionar varios archivos para la categoría seleccionada.</p> <p>6.- Carga al sistema los archivos seleccionados mostrándolos en una tabla.</p>

		7.-El caso de uso termina cuando los archivos seleccionados han sido importados al sistema.
Flujos alternos	Ninguno	
Post - condición	Archivos para el entrenamiento por categorías importados al sistema.	
Formulario	 	

AUTORES: AGUAL, MENDOZA,2016.

Tabla 14. Importar carpeta de archivos de entrenamiento.

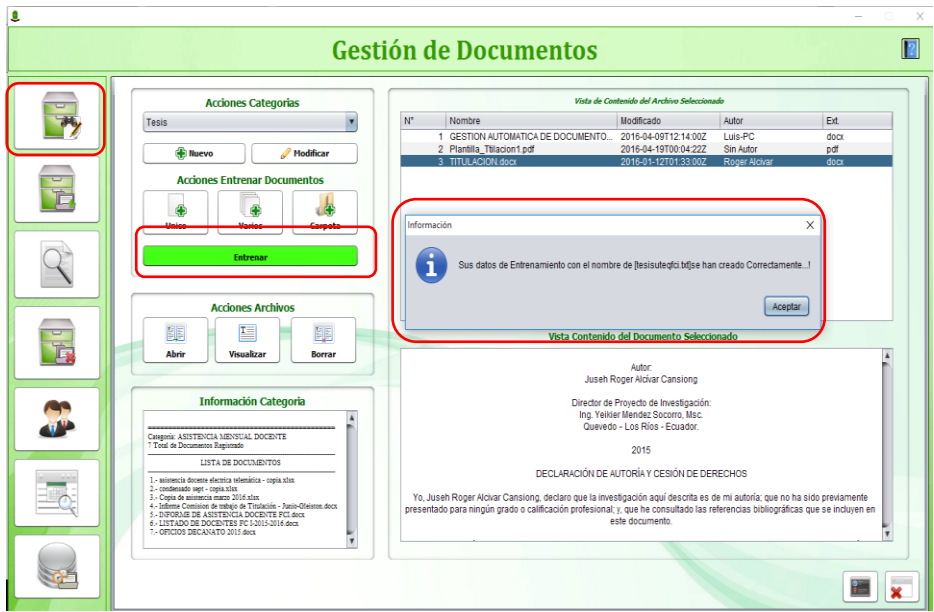
Caso de uso	Importar carpeta de archivos de entrenamiento.	
Código	Cu008	
Actores	Usuario (Administrador/Invitado)	
Propósito	Importar una carpeta de archivos para el entrenamiento.	
Resumen	El usuario importa una carpeta de archivos por cada categoría para entrenamiento de los documentos.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	La aplicación debe estar iniciada y haber iniciado sesión el usuario.	
Flujo normal de eventos	<p>Acción del actor</p> <p>1.-El caso de uso inicia cuando el usuario selecciona el módulo “Importar Archivo de Entrenamiento”.</p> <p>3.-El usuario elige una categoría y selecciona la opción “Carpeta”.</p> <p>5.-El usuario selecciona la carpeta de archivos.</p>	<p>Respuesta del sistema</p> <p>2.-El sistema muestra la interfaz de “Importar Archivo de Entrenamiento”.</p> <p>4.-El sistema muestra el formulario para seleccionar la carpeta de archivos para la categoría seleccionada.</p> <p>6.-El sistema carga los archivos de la carpeta mostrándolos en una tabla.</p> <p>7.-El sistema muestra el contenido del documento en la parte inferior.</p>

		8.- El caso de uso termina cuando los archivos de la carpeta seleccionada han sido importados al sistema.
Flujos alternos	Ninguno.	
Post - condición	Archivos para el entrenamiento por categorías importados al sistema.	
Formulario	 	

AUTORES: AGUAL; MENDOZA,2016.

Tabla 15. Entrenar archivos por categoría.

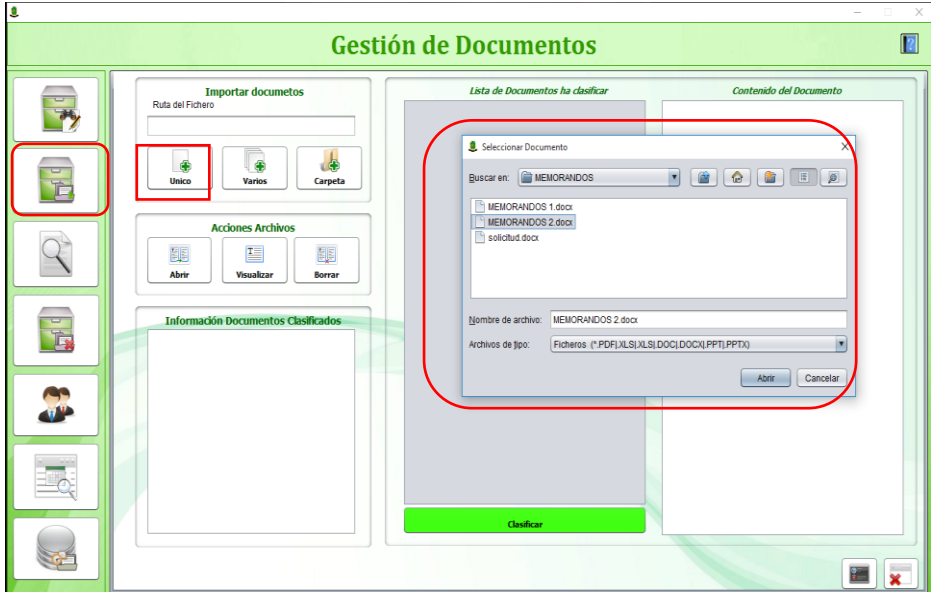
Caso de uso	Entrenar archivos por categoría.	
Código	Cu009	
Actores	Usuario (Administrador/Invitado)	
Propósito	Generar el modelo de aprendizaje por cada categoría.	
Resumen	El usuario procede a generar el modelo de aprendizaje a partir de los archivos importados por cada categoría.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	El usuario debe haber iniciado sesión y los archivos de entrenamiento importados al sistema.	
Flujo normal de eventos	<p>Acción del actor</p> <p>1.-El caso de uso inicia cuando el usuario selecciona el módulo “Importar Archivo de Entrenamiento”.</p> <p>3.-El usuario elije una categoría.</p> <p>5.-El usuario selecciona un archivo importado y da clic en la opción “Entrenar”.</p>	<p>Respuesta del sistema</p> <p>2.-El sistema muestra la interfaz de “Importar Archivo de Entrenamiento”.</p> <p>4.-El sistema muestra los archivos importados”.</p> <p>6.-El sistema muestra un mensaje de confirmación expresando que sus datos de entrenamiento han sido generados.</p> <p>7.-El caso de uso termina cuando los modelos de aprendizaje por cada categoría son generados.</p>

Flujos alternos	Ninguno.
Post - condición	Modelos de aprendizaje generados por cada categoría.
Formulario	

AUTORES: AGUAL; MENDOZA,2016.

Tabla 16. Importar archivo único.

Caso de uso	Importar archivo único.	
Código	Cu010	
Actores	Usuario (Administrador/Invitado)	
Propósito	Importar un archivo para su clasificación automática.	
Resumen	El usuario importa un archivo para la clasificación automática de documentos.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	El usuario debe haber iniciado sesión y los modelos de aprendizaje deben haberse generado.	
Flujo normal de eventos	Acción del actor 1.-El caso de uso inicia cuando el usuario selecciona el modulo “Clasificar documentos”.	Respuesta del sistema

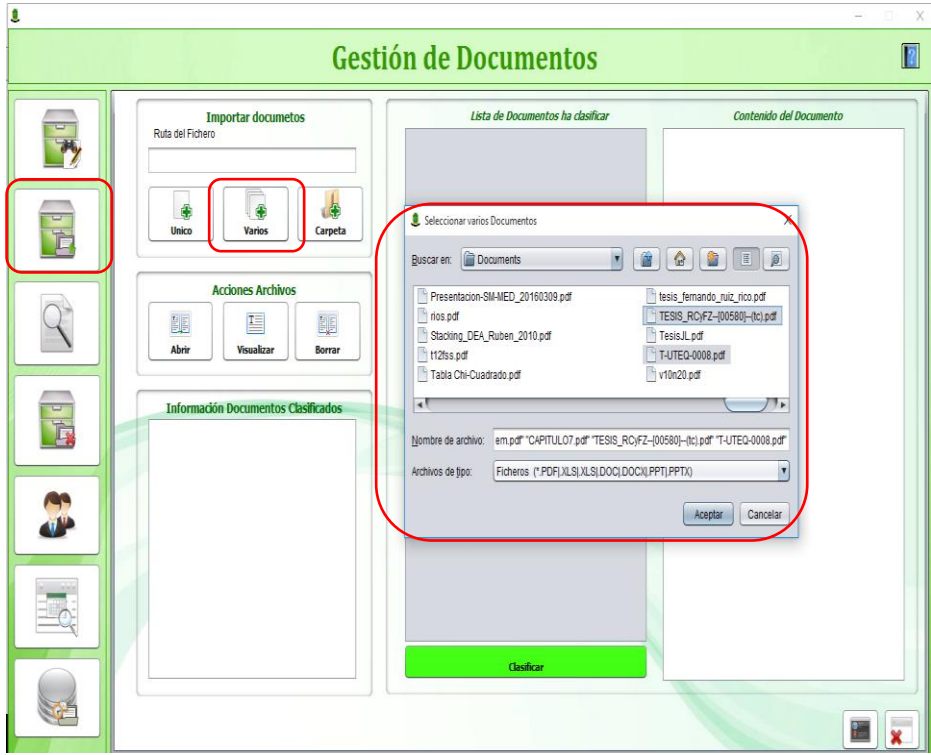
	<p>3.-El usuario selecciona la opción “Único”.</p> <p>5.-El usuario selecciona el archivo.</p>	<p>2.-El sistema muestra la interfaz de “Clasificar documentos”.</p> <p>4.-El sistema muestra el formulario para seleccionar un único archivo para clasificar.</p> <p>6.-El sistema carga al sistema el archivo seleccionado en una tabla.</p> <p>7.-El caso de uso termina cuando el archivo para la clasificación es importado al sistema correctamente.</p>
Flujos alternos	Línea 6.- El administrador puede visualizar el contenido del documento en la parte izquierda, dando doble clic sobre un archivo.	
Post - condición	Los documentos digitales son importados al sistema para su respectiva clasificación.	
Formulario		

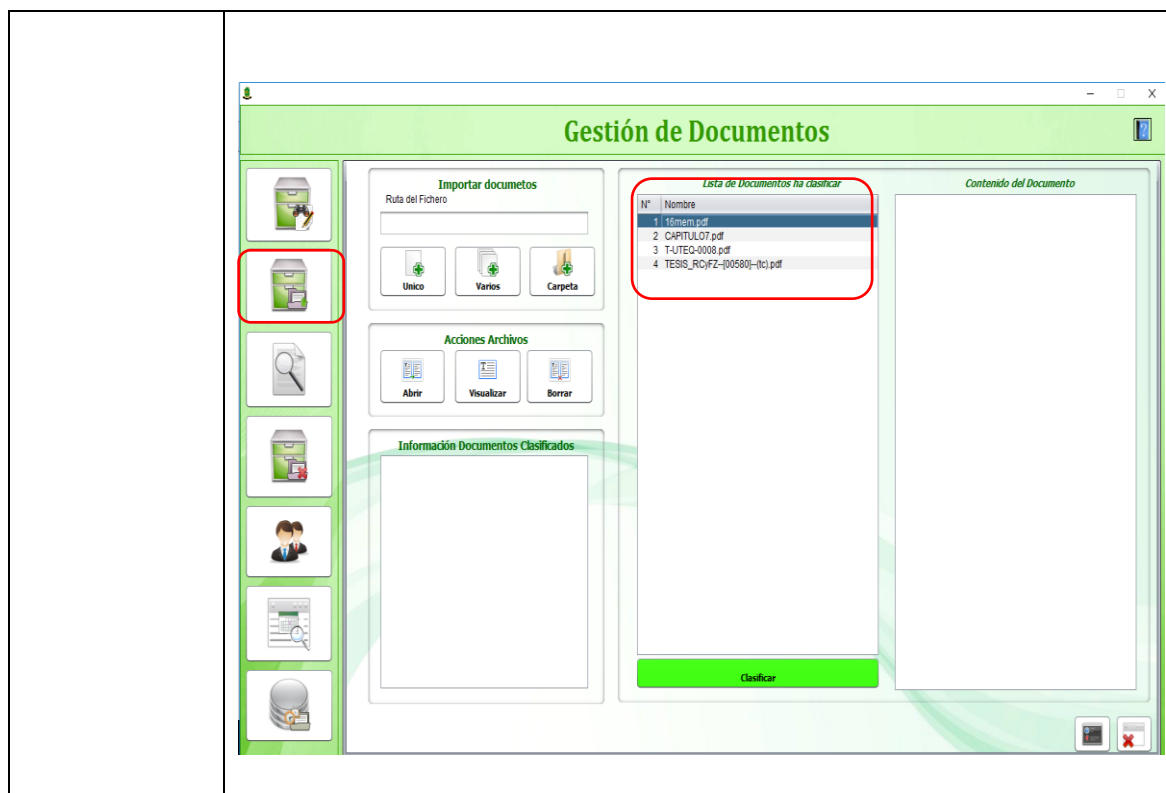


AUTORES: AGUAL, MENDOZA, 2016.

Tabla 17. Importar Varios archivos.

Caso de uso	Importar Varios archivos.	
Código	Cu011	
Actores	Usuario (Administrador/Invitado)	
Propósito	Importar varios archivos para su clasificación automática.	
Resumen	El usuario importa varios archivos para la clasificación automática de documentos.	
Prioridad	Esencial.	
Tipo	Primario.	
Pre-condición	El usuario debe haber iniciado sesión y los modelos de aprendizaje deben haberse generado.	
Flujo normal de eventos	Acción del actor 1.-El caso de uso inicia cuando el usuario inicia el modulo “Clasificar documentos”. 3.-El usuario selecciona la opción “Varios”.	Respuesta del sistema 2.-El sistema muestra la interfaz de “Clasificar documentos”.

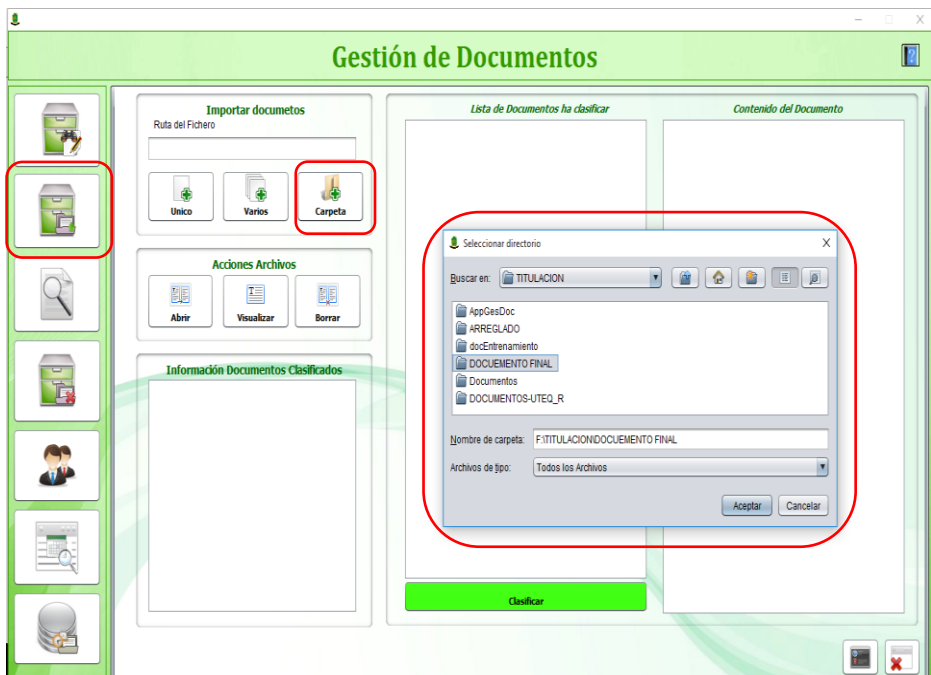
	<p>5.-El usuario selecciona los archivos que desea importar.</p>	<p>4.-El sistema muestra el formulario para seleccionar varios archivos para clasificar.</p> <p>6.-Carga al sistema los archivos seleccionados en una tabla.</p> <p>7.-El caso de uso termina cuando los archivos para la clasificación son importados al sistema correctamente.</p>
Flujos alternos	Línea 6.- El usuario puede visualizar el contenido del documento en la parte izquierda, dando doble clic sobre un archivo.	
Post - condición	Los documentos digitales son importados al sistema para su clasificación.	
Formulario		

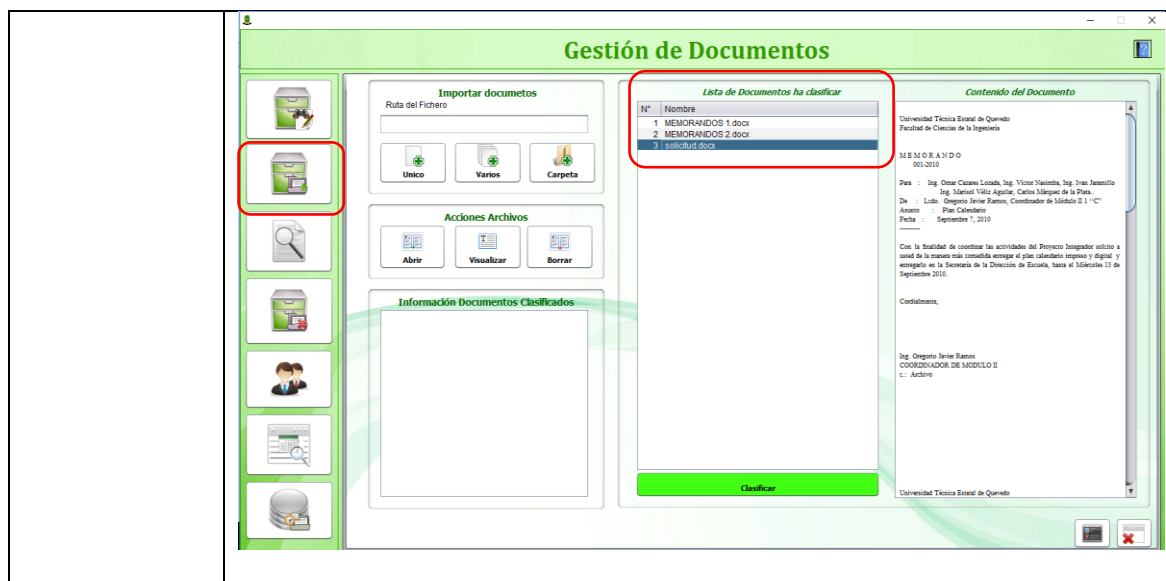


AUTORES: AGUAL, MENDOZA,2016.

Tabla 18. Importar carpeta de archivos.

Caso de uso	Importar carpeta de archivos.	
Código	Cu012	
Actores	Usuario (Administrador/Invitado).	
Propósito	Importar carpeta de archivos para su clasificación automática.	
Resumen	El usuario importa una carpeta de archivos para la clasificación automática de documentos.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	El usuario debe haber iniciado sesión y los modelos de aprendizaje deben haberse generado.	
Flujo normal de eventos	Acción del actor 1.-El caso de uso inicia cuando el usuario selecciona el modulo “Clasificar documentos”.	Respuesta del sistema 2.-El sistema muestra la interfaz de “Clasificar documentos”.

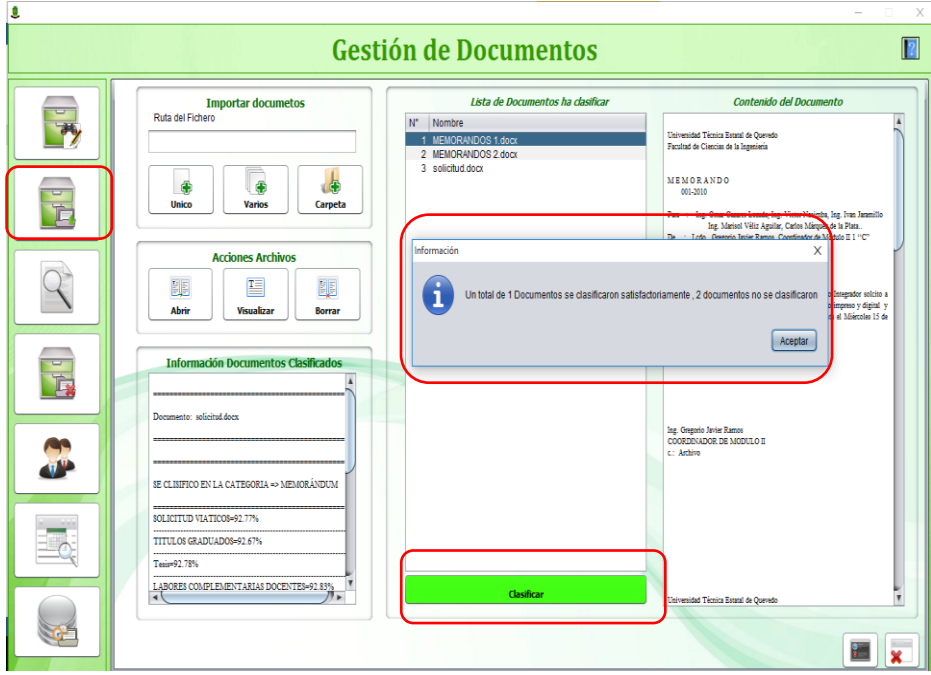
	<p>3.-El usuario selecciona la opción “Carpeta”.</p> <p>5.-El usuario selecciona la carpeta de archivos.</p>	<p>4.-El sistema muestra el formulario para seleccionar la carpeta de archivos para clasificar.</p> <p>5.-Carga al sistema los archivos seleccionados en una tabla.</p> <p>6.-El caso de uso termina cuando los archivos de la carpeta para la clasificar son importados al sistema correctamente.</p>
Flujos alternos	Línea 6.- El administrador puede visualizar el contenido del documento en la parte izquierda, dando doble clic sobre un archivo.	
Post - condición	Los documentos digitales son importados al sistema para su clasificación.	
Formulario		



AUTORES: AGUAL, MENDOZA,2016.

Tabla 19. Clasificar.

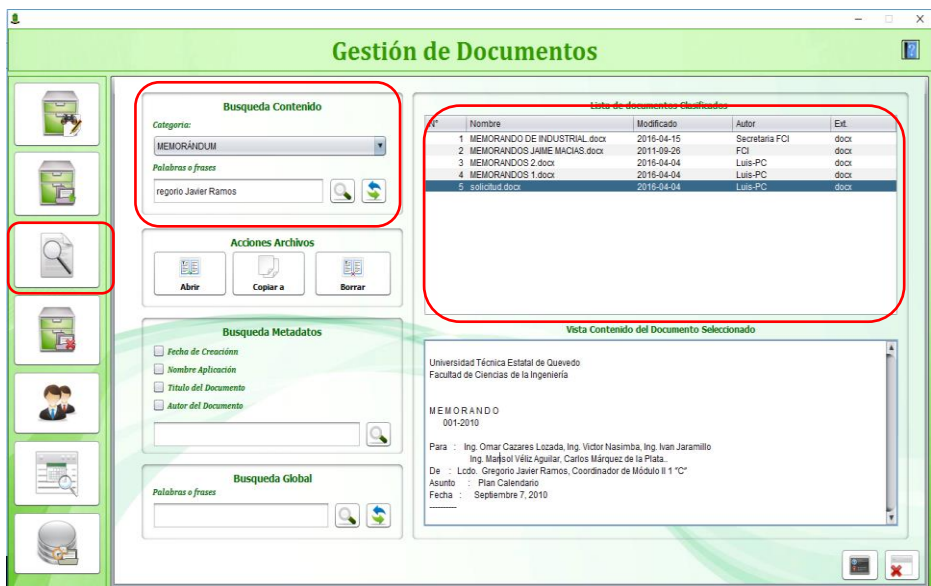
Caso de uso	Clasificar.	
Código	Cu013	
Actores	Usuario (Administrador/Invitado)	
Propósito	Clasificar automáticamente los documentos importados al sistema.	
Resumen	El usuario procede a clasificar todos los documentos importados al sistema.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	El usuario debe haber iniciado sesión, los modelos de aprendizaje generados y los archivos deben previamente importados al sistema.	
Flujo normal de eventos	Acción del actor 1.-El caso de uso inicia cuando el usuario selecciona el modulo “Clasificar documentos”. 3.-El usuario importa los documentos a clasificar (único, varios, carpeta).	Respuesta del sistema 2.-El sistema muestra la interfaz de “Clasificar documentos”.

	<p>4.-El usuario selecciona la opción “Clasificar”.</p>	<p>5.- El sistema clasifica los archivos, mostrados un mensaje, indicando a que categoría fue asignada automáticamente.</p> <p>6.-El caso de uso termina cuando los archivos son clasificados correctamente.</p>
Flujos alternos	<p>Línea 3.- Los documentos para clasificar importados pueden ser quitados dando clic derecho sobre ellos.</p> <p>Línea 5.- En caso que los documentos ya se encuentren en la base de datos como clasificado, el sistema mostrara un mensaje diciendo que los documentos ya están en la base de datos</p>	
Post - condición	Los documentos digitales son automáticamente clasificados.	
Formulario		

AUTORES: AGUAL; MENDOZA,2016.

Tabla 20. Búsqueda por categoría.

Caso de uso	Búsqueda por Contenido.	
Código	Cu014	
Actores	Usuario (Administrador/Invitado)	
Propósito	Buscar los documentos clasificados por categoría.	
Resumen	El usuario procede buscar documento por categoría, escribiendo un criterio de búsqueda.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	El usuario debe haber iniciado sesión y los documentos deben haber sido clasificados.	
Flujo normal de eventos	Acción del actor 1.-El caso de uso inicia cuando el usuario selecciona el modulo “Buscar documentos”. 3.-El usuario selecciona una categoría. 4.-El usuario escribe el criterio de búsqueda. 5.-El usuario presiona la tecla “enter” o da un clic en botón con icono de lupa para buscar.	Respuesta del sistema 2.-El sistema muestra la interfaz de “Buscar documentos”. 6.-El sistema muestra los documentos que han sido encontrados en una tabla, con el criterio proporcionado por el usuario.

	<p>7.-El usuario selecciona un documento de la lista de encontrados dando doble clic.</p> <p>9.-El usuario da clic en “Abrir” para visualizar el documento.</p>	<p>8.-Muestra el contenido el documento seleccionado en la parte superior.</p> <p>10.-El caso de uso termina cuando el usuario obtiene el documento requerido.</p>
Flujos alternos	<p>Línea 6.- En caso que no existan documentos con el criterio de búsqueda, el sistema mostrará un mensaje.</p> <p>Línea 7.-También puede enviar documento un clasificado a borradores seleccionando mediante la opción “borrar” o dando clic derecho sobre el documento.</p> <p>Línea 9.- El administrador puede copiar el archivo encontrado un lugar específico dando clic sobre el botón “copiar a”.</p>	
Post - condición	Documentos encontrados.	
Formulario		

AUTOR: AGUAL; MENDOZA,2016.

Tabla 21. Búsqueda por Metadatos.

Caso de uso	Búsqueda por Metadatos.	
Código	Cu015	
Actores	Usuario (Administrador/Invitado).	
Propósito	Buscar los documentos clasificados por metadatos.	
Resumen	El usuario procede buscar documento por metadatos que contiene el mismo, escribiendo un criterio de búsqueda.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	El usuario debe haber iniciado sesión y los documentos deben haber sido clasificados.	
Flujo normal de eventos	<p>Acción del actor</p> <p>1.-El caso de uso inicia cuando el usuario selecciona el modulo “Buscar documentos”.</p> <p>3.-El usuario selecciona una opción de los metadatos:</p> <ul style="list-style-type: none"> a) Fecha de creación. b) Nombre de la aplicación. c) Título del documento. d) Autor del documento. <p>4.- El usuario escribe el criterio de búsqueda, según los metadatos seleccionado y presiona la tecla “enter” o da un clic en botón con icono de lupa para buscar.</p>	<p>Respuesta del sistema</p> <p>2.-El sistema muestra la interfaz de “Buscar documentos”.</p> <p>5.-El sistema muestra los documentos que han sido encontrados en una tabla, con el</p>

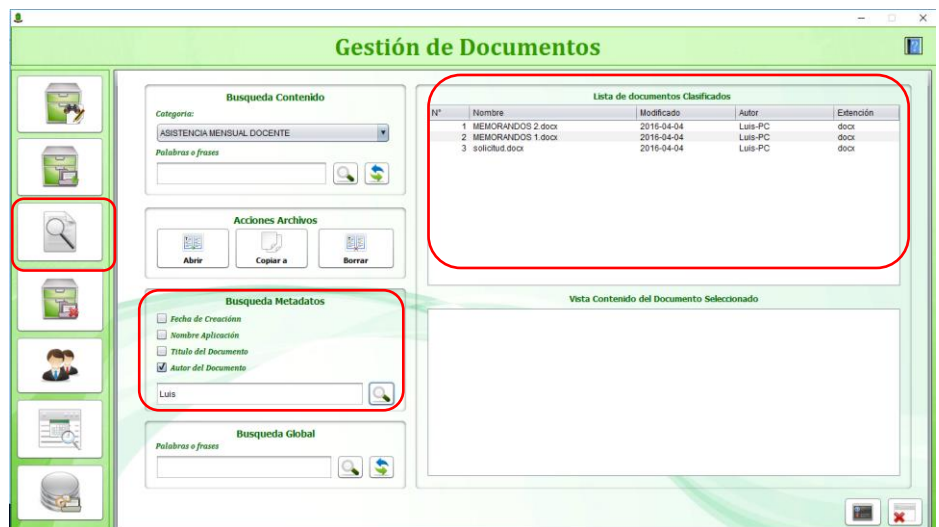
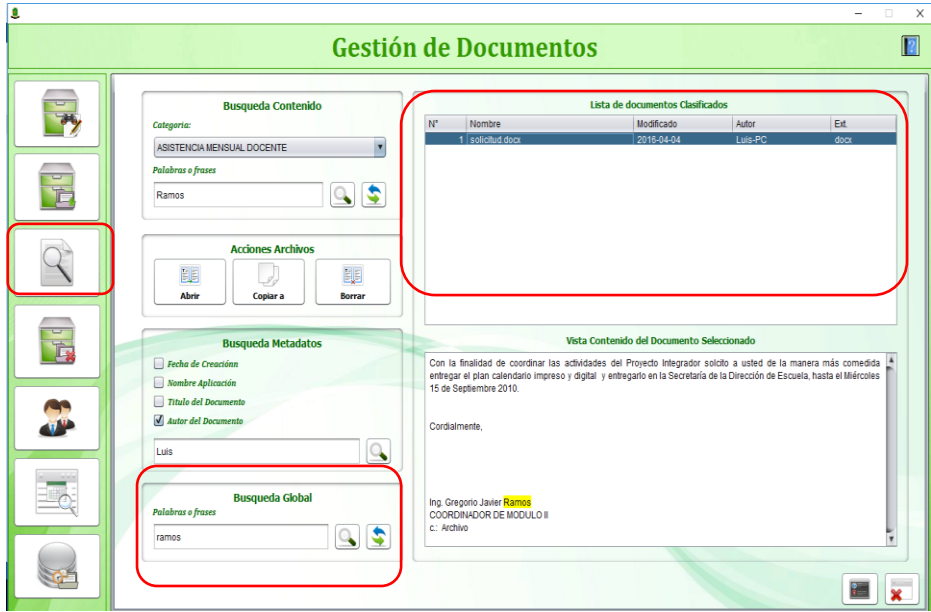
	<p>6.- Selecciona un documento de la lista de encontrados dando doble clic.</p> <p>8.-El administrador da clic en “Abrir” para visualizar el documento.</p>	<p>criterio proporcionado por el usuario.</p> <p>7.-El sistema muestra el contenido el documento seleccionado en la parte superior.</p> <p>9.-El caso de uso termina cuando el usuario obtiene el documento requerido.</p>																				
Flujos alternos	<p>Línea 6.- En caso que no existan documentos con el criterio de búsqueda, el sistema mostrará un mensaje.</p> <p>Línea 7.-También puede enviar un documento clasificado a borradores seleccionando mediante la opción “borrar” o dando clic derecho sobre el documento.</p> <p>Línea 8.- El usuario puede copiar el archivo encontrado un lugar específico dando clic sobre el botón “copiar a”.</p>																					
Post - condición	Documentos encontrados.																					
Formulario	 <p>The screenshot shows the 'Gestión de Documentos' application. It features a sidebar with icons for document management. The main area includes a 'Busqueda Contenido' section with a category dropdown set to 'ASISTENCIA MENSUAL DOCENTE' and a search bar. Below this are 'Acciones Archivos' buttons: 'Abrir', 'Copiar a', and 'Borrar'. There is also a 'Busqueda Metadatos' section with checkboxes for 'Fecha de Creación', 'Nombre Aplicación', 'Título del Documento', and 'Autor del Documento' (which is checked and has 'Luis' entered in the search bar). At the bottom is a 'Busqueda Global' section. On the right, a table titled 'Lista de documentos Clasificados' displays a list of documents with columns for N°, Nombre, Modificado, Autor, and Extensión. The table contains three rows of data.</p> <table><tr><th>N°</th><th>Nombre</th><th>Modificado</th><th>Autor</th><th>Extensión</th></tr><tr><td>1</td><td>MEMORANDOS 2.docx</td><td>2016-04-04</td><td>Luis-PC</td><td>docx</td></tr><tr><td>2</td><td>MEMORANDOS 1.docx</td><td>2016-04-04</td><td>Luis-PC</td><td>docx</td></tr><tr><td>3</td><td>solicitud.docx</td><td>2016-04-04</td><td>Luis-PC</td><td>docx</td></tr></table>		N°	Nombre	Modificado	Autor	Extensión	1	MEMORANDOS 2.docx	2016-04-04	Luis-PC	docx	2	MEMORANDOS 1.docx	2016-04-04	Luis-PC	docx	3	solicitud.docx	2016-04-04	Luis-PC	docx
N°	Nombre	Modificado	Autor	Extensión																		
1	MEMORANDOS 2.docx	2016-04-04	Luis-PC	docx																		
2	MEMORANDOS 1.docx	2016-04-04	Luis-PC	docx																		
3	solicitud.docx	2016-04-04	Luis-PC	docx																		

Tabla 22. Búsqueda Global.

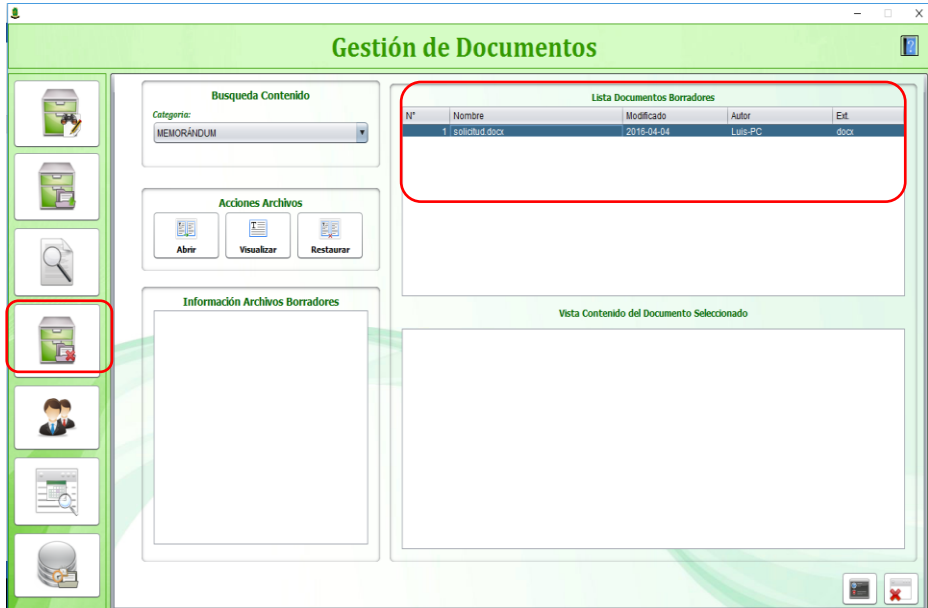
Caso de uso	Búsqueda Global.	
Código	Cu016	
Actores	Usuario (Administrador/Invitado)	
Propósito	Buscar los documentos clasificados por todas las categorías disponibles.	
Resumen	El usuario procede buscar los documentos, por todas las categorías, escribiendo un criterio de búsqueda.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	El usuario debe haber iniciado sesión y los documentos deben haber sido clasificados.	
Flujo normal de eventos	<p>Acción del actor</p> <p>1.-El caso de usos inicia cuando el usuario selecciona el modulo “Buscar documentos”</p> <p>3.-El usuario selecciona una opción “Búsqueda Global”.</p> <p>4.-El usuario escribe el criterio de búsqueda.</p> <p>5.-El usuario presiona la tecla “enter” o da un clic en botón con icono de lupa para buscar.</p>	<p>Respuesta del sistema</p> <p>2.-El sistema muestra la interfaz de “Buscar documentos”.</p> <p>6.-El sistema muestra los documentos que han sido encontrados en una tabla, con el criterio proporcionado.</p>

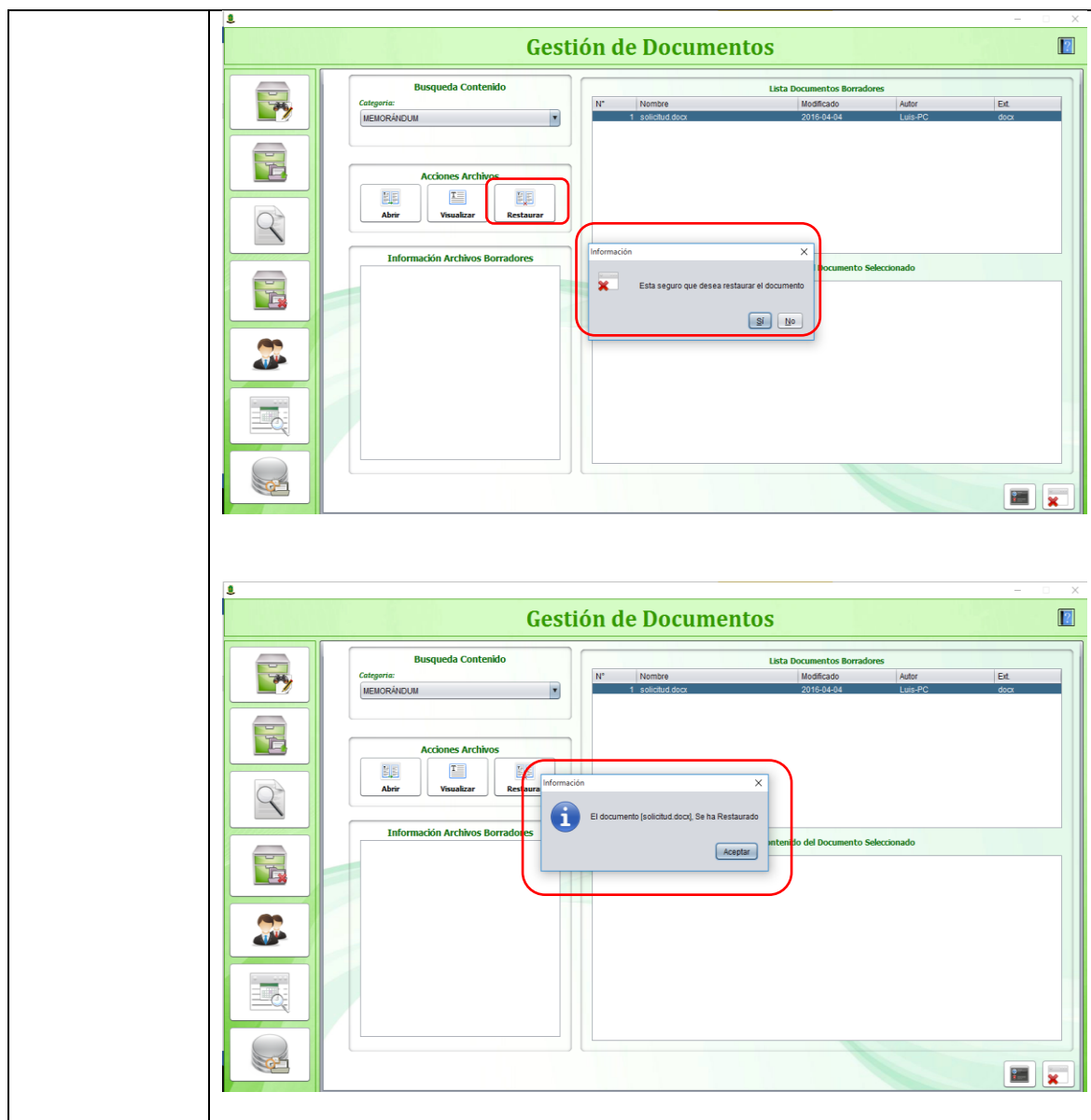
	<p>7.- Selecciona un documento de la lista de encontrados dando doble clic.</p> <p>9.-El administrador da clic en “Abrir” para visualizar el documento.</p>	<p>8.-El sistema muestra el contenido el documento seleccionado en la parte superior.</p> <p>10.-El caso de uso termina cuando el usuario obtiene el documento requerido.</p>
Flujos alternos	<p>Línea 6.- En caso que no existan documentos con el criterio de búsqueda, el sistema mostrará un mensaje.</p> <p>Línea 7.-También puede enviar un documento clasificado a borradores seleccionando mediante la opción “borrar” o dando clic derecho sobre el documento.</p> <p>Línea 9.- El usuario puede copiar el archivo encontrado un lugar específico dando clic sobre el botón “copiar a”.</p>	
Post - condición	Documentos encontrados.	
Formulario		

AUTORES: AGUAL; MENDOZA,2016.

Tabla 23. Restaurar documentos.

Caso de uso	Restaurar documentos.	
Código	Cu017	
Actores	Usuario (Administrador/Invitado).	
Propósito	Restaurar los documentos clasificados eliminados	
Resumen	El usuario restaura los documentos clasificados que han sido eliminados por el usuario en la opción de búsqueda de documentos.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	El usuario debe haber iniciado sesión, los documentos deben haber borrados de la lista de clasificados.	
Flujo normal de eventos	Acción del actor 1.-El caso de uso inicia cuando el usuario selecciona el modulo “Borradores de documentos” 3.-El usuario selecciona una categoría. 5.-El usuario selecciona un documento de la lista. 6.-El usuario da clic en el botón “Restaurar”. 8.-El usuario da clic en el botón “Si”.	Respuesta del sistema 2.-El sistema muestra la interfaz de “Borradores de documentos”. 4.-El sistema muestra la lista de documentos en una tabla. 7.-El sistema muestra un mensaje de diálogo preguntado “si está seguro de restaurar el documento “.

		<p>9.-El sistema muestra un mensaje de confirmación que se ha restaurado el documento.</p> <p>10.-El caso de uso termina cuando el documento seleccionado ha sido restaurado.</p>
Flujos alternos	<p>Línea 4.- Si la categoría seleccionada no posee documentos enviados a borradores, la tabla no se reflejará.</p> <p>Línea 6.- El usuario puede abrir el documento seleccionado dando clic en botón “Abrir”.</p> <p>Línea 6.- El usuario poder visualizar el contenido del documento seleccionado en la parte inferior, dando clic en el “Visualizar”.</p> <p>Línea 8.- Si no requiere restaurar el documento el usuario procede a dar clic en “No”.</p>	
Post - condición	Los documentos que el usuario eliminó en la opción búsqueda documentos, son restaurados.	
Formulario		

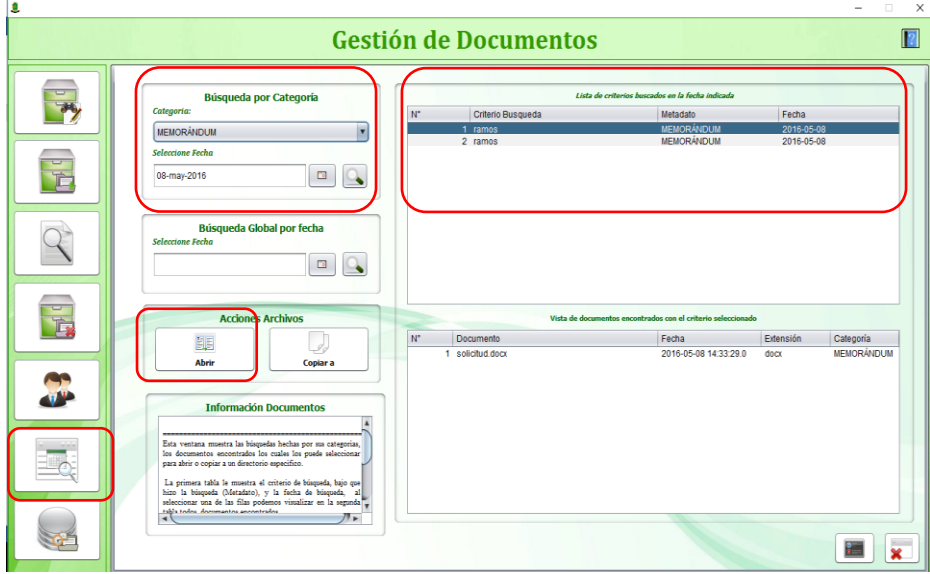


AUTORES: AGUAL; MENDOZA,2016.

Tabla 24. Búsqueda por categoría.

Caso de uso	Búsqueda por categoría.
Código	Cu018
Actores	Usuario (Administrador/Invitado).
Propósito	Buscar documentos por categorías a partir del historial de búsqueda.
Resumen	El usuario realiza una búsqueda de documentos seleccionando cada una de las categorías, mediante el historial de búsquedas.
Prioridad	Esencial.
Tipo	Primario

Pre-condición	El usuario debe haber iniciado sesión y realizadas varias búsquedas por categorías.	
Flujo normal de eventos	<p>Acción del actor</p> <p>1.-El caso de uso inicia cuando el usuario selecciona el modulo “Registro de Búsquedas”.</p> <p>3.-El usuario selecciona la opción “Búsqueda por categoría”.</p> <p>4.-El usuario selecciona una categoría.</p> <p>5.-El usuario selecciona una fecha en el calendario y da clic en botón buscar.</p> <p>7.-El usuario selecciona un elemento de la tabla con doble clic.</p> <p>9.-El usuario procede abrir el documento seleccionado mediante el botón “Abrir”.</p>	<p>Respuesta del sistema</p> <p>2.-El sistema muestra la interfaz de “Registro de Búsquedas”.</p> <p>6.-El sistema muestra la lista de elementos en una tabla con los criterios de búsqueda realizados.</p> <p>8.-El sistema muestra en la parte inferior el nombre del documento y la fecha en que se realizó la búsqueda.</p> <p>10.-El sistema abre el documento seleccionado.</p>

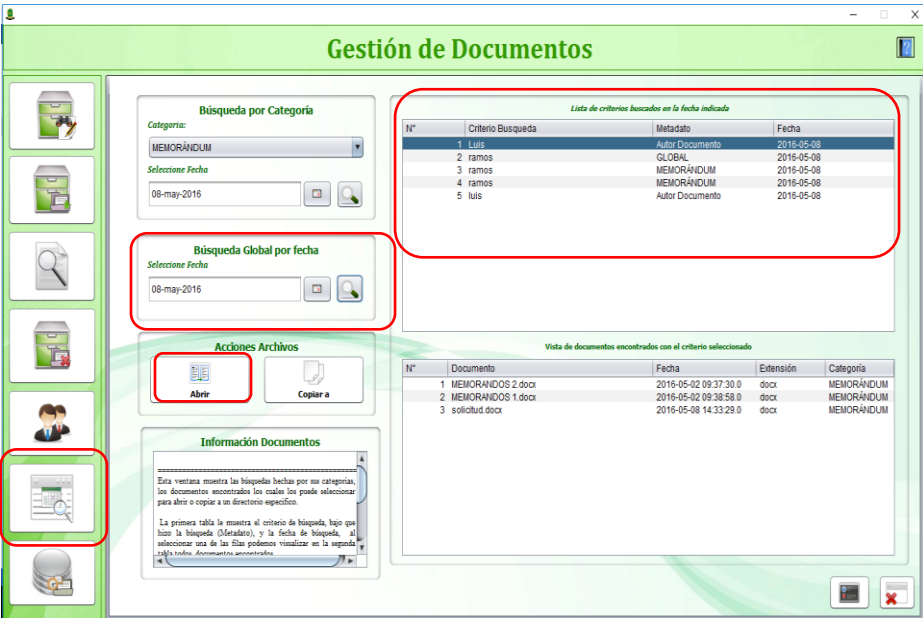
		11.-El caso de uso termina cuando el documento es encontrado a través del historial de búsquedas.
Flujos alternos	<p>Línea 4.-Si la categoría seleccionada no posee registro de búsqueda de documentos, la tabla no se reflejará.</p> <p>Línea 10.- El usuario puede copiar el documento seleccionado en lugar específico, con la opción “Copiar a “.</p>	
Post - condición	Documentos encontrados a partir del historial de búsquedas.	
Formulario		

AUTORES: AGUAL; MENDOZA,2016.

Tabla 25. Búsqueda global por fecha.

Caso de uso	Búsqueda global por fecha.
Código	Cu019
Actores	Usuario (Administrador/Invitado).
Propósito	Buscar documentos por categorías a partir del historial de búsqueda.
Resumen	El usuario realiza una búsqueda de documentos de manera global por todas las categorías, mediante el historial de búsquedas.
Prioridad	Esencial.
Tipo	Primario
Pre-condición	El usuario debe haber iniciado sesión y realizadas varias búsquedas por categorías.

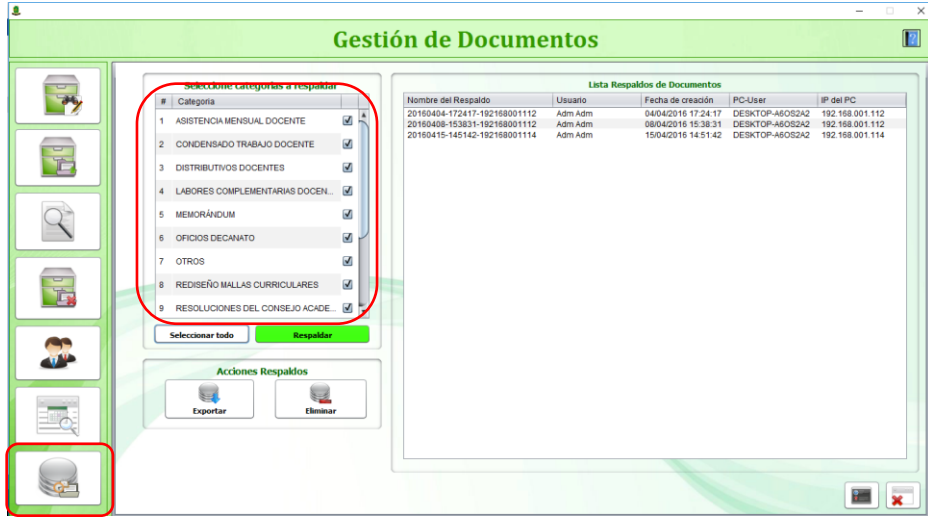
	Acción del actor	Respuesta del sistema
Flujo normal de eventos	<p>1.-El caso uso inicia cuando el usuario selecciona el modulo “Registro de Búsquedas”.</p> <p>3.-El usuario selecciona la opción “Búsqueda global por fecha”.</p> <p>4.-El usuario selecciona una fecha en el calendario.</p> <p>5.-El usuario da clic en botón buscar.</p>	<p>2.-El sistema muestra la interfaz de “Registro de Búsquedas”.</p>
	<p>7.-El usuario selecciona un elemento de la tabla con doble clic.</p> <p>9.-El usuario procede abrir el documento seleccionado mediante el botón “Abrir”.</p>	<p>6.-El sistema muestra la lista de elementos en una tabla con los criterios de búsqueda realizados.</p> <p>8.-El sistema muestra en la parte inferior el nombre del documento y la fecha en que se realizó la búsqueda.</p> <p>10.-El sistema abre el documento seleccionado.</p> <p>11.-El caso de uso termina cuando el documento es encontrado a través del historial de búsquedas.</p>

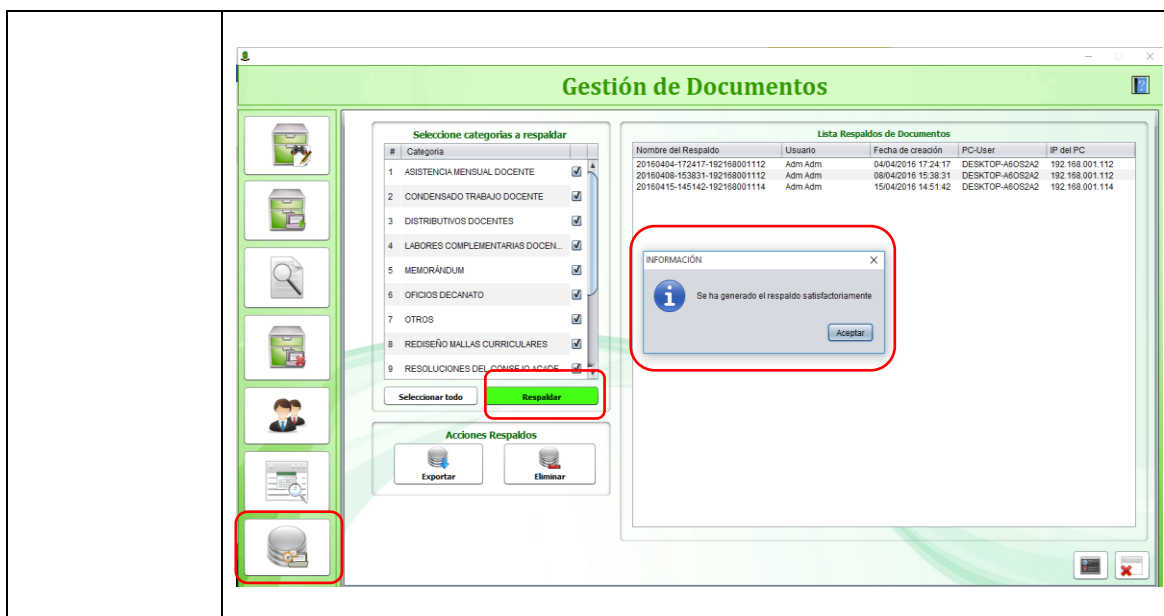
Flujos alternos	<p>Línea 6.-Si la categoría seleccionada no posee registro de búsqueda de documentos, la tabla no se reflejará.</p> <p>Línea 9.- El usuario puede copiar el documento seleccionado en lugar específico, con la opción “Copiar a “.</p>
Post - condición	Documentos encontrados a partir del historial de búsquedas.
Formulario	

AUTORES: AGUAL, MENDOZA,2016.

Tabla 26. Respaldar Información.

Caso de uso	Respaldar Información.	
Código	Cu020	
Actores	Usuario (Administrador/Invitado).	
Propósito	Respaldar todos los documentos que posee el sistema.	
Resumen	El usuario respalda todos los documentos que posee el sistema.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	El usuario debe haber iniciado sesión y los documentos haber sido clasificados.	
Flujo normal de eventos	Acción del actor	Respuesta del sistema

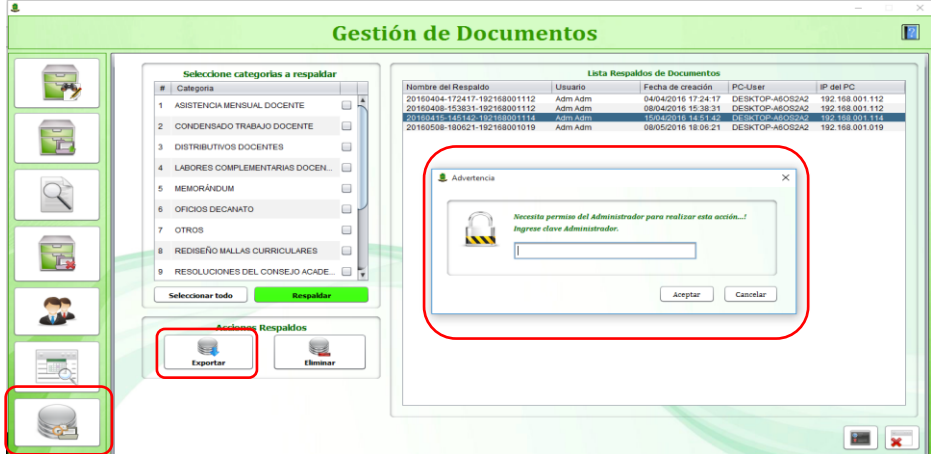
	<p>1.-El caso de uso inicia cuando el usuario selecciona el modulo “Respaldo Información”.</p> <p>3.-El usuario selecciona las categorías que desea respaldar o seleccionar todas las categorías, dando clic en el botón “Seleccionar todos”.</p> <p>4.-El usuario procede a respaldar la información, dando el clic en “Respaldo”.</p>	<p>2.-El sistema muestra la interfaz de “Respaldo Información”.</p> <p>5.-El sistema muestra un mensaje de confirmación que se ha respaldado la información.</p> <p>6.-El caso de uso termina cuando se ha generado el respaldo de información correctamente.</p>																				
Flujos alternos	Ninguna.																					
Post - condición	Respaldo de los documentos clasificados generados correctamente.																					
Formulario	 <p>The screenshot displays the 'Gestión de Documentos' application. On the left, a sidebar contains various icons. The main area is divided into two panels. The left panel, titled 'Seleccione categorías a respaldar', lists nine categories with checkboxes, all of which are checked. A red circle highlights the 'Seleccionar todas' button at the bottom of this list. The right panel, titled 'Lista Respaldos de Documentos', contains a table with the following data:</p> <table><tr><th>Nombre del Respaldo</th><th>Usuario</th><th>Fecha de creación</th><th>PC-User</th><th>IP del PC</th></tr><tr><td>20160404-172417-192168001112</td><td>Adm Adm</td><td>04/04/2016 17:24:17</td><td>DESKTOP-A60S242</td><td>192.168.001.112</td></tr><tr><td>20160408-153831-192168001112</td><td>Adm Adm</td><td>08/04/2016 15:38:31</td><td>DESKTOP-A60S242</td><td>192.168.001.112</td></tr><tr><td>20160415-145142-192168001114</td><td>Adm Adm</td><td>15/04/2016 14:51:42</td><td>DESKTOP-A60S242</td><td>192.168.001.114</td></tr></table> <p>Below the table, there are buttons for 'Exportar' and 'Eliminar'. A red circle also highlights the 'Respaldo' button in the bottom left corner of the application window.</p>		Nombre del Respaldo	Usuario	Fecha de creación	PC-User	IP del PC	20160404-172417-192168001112	Adm Adm	04/04/2016 17:24:17	DESKTOP-A60S242	192.168.001.112	20160408-153831-192168001112	Adm Adm	08/04/2016 15:38:31	DESKTOP-A60S242	192.168.001.112	20160415-145142-192168001114	Adm Adm	15/04/2016 14:51:42	DESKTOP-A60S242	192.168.001.114
Nombre del Respaldo	Usuario	Fecha de creación	PC-User	IP del PC																		
20160404-172417-192168001112	Adm Adm	04/04/2016 17:24:17	DESKTOP-A60S242	192.168.001.112																		
20160408-153831-192168001112	Adm Adm	08/04/2016 15:38:31	DESKTOP-A60S242	192.168.001.112																		
20160415-145142-192168001114	Adm Adm	15/04/2016 14:51:42	DESKTOP-A60S242	192.168.001.114																		

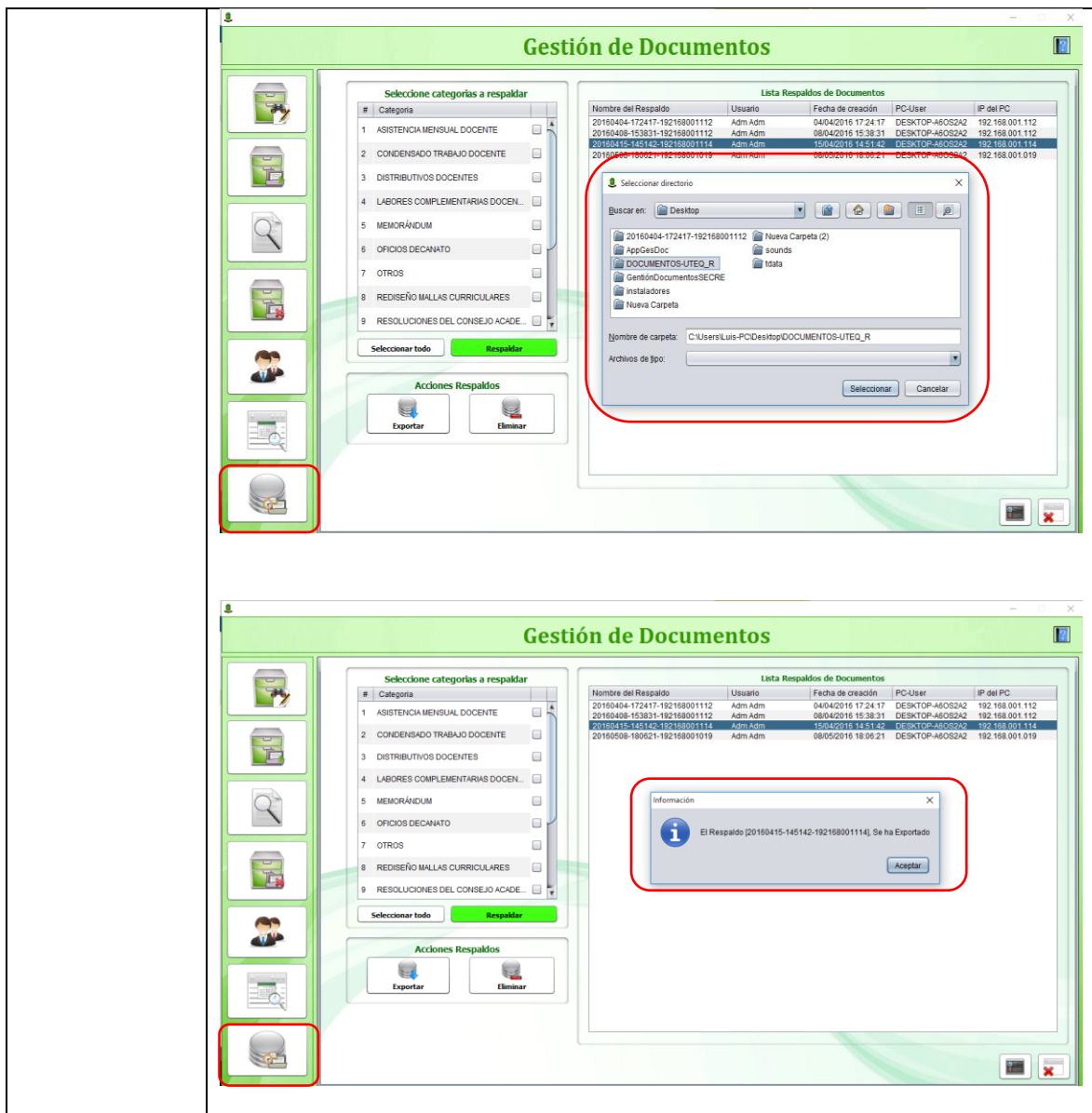


AUTORES: AGUAL; MENDOZA, 2016.

Tabla 27. Exportar respaldo.

Caso de uso	Exportar respaldo.	
Código	Cu021	
Actores	Usuario (Administrador/Invitado)	
Propósito	Exportar el respaldo generado de todos los documentos que posee el sistema, a un lugar específico.	
Resumen	El usuario exporta el respaldo de información generada a una ubicación específica.	
Prioridad	Esencial.	
Tipo	Primario	
Pre-condición	El usuario debe haber iniciado sesión y generado el respaldo de los documentos.	
Flujo normal de eventos	Acción del actor 1.-El caso de uso inicia cuando el usuario selecciona el modulo “Respaldo Información”.	Respuesta del sistema 2.-El sistema muestra la interfaz de “Respaldo Información” y muestra el listado de respaldos realizados en una tabla.

	<p>3.-El usuario selecciona un respaldo generado.</p> <p>4.-El usuario procede a exportar el respaldo, dando el clic en el botón “Exportar”.</p> <p>6.-El usuario ingresa la clave.</p> <p>8.-El usuario selecciona una ubicación y da clic en “Guardar”.</p>	<p>5.-El sistema muestra un cuadro de diálogo solicitando la clave del administrador.</p> <p>7.-El sistema muestra la interfaz de navegación de archivos.</p> <p>9.-El sistema muestra un mensaje de confirmación.</p> <p>10.-El caso de uso termina cuando se ha exportado el respaldo correctamente.</p>
Flujos alternos	Línea 6.- El usuario ingresa la clave errónea, el sistema mostrará un mensaje de error ingresó de datos.	
Post - condición	Respaldo de información generada exportada a un lugar específico.	
Formulario		

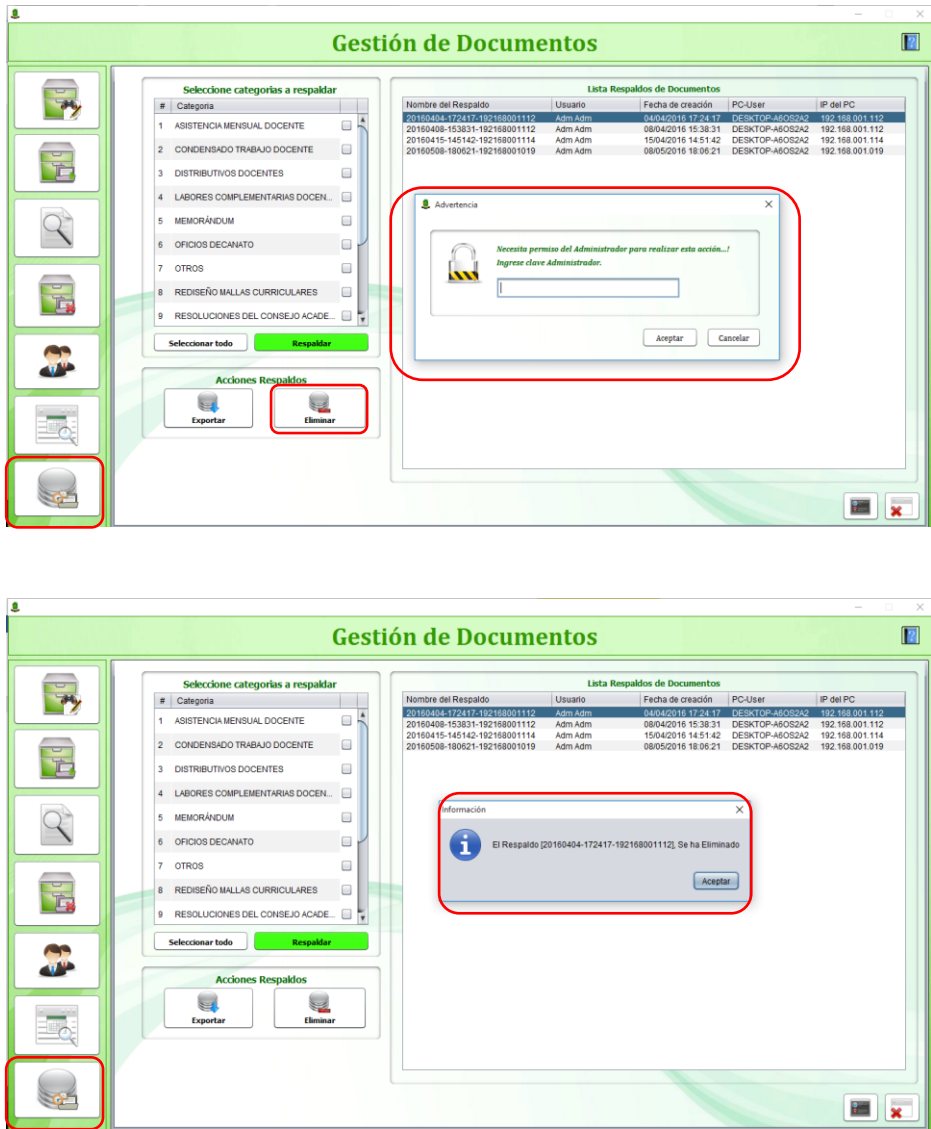


AUTORES: AGUAL; MENDOZA, 2016.

Tabla 28. Eliminar respaldo.

Caso de uso	Eliminar respaldo.
Código	Cu022
Actores	Usuario (Administrador/Invitado).
Propósito	Eliminar los respaldos que se han generado.
Resumen	El usuario procede a eliminar los respaldos que los hizo de manera incorrecta.
Prioridad	Esencial.
Tipo	Primario

Pre-condición	Debe estar la aplicación iniciada y deben existir respaldos generados.	
Flujo normal de eventos	<p>Acción del actor</p> <p>1.-El caso de uso inicia cuando el usuario selecciona el modulo “Respaldo Información”</p> <p>4.-El usuario selecciona un respaldo generado.</p> <p>5.-El usuario procede a eliminar el respaldo, dando el clic en el comando “Eliminar”.</p> <p>7.-El usuario ingresa la clave.</p>	<p>Respuesta del sistema</p> <p>2.-El sistema muestra la interfaz de “Respaldo Información”, mostrando el listado de respaldos en una tabla.</p> <p>6.-El sistema muestra un cuadro de diálogo solicitando la clave del administrador.</p> <p>8.-El sistema muestra la interfaz de navegación de archivos.</p> <p>9.-El sistema muestra un mensaje de confirmación que se ha eliminado correctamente el respaldo seleccionado.</p> <p>10.-El caso de uso termina cuando el respaldo ha sido eliminado correctamente.</p>

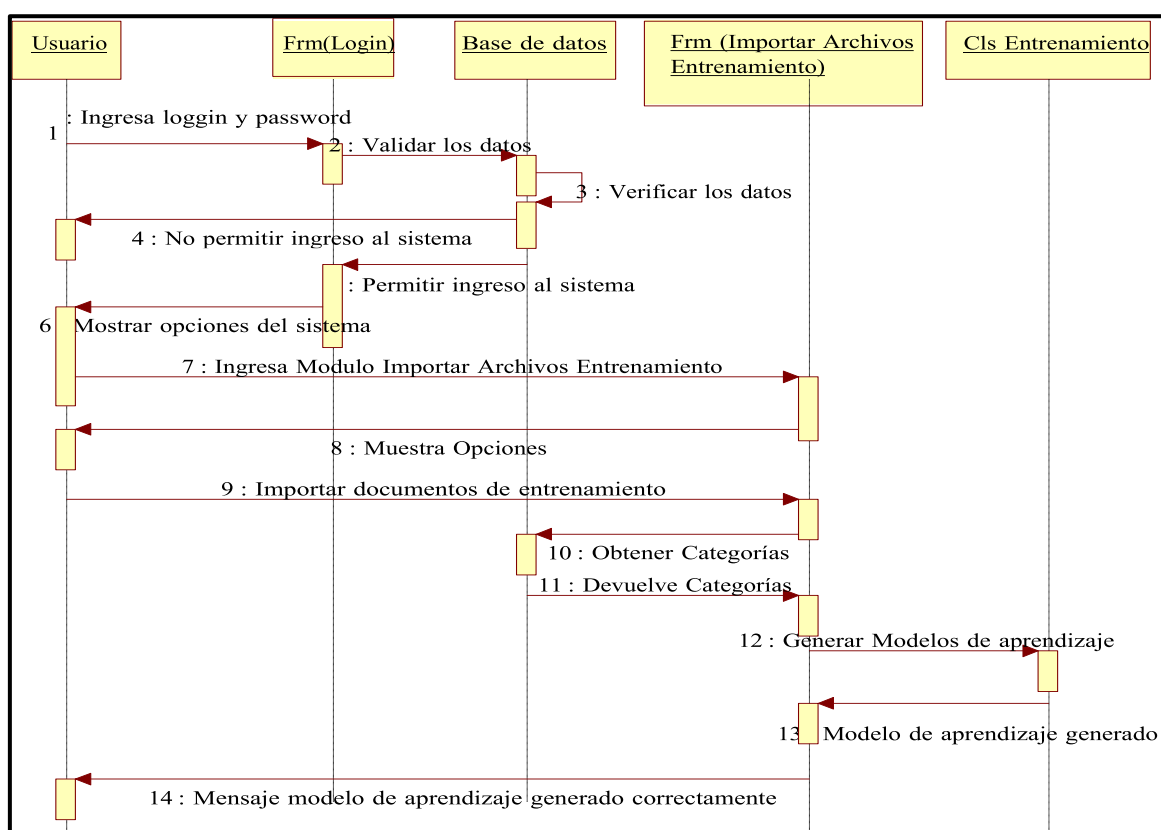
Flujos alternos	Línea 9.- El usuario ingresa la clave errónea, el sistema mostrará un mensaje de error.
Post - condición	Respaldo de información eliminada.
Formulario	 <p>The application 'Gestión de Documentos' features a sidebar with document categories and a main area with a table of backups. The categories include: 1. ASISTENCIA MENSUAL DOCENTE, 2. CONDENSADO TRABAJO DOCENTE, 3. DISTRIBUTIVOS DOCENTES, 4. LABORES COMPLEMENTARIAS DOCEN..., 5. MEMORÁNDUM, 6. OFICIOS DECANATO, 7. OTROS, 8. REDISEÑO MALLAS CURRICULARES, and 9. RESOLUCIONES DEL CONSEJO ACAD... The backup table lists details such as 'Nombre del Respaldo', 'Usuario', 'Fecha de creación', 'PC-User', and 'IP del PC'. The 'Advertencia' dialog box appears when an administrator action is required, and the 'Información' dialog box confirms the deletion of a specific backup.</p>

AUTORES: AGUAL; MENDOZA, 2016.

3.8.2.2.Diagramas de secuencias.

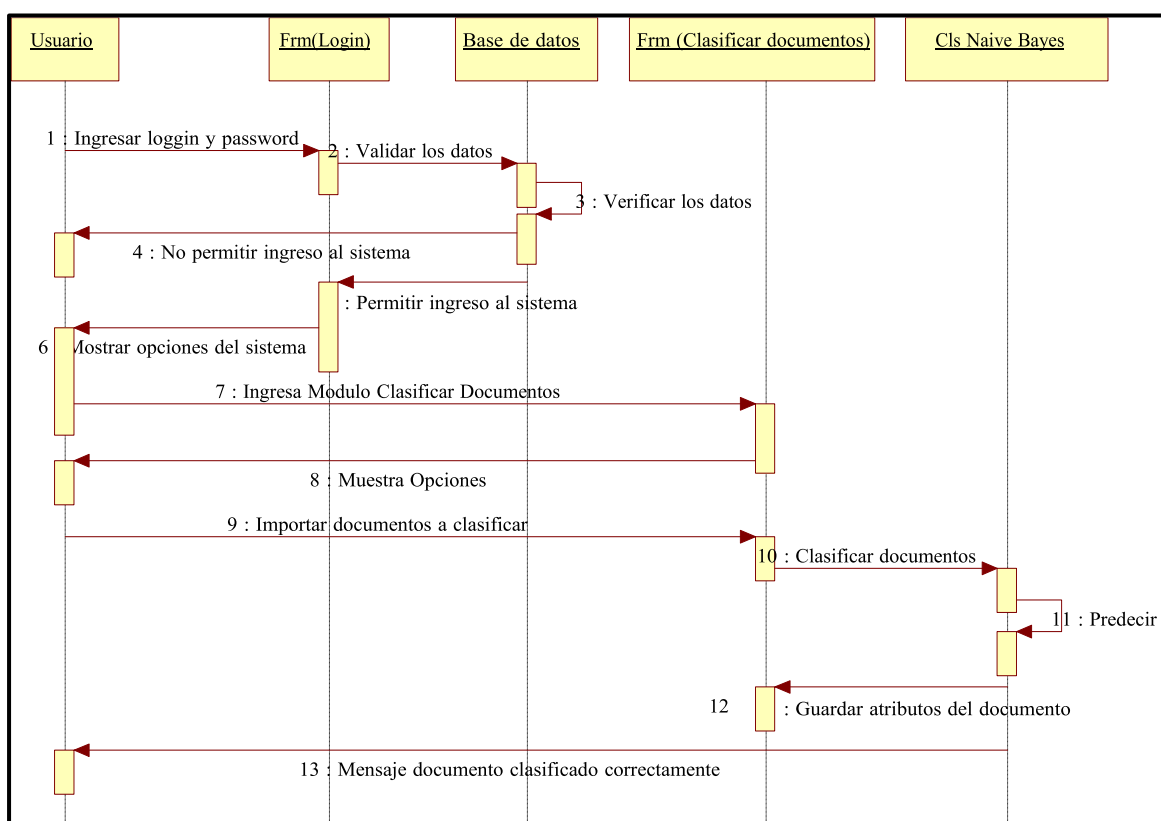
Los diagramas de secuencia nos permiten conocer la interacción de los diferentes objetos que interactúan en el sistema, a continuación, se podrá apreciar los diagramas de los procesos más importantes de la aplicación:

Figura.- 12. Diagrama de secuencia "Obtener el modelo de aprendizaje "



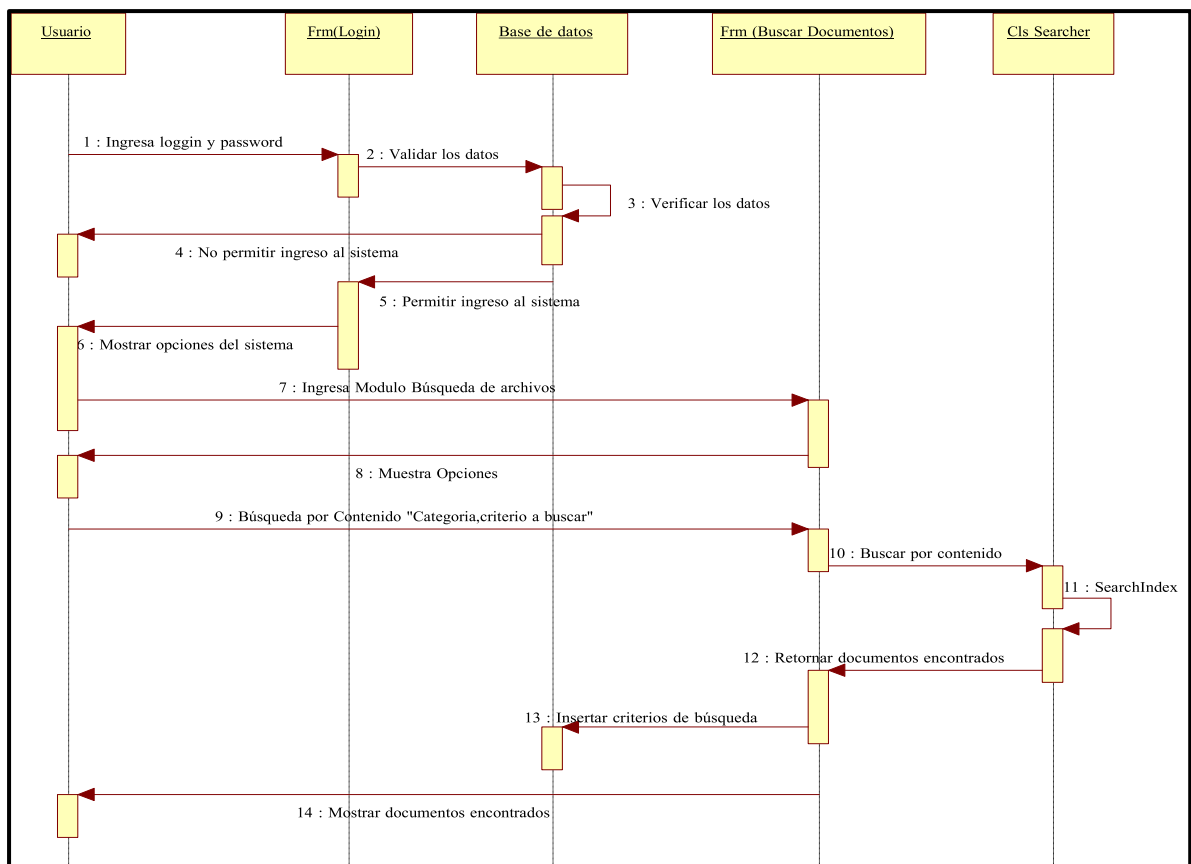
AUTORES: AGUAL; MENDOZA,2016.

Figura.- 13. Diagrama de secuencia " Clasificar documentos "



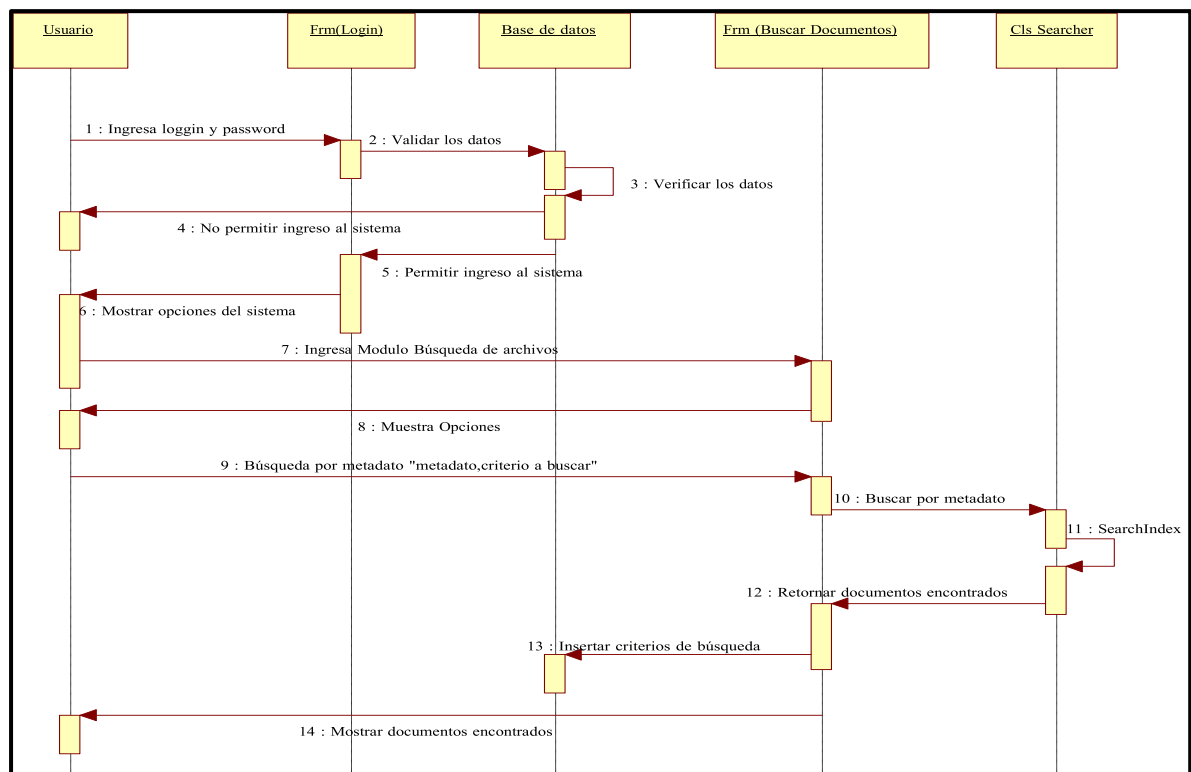
AUTORES: AGUAL; MENDOZA, 2016.

Figura.- 14. Diagrama de secuencia " Búsqueda de documentos por contenido"



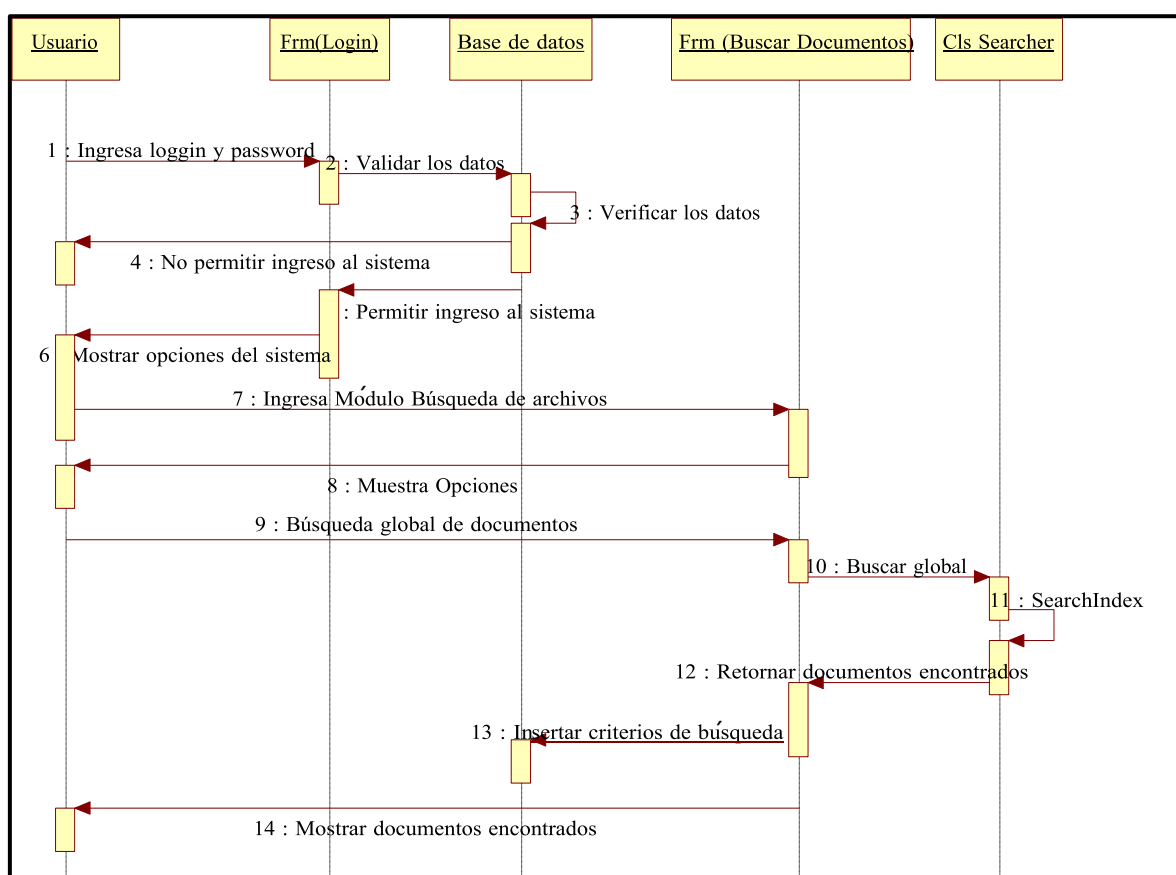
AUTORES: AGUAL; MENDOZA,2016.

Figura.- 15. Diagrama de secuencia " Búsqueda de documentos por metadatos"



AUTORES: AGUAL; MENDOZA,2016.

Figura.- 16. Diagrama de secuencia "Búsqueda Global de documentos"



AUTORES: AGUAL; MENDOZA,2016.

3.8.3. Codificación del proyecto.

En esta parte se define los procesos principales elaborados en la aplicación gestión automática de documentos digitales basados en el contenido textual, para facilitar el acceso a la información de la Facultad de Ciencias de la Ingeniería, cada proceso desarrollado se describe a continuación:

3.8.3.1. Definición de las categorías

Mediante las entrevistas realizadas al personal a cargo del manejo de información documental en la Secretaría de la FCI, se obtuvo una lista de categorías con el cual son diferenciados los documentos digitales que se generan en diferentes oficinas de la Facultad de Ciencias de la Ingeniería tales como el decanato, coordinación de carreras, entre otros., los mismos que llegan a manos de la persona encargada de la Secretaría para su posterior manipulación.

Ente las categorías mencionadas tenemos:

- Asistencia mensual docente.

- Condensado trabajo docente.
- Distributivos docentes.
- Labores complementarias docentes.
- Memorandos.
- Oficios decanato.
- Rediseño de proyectos curriculares
- Resoluciones del consejo académico de Facultad.
- Solicitud de estancias.
- Solicitud de permiso.
- Solicitud viáticos.
- Títulos graduados.
- Otros.

Además, se tiene la opción de crear una nueva categoría, en caso que la FCI genere un nuevo trámite académico.

3.8.3.2. Entrenamiento por categoría.

Una vez definidas todas las categorías, se procedió a obtener los documentos por cada una de la categoría mencionada para realizar el proceso de entrenamiento, es decir obtener el aprendizaje de los textos previamente categorizados, cuyo objetivo es que, a partir de un conjunto de documentos de entrenamiento, el algoritmo cree un modelo de aprendizaje para su posterior clasificación.

3.8.3.3. Pre – procesamiento en lenguaje natural de los archivos de entrenamiento

En el proceso de entrenamiento de los documentos por categoría, se procede a obtener toda la información textual del documento a entrenar utilizando herramientas de software libre (*poi.apache.org* y *tm-extractors*), tal como se logra apreciar en la *tabla 29*, almacenándolos en archivos de texto plano por cada categoría, para posteriormente realizar el pre-procesamiento de información utilizando la biblioteca Apache Lucene, que nos proporcionan los diferentes procedimientos para el tratamiento de la información.

Tabla 29. Información textual extraída de una resolución académica.

UNIVERSIDAD	TECNICA	ESTATAL	DE	QUEVEDO
FACULTAD	DE	CIENCIAS	DE	LA
DECANATO				INGENIERÍA

RESOLUCIONES DE LA COMISIÓN ACADÉMICA, DE LA FACULTAD DE CIENCIAS DE LA INGENIERIA DE LA UNIVERSIDAD TECNICA ESTATAL DE QUEVEDO CELEBRADA EL DÍA JUEVES 3 DE SEPTIEMBRE DEL AÑO DOS MIL QUINCE. PRIMERA

RESOLUCIÓN.-

Fijar el día Jueves 3 de septiembre del 2015, a las 15H00, como fecha para que el egresado de la carrera de Ingeniería Mecánica señor BARZOLA SUÁREZ HAMILTON ESTIVEN , lleve a cabo, en el Auditorio de la Facultad, el acto de Sustentación y Defensa de su Tesis de Grado titulada “EVALUACIÓN DEL RECURSO ENERGÉTICO PARA SELECCIÓN Y PROPUESTA DE UN AEROGENERADOR DE EJE VERTICAL, QUEVEDO-AÑO 2015”, Resolución que se adopta en virtud al Informe de fecha 19 de agosto del 2015, suscrita por el Ing. Gabriel Pico Saltos, Director de Tesis y acorde a lo que establece el Art. 27 del Reglamento General de Graduación de Pregrado de la Universidad Técnica Estatal de Quevedo. Dispónese que el egresado autor de la investigación cumpla con los requisitos exigidos para estos casos de acuerdo con la normativa interna de la Facultad.

AUTORES: AGUAL; MENDOZA, 2016.

3.8.3.4. Eliminar o suprimir los signos de puntuación.

Los signos de puntuación son los que dan sentido a las oraciones representados en un documento, pero en el tratamiento de la información para la clasificación de documentos, se procede a eliminar tales signos de puntuación como las:

- La coma [,].
- El punto y coma [;].
- El punto [.].
- Los dos puntos [:].
- Puntos suspensivos [...].

- Las comillas ["].

Por medio del siguiente código:

Figura.- 17. Código de supresión de los signos de puntuación.

```
public final TokenStream tokenStream(String fieldName, Reader reader) {
    TokenStream result = new StandardTokenizer(LUCENE_VERSION, reader);
    result = new StandardFilter(LUCENE_VERSION, result);
    result = new LowerCaseFilter(LUCENE_VERSION, result);
    result = new ISOLatin1AccentFilter(result);
    result = new StopFilter(LUCENE_VERSION, result, SPANISH_STOP_WORDS_SET);
    result = new PorterStemFilter(result);
    result = new SpanishStemFilter(result);
    System.out.println("Token stream" + result);
    return result;
}
```

AUTOR: AGUAL; MENDOZA,2016.

Tabla 30. Supresión de los signos de puntuación.

8UNIVERSIDAD	TECNICA	ESTATAL	DE	QUEVEDO
FACULTAD	DE	CIENCIAS	DE	INGENIERÍA
DECANATO RESOLUCIONES DE LA COMISIÓN ACADÉMICA DE LA FACULTAD				
DE CIENCIAS DE LA INGENIEIRA DE LA UNIVERSIDAD TECNICA ESTATAL DE				
QUEVEDO CELEBRADA EL DÍA JUEVES 3 DE SEPTIEMBRE DEL AÑO DOS MIL				
QUINCE.				
PRIMERA				
RESOLUCIÓN				
Fijar el día Jueves 3 de septiembre del 2015 a las 15H00 como fecha para que el egresado				
de la carrera de Ingeniería Mecánica señor BARZOLA SUÁREZ HAMILTON ESTIVEN				
lleve a cabo en el Auditorio de la Facultad el acto de Sustentación y Defensa de su Tesis				
de Grado titulada EVALUACIÓN DEL RECURSO ENERGÉTICO PARA SELECCIÓN Y				
PROPUESTA DE UN AEROGENERADOR DE EJE VERTICA QUEVEDO AÑO 2015				
Resolución que se adopta en virtud al Informe de fecha 19 de agosto del 2015 suscrita por				
el Ing Gabriel Pico Saltos Director de Tesis y acorde a lo que establece el Art 27 del				
Reglamento General de Graduación de Pregrado de la Universidad Técnica Estatal de				
Quevedo Dispónese que el egresado autor de la investigación cumpla con los requisitos				
exigidos para estos casos de acuerdo con la normativa interna de la Facultad				

AUTORES: AGUAL, MENDOZA,2016

3.8.3.5. Conversión de las palabras a minúsculas

Para permitir una adecuada indexación, es necesario convertir el texto a minúsculas, de esta manera tenemos nuestro texto normalizado, sin prioridad de escritura en las palabras del conjunto de entrenamiento. El código implementado en el sistema para la conversión de palabras de mayúscula a minúscula es el siguiente:

Figura.- 18. Código de conversión de las palabras a minúsculas.

```
public final TokenStream tokenStream(String fieldName, Reader reader) {  
  
    TokenStream result = new StandardTokenizer(LUCENE_VERSION, reader);  
    result = new StandardFilter(LUCENE_VERSION, result);  
    result = new LowerCaseFilter(LUCENE_VERSION, result);  
    result = new ISOLatin1AccentFilter(result);  
    result = new StopFilter(LUCENE_VERSION, result, SPANISH_STOP_WORDS_SET);  
    result = new PorterStemFilter(result);  
    result = new SpanishStemFilter(result);  
    System.out.println("Token stream" + result);  
    return result;  
}
```

AUTORES: AGUAL; MENDOZA, 2016.

Tabla 31. Conversión de las palabras a minúsculas.

8universidad tecnica estatal de quevedo facultad de ciencias de la ingeniería decanato resoluciones de la comisión académica de la facultad de ciencias de la ingenieira de la universidad tecnica estatal de quevedo celebrada el día jueves 3 de septiembre del año dos mil quince primera resolución fijar el día jueves 3 de septiembre del 2015 a las 15h00 como fecha para que el egresado de la carrera de ingeniería mecánica señor barzola suárez hamilton estiven lleve a cabo en el auditorio de la facultad el acto de sustentación y defensa de su tesis de grado titulada evaluación del recurso energético para selección y propuesta de un aerogenerador de eje vertical quevedo año 2015 resolución que se adopta en virtud al informe de fecha 19 de agosto del 2015 suscrita por el ing gabriel pico saltos director de tesis y acorde a lo que establece el art 27 del reglamento general de graduación de pregrado de la universidad técnica estatal de quevedo dispónese que el egresado autor de la investigación cumpla con los requisitos exigidos para estos casos de acuerdo con la normativa interna de la facultad

AUTORES: AGUAL; MENDOZA,2016.

3.8.3.6. Eliminación de los signos de Acentuación de las palabras.

Realizar la normalización de textos, es tratar homogeneizar todo el texto del documento en que se va a trabajar como conjunto entrenamiento, en este proceso se procede a la eliminación de los signos de acentuación como las tildes (´) en las palabras, a través del siguiente código.

Figura.- 19. Código de eliminación de Acentuación.

```
public final TokenStream tokenStream(String fieldName, Reader reader) {  
    TokenStream result = new StandardTokenizer(LUCENE_VERSION, reader);  
    result = new StandardFilter(LUCENE_VERSION, result);  
    result = new LowerCaseFilter(LUCENE_VERSION, result);  
    result = new ISOLatin1AccentFilter(result);  
    result = new StopFilter(LUCENE_VERSION, result, SPANISH_STOP_WORDS_SET);  
    result = new PorterStemFilter(result);  
    result = new SpanishStemFilter(result);  
    System.out.println("Token stream" + result);  
    return result;  
}
```

AUTORES: AGUAL; MENDOZA,2016.

Tabla 32. Eliminación de los signos de Acentuación.

8universidad tecnica estatal de quevedo facultad de ciencias de la ingeniería decanato resoluciones de la comisión académica de la facultad de ciencias de la ingeniería de la universidad tecnica estatal de quevedo celebrada el día jueves 3 de septiembre del año dos mil quince primera resolución fijar el día jueves 3 de septiembre del 2015 a las 15h00 como fecha para que el egresado de la carrera de ingeniería mecánica señor barzola suarez hamilton estiven lleve a cabo en el auditorio de la facultad el acto de sustentación y defensa de su tesis de grado titulada evaluación del recurso energético para selección y propuesta de un aerogenerador de eje vertical quevedo año 2015 resolución que se adopta en virtud al informe de fecha 19 de agosto del 2015 suscrita por el ing gabriel pico saltos director de tesis y acorde a lo que establece el art 27 del reglamento general de graduación de pregrado de la universidad tecnica estatal de quevedo disponese que el egresado autor de la investigación cumpla con los requisitos exigidos para estos casos de acuerdo con la normativa interna de la facultad

AUTORES: AGUAL; MENDOZA,2016.

3.8.3.7. Filtro de Palabras Vacías (StopWords).

El filtrado de palabras vacías, se realiza mediante la supresión de las palabras que no aportarían valor significativo para la clasificación, la eliminación de los *stopwords* hace la

siguiente operación:

- Analiza el texto procesado, con un conjunto de palabras vacías ya definidas en un Datasets, así de esta manera se va comparando el texto plano, con el *dataset* encontrando las palabras vacías y suprimiéndolas, con lo que al final el documento quedaría solo con las palabras que tienen un valor significativo para el proceso de clasificación.

Figura.- 20. Filtrado de Palabras Vacías.

AUTORES: AGUAL; MENDOZA,2016.

```
public final TokenStream tokenStream(String fieldName, Reader reader) {  
  
    TokenStream result = new StandardTokenizer(LUCENE_VERSION, reader);  
    result = new StandardFilter(LUCENE_VERSION, result);  
    result = new LowerCaseFilter(LUCENE_VERSION, result);  
    result = new ISOLatinAccentFilter(result);  
    result = new StopFilter(LUCENE_VERSION, result, SPANISH_STOP_WORDS_SET);  
    result = new PorterStemFilter(result);  
    result = new SpanishStemFilter(result);  
    System.out.println("Token stream" + result);  
    return result;  
}
```

Tabla 33. DataSet de palabras en español.

mucho, tuviese, míos, seremos, del, serán, vuestros, hube, tus, desde, habríamos, les, está, estuvierais, hubieras, durante, otras, mí, habido, esté, estarán, fuera, algunas, tendrías, estarás, habréis, al, tengan, quien, son, habrían, tengas, habrías, estar, estas, sobre, estaréis, unos, nosotros, tuvieron, mía, estuviste, estarían, seáis, habíamos, todo, tiene, estarías, tuyas, tenemos, hubiéramos, mío, sus, se, algo, tendrán, mis, por, serías, sea, estuviéramos, ella, vuestras, entre, tendrás, a, su, teniendo, esas, fuiste, tuvieseis, o, te, ti, habrán, tendremos, tuve, y, habrás, hubimos, fueran, tu, fuese, tendré, estáis, sois, hasta, suyos, seré, fueras, nosotras, tuvieran, tendrían, tuvieras, estaban, han, habidas, tuviésemos, estabas, has, fueron, tenían, estuve, hay, tendrá, tenías, él, el, estuvo, en, habéis, poco, es, estés, habían, estén, tengáis, habías, serás, hubisteis, otros, hubiera, que, vosotros, seréis, era, tuyas, e, tuviesen, también, hayan, nuestra, tuvieses, hayas, otra, antes, vuestra, sin, sí, fueseis, nuestro, esta, más, eras, contra, otro, hubiese, vuestro, tenidos, una, estada, cuando, estuvisteis, todos, estuvieron, muchos, pero, hemos, uno, habidos, habrá, ha, tuvo, tuvimos, he, estuviesen, habré, donde, tú, tienen, ya, fueses, estemos, tienes, tenéis, como, esos, estad, fuisteis, habremos, yo, nuestros, seamos, estaba,

siendo, seríamos, fuimos, estos, será, fuesen, estadas, tendríais, hayáis, tuviéramos, estuvimos, somos, estábamos, teníais, hubieses, estuvieran, hubo, un, estuvieras, estaremos, qué, habíais, tuya, hubiésemos, fuerais, hubieseis, nos, los, estéis, seríais, tuvierais, tuyo, estuviésemos, quienes, había, ante, no, hubieron, sido, nuestras, estado, tenida, tenido, nada, estuviera, esa, sería, la, fue, ese, le, serían, tened, habida, estuvieses, estuvieseis, teníamos, eso, con, lo, vosotras, tuviera, fui, tuviste, estará, estaría, erais, porque, hubiste, estaré, estuviese, estamos, ellas, me, eran, de, mi, soy, las, este, cual, estaríamos, tenga, esto, hubieran, hayamos, están, mías, estás, ni, tendríamos, tendría, haya, tengo, estabais, fuésemos, habríais, estados, tenidas, muy, éramos, seas, eres, algunos, tuvisteis, tanto, habría, ellos, para, suya, os, fuéramos, tendréis, hubierais, hubiesen, estoy, tenía, habiendo, estaríais, tengamos, tuyos, suyo, estando, sean, i , universidad, técnica, tecnica, estatal, quevedo, ingeniería, ing, ingeniero, lcda, lcdo, facultad, ciencias, ingeniería, fci, uteq, phd, ab, lic, arq, carrera, egdo, egda, señor, sr, atentamente, m.sc, eco, docente, escuela, decano

AUTORES: AGUAL, MENDOZA,2016.

Tabla 34. Proceso de filtrado de palabras vacías.

8universidad decanato resoluciones comision academica celebrada día jueves 3 septiembre año dos mil quince primera resolucion fijar día jueves 3 septiembre 2015 15h00 fecha egresado carrera mecanica barzola suarez hamilton estiven lleve cabo auditorio facultad acto sustentacion defensa tesis grado titulada evaluacion recurso energetico seleccion propuesta aerogenerador eje vertical año 2015 resolucion adopta virtud informe fecha 19 agosto 2015 suscrita gabriel pico saltos director tesis acorde establece art 27 reglamento general graduacion pregrado disponese egresado autor investigacion cumpla requisitos exigidos acuerdo normativa interna

AUTORES: AGUAL; MENDOZA,2016.

3.8.3.8. Stemming – Lematización de términos.

En el pre - procesamiento de los documentos, el *stemming* o lematización de términos, consiste en ir identificando, reemplazando y eliminando los sufijos de los términos que se encuentren el texto plano, dejándolos a las palabras a su raíz.

Figura.- 21. Lematización de Términos.

```
public final TokenStream tokenStream(String fieldName, Reader reader) {

    TokenStream result = new StandardTokenizer(LUCENE_VERSION, reader);
    result = new StandardFilter(LUCENE_VERSION, result);
    result = new LowerCaseFilter(LUCENE_VERSION, result);
    result = new ISOLatin1AccentFilter(result);
    result = new StopFilter(LUCENE_VERSION, result, SPANISH_STOP_WORDS_SET);
    result = new PorterStemFilter(result);
    result = new SpanishStemFilter(result);
    System.out.println("Token stream" + result);
    return result;
}
```

AUTORES: AGUAL; MENDOZA,2016.

Tabla 35. Proceso de stemming o lematización de términos.

8universidad decanat resolucio comision academic celebrad día juev septiembr año dos mil quinc primer resolucio fijar día juev septiembr 2015 15h00 fech egresad mecanic barzol suarez hamilton estiven llev cabo auditori acto sustentacion defens tesis grad titulad evaluacion recurs energetic seleccion propuest aerogenerador eje vertical año 2015 resolucio adopto virtud inform fech 19 agost 2015 suscrit gabriel pico salt director tesis acord establec art 27 reglament general graduacion pregrad dispones egresad autor investigacion cumpl requisit exigid acuerd normativ intern

AUTORES: AGUAL, MENDOZA,2016.

3.8.3.9. Creación de Datasets por categorías.

Terminado el pre - procesamiento de los documentos a entrenar, como resultado el texto plano queda tal como se muestra en la Tabla 36, todo el proceso de entrenamiento se la debe realizar por cada categoría definida en el sistema, obteniendo así nuestro *dataset*.

Tabla 36. Dataset generado..

8universidad decanat resolucio comision academic celebrad día juev septiembr año dos mil quinc primer resolucio fijar día juev septiembr 2015 15h00 fech egresad mecanic barzol suarez hamilton estiven llev cabo auditori acto sustentacion defens tesis grad titulad evaluacion recurs energetic seleccion propuest aerogenerador eje vertical año 2015 resolucio adopto virtud inform fech 19 agost 2015 suscrit gabriel pico salt director tesis acord establec art 27 reglament general graduacion pregrad dispones egresad autor

AUTORES: AGUAL, MENDOZA, 2016.

3.8.3.10. Clasificación de documentos aplicando el Algoritmo Naive Bayes.

Una vez obtenido los datasets por categorías a través del pre – procesado de los documentos de entrenamiento, se procede aplicar el algoritmo Naive Bayes, para generar el modelo de aprendizaje y así clasificar los documentos nuevos que el usuario ingrese al sistema. La secuencia de pasos que se realiza para obtener en modelo de aprendizaje es el siguiente:

- Obtener los datasets de entrenamiento por categorías establecidas en el sistema.
- Cargar a memoria todos los datasets de entrenamiento.
- Determinar el *setChisquareCriticalValue* (valor crítico del Chi Cuadrado) para evaluar el valor predictivo de los atributos, en la aplicación se estableciendo que el valor crítico del Chi cuadrado sea el 0,001.
- Modelo de aprendizaje generado para clasificación de documentos.

Por cada documento nuevo que el usuario ingrese al sistema, mediante el modelo de aprendizaje, el algoritmo nos predice a que categoría va a pertenecer.

3.8.3.11. Indexación de los documentos.

Para realizar la búsqueda de documentos de la manera rápida, se procede a indexar el texto pre-procesado y transformarlo a un formato que nos permita realizar búsquedas rápidamente. La Indexación es como una estructura de datos que permite acceso aleatorio a las palabras que almacena los documentos en texto plano.

3.8.3.12. Búsqueda por categoría

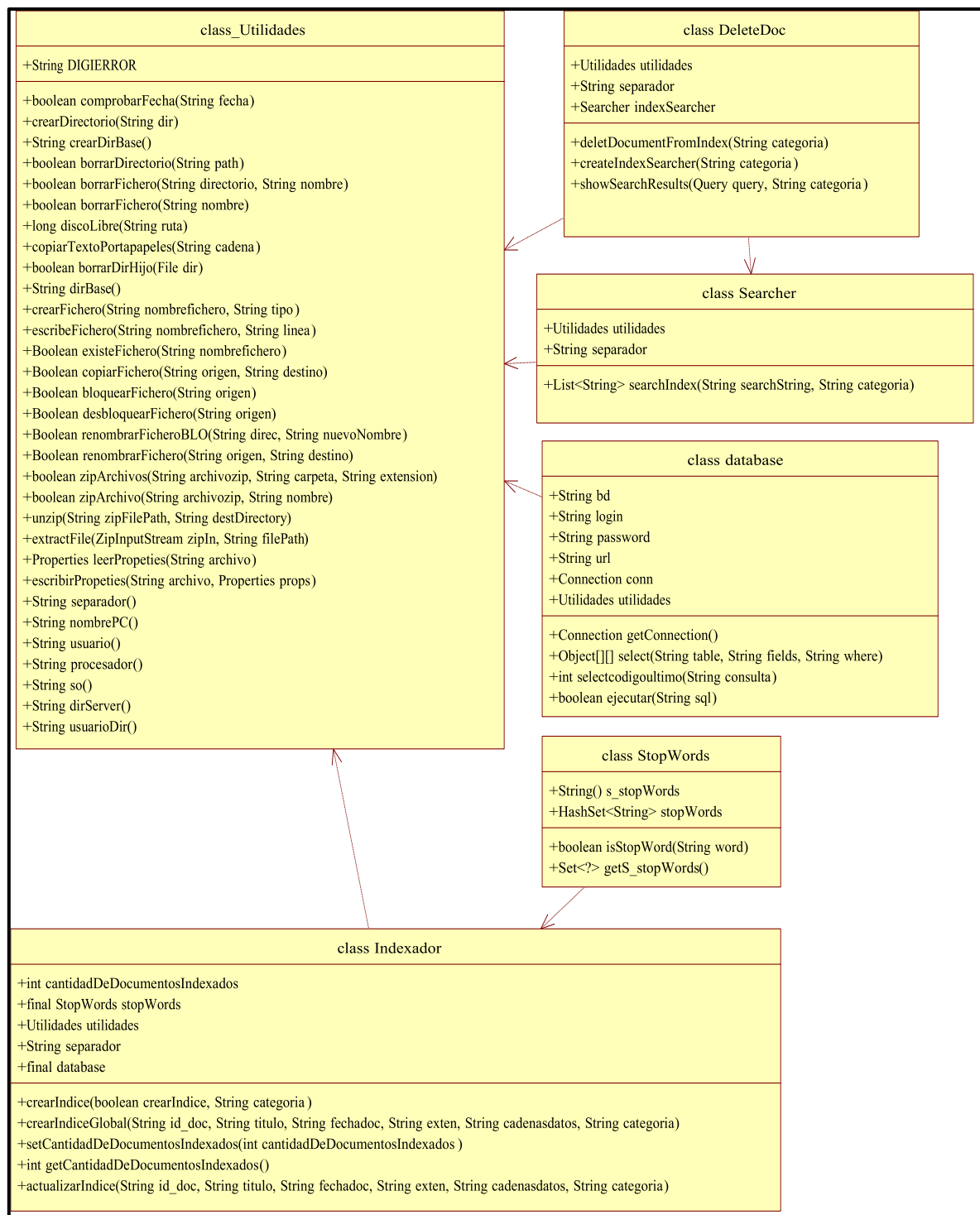
Una vez que se ha generado el índice de documentos se procede a la búsqueda a través de las categorías definidas, mostrándonos como resultado una lista de documentos que contenga el criterio de búsqueda que ingresó el usuario.

3.8.3.13. Búsqueda global

La búsqueda global consiste en encontrar un documento por medio de todas las categorías, es decir la biblioteca Apache Lucene se encarga de localizar el documento por medio del índice global.

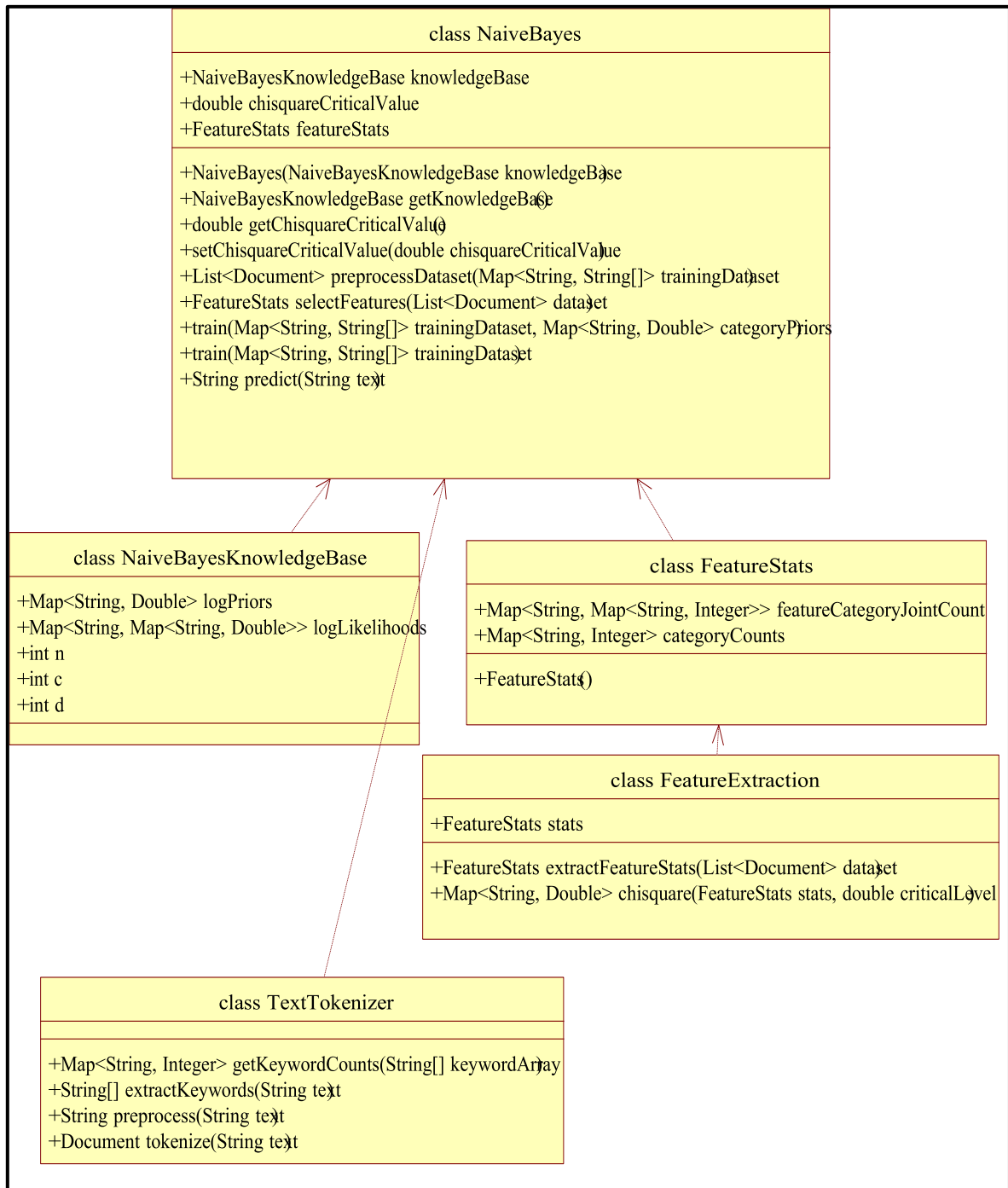
3.8.3.14. Diagramas de Clase

Figura.- 22. Diagramas de clase Utilidades, DeleteDoc, Searcher, database, StopWords, Indexador.



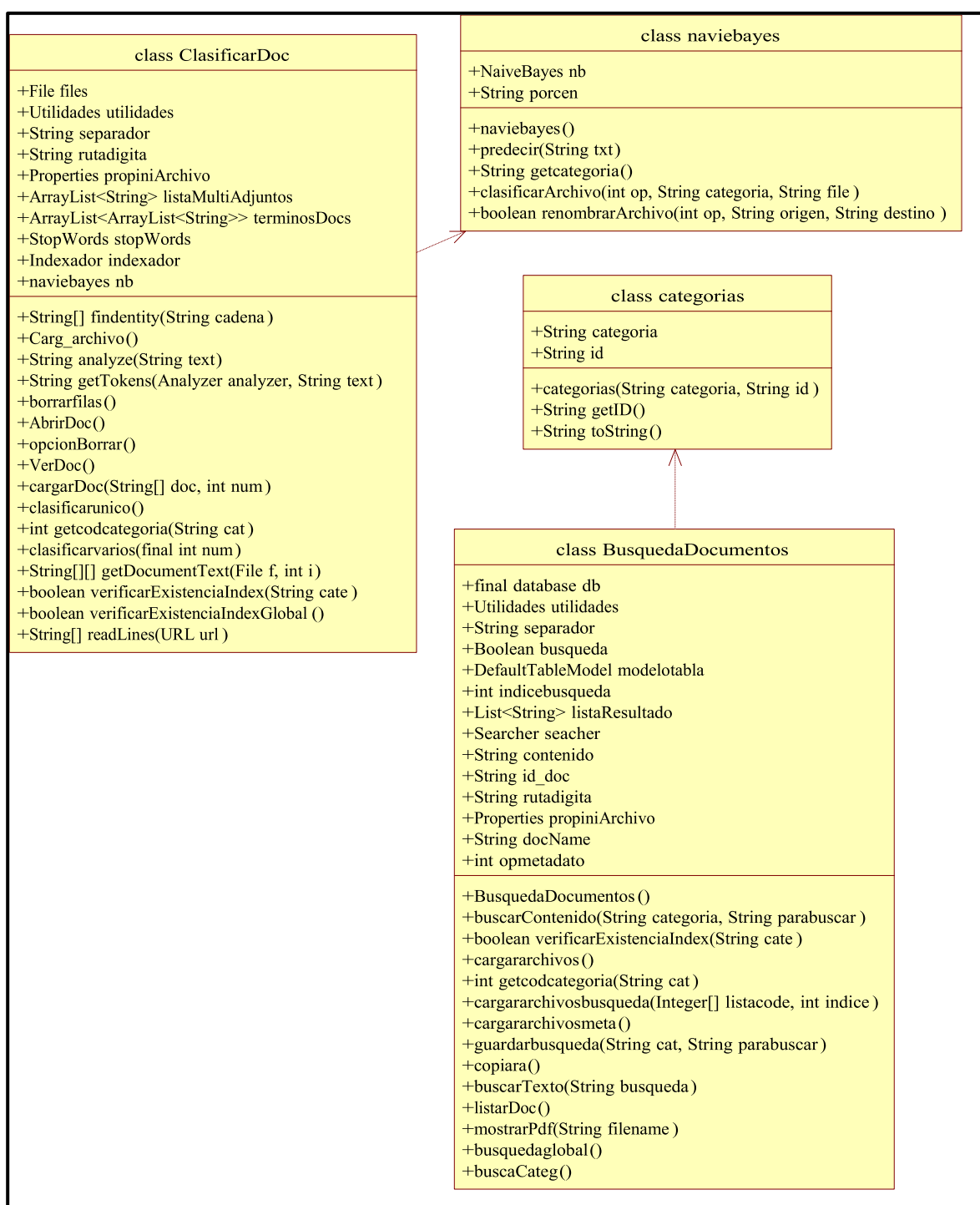
AUTORES: AGUAL; MENDOZA,2016.

Figura.- 23. Diagramas de clase NaiveBayes, NaiveBayesKnowledgeBase, FeatureStats, FeatureExtraction, TextTokenizer



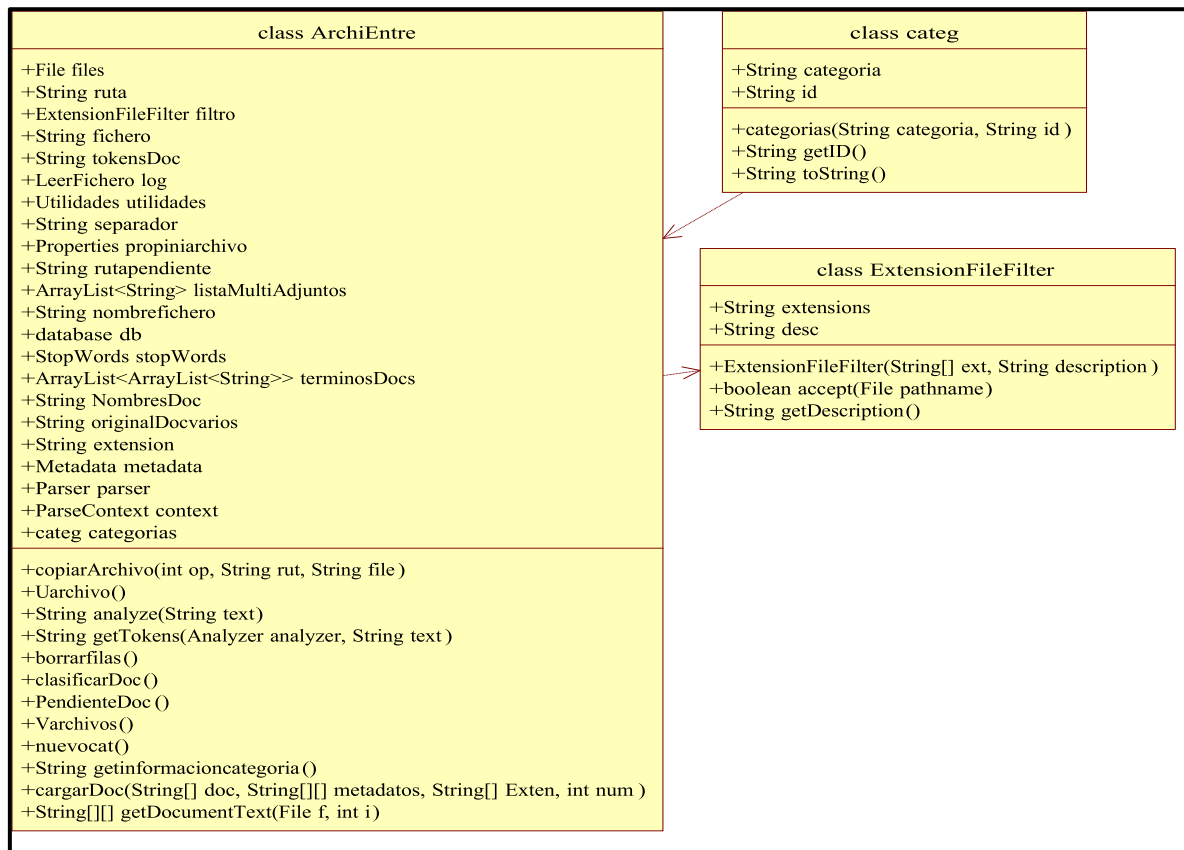
AUTORES: AGUAL; MENDOZA,2016.

Figura.- 24. Diagrama de Clases ClasificarDoc, naivebayes,categorias,BusquedaDocumentos



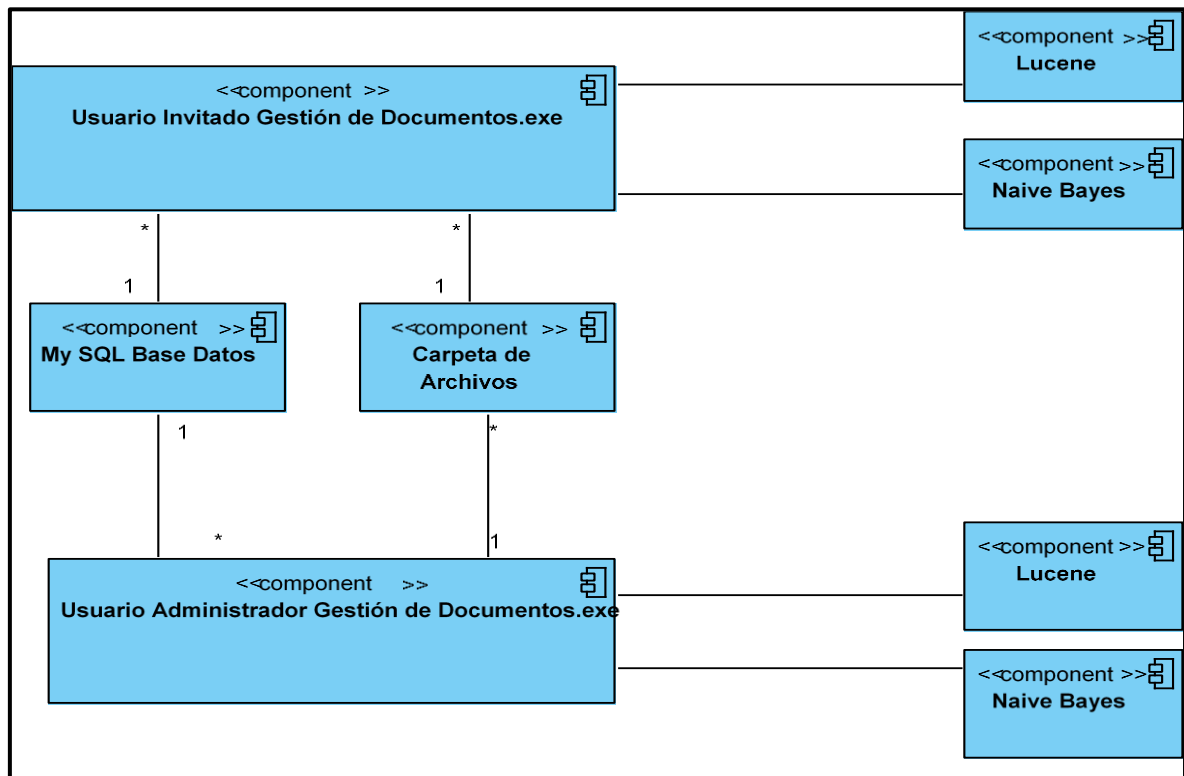
AUTORES: AGUAL; MENDOZA,2016.

Figura.- 25. Diagramas de clase ArchiEntren, categ, ExtensionFileFiltrer



AUTORES: AGUAL; MENDOZA,2016.

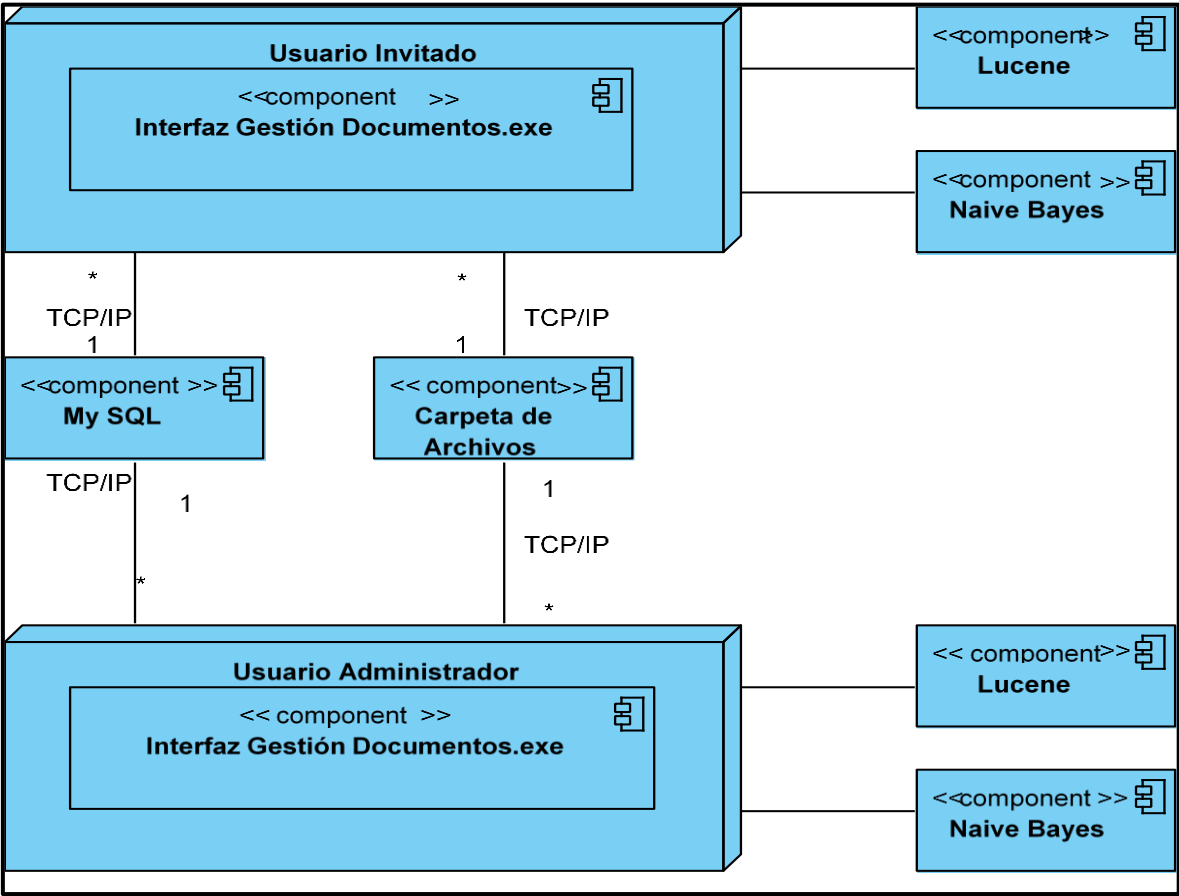
Figura.- 26. Diagrama de paquetes



AUTORES: AGUAL; MENDOZA,2016.

3.8.4. Pruebas del Sistema

Figura.- 27. Diagrama de Despliegue



AUTORES: AGUAL; MENDOZA,2016.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. Resultados

A continuación, el detalle del análisis de la gestión de documentos digitales en la FCI en la actualidad y haciendo uso de la aplicación desarrollada.

4.1.1. Gestión de documentos digitales en la Secretaría de la FCI.

En la gestión de documentos de la FCI, la secretaria realiza los siguientes procesos de forma manual:

Para organización de los archivos:

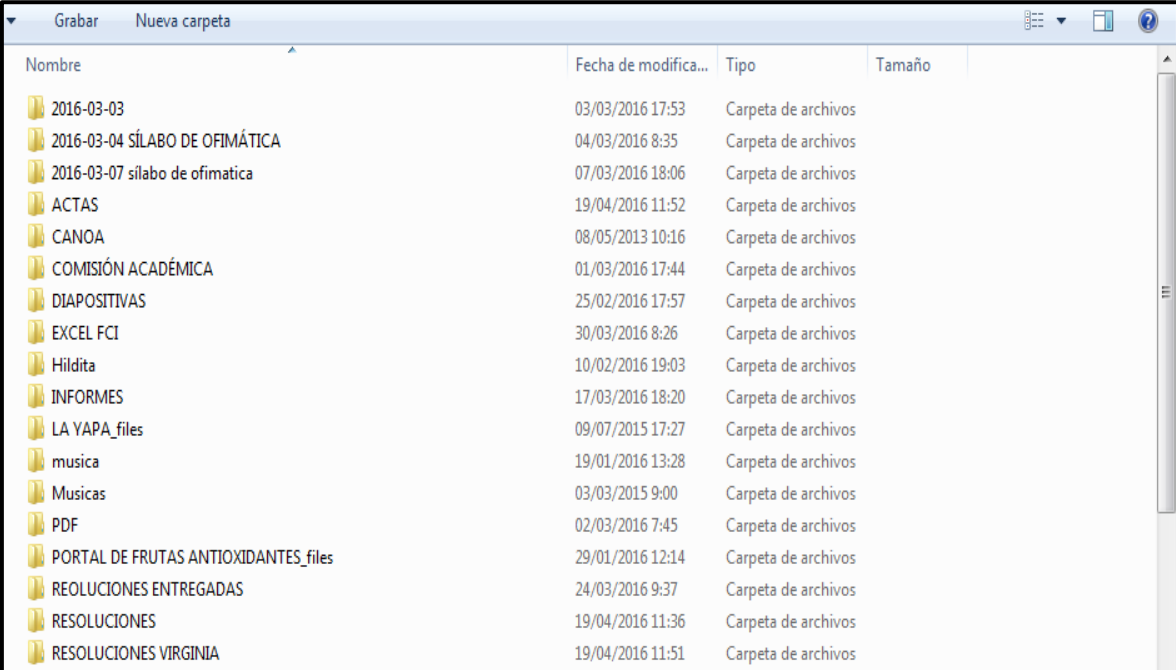
1. Recolectar los documentos que se generan en los diferentes departamentos de la FCI.
2. Revisar el contenido de los archivos, para diferenciar a que trámite pertenece.
3. Crear carpetas según los trámites generados, para el almacenamiento los archivos diferenciados.
4. Guardar los archivos, en las carpetas creadas.

Para la búsqueda de los archivos:

1. Abrir la carpeta, que según el criterio de requerimiento del usuario, puede contener el archivo deseado.
2. Abrir el archivo y revisar si la información que requiere el usuario está en el mismo.
3. En caso de que el archivo revisado, no se encuentre la información del usuario, la secretaría debe proceder a revisar las demás carpetas hasta encontrar el documento deseado.

Como se puede apreciar en la *Figura.- 17*, no existe una organización adecuada de los trámites que se genera en la FCI, además existe un considerado consumo recursos creando carpetas para los documentos que en ciertos casos puede existir duplicaciones de archivos, aumentando de esta manera que las probabilidades de búsqueda de un documento sean mínimas.

Figura.- 28. Situación actual de organizar los documentos en la Secretaría de la FCI.



Nombre	Fecha de modifica...	Tipo	Tamaño
2016-03-03	03/03/2016 17:53	Carpeta de archivos	
2016-03-04 SÍLABO DE OFIMÁTICA	04/03/2016 8:35	Carpeta de archivos	
2016-03-07 sílabo de ofimatica	07/03/2016 18:06	Carpeta de archivos	
ACTAS	19/04/2016 11:52	Carpeta de archivos	
CANOA	08/05/2013 10:16	Carpeta de archivos	
COMISIÓN ACADÉMICA	01/03/2016 17:44	Carpeta de archivos	
DIAPPOSITIVAS	25/02/2016 17:57	Carpeta de archivos	
EXCEL FCI	30/03/2016 8:26	Carpeta de archivos	
Hildita	10/02/2016 19:03	Carpeta de archivos	
INFORMES	17/03/2016 18:20	Carpeta de archivos	
LA YAPA_files	09/07/2015 17:27	Carpeta de archivos	
musica	19/01/2016 13:28	Carpeta de archivos	
Musicas	03/03/2015 9:00	Carpeta de archivos	
PDF	02/03/2016 7:45	Carpeta de archivos	
PORTAL DE FRUTAS ANTIOXIDANTES_files	29/01/2016 12:14	Carpeta de archivos	
RESOLUCIONES ENTREGADAS	24/03/2016 9:37	Carpeta de archivos	
RESOLUCIONES	19/04/2016 11:36	Carpeta de archivos	
RESOLUCIONES VIRGINIA	19/04/2016 11:51	Carpeta de archivos	

FUENTE: SECRETARIA DE LA FACULTAD DE CIENCIAS DE LA INGENIERÍA.

AUTORES: AGUAL; MENDOZA,2016.

4.1.2. Gestión de documentos digitales en la FCI, utilizando el aplicativo elaborado.

El sistema informático desarrollado permite la gestión de los documentos digitales de una manera más sencilla:

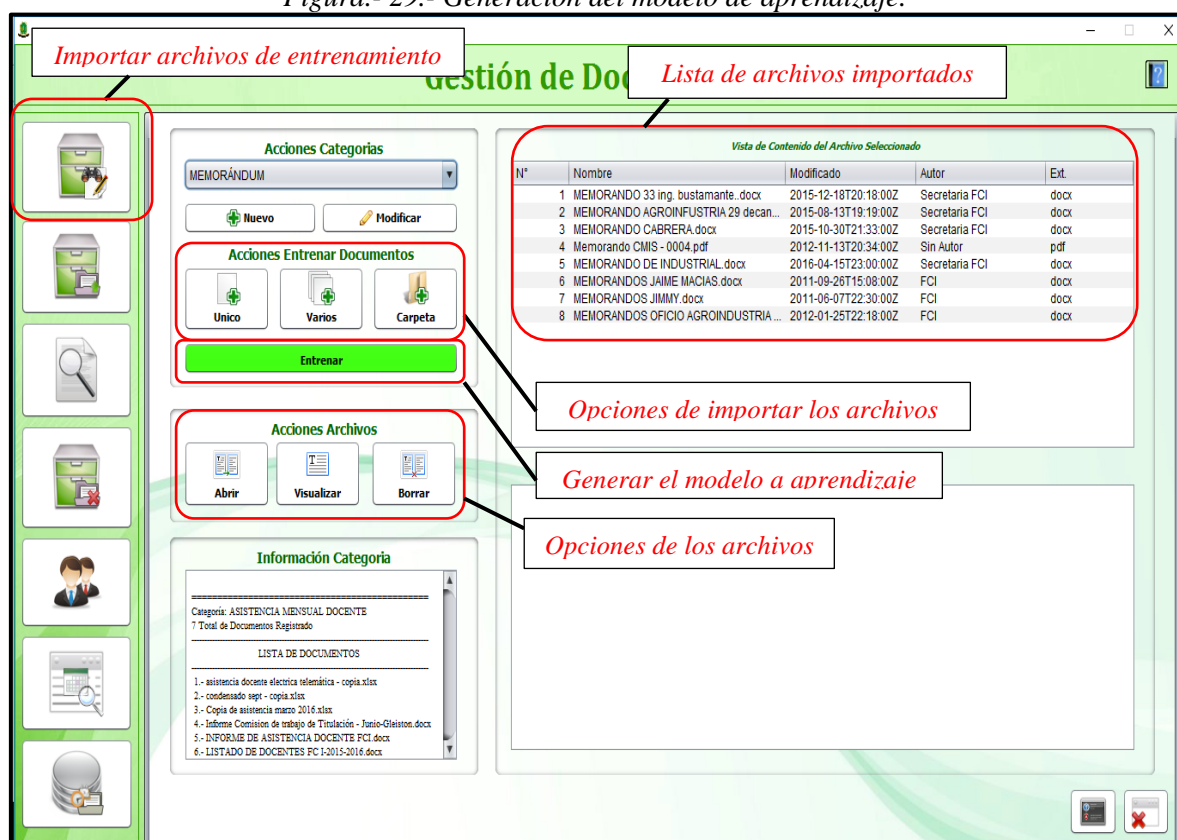
- Organizando los archivos de manera automática, lo que implica un menor consumo de recursos y tiempo,
- Encontrar un documento de manera rápida, sin tener que abrir las carpetas en las que se encuentran almacenadas.
- Respalidar la información, para evitar pérdida alguna de los archivos.

La gestión de documentos haciendo el uso del sistema se realiza a través de los siguientes pasos:

1. **Obtención del modelo de aprendizaje** por cada una de las categorías que dispone la aplicación, seleccionando la opción *“Importar archivos de entrenamiento”*, posteriormente se procede a importar al sistema los documentos que nos van servir como base para que el algoritmo aprenda la forma como están estructurados los archivos, se puede subir al sistema los documentos de entrenamiento mediante las

opciones *uno archivo*, *varios archivos* o *una carpeta de archivos* por cada una de las categorías, todos los documentos que se han importado al aplicativo se los puede visualizar en una tabla (*lista de archivos Importados*) los mismos que se tienen las diferentes opciones *de abrir*, *visualizar el contenido del archivo* o *eliminarlos* si el archivo es incorrecto o no presenta suficiente información que pueda servir para el entrenamiento, una vez que se importaron los archivos y se verificaron si son adecuados para el entrenamiento se procede a *generar el modelo de aprendizaje*, presionando la opción *Entrenar*.

Figura.- 29.- Generación del modelo de aprendizaje.



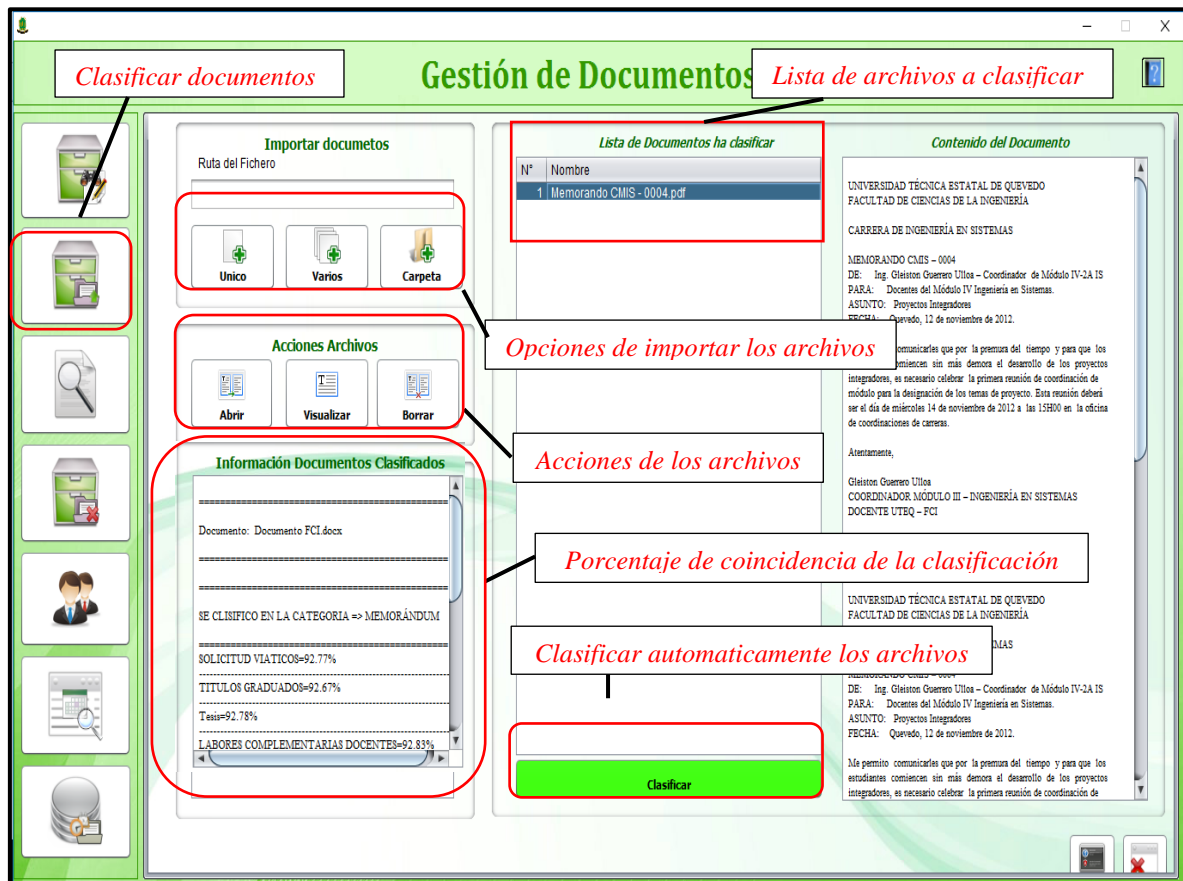
FUENTE: SISTEMA PARA GESTIÓN AUTOMÁTICA DE DOCUMENTOS DIGITALES.

AUTORES: AGUAL; MENDOZA,2016.

2. **Clasificación de los documentos** forma automática a través del modelo de aprendizaje generado, seleccionando la opción “*Clasificar documentos*” de la aplicación, posteriormente se procede a importar a la aplicación los documentos a clasificar mediante las opciones *uno archivo*, *varios archivos* o *una carpeta de archivos* todos los documentos que se han importado al aplicativo se los puede visualizar en una tabla (*Lista de archivos Importados*) los mismos que tienen las diferentes opciones *de abrir*, *visualizar el contenido del archivo* o *eliminarlos* si fuera el caso, una vez que se importaron los archivos se procede a *Clasificar*

automáticamente los archivos mostrando en la parte izquierda a que categoría fue clasificado y almacenado, además se puede apreciar el porcentaje de coincidencia con relación a las demás categorías, de esta manera la aplicación diferencia a que trámite pertenece los archivos, direccionándolos y almacenándolos en sus categorías respectivas.

Figura.- 30. Clasificación automática de los archivos.



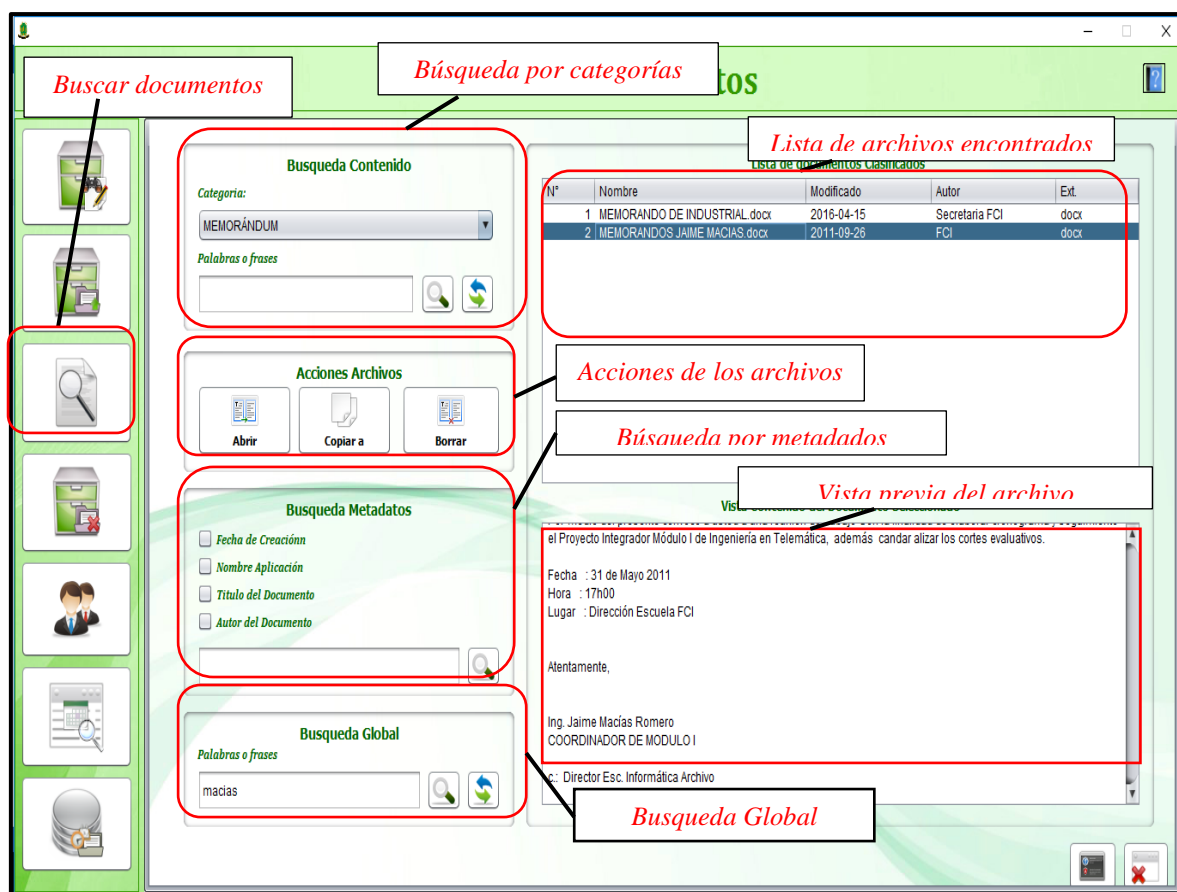
FUENTE: SISTEMA PARA GESTIÓN AUTOMÁTICA DE DOCUMENTOS DIGITALES.

AUTORES: AGUAL; MENDOZA,2016.

3. **Búsqueda de los documentos** que requiere el usuario mediante la exploración contenido textual del archivo, seleccionando la opción “*Buscar documento*” de la aplicación, las búsquedas de los archivos se puede realizar de diferentes maneras sea por “*Búsqueda por categorías*” seleccionando una categoría y escribiendo el criterio a buscar, “*Búsqueda por metadatos*” seleccionando uno de los metadatos disponible como *Fecha de creación*, *Nombre de la aplicación*, *Titulo del documento*, *Autor del documento* y definiendo el criterio a buscar según el metadato y por último la “*Búsqueda global*” la misma que realiza una búsqueda completa por todas las categorías disponibles, todos los documentos encontrados a través los diferentes

métodos de búsqueda los muestra en una tabla (*Lista de archivos encontrados*), para la manipulación de los documentos se tiene las opciones de *Abrir*, *Copiar a*, *Borrar* y una visualización previa del archivo. Así de esta manera se reduce el tiempo de búsqueda un archivo.

Figura.- 31. Búsqueda de los documentos.



FUENTE: SISTEMA PARA GESTIÓN AUTOMÁTICA DE DOCUMENTOS DIGITALES.

AUTORES: AGUAL; MENDOZA,2016.

En la *Figura.- 32*, se puede apreciar las categorías en donde los archivos son clasificados de manera automática, de esta manera se reduce el consumo de almacenamiento del disco duro, y a su vez agiliza el proceso de búsqueda de los archivos.

Figura.- 32. Organización de los archivos mediante el sistema.

Nombre	Fecha de modifica...	Tipo	Tamaño
ASISTENCIA MENSUAL DOCENTE	15/04/2016 2:52	Carpeta de archivos	
CONDENSADO TRABAJO DOCENTE	15/04/2016 2:52	Carpeta de archivos	
DISTRIBUTIVOS DOCENTES	15/04/2016 2:52	Carpeta de archivos	
LABORES COMPLEMENTARIAS DOCENTES	15/04/2016 2:52	Carpeta de archivos	
MEMORÁNDUM	15/04/2016 2:52	Carpeta de archivos	
OFICIOS DECANATO	15/04/2016 2:52	Carpeta de archivos	
OTROS	15/04/2016 2:52	Carpeta de archivos	
REDISEÑO MALLAS CURRICULARES	15/04/2016 2:52	Carpeta de archivos	
RESOLUCIONES DEL CONSEJO ACADEMI...	15/04/2016 2:52	Carpeta de archivos	
SOLICITUD DE ESTANCIAS	15/04/2016 2:52	Carpeta de archivos	
SOLICITUD DE PERMISO	15/04/2016 2:52	Carpeta de archivos	
SOLICITUD VIATICOS	15/04/2016 2:52	Carpeta de archivos	
TITULOS GRADUADOS	15/04/2016 2:52	Carpeta de archivos	

FUENTE: SISTEMA PARA GESTIÓN AUTOMÁTICA DE DOCUMENTOS DIGITALES.

AUTORES: AGUAL; MENDOZA,2016.

Mediante el aplicativo desarrollado para la gestión de documentos digitales se tiene una organización adecuada de cada trámite que se genera en la FCI a través de carpetas únicas que no permite que exista un archivo duplicado, no permite que se manipulen las carpetas creadas por personal no autorizado, porque se encuentran de manera oculta en una ubicación específica del computador garantizado así la confidencialidad de los archivos.

La búsqueda de documentos se realiza de manera rápida por medio de un módulo creado para tal efecto, el cual hace una búsqueda por el contenido textual ya sea por medio de cada categoría o por una búsqueda global de todas las categorías que dispone la aplicación.

4.2. Pruebas y resultados.

Para desarrollar la aplicación de gestión automática de documentos se implementó el algoritmo Naive – Bayes que permitió organizar los documentos, de la misma manera se utilizó la biblioteca Apache Lucene de apache para minimizar el tiempo de búsqueda de los archivos digitales. Se hicieron varias pruebas tomando el tiempo que se consume en realizar las tareas de organización y búsqueda de los mismos, de la forma que se viene utilizando actualmente (manual), en comparación con la nueva metodología implementada (automática). A continuación, podemos apreciar la tabla comparativa con las ventajas y desventajas de los algoritmos analizados el proyecto.

Tabla 37. Ventajas y desventajas de los algoritmos estudiados.

Algoritmo	Ventajas	Desventajas
k-NN (k-Nearest Neighbors)	<ul style="list-style-type: none"> -Buena precisión en la clasificación. -Requiere una pequeña cantidad de datos de entrenamiento. 	<ul style="list-style-type: none"> -El tiempo de clasificación de un nuevo documento es alto. -Requiere gran capacidad de almacenamiento.
Árboles de Decisión	<ul style="list-style-type: none"> -Fáciles de interpretar. -Es posible derivar reglas. 	<ul style="list-style-type: none"> -El tiempo de entrenamiento es elevado. -Un documento sólo puede estar conectado por una rama y error en una rama llevar a errores en las ramas inferiores.
Naive Bayes	<ul style="list-style-type: none"> -Implementación sencilla -Requiere una pequeña cantidad de datos de entrenamiento -Consumo mínimo tanto en la CPU y la memoria. 	
Máquina de vectores de soporte	<ul style="list-style-type: none"> -Captura bien las características inherentes de los datos. 	<ul style="list-style-type: none"> -Requiere grandes cantidades de tiempo y memoria computacional.
Algoritmo K-Means.	<ul style="list-style-type: none"> -Implementación sencilla -Agrupación por clusters 	<ul style="list-style-type: none"> -Coste computacional y el espacio en memoria alto. -Realizar sucesivas ocasiones el algoritmo para tener en el resultado más óptimo.

AUTORES: AGUAL; MENDOZA,2016.

- **Prueba 1** Creación de las carpetas por categorías.

Tabla 38. Prueba 1 Creación de carpetas.

	Tiempo utilizado
Creación manual	1 minuto 30 segundos
Creación automática	15 segundos

AUTORES: AGUAL; MENDOZA,2016.

- **Prueba 2** Organización de los archivos por categorías.

Tabla 39. Prueba 2 Organización de los archivos.

	Tiempo utilizado
Organización manual	3 minutos
Organización automática	30 segundos

AUTORES: AGUAL; MENDOZA,2016.

- **Prueba 3** Búsqueda de los documentos.

Tabla 40. Prueba 3 Búsqueda de los documentos.

	Tiempo utilizado
Búsqueda manual	1 minutos
Búsqueda automática	10 segundos

AUTORES: AGUAL; MENDOZA,2016.

4.3. Discusión.

El objetivo del proyecto de M. Alicia Pérez Abelleira y Carolina A. Cardoso(2010), es implementar un buscador semántico y la clasificación automática de los documentos utilizando para tal efecto el algoritmo no supervisado **SMO (máquinas de vectores soporte)**, este proyecto el algoritmo seleccionado brinda resultados aceptables pero requiere un tiempo de considerado para su procesamiento y a su vez requiere un alto consumo de memoria, lo que se implementó en el proyecto de la Universidad Técnica Estatal de Quevedo específicamente en la Secretaría de la FCI, fue la clasificación e indexación de documentos utilizando el algoritmo **Naive Bayes**, el cual no necesita mucho consumo de memoria y su procesamiento es ágil, agilitando de esta manera los procesos y minimizando el consumo los recursos.

En cuanto a la búsqueda de documentos el proyecto de M. Alicia Pérez Abelleira y Carolina A. Cardoso(2010), se basa en la extracción de ciertos atributos como, nombres, títulos, años,

entidades, etc., para así de esta manera buscar los archivos, lo que se implementó en el proyecto de la Universidad Técnica Estatal de Quevedo fue la búsqueda extensa por todo el contenido textual de los documentos clasificados aprovechando la funcionalidades que nos ofrece la biblioteca **Apache Lucene** y otras herramientas de código abierto tales; **Tika,Tm-extractors, OpenNLP** que sirvieron para la identificación de los diferentes formatos que poseen los documentos digitales y el procesamiento de los mismos, logrando de esta manera analizar todo el contenido textual de los documentos.

CAPITULO V
CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones.

El trabajo elaborado es una propuesta de un sistema de gestión automática de documentos digitales basado en el contenido textual, para facilitar el acceso a la información de la Facultad de Ciencias de la Ingeniería, utiliza como eje principal un corpus de documentos que se tramitan en la Secretaría de la FCI, a continuación, se explica las conclusiones alcanzadas.

- Se logró analizar la fuente de información documental brindada por la Secretaría de la FCI, mediante el estudio de las técnicas y modelos dentro del área de la clasificación de documentos, se encontró que el algoritmo de clasificación “Naive Bayes” es la más utilizada y brinda resultados aceptables, por el tiempo requerido para su uso es menor comparado con los demás algoritmos existentes [38], de esta manera se procedió a organizar los documentos digitales, reduciendo así consumo de recursos.
- La garantía de los documentos digitales consiste en que no se refleje duplicaciones en los archivos y de la misma manera que no se elimine ningún tipo de documento, mediante la realización de un respaldo en la información que se encuentran en la secretaria de la FCI.
- Mediante la utilización de las herramientas que nos proporciona la inteligencia artificial se logró agilizar el proceso de búsqueda de los archivos en la gestión de documentos digitales en la FCI, aplicando la biblioteca Apache Lucene [39], reduciendo el tiempo de encontrar los documentos, mediante la búsqueda por categorías y la búsqueda global a través del contenido textual de los archivos.

5.2. Recomendaciones.

- Desarrollar el sistema en un entorno web, para poder utilizar las ventajas que nos ofrecen las aplicaciones web.
- Implementar un módulo que realice los respaldos de información periódicamente (con o sin pedir autorización al usuario) y almacenarlo en la nube (internet).
- Realizar un módulo de seguimiento de los trámites que los estudiantes realicen, para que conozcan el proceso de los mismos.

CAPITULO VI

BIBLIOGRAFÍA

6.1. Bibliografía

- [1] N. J. P. Daza, *La Gestión de Documentos modelos y características*, España: Fundacion, 2011.
- [2] M. G. M. R. L. José Alonso, *La norma ISO 15489: un marco sistemático de buenas práctica de gestión documental en las organizaciones.*, España, 2008.
- [3] D. A. Escribano, *Calsificaciòn Automática de Videos*, España, 2009.
- [4] R. M. Coyotl, *Clasificación Automática de Textos considerando el estilo de Redacción*, Tonantzintla, 2007.
- [5] J. L. A. B. A. F. Z. Carlos Figueroa, «OpenCourseWare de la Universidad de Salamanca,» [En línea]. Available: <http://ocw.usal.es/enseñanzas-tecnicas/recuperacion-avanzada-de-la-informacion/contenidos/clasificacion.pdf>.
- [6] J. E. R. R. Leslie del Carmen González Díaz, «Prototipo de software para la Clasificación de datos mediante el método Bayesiano Tan– “udTAN”,» *Vinculos Ciencia Tecnología y Sociedad: Un enlace hacia el futuro*, vol. 3, nº 1, p. 19, 2006.
- [7] Datumbox.com, «Datumbox,» [En línea]. Available: <http://blog.datumbox.com/machine-learning-tutorial-the-naive-bayes-text-classifier/>.
- [8] J. A. B. R. R. Carlos Figuerola, «Algunas Técnicas de Clasificación Automática de Documentos,» *Researchgate*, vol. 13, p. 8, 2004.
- [9] R. Günther, *Aceleracion del Algortimo K-NN*, 2009.
- [10] S. P. L. M. F. Andrés Zapata Tapasco, «Método basado en clasificadores k-NN parametrizados con algoritmos genéticos y la estimación de la reactancia para localización de fallas en sistemas de distribución.,» *Sistema de Información Científica Redalyc Red de Revistas Científicas de América Latina y el Caribe, España y Portugal*, nº 70, p. 13, 2014.
- [11] G. J. Hernández M, «Aplicaciones de Procesamiento de Lenguaje Natural,» *Revista Politécnica*, vol. 32, nº 1, p. 10, 2013.
- [12] Ecured, «Ecured- Conocimiento con todos y para todos,» [En línea]. Available: http://www.ecured.cu/M%C3%A1quinas_de_Soporte_Vectorial.
- [13] D. A. T. Silva, *Diseño y aplicación de un metodología para el análisis de noticias policiales utilizando minería de textos*, Santiago de Chile, 2013.
- [14] G. E. S. K. D. M. R. Raúl Benítez, *Inteligencia artificial avanzada*, Barcelona: Editorial UOC, 2014.
- [15] S. R. C. Paula Andrea Benavides Cañón, *Procesamiento del Lenguaje Natural en la Recuperación de Información*, Colombia, 2001.

- [16] A. S. E. Aguirre García Nerea, *Sistema de ayuda al diagnóstico clínico: clasificación de diagnósticos clínicos*, 2013.
- [17] F. R. ., S. S. Concepción Pérez de Celis, «Creación y uso de una ontología relacionada con genes, síndromes, síntomas y enfermedades para la clasificación de textos biomédicos,» *Revista signos*, vol. 47, nº 84, p. 22, 2014.
- [18] F. F.-R. F. D. J. M. C. José R. Méndez, «Sistemas inteligentes para la detección y filtrado de correo spam: una revisión,» *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, vol. 34, p. 19, 2007.
- [19] A. M. Santiago Coteló, *Búsqueda de documentos utilizando criterios semánticos*, Montevideo, Uruguay, 2012.
- [20] C. B. ., F. Andrés Vuotto, «Aplicación del factor TF-IDF en el análisis semántico de una colección documental.,» *Biblios, Revista de Bibliotecología y Ciencias de la Información*, vol. 60, p. 13, 2015.
- [21] F. T. R. Montejó, *Mètode para la evaluacion automàtica de la organizaciòn de textos argumentativos*, Bogota: Universidad Nacional de Colombia, 2014.
- [22] J. R. L. Fuentes, *Desarrollo de Software Ágil: Extremme Programming y Scrum*. 2ª Edición, IT Campus Academy, 2015.
- [23] A. Q. D. D. M. G. Alveiro Rosado Gómez, «Desarrollo ágil de software aplicando programación extrema,» *Revista Ingenio*, vol. 5, nº 1, p. 3, 2012.
- [24] D. de, «Definicion de,» 2012. [En línea]. Available: <http://definicion.de/java/>.
- [25] N. Community, «NetBeans IDE,» 2016. [En línea]. Available: <https://netbeans.org>.
- [26] A. L. Core, «Apache Lucene,» 2016. [En línea]. Available: <https://lucene.apache.org/core/>.
- [27] K. T. -. L. a. S. consultant, «LuceneTutorial.com,» [En línea]. Available: <http://www.lucenetutorial.com/basic-concepts.html>.
- [28] A. S. Foundation, «Técnicas avanzadas de recuperación de información: procesos, técnicas y métodos,» Lucene analyzers-common API, [En línea]. Available: https://lucene.apache.org/core/4_4_0/analyzers-common/overview-summary.html.
- [29] O. Corporation, «MySQL,» [En línea]. Available: <https://www.mysql.com/>.
- [30] T. A. S. Foundation, «Apache Tika,» [En línea]. Available: <https://tika.apache.org/>.
- [31] T. A. S. Foundation, «The Apache Software Foundation, Apache POI,» [En línea]. Available: <https://poi.apache.org/>.
- [32] T. A. S. Foundation, «OpenNLP,» [En línea]. Available: <https://opennlp.apache.org/>.
- [33] R. Venegas, «Clasificación de textos académicos en función de su contenido léxico-semántico,» *Revista signos*, vol. 40, nº 63, 2007.

- [34] E. M. J. H. M. Jose María Quinteiro González, «Clasificación de textos en Leguaje natural usando la Wikipedia,» *Revista Ibérica de Sistemas e Tecnologías de Información*, nº 8, p. 14, 2011.
- [35] G. P. ., X. O. Javier Caicedo, *Marco de trabajo para indexación, clasificación y recopilación automática de documentos digitales.*, Guayaquil, 2009.
- [36] V. U. C. B. Martha Varguez, *Clasificacion de documentos usando Màquinas de Vectores de Apoyo*, Mexico, 2014, p. 12.
- [37] C. A. C. M. Alicia Pérez Abelleira, *Minería de texto para la categorización automática de documentos.*, España, 2010.
- [38] I. O. ., V. R. Marcelo Mendoza, «Categorización de texto en bases documentales a partir de modelos computacionales livianos,» *SciELO*, vol. 44, nº 77, 2011.
- [39] M. C. Lorient Yañez, *Comparación de las posibilidades de dos máquinas de búsqueda (Lucene y MG4J) sobre textos XML*, 2013.

CAPITULO VII

ANEXOS

Manual de Usuario del aplicativo “Gestión automática de documentos digitales basado en el contenido textual, para facilitar el acceso a la información de Facultad de Ciencias de la Ingeniería”

El propósito de este manual es facilitar al usuario la operación de las diferentes pantallas la que dispone el Sistema informático de gestión automática de documentos.

1. IMPLEMENTACIÓN DEL SISTEMA

a) Requerimientos de hardware

Contar con:

- Computadora personal.
- Conexión a Internet.

b) Requerimientos de software

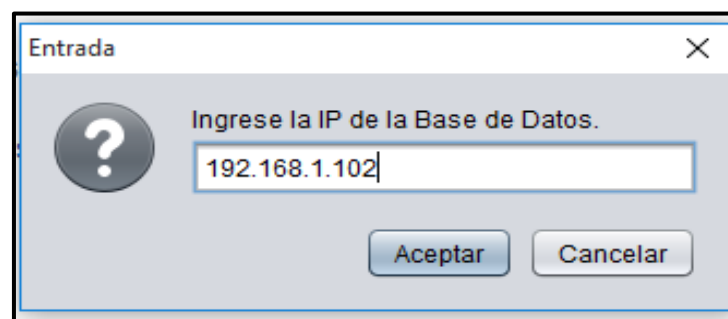
Contar con:

- Sistema operativo Windows.
- Máquina virtual de Java.
- Gestor de Base Datos para el Servidor (MySQL)

2.-INGRESANDO AL SISTEMA

Una vez localizado el nombre de la aplicacion *Gestion de docuentos* en el escritorio, dar doble clic para que se ejecute el aplicativo, inmediatamente el sistema presenta un formulario que nos solicita la direccion ip de la Base de Datos (servidor).

Figura.- 33. Funcionamiento del Sistema - Ingreso de la dirección ip de la Base de datos.



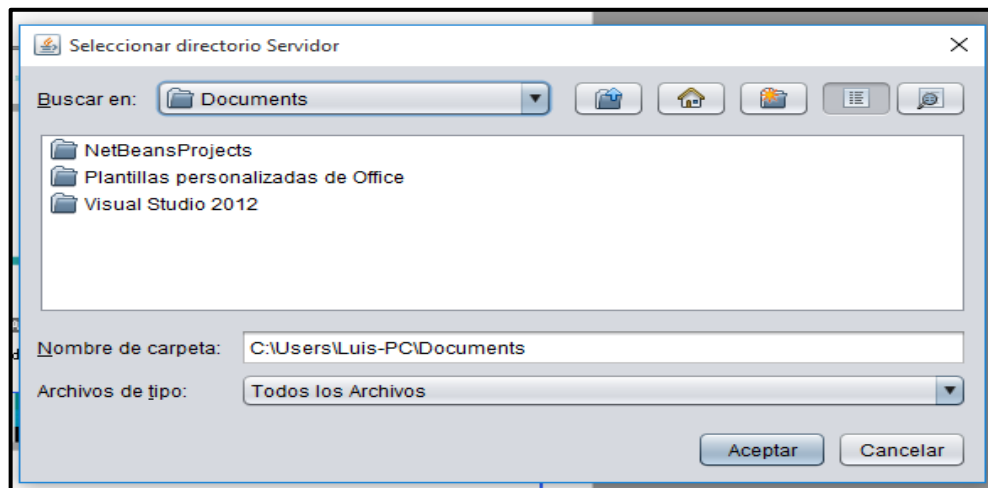
Inmediatamente se nos muestra el formulario de Inicio de sesión, el cual solicita el nombre de usuario registrado y su respectiva contraseña.

Figura.- 34. Funcionamiento del sistema - Inicio de sesión.



Verificado que los datos ingresados por el usuario son los correctos, el sistema solicitará la ubicación de la carpeta servidor de archivos.

Figura.- 35. Funcionamiento del sistema - Seleccionar el directorio de archivos del servidor.



Seleccionado la carpeta de archivos del servidor, el sistema nos muestra la interfaz principal de la aplicación, la misma que se cargará dependiendo del rol de usuario. Entre los roles que dispone el sistema se tiene a:

Administrador: Posee todos los privilegios dentro del sistema, generalmente son sus creadores y/o personas capaces de manejar toda su información y en caso de que fuera necesario la modificación del mismo.

Invitado: Posee ciertos privilegios dentro del sistema, los cuales son proporcionados por el usuario Administrador.

Figura.- 36. Funcionamiento del sistema - Interfaz principal (Usuario Administrador).



Figura.- 37. Funcionamiento del sistema - Interfaz principal (Usuario Invitado).



3.- FUNCIONALIDAD GENERAL

Las pantallas del sistema se dividen en siete módulos:

- Módulo “Importar Archivo de Entrenamiento”.
- Módulo “Clasificar Documentos”.
- Módulo “Buscar Documentos”.

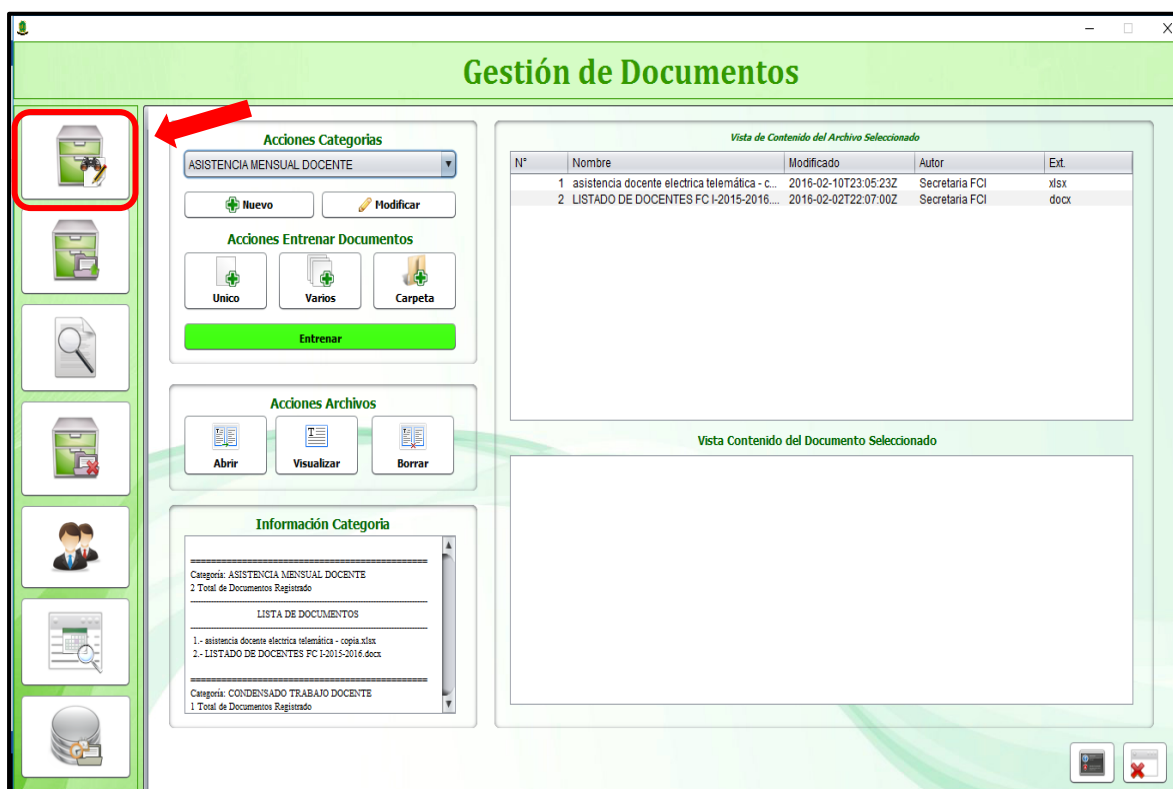
- Módulo “Borradores de Documentos”.
- Módulo “Administrador de Usuarios”.
- Módulo “Vista Registro de Búsquedas”.
- Módulo “Respaldo Información”.

4.-MÓDULOS QUE INTEGRAN EL SISTEMA

MÓDULO IMPORTAR ARCHIVO DE ENTRENAMIENTO

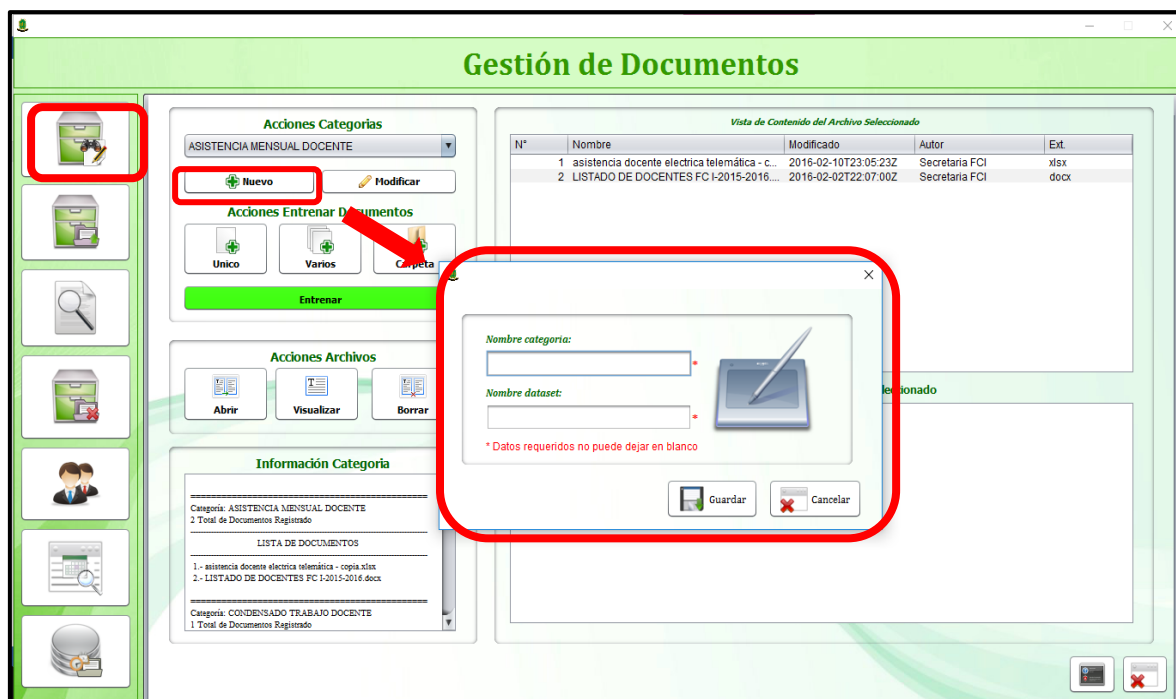
Este módulo permite crear nuevas categorías para la administración de los archivos digitales, importar los archivos de entrenamiento, los mismos que nos permitirá obtener los diferentes modelos de aprendizaje para posteriormente clasificar los documentos.

Figura.- 38. Funcionamiento del sistema - Importar Archivo de Entrenamiento.



- Nueva categoría.

Haga clic *Nuevo*, para crear una nueva categoría, digite el nombre respectivo de la categoría y un nombre adicional para identificar el conjunto de datos dentro del sistema.

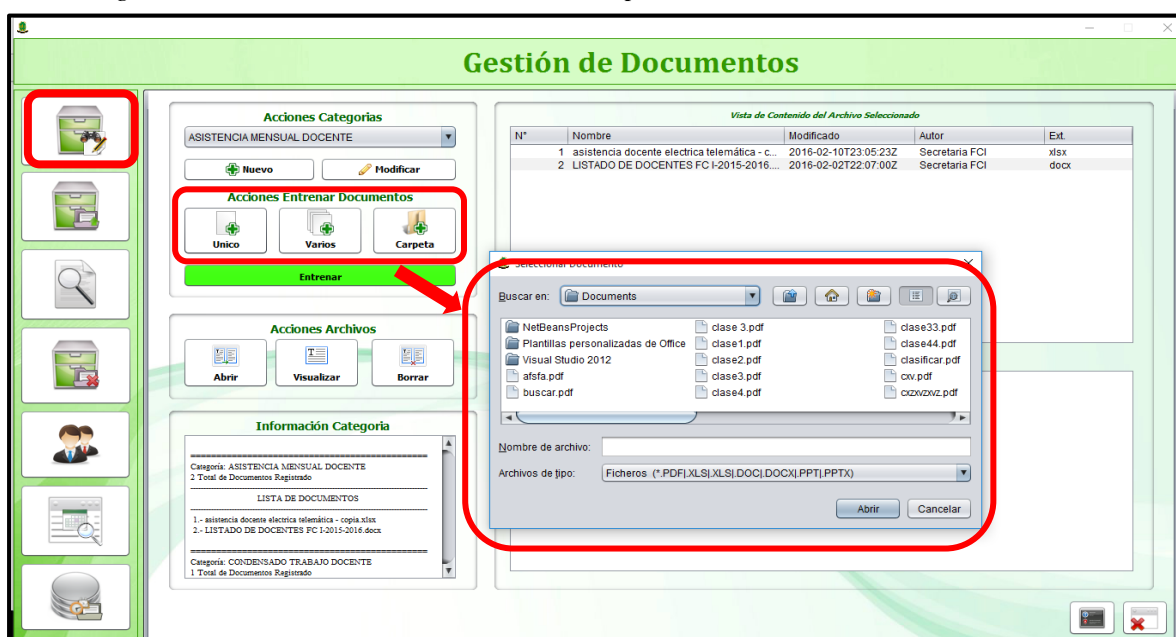


• Importar archivos

La obtención de los modelos de aprendizaje por cada categoría que dispone el sistema, se las realiza mediante la importación de archivos que pueden ser de diferentes formatos (pdf, xlsx, docx,ppt), a través de las opciones:

- **Único:** Importa solo un archivo de entrenamiento al sistema.
- **Varios:** Importa varios archivos de entrenamiento al sistema.
- **Carpeta:** Importa todos los archivos que contiene una carpeta.

Figura.- 39. Funcionamiento del sistema - Importación de archivos de entrenamiento.

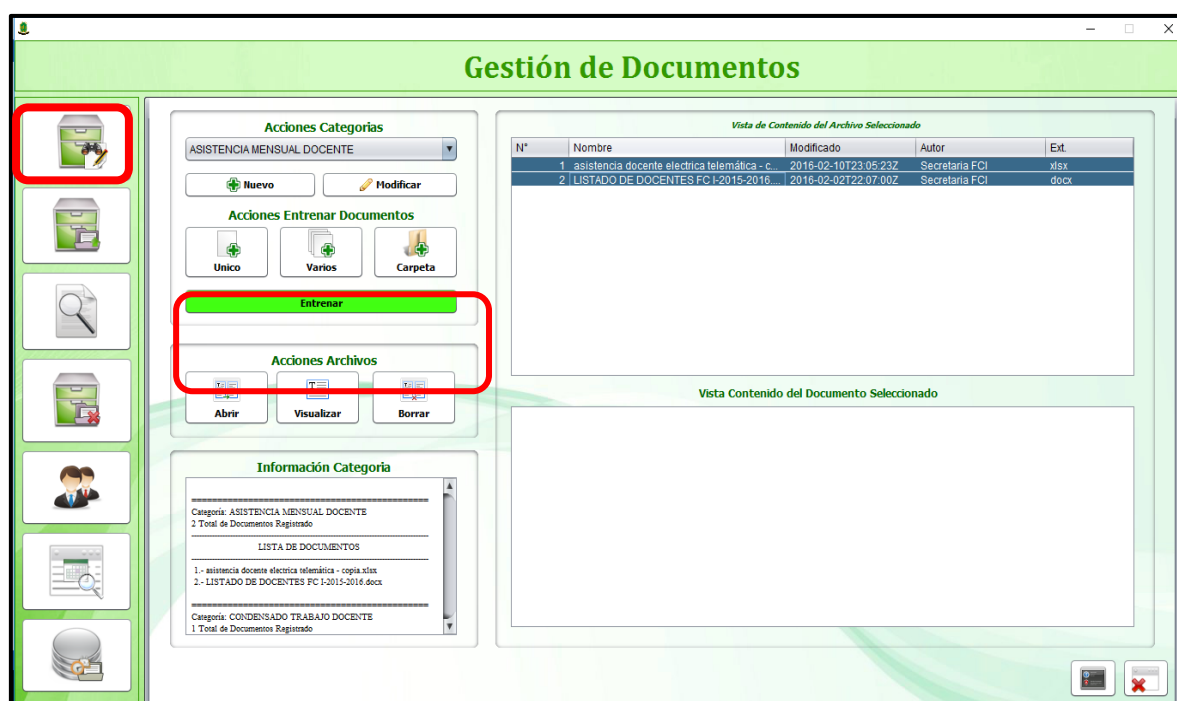


Manipular los archivos.

Todos los archivos que se importen al sistema, se puede ver una vista previa del documento, abrir el documento completo y borrar en caso que no sea el adecuado para la respectiva categoría a través de los siguientes comandos:

- **Abrir:** Abre el documento seleccionado de la lista de importados.
- **Visualizar:** Muestra una vista previa del documento seleccionado en la parte inferior del sistema.
- **Borrar:** Borra los documentos que el usuario selecciona de la lista de importados.

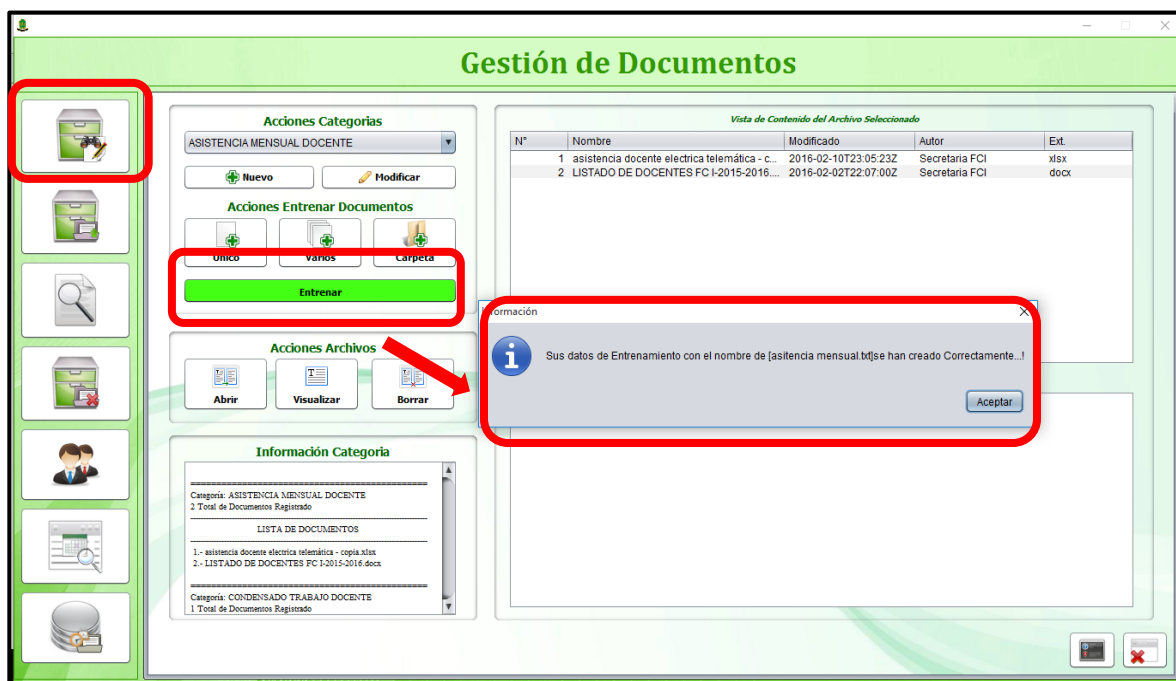
Figura.- 40. Funcionamiento del sistema - Manipulación de los archivos



• Obtener el modelo de aprendizaje.

Cuando todos los archivos para el entrenamiento han sido importados correctamente al sistema, se debe generar el modelo de aprendizaje por cada categoría, dando click en el botón “Entrenar”. El sistema nos mostrará un mensaje de confirmación mencionando que el modelo de aprendizaje ha sido generado de manera satisfactoria.

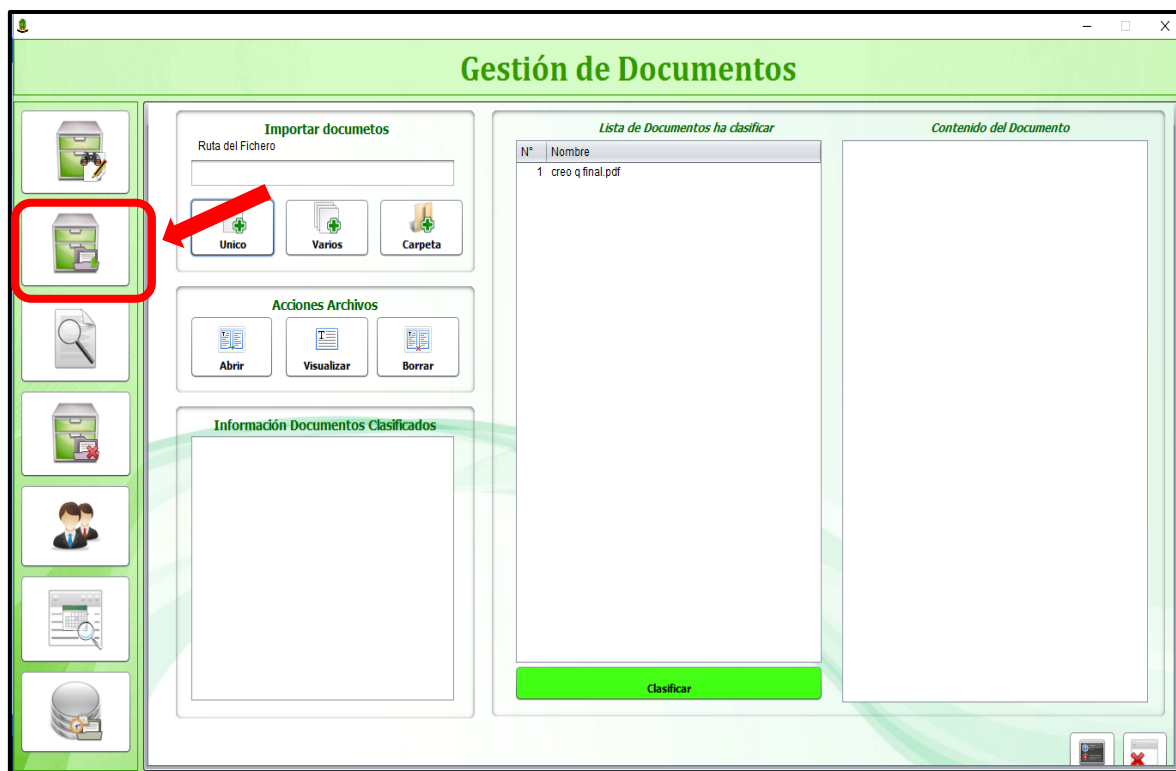
Figura.- 41. Funcionamiento del sistema - Generar el modelo de aprendizaje.



MÓDULO CLASIFICAR DOCUMENTOS.

Este módulo permite la clasificación de documentos de manera automática, todos los archivos que el usuario importe al sistema son categorizados automáticamente, por medio del modelo de aprendizaje generado en el módulo “Importar archivos de entrenamiento”.

Figura.- 42. Funcionamiento del sistema - Clasificar documentos.

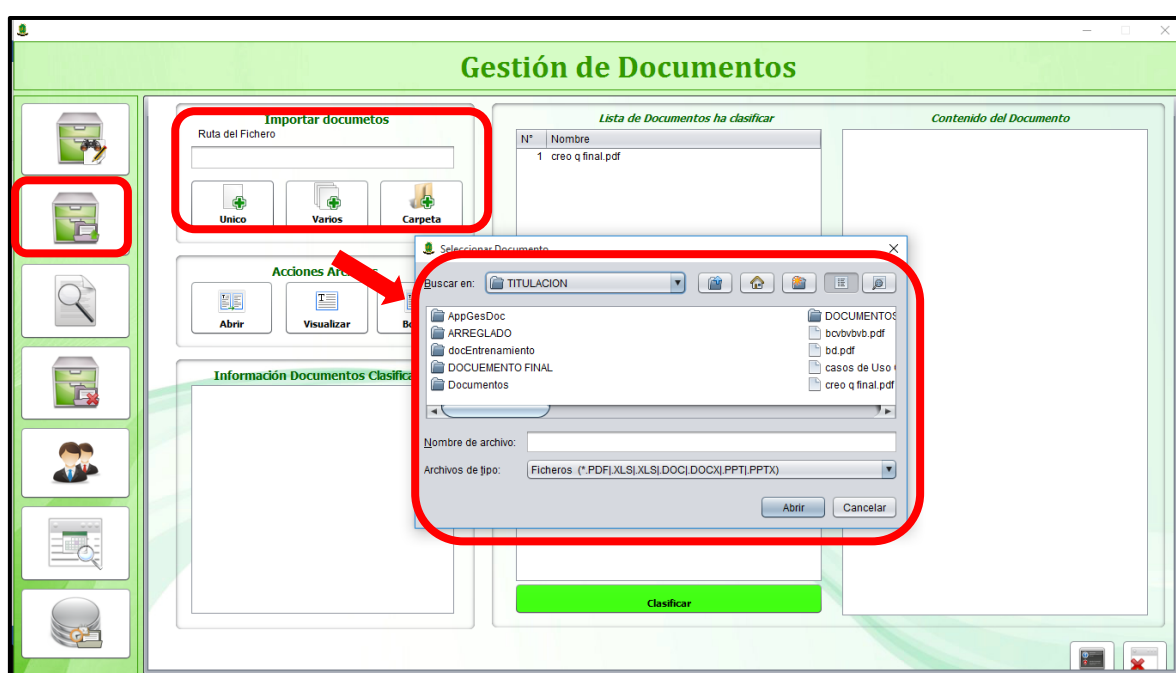


- **Importar documentos.**

El primer paso para la clasificación de los documentos, es importar todos los archivos que desee organizar, los archivos pueden ser: pdf, xlsx, docx, ppt. La importación de los documentos se la realiza por medio de las opciones:

- **Único:** Importa solo un archivo al sistema.
- **Varios:** Importa varios archivos al sistema.
- **Carpeta:** Importa todos los archivos que contiene una carpeta

Figura.- 43. Funcionamiento del sistema - Importar documentos a clasificar .

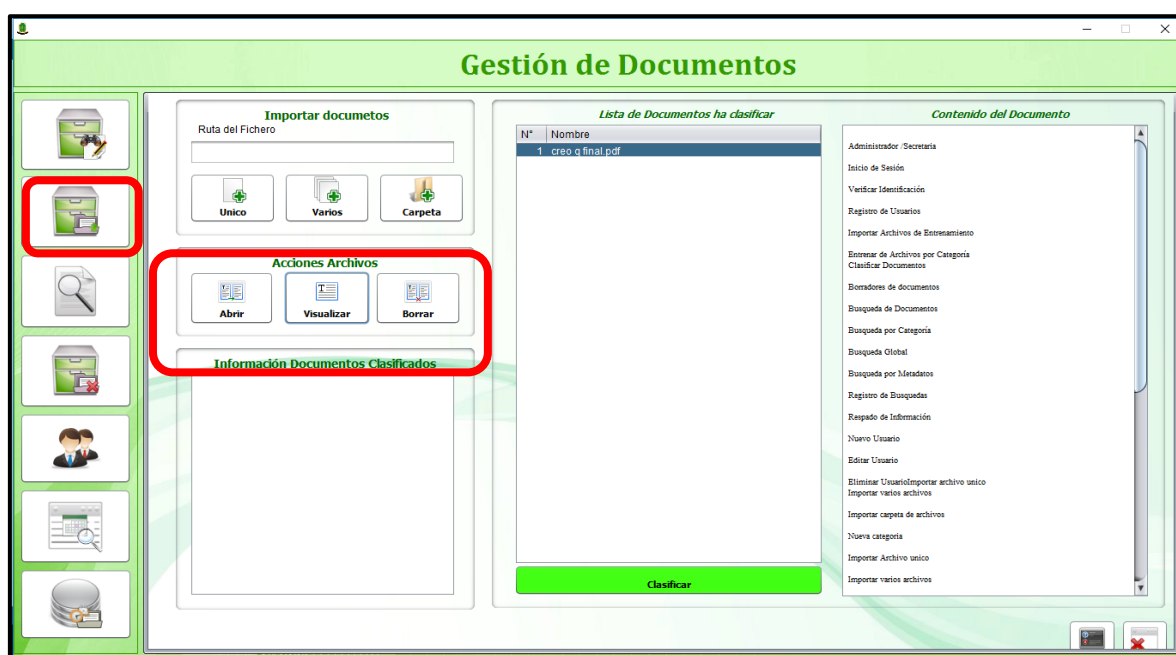


Manipular los archivos.

Todos los archivos que se importen al sistema, se puede hacer una vista previa del documento, abrir el documento completo y borrar en caso que no sea el adecuado, estas operaciones se la realizan a través de los siguientes comandos:

- **Abrir:** Abre el documento seleccionado de la lista de importados.
- **Visualizar:** Muestra una vista previa del documento seleccionado en la parte derecha de esta pantalla.
- **Borrar:** Borra los documentos que el usuario selecciona de la lista de importados.

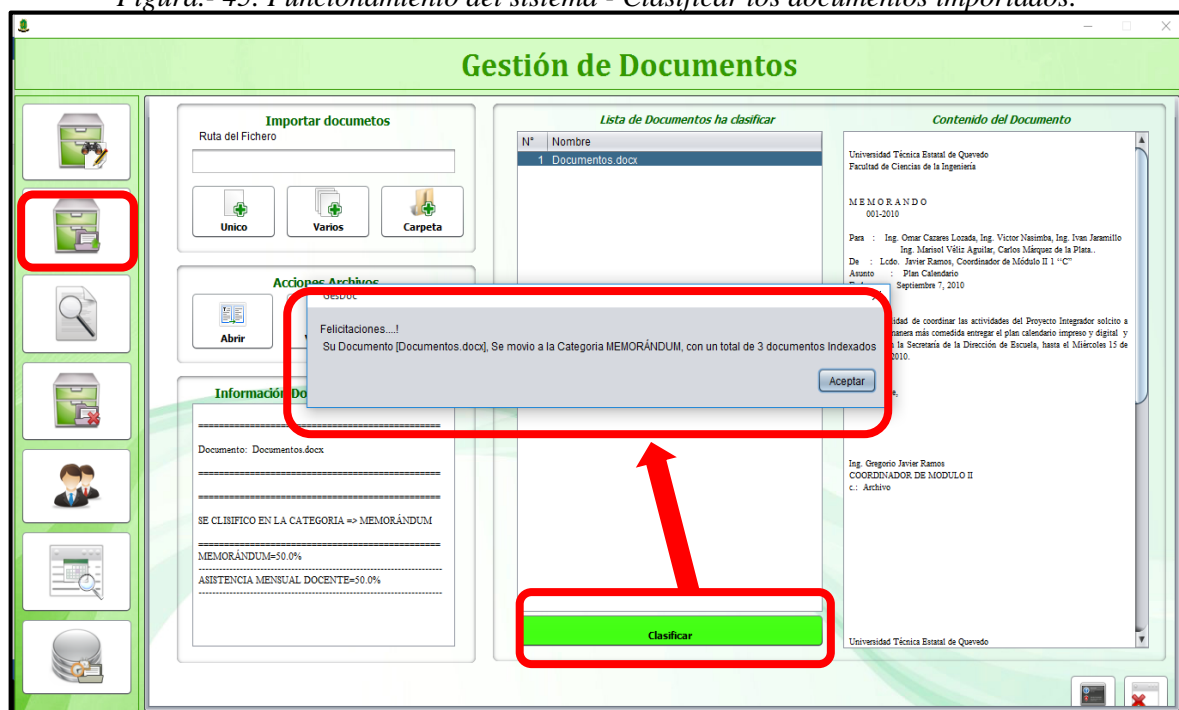
Figura.- 44. Funcionamiento del sistema - Operaciones con los archivos importados.



- **Clasificar los documentos.**

Cuando todos los archivos para clasificar han sido importados al sistema, se realiza su respectiva clasificación automática, dando click en el botón “Clasificar”. El sistema nos mostrará un mensaje de confirmación mencionando que los documentos han sido clasificados a sus respectivas categorías y en la parte izquierda nos detalla a que categoría fue clasificado cada documento.

Figura.- 45. Funcionamiento del sistema - Clasificar los documentos importados.



MÓDULO BUSQUEDA DE DOCUMENTOS

En este módulo se realiza la búsqueda de todos los documentos que han sido clasificados, por medio de criterios específicos que requiera el usuario.

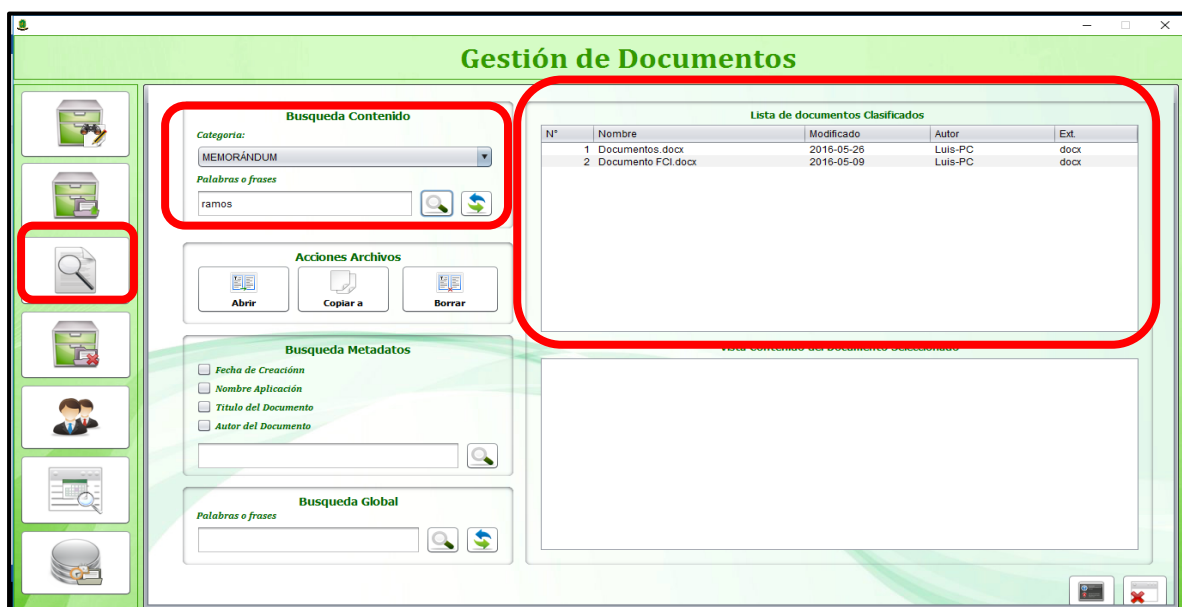
Figura.- 46. Funcionamiento del sistema - Búsqueda de documentos.



El sistema posee diferentes opciones de búsqueda de archivos, los cuales tienen un objetivo en común “encontrar los documentos que requiera el usuario “entre las opciones se tiene a:

- **Búsqueda por categoría:** El usuario selecciona una categoría, escribe el criterio de búsqueda y hace clic en el comando buscar (icono de lupa), inmediatamente en la parte derecha nos muestra una lista con todos los documentos encontrados que contienen el criterio de búsqueda proporcionado.

Figura.- 47. Funcionamiento del sistema - Búsqueda por categoría.

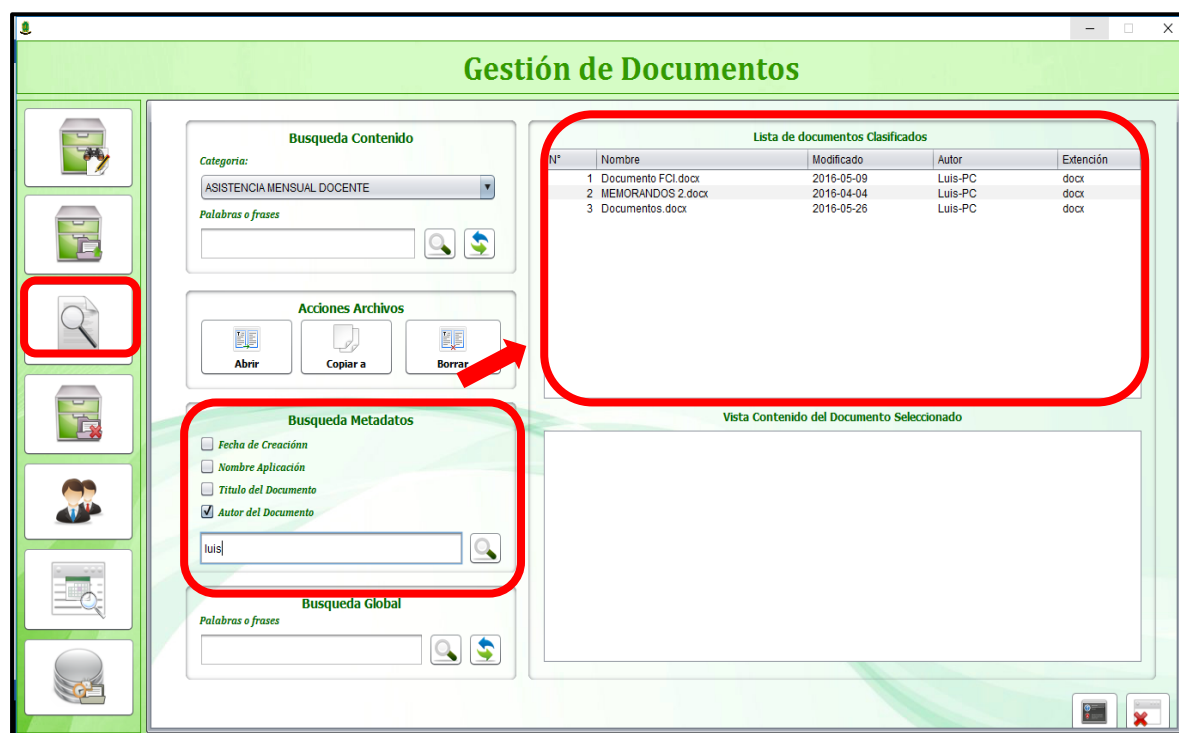


- **Búsqueda por metadatos:** La búsqueda de documentos por metadatos consiste en localizar los archivos por medio de las propiedades de los archivos, entre los cuales se tiene en el sistema los siguientes atributos:

- **Fecha de creación.**
- **Nombre de la aplicación.**
- **Título del documento.**
- **Autor del documento.**

Se selecciona un metadato, ingresa el criterio respectivo y hace clic en buscar (ícono de lupa), inmediatamente nos muestra en la parte derecha la lista de los archivos encontrados.

Figura.- 48. Funcionamiento del sistema - Búsqueda por metadatos.



- **Búsqueda Global:** La búsqueda global consiste en hacer una extensiva localización de los archivos por todas las categorías respectivas, ingresando su criterio de búsqueda y dando clic en buscar (ícono de lupa), la aplicación ejecuta la búsqueda y muestra todos los resultados encontrados en la parte derecha del sistema.

Figura.- 49. Funcionamiento del sistema - Búsqueda Global.



Manipular los archivos.

Todos los archivos que han sido encontrados, se puede hacer abrir el documento completo copiar a una dirección elegida por el usuario o enviarlos a borradores en caso que no sea el adecuado, estas operaciones se la realizan a través de los siguientes comandos:

- **Abrir:** Abre el documento seleccionado de la lista de encontrados.
- **Copiar a:** Copia documento seleccionado a una dirección específica.
- **Borrar:** Envía los archivos a “Borradores de documentos”, en caso que el/los documentos, según la opinión del usuario no tiene relevancia alguna.

Figura.- 50.-Funcionamiento del sistema - Manipulación de archivos.



MÓDULO BORRADORES DE DOCUMENTOS

En este módulo se podrá apreciar todos los documentos que se enviaron a borradores en el módulo “Búsqueda de documentos”. Los archivos que se encuentran en borradores, se los puede restaurar a su localización original.

Figura.- 51. Funcionamiento del sistema – Borradores de documentos.



Manipular archivos.

Todos los archivos que han sido enviados a borradores, se puede hacer una vista previa del documento, abrir el documento completo y restaurarlos a su localización original, estas operaciones se la realizan a través de los siguientes comandos:

- **Abrir:** Abre el documento seleccionado de la lista de borradores.
- **Visualizar:** Muestra una vista previa del archivo seleccionado en la parte inferior de la aplicación.
- **Restaurar:** Restaura el documento seleccionado a su localización original en la carpeta de archivos.

Figura.- 53. Funcionamiento del sistema - Manipular archivos .

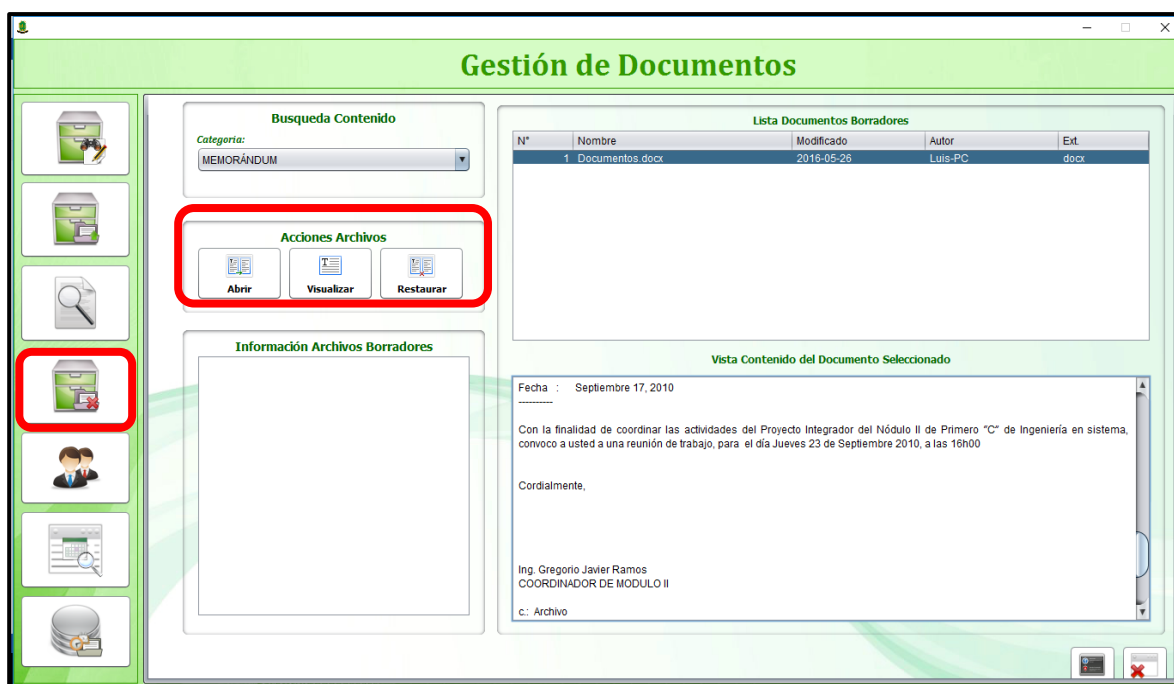
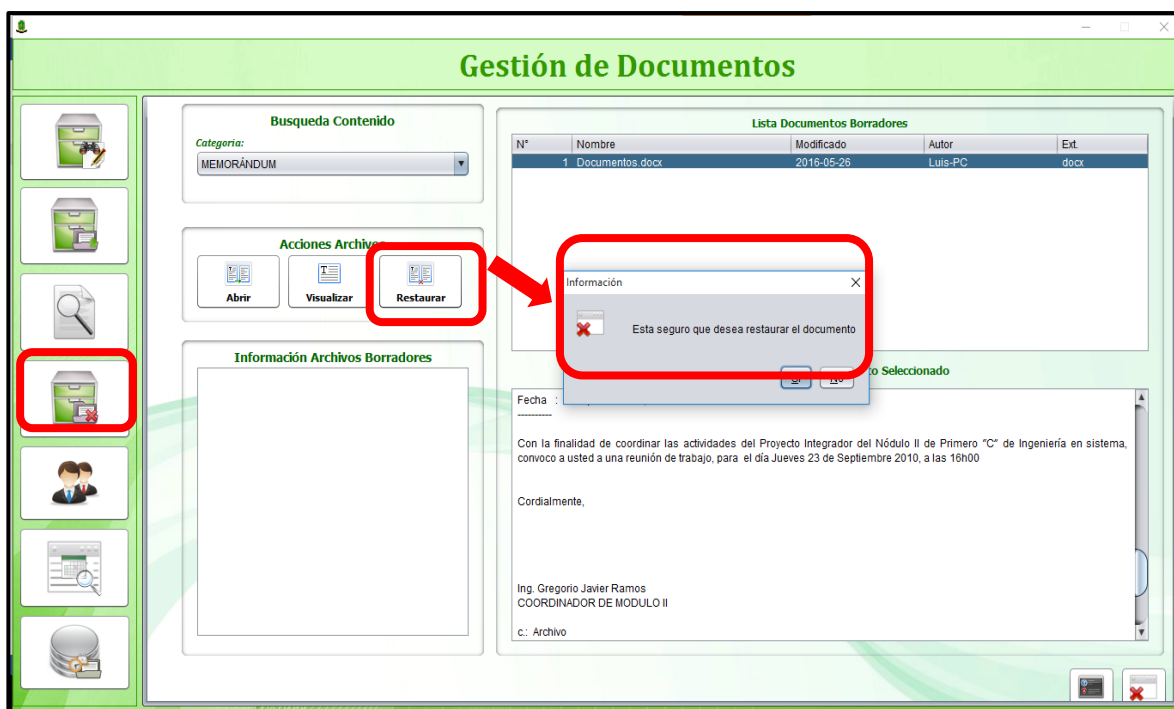


Figura.- 52. Funcionamiento del sistema - Restaurar archivo.



MÓDULO ADMINISTRADOR DE USUARIOS

Para acceder al sistema el administrador debe agregar usuarios (invitados), es en este módulo que se realiza la creación, edición, eliminación de usuarios y dar permisos a las diferentes funcionalidades que dispone el sistema. Para acceder a este módulo se debe dar doble clic sobre la opción “Administrador de usuarios”.

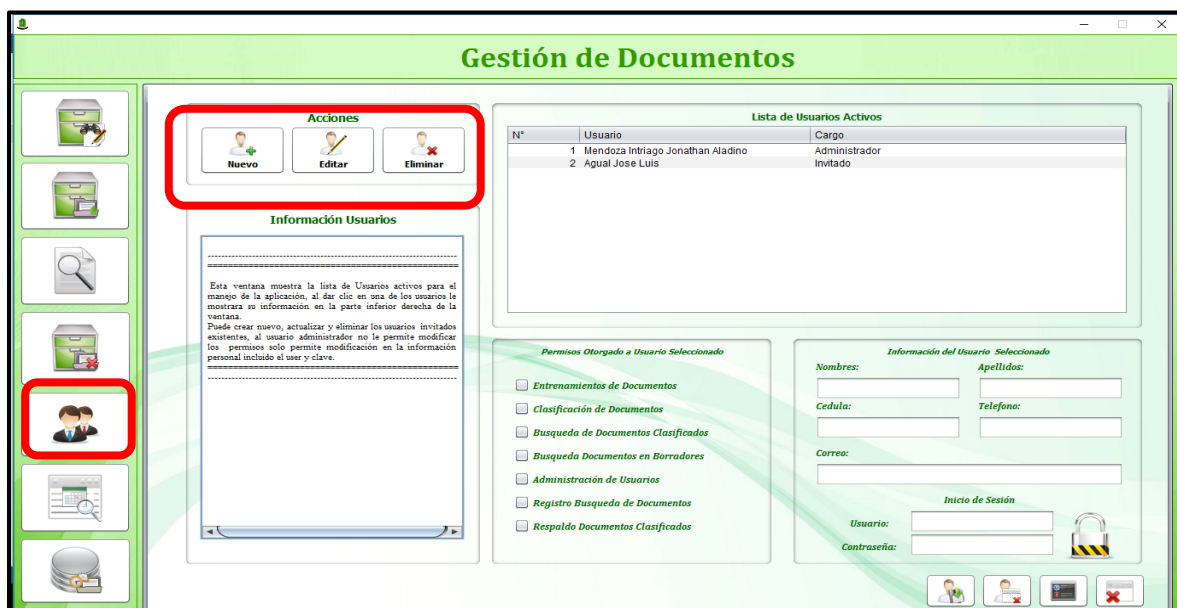
Figura.- 54. Funcionamiento del sistema -Administrador de usuarios.



En la pantalla nos mostrará las diferentes opciones:

- **Nuevo:** Permite crear un nuevo usuario y asignar los permisos adecuados para la manipulación del sistema.
- **Editar:** Permite editar los datos del usuario al mismo tiempo que puede asignar o revocar los permisos respectivos.
- **Eliminar:** Permite la eliminación de un usuario del sistema.

Figura.- 55. Funcionamiento del sistema - Opciones de Administrador de usuarios.



• Nuevo Usuario

El usuario (administrador) debe presionar el botón “Nuevo”, para crear un nuevo usuario Invitado para la manipulación del sistema, procediendo a llenar la información básica del usuario y asignándole los permisos respectivos. Completado los campos requeridos el usuario (administrador) debe guardar toda la información dando clic en el botón “Guardar”.

Figura.- 56. Funcionamiento del sistema - Crear nuevo usuario (invitado.)

Permisos Otorgado a Usuario Seleccionado

- ☐ Entrenamientos de Documentos
- ☐ Clasificación de Documentos
- ☒ Busqueda de Documentos Clasificados
- ☐ Busqueda Documentos en Borradores
- ☐ Administración de Usuarios
- ☐ Registro Busqueda de Documentos
- ☒ Respaldo Documentos Clasificados

Información del Usuario Seleccionado

Nombres: Jose Luis Apellidos: Aguilera

Cedula: 1201194938 Telefono: 0997309970

Correo: pepelucho@hotmail.com

Inicio de Sesión

Usuario: luis Contraseña: ****

Guardar

• Editar Usuario

En esta sección de administración de usuarios, para editar los datos de un usuario (invitado), el administrador debe seleccionar un usuario creado y dar clic sobre el botón “Editar”, el cual nos preguntará si “Desea modificar el usuario”.

Figura.- 57. Funcionamiento del sistema - Editar usuario.

Acciones

Nuevo Editar Eliminar

Lista de Usuarios Activos

N°	Usuario	Cargo
1	Mendoza Intriago Jonathan Aladino	Administrador
2	Aguilera Jose Luis	Invitado

Información

Desea modificar el Usuario [Aguilera Jose Luis]

Sí No

Al aceptar modificar los datos del usuario, el sistema procede a cargar la información del usuario y los permisos asignados, a continuación, el administrador puede actualizar la información que desee, una vez realizado los cambios necesarios procede a guardar los datos

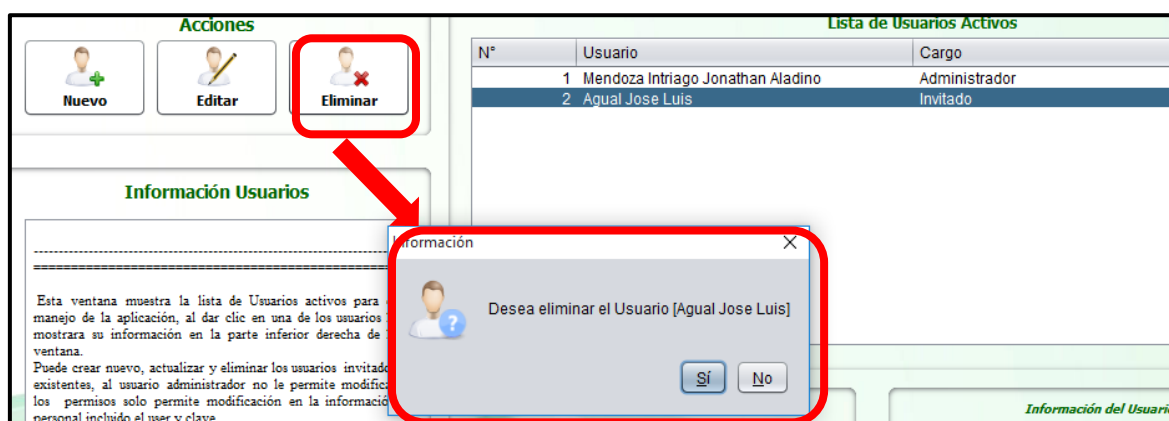
dando clic en el botón “Guardar”.

The screenshot displays a user management interface with two main sections. The left section, titled "Permisos Otorgado a Usuario Seleccionado", contains a list of permissions with checkboxes: "Entrenamientos de Documentos" (unchecked), "Clasificación de Documentos" (checked), "Busqueda de Documentos Clasificados" (checked), "Busqueda Documentos en Borradores" (unchecked), "Administración de Usuarios" (unchecked), "Registro Busqueda de Documentos" (unchecked), and "Respaldo Documentos Clasificados" (checked). The right section, titled "Información del Usuario Seleccionado", contains fields for "Nombres" (Jose Luis), "Apellidos" (Aguil), "Cedula" (1201194938), "Telefono" (0997309970), and "Correo" (pepelucho@hotmail.com). Below these fields is a login section titled "Inicio de Sesión" with fields for "Usuario" (luis) and "Contraseña" (****). At the bottom right, there are four icons: a user icon (highlighted with a red box), a document icon, a lock icon, and a delete icon.

• Eliminar Usuario

Para eliminar un usuario del sistema, se debe seleccionar un usuario de la lista de registrados y dar click sobre el botón “Eliminar”, el cual mostrará un mensaje preguntado “Desea eliminar el usuario”, al confirmar el mensaje se eliminará el usuario seleccionado.

Figura.- 58. Funcionamiento del sistema - Eliminar usuario.



MÓDULO VISTA REGISTRO DE BUSQUEDA

Este módulo nos ayuda a buscar la información que se han realizado en el módulo “Búsqueda de documentos”, a través del historial de búsqueda, es decir encontrar los documentos que ya han sido encontrados, seleccionando las fechas en los que se realizó la búsqueda.

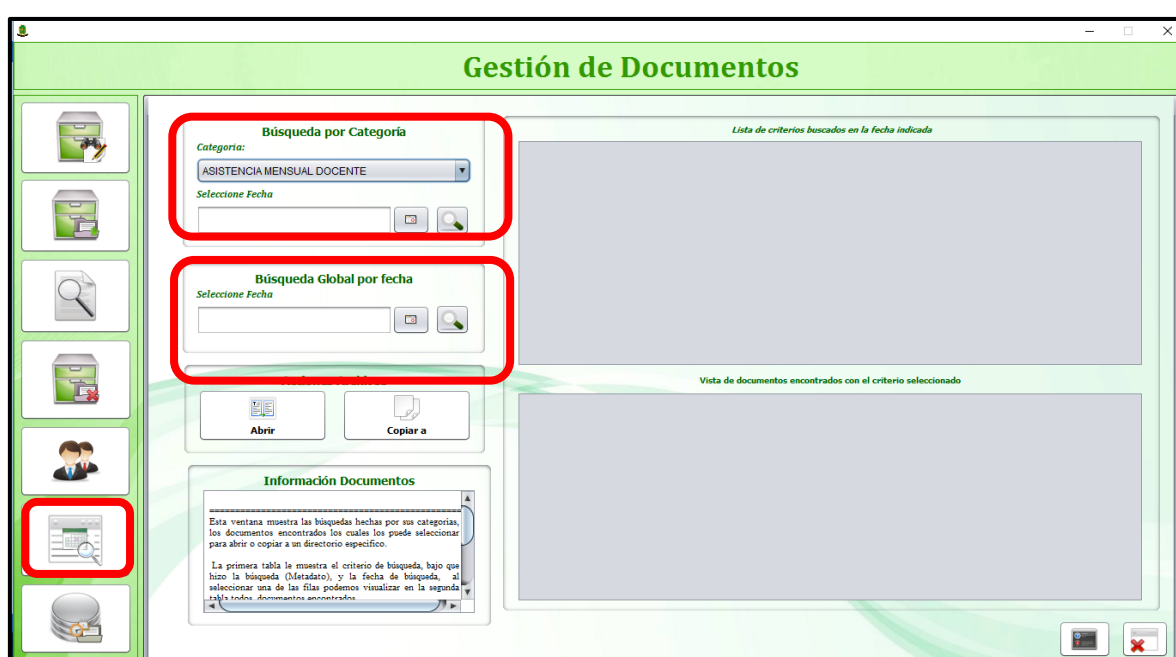
Figura.- 59. Funcionamiento del sistema - Vista registro de búsquedas.



El formulario nos mostrará dos opciones para realizar la búsqueda a través del historial, las cuales son:

- **Búsqueda por Categoría:** Nos permite encontrar por categorías los documentos que han sido encontrados con anterioridad.
- **Búsqueda global por fecha:** Realiza una búsqueda por todo el historial, seleccionando una fecha específica.

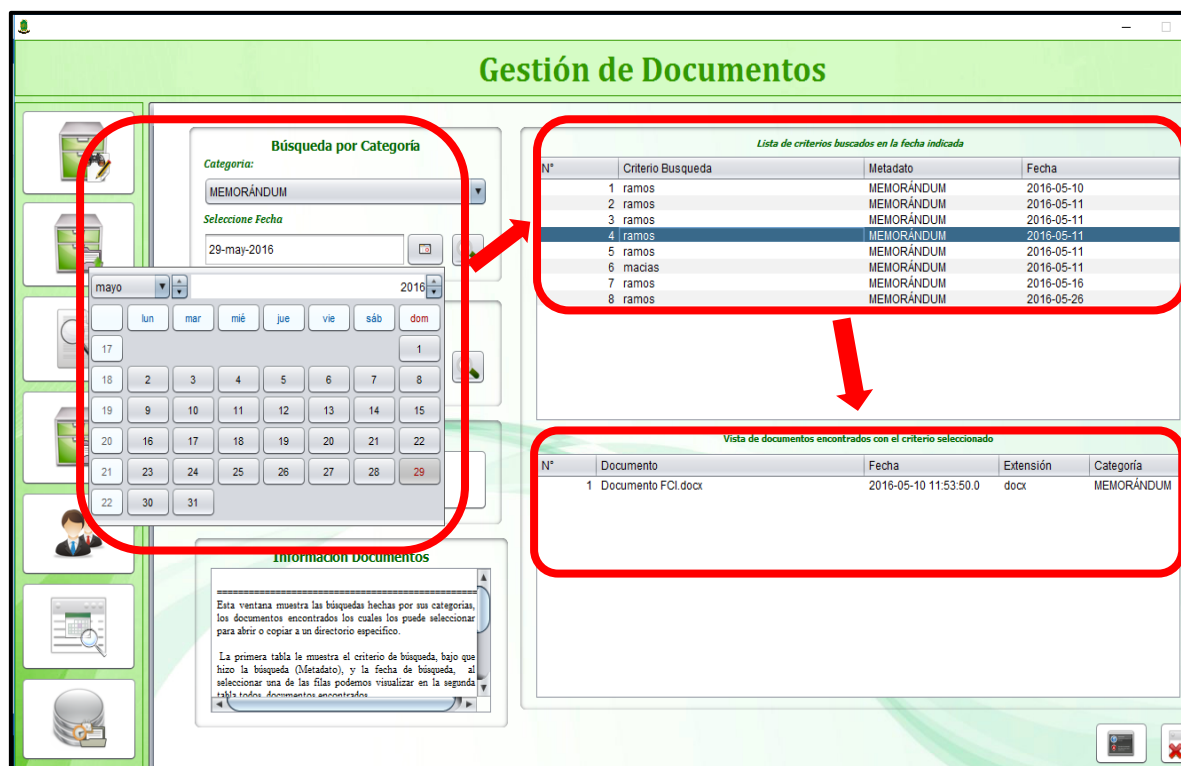
Figura.- 60. Funcionamiento del sistema - Formulario Vista registro de búsquedas.



• Búsqueda por Categoría

En esta sección nos permite encontrar documentos por categorías por medio del historial, seccionando una fecha en la que se haya realizado la búsqueda y presionando el botón buscar (ícono de lupa). Los resultados nos muestran en la parte superior derecha, una lista de criterios con los que se ha realizado la búsqueda en la fecha indicada y en la parte inferior derecha los documentos que se han encontrado con ese criterio de búsqueda.

Figura.- 61. Funcionamiento del sistema –Búsqueda por categoría.

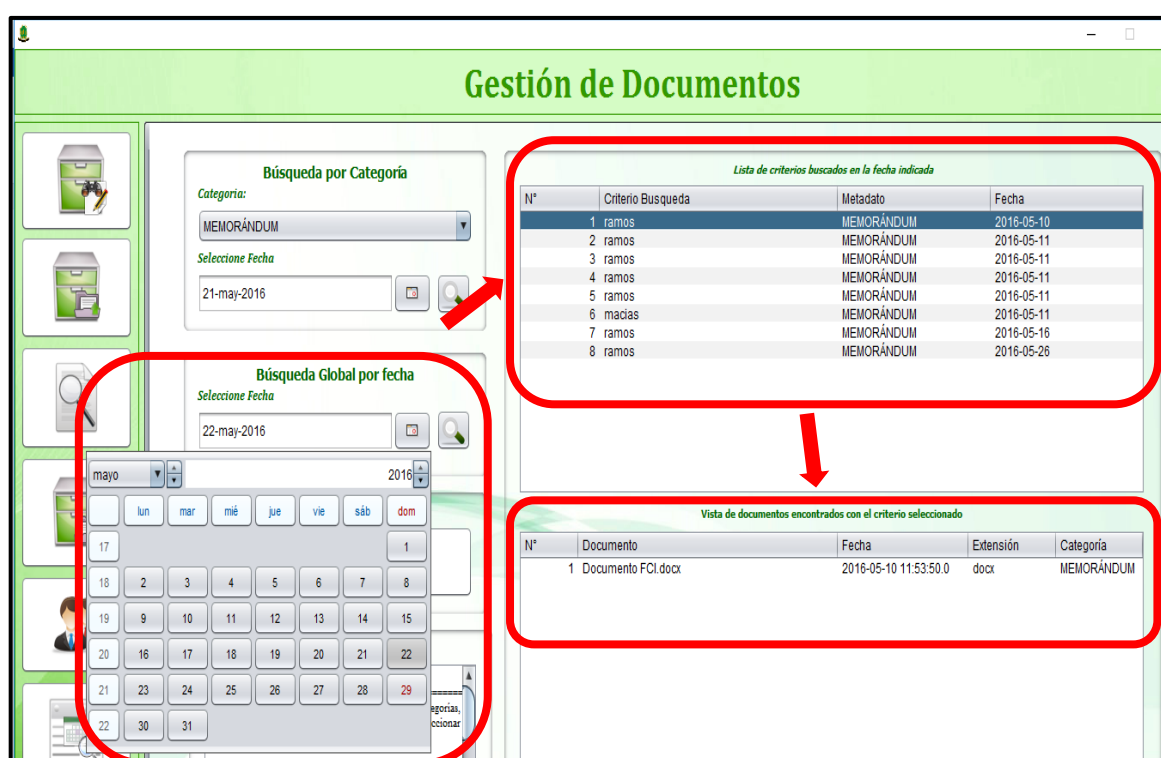


• Búsqueda global por fecha

La búsqueda global de los documentos ya encontrados a través del historial, se selecciona una fecha en la que se hayan realizado unas búsquedas y presionando el botón buscar (ícono de lupa).

Los resultados nos presentan en la parte superior derecha, una lista de criterios con los que se ha realizado la búsqueda en la fecha indicada y en la parte inferior derecha los documentos que se han encontrado con ese criterio de búsqueda.

Figura.- 62. Funcionamiento del sistema - Búsqueda global por fecha..



Acciones de archivos

Los archivos encontrados se los puede manipular con las siguientes acciones:

- **Abrir:** Nos permite abrir el documento seleccionado.
- **Copiar a:** Nos permite copiar el archivo seleccionado a un lugar específico.

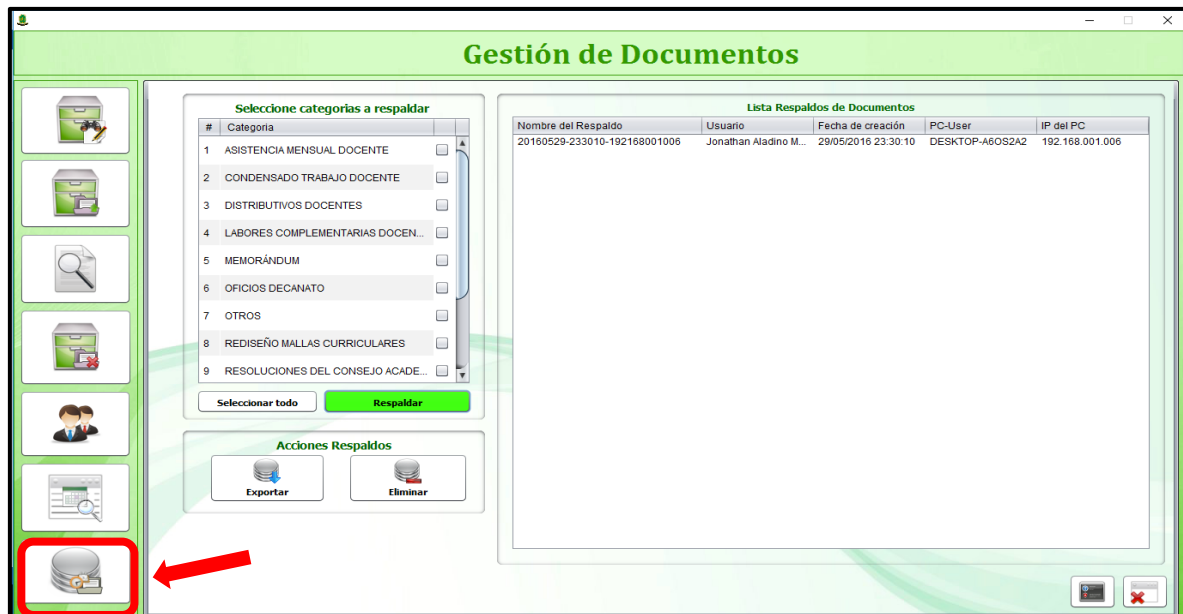
Figura.- 63. Funcionamiento del sistema – Manipulación de los archivos.



MÓDULO RESPALDO DE INFORMACIÓN

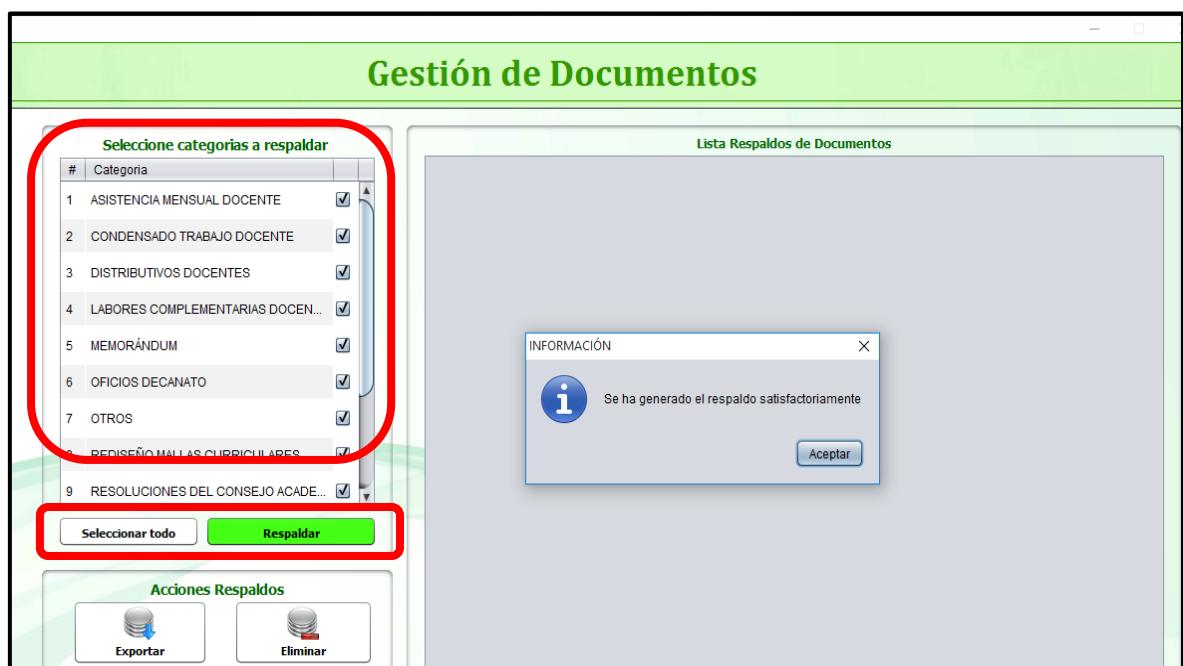
Este módulo permite realizar un respaldo de todos los documentos que se haya importado y clasificado en el sistema.

Figura.- 64. Funcionamiento del sistema - Respaldo de información.



El proceso de realizar un respaldo de información es sencillo, se puede seleccionar una por una las categorías que requiera respaldar o seleccionar todas mediante el comando “*Seleccionar todo*” y a continuación presionar el botón “*Respaldo*”, el cual nos presentará un mensaje de confirmación que “se ha generado el respaldo satisfactoriamente”.

Figura.- 65. Funcionamiento del sistema - Generar respaldo.



La lista de respaldos generados se puede visualizar en la parte derecha, mostrándonos los siguientes atributos:

- **Nombre del respaldo:** Nombre del respaldo generado.
- **Usuario:** Nombre del usuario que generó el respaldo.
- **Fecha de Creación:** Fecha en la que se creó el respaldo.
- **PC-User:** Nombre de la máquina desde que se realizó el respaldo.
- **IP del PC:** La dirección IP de la máquina en la que se realizó el respaldo.

Figura.- 66. Funcionamiento del sistema - Lista de respaldos generados.

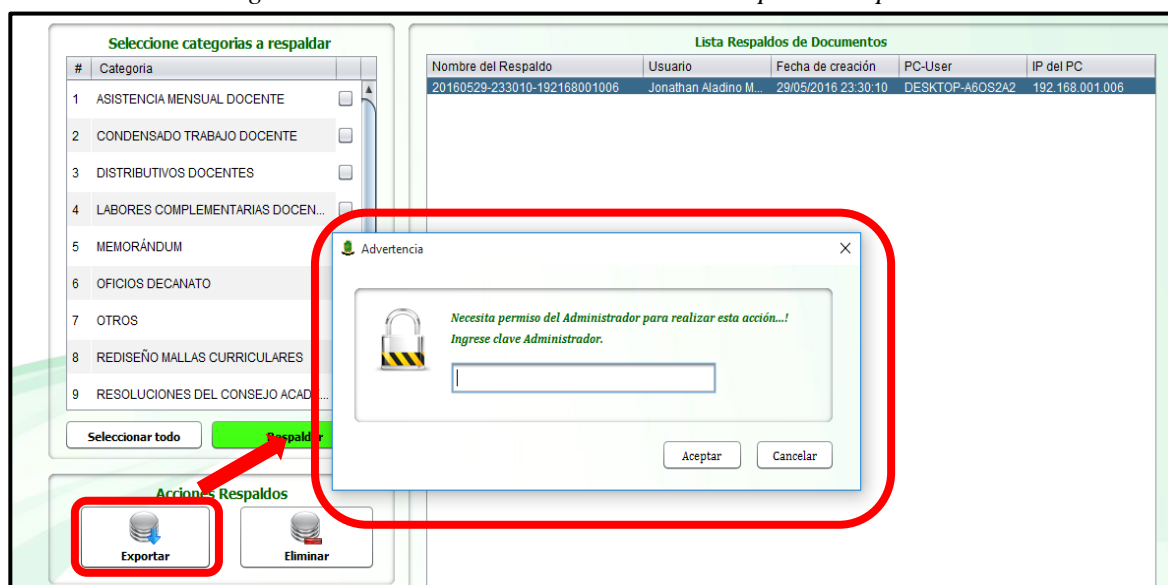
Lista Respaldos de Documentos				
Nombre del Respaldo	Usuario	Fecha de creación	PC-User	IP del PC
20160529-233010-192168001006	Jonathan Aladino M...	29/05/2016 23:30:10	DESKTOP-A6OS2A2	192.168.001.006

Acciones de los respaldos

Los respaldos que han sido generados se los puede manipular con las siguientes opciones:

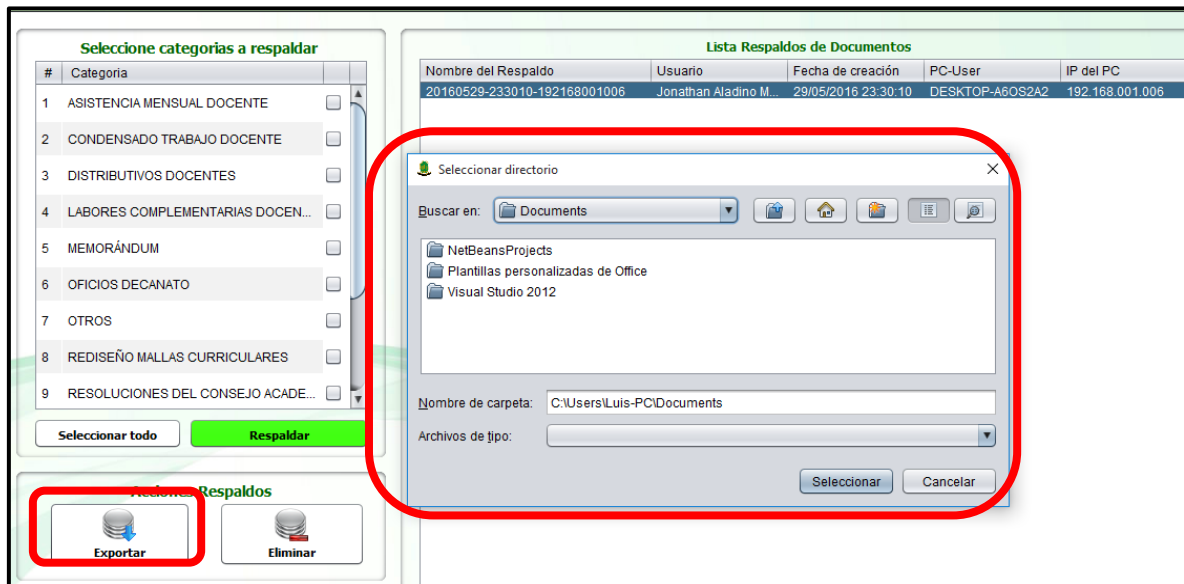
- **Exportar:** Permite exportar un respaldo generado aun lugar específico, dando clic sobre el comando “Exportar” nos presentará un formulario donde solicita la clave del administrador.

Figura.- 67. Funcionamiento del sistema - Exportar respaldo.



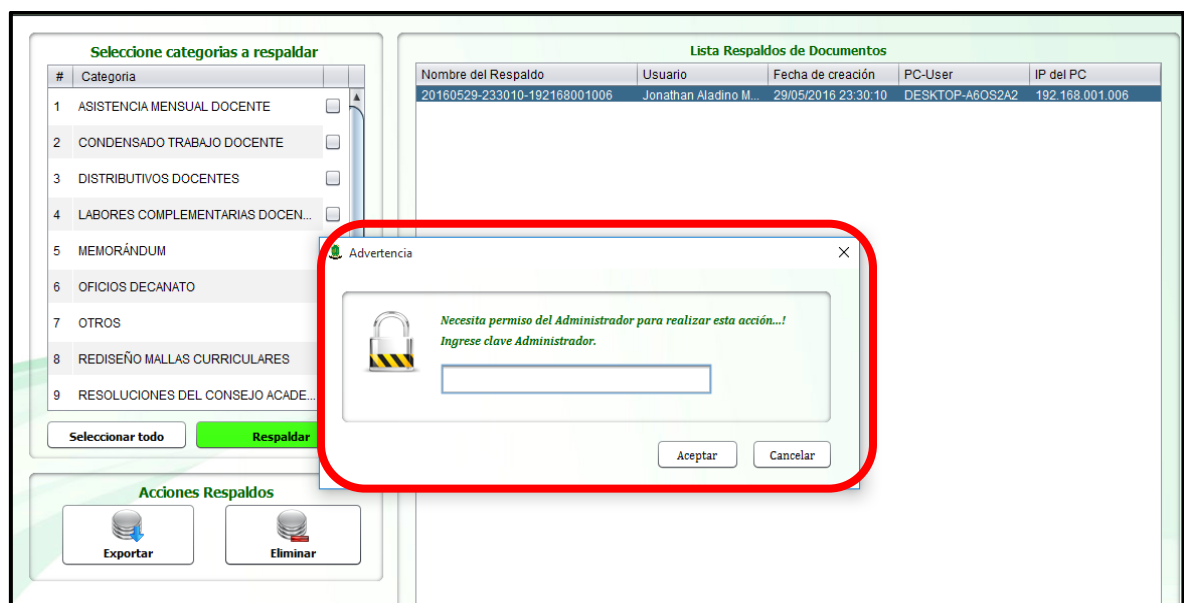
Una vez que el sistema comprueba la clave del administrador, se debe elegir el lugar donde desee exportar el archivo generado y presionar el botón “seleccionar”, inmediatamente se guardará el respaldo en la dirección elegida mostrando un mensaje de confirmación.

Figura.- 68. Funcionamiento del sistema. Elegir dirección donde exportar el respaldo.



- **Eliminar respaldo:** Para eliminar un respaldo generado, primero se debe seleccionar un respaldo y presionar el botón “eliminar”, el sistema presentara un formulario donde solicita la clave del administrador.

Figura.- 69. Funcionamiento del sistema - Eliminar respaldo.



Una vez que el sistema comprueba la clave del administrador, se elimina el respaldo seleccionado mostrando un mensaje de confirmación.

Figura.- 70. Situación actual de la Gestión de documentos de la FCI.

Grabar Nueva carpeta			
Nombre	Fecha de modifica...	Tipo	Tamaño
2016-03-03	03/03/2016 17:53	Carpeta de archivos	
2016-03-04 SÍLABO DE OFIMÁTICA	04/03/2016 8:35	Carpeta de archivos	
2016-03-07 sílabo de ofimatica	07/03/2016 18:06	Carpeta de archivos	
ACTAS	19/04/2016 11:52	Carpeta de archivos	
CANOA	08/05/2013 10:16	Carpeta de archivos	
COMISIÓN ACADÉMICA	01/03/2016 17:44	Carpeta de archivos	
DIAPPOSITIVAS	25/02/2016 17:57	Carpeta de archivos	
EXCEL FCI	30/03/2016 8:26	Carpeta de archivos	
Hildita	10/02/2016 19:03	Carpeta de archivos	
INFORMES	17/03/2016 18:20	Carpeta de archivos	
LA YAPA_files	09/07/2015 17:27	Carpeta de archivos	
musica	19/01/2016 13:28	Carpeta de archivos	
Musicas	03/03/2015 9:00	Carpeta de archivos	
PDF	02/03/2016 7:45	Carpeta de archivos	
PORTAL DE FRUTAS ANTIOXIDANTES_files	29/01/2016 12:14	Carpeta de archivos	
RESOLUCIONES ENTREGADAS	24/03/2016 9:37	Carpeta de archivos	
RESOLUCIONES	19/04/2016 11:36	Carpeta de archivos	
RESOLUCIONES VIRGINIA	19/04/2016 11:51	Carpeta de archivos	

Figura.- 71. Organización de los documentos mediante el sistema desarrollado.

Nombre	Fecha de modifica...	Tipo	Tamaño
ASISTENCIA MENSUAL DOCENTE	25/05/2016 16:25	Carpeta de archivos	
CONDENSADO TRABAJO DOCENTE	25/05/2016 16:25	Carpeta de archivos	
DISTRIBUTIVOS DOCENTES	25/05/2016 16:25	Carpeta de archivos	
LABORES COMPLEMENTARIAS DOCENTES	25/05/2016 16:25	Carpeta de archivos	
MEMORÁNDUM	26/05/2016 3:20	Carpeta de archivos	
OFICIOS DECANATO	25/05/2016 16:25	Carpeta de archivos	
OTROS	25/05/2016 16:25	Carpeta de archivos	
REDISEÑO MALLAS CURRICULARES	25/05/2016 16:25	Carpeta de archivos	
RESOLUCIONES DEL CONSEJO ACADEMI...	25/05/2016 16:25	Carpeta de archivos	
SOLICITUD DE ESTANCIAS	25/05/2016 16:25	Carpeta de archivos	
SOLICITUD DE PERMISO	25/05/2016 16:25	Carpeta de archivos	
SOLICITUD VIATICOS	25/05/2016 16:25	Carpeta de archivos	
TITULOS GRADUADOS	25/05/2016 16:25	Carpeta de archivos	

Figura.- 72. Solicitud de permiso para proporcionarnos la documentación de la Secretaría de la FCI

Quevedo 26 de febrero de 2016

Señor (a)

Ing. Jorge Murillo Oviedo


DECANO DE LA FACULTAD DE CIENCIAS DE LA INGENIERIA

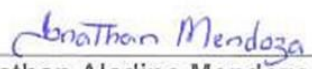
De mi consideración:


Por medio de la presente, los egresados de la Carrera de Ingeniera en Sistemas: **José Luis Agual Espín y Jonathan Aladino Mendoza Intriago**, nos dirigimos a usted para solicitarle de la manera más comedida documentos digitales de la FCI los cuales están a cargo de la Lcda. Virginia Zambrano Menéndez, los mismos que serán utilizados para el proceso de Titulación Especial - Modalidad Proyecto de Investigación como pruebas y evidencia para el debido proceso de titulación.

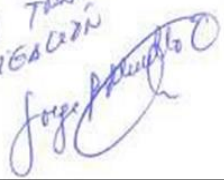
En espera de poder ser atendido lo más pronto posible, quedo de usted:

Atentamente,


José Luis Agual Espín
CI: 120600085-1


Jonathan Aladino Mendoza Intriago
CI: 131119493-8



*Autorizado
Sra. Virginia Zambrano
para el apoyo necesario
para el trabajo de
Investigación*


26 FEB 2016
