



UNIVERSIDAD CENTRAL DE VENEZUELA

Facultad de Ciencias

Escuela de Computación

Centro de Ingeniería de Software y Sistemas (ISYS)

Laboratorio de Inteligencia Artificial

Opción Profesional: Inteligencia Artificial



Minería de textos para la asignación automática de jurados a Trabajos Especiales de Grado

Trabajo Especial de Grado presentado ante la Ilustre

Universidad Central de Venezuela

Por la Bachiller

Dayana Raquel Torres Navarro

C.I. V-18.708.915

Para optar al título de Licenciado en Computación

Tutores:

Prof.^a Esmeralda Ramos

Prof.^a Haydemar Nuñez

Caracas, julio de 2011

Universidad Central de Venezuela
Facultad de Ciencias
Escuela de Computación



ACTA DEL VEREDICTO

Quienes suscriben, miembros del jurado designado por el Consejo de la Escuela de Computación, para dictaminar sobre el Trabajo Especial de Grado titulado: "**Minería de textos para la asignación automática de jurados a Trabajos Especiales de Grado**" y presentado por la bachiller *Dayana Raquel Torres Navarro*, Cédula de Identidad V-18.708.915, para optar al título de Licenciado en Computación, dejan constancia de lo siguiente:

Leído como fue, dicho trabajo por cada uno de los miembros del jurado, se fijó el día 14 de julio de 2011 a las 04:00 p.m., para que su autora lo defendiera en forma pública, lo que hizo en el aula PA-I de la Escuela de Computación, mediante una presentación oral del contenido del Trabajo Especial de Grado, luego de lo cual respondió a las preguntas formuladas. Finalizada la defensa pública del Trabajo Especial de Grado, el jurado decidió **APROBARLO** con una calificación de **20** puntos.

En fe de lo cual se levanta la presente Acta, en la Ciudad Universitaria de Caracas a los catorce días del mes de julio del año dos mil once, dejándose también constancia de que actuó como Coordinador del Jurado la Profesora Esmeralda Ramos.

Jurado Principal

Prof.^a Esmeralda Ramos
(Tutora)

Prof.^a Haydemar Nuñez
(Tutora)

Prof. Rhadamés Carmona
(Jurado)

Prof. Marcel Castro
(Jurado)

Prof. Otilio Rojas
(Suplente)

Prof. Eugenio Scalise
(Suplente)

MINERÍA DE TEXTOS PARA LA ASIGNACIÓN AUTOMÁTICA DE JURADOS A TRABAJOS ESPECIALES DE GRADO

RESUMEN

Este trabajo tiene como objetivo el desarrollo de un prototipo basado en minería de textos para la asignación automática de jurados a Trabajos Especiales de Grado (TEG), a través de la clasificación de los documentos dentro de las áreas profesionales de la Escuela de Computación. Para la creación del clasificador se aplicó el proceso de minería de textos sobre una recopilación de documentos de TEG, lo cuales fueron preprocesados, eliminando signos de puntuación, palabras no informativas y aplicando *word stemming*, y representados mediante el modelo de espacio vectorial con diferentes técnicas de indexación. Asimismo, se probaron diferentes medidas de relevancia para la reducción de la dimensionalidad del espacio de atributos. Para la estimación del modelo de clasificación se realizaron pruebas aplicando el algoritmo k vecinos más cercanos con diferentes parámetros, alcanzando un 82% de clasificaciones correctas. A través de la recomendación de hasta dos áreas, se logró incrementar el porcentaje de clasificaciones correctas a 95%. Este módulo de recomendación de áreas se integró con el módulo de asignación, con el fin de realizar la asignación de los jurados, considerando a los profesores pertenecientes al o las áreas recomendadas, en base a criterios de elección previamente definidos. En conclusión, fue posible obtener un buen rendimiento de clasificación de documentos de TEG aplicando minería de textos y se logró automatizar el proceso de asignación de jurados a TEG, lo que a su vez permite llevar un registro histórico de las asignaciones, contribuyendo además con la automatización de procesos de la Escuela de Computación.

Palabras claves:

Minería de textos, modelos de clasificación, K vecinos más cercanos, algoritmo de Porter *Stemming*, Weka.

Autor:

Dayana Raquel Torres Navarro.
dayana.torres@ciens.ucv.ve

Tutores:

Prof.^a Esmeralda Ramos.
esmeralda.ramos@ciens.ucv.ve

Prof.^a Haydemar Nuñez.
haydemar.nunez@ciens.ucv.ve

Fecha: julio 2011

DEDICATORIA

A mis amados padres,
mi adorada hermana,
mi querida Universidad Central de Venezuela,
mi segunda casa, la Facultad de Ciencias,
y mi Escuela de Computación.

AGRADECIMIENTOS

Agradezco a Dios por permitirme ser todo lo que soy hoy en día, por siempre estar presente en mi vida y la de mi familia, cuidándonos y guiándonos, por nunca desampararnos y por darnos salud y fuerzas para seguir adelante y lograr nuestras metas.

Muchas gracias a mis padres y a mi hermana, que son la razón de mi existencia y a quienes dedico todos mis logros, por su enorme colaboración, sus consejos, su preocupación, su interés, su cariño y amor, por darme ánimos y fuerzas para alcanzar todas mis metas y ser quienes siempre han estado conmigo apoyándome en todo. ¡Los amo!

A mi amado novio, César Cortés, le agradezco por toda su comprensión, apoyo, ayuda, cariño y consejos brindados, por no solo estar en los momentos de alegría y disfrute sino también en aquellos más duros y de mayor sacrificio, por siempre estar pendiente y preocupado por mi y por los muchos gratos y hermosos momentos compartidos en la segunda casa de los dos, nuestra Facultad de Ciencias.

A mi Cuñis lendo Edsel Hernández y su linda familia, por estar pendiente y ofrecer su ayuda y colaboración, como siempre. Se les quiere.

A mis tutoras Esmeralda Ramos y Haydemar Nuñez les doy las gracias por guiarme a lo largo de este trabajo, por el tiempo y esfuerzo invertido en ello, y por sus enseñanzas y consejos. En especial, le agradezco a Esmeralda por su cariño de larga data y espero sepas que es totalmente recíproco hacia tí y tu familia.

Igualmente, gracias a todas las personas que de alguna manera colaboraron en la realización de este trabajo: mis jurados, Marcel Castro y Rhadamés Carmona, los profesores Sergio Rivas, Concettina Di Vasta, María Elena Villapol, David Pérez, Karima Velásquez y Daniel Villavicencio, la directora y el personal de la Biblioteca Alonso Gamero, y a todas las personas que estuvieron pendientes de mi durante este proceso.

Por último, pero no menos importante, agradezco a mis profesores y prepas de la Facultad por los conocimientos y enseñanzas transmitidas y, por supuesto, a las bonitas amistades que surgieron o continuaron a lo largo de mi carrera por todo su cariño, en especial: mi mejor amiga Adri, mis panitas Piolito, Hector, Fran, Chris (Cags), mi amorcita Shirly, mi mallikis Sheila, mi Chinita hermosa, Michellita, mi hermanito Jon, Cesarcño huesitos, Anita, Roni, Grey, Nevi, mi tocapita, Joe, Ant, Jon U., Nacho, Dani G., Luis S., Cyborg, Carlitos Izquierdo, Pascual, Arcadio, Malú, Adriana, Drau, el primo Jesús, Oscarcito, El Fanton, Deivid, Jon Chino, Rocío, Yaina, Tomy, Zulma, Massi, Musta (Q.E.P.D.), MB (Chu), la prepa, los gemelitos, las gemelitas, Gabito Vidal, Yisus, Kata, Ary, y a aquellos que me quedo sin nombrar, a todos los llevaré en mi memoria y en mi corazón y les deseo lo mejor.

TABLA DE CONTENIDO

RESUMEN	III
ÍNDICE DE TABLAS.....	IX
ÍNDICE DE FIGURAS.....	X
ÍNDICE DE ANEXOS	XI
INTRODUCCIÓN	1
CAPÍTULO 1. MARCO TEÓRICO	3
1.1 CONCEPTOS BÁSICOS DE LA MINERÍA DE DATOS	3
1.1.1. Tareas de minería de datos	3
1.2 EL PROCESO DE MINERÍA DE DATOS	5
1.2.1. Preparación de los datos	6
1.2.2. Minería de datos	7
1.3 MINERÍA DE TEXTOS.....	10
1.3.1. Proceso de minería de textos.....	12
1.4 APLICACIONES Y TRABAJOS PREVIOS	19
CAPÍTULO 2. MARCO APLICATIVO	21
2.1 PLANTEAMIENTO DEL PROBLEMA	21
2.2 PROPUESTA DE SOLUCIÓN	22
2.3 OBJETIVOS	23
2.3.1. Objetivo General.....	23
2.3.2. Objetivos Específicos	23
2.4 PROCESO DE MINERÍA DE TEXTOS	23
2.4.1. Recolección de los datos	24
2.4.2. Preparación de los textos	25
2.4.3. Estimación y evaluación del modelo de clasificación.....	30
2.5 DESARROLLO DEL PROTOTIPO	36
2.5.1. Módulo de recomendación de áreas.....	36
2.5.2. Módulo de asignación de jurados.....	38
2.5.3. Módulo de consulta	39

2.5.4. Módulo de mantenimiento	39
2.5.5. Tecnologías utilizadas	40
2.5.6. Diseño de la interfaz.....	42
2.5.7. Pruebas y resultados	54
CONCLUSIONES Y RECOMENDACIONES	58
REFERENCIAS.....	61
ANEXOS.....	65

ÍNDICE DE TABLAS

Tabla 1.	Principales propuestas de pesado para la indexación de términos.	14
Tabla 2.	Principales funciones propuestas para medir la relevancia de los términos.....	15
Tabla 3.	Cantidad de documentos recopilados para las áreas profesionales.	25
Tabla 4.	Diccionario de palabras no informativas del español.....	26
Tabla 5.	Selección final de atributos para los tres tipos de indexación aplicados.....	29
Tabla 6.	Resultados del clasificador k-NN con la selección por entropía.	31
Tabla 7.	Resultados del clasificador k-NN con la selección por pesado ltc.	32
Tabla 8.	Resultados del clasificador k-NN con la selección por pesado tfc.	33
Tabla 9.	Mejor resultado de k-NN para la selección de la indexación con pesado por entropía.....	35
Tabla 10.	Actores que participan en el prototipo.	44
Tabla 11.	Detalle de los casos de uso del nivel 1.....	45
Tabla 12.	Detalle de los casos de uso del nivel 2.....	48
Tabla 13.	Distribución de los documentos de prueba por áreas.	55
Tabla 14.	Resultados de las pruebas realizadas al prototipo.	55
Tabla 15.	Comparación de los resultados obtenidos al estimar el modelo y los alcanzados por el prototipo.	56
Tabla 16.	Estado de los profesores de las áreas recomendadas para el TEG de prueba Nro. 12.....	56
Tabla 17.	Jurados asignados al TEG de prueba Nro. 12.....	56

ÍNDICE DE FIGURAS

Figura 1.	Pasos del proceso de MD.	6
Figura 2.	Tareas del procesamiento de textos.	10
Figura 3.	Matriz de decisión para la clasificación de textos.....	11
Figura 4.	Matriz de representación de los documentos mediante indexación de términos.....	13
Figura 5.	Conducta de clasificación de los clasificadores como k-NN.	18
Figura 6.	Prototipo de sistema de asignación de jurados a TEG propuesto.	22
Figura 7.	Matriz de representación de los documentos recopilados.	27
Figura 8.	Medidas para la evaluación del clasificador.	34
Figura 9.	Módulo de recomendación de áreas del prototipo.....	37
Figura 10.	Estructura del patrón de diseño MVC.....	43
Figura 11.	Casos de uso del prototipo - Nivel 0.....	44
Figura 12.	Casos de uso del prototipo - Nivel 1.....	45
Figura 13.	Casos de uso del prototipo - Nivel 2.....	47
Figura 14.	Interfaz del módulo de recomendación de áreas – Paso 1.....	51
Figura 15.	Interfaz del módulo de recomendación de áreas – Paso 2.....	52
Figura 16.	Interfaz del módulo de asignación de jurados – Paso 3.....	52
Figura 17.	Interfaz del módulo de consulta – Paso 1.....	53
Figura 18.	Interfaz del módulo de consulta – Paso 2.....	54

ÍNDICE DE ANEXOS

Anexo 1.	Descripción del formato de archivos “.arff” (García, D.).....	65
Anexo 2.	Subgrupos de atributos obtenidos por cada algoritmo de medición de relevancia de términos para el archivo de indexación mediante pesado por entropía.....	67
Anexo 3.	Subgrupos de atributos obtenidos por cada algoritmo de medición de relevancia de términos para el archivo de indexación mediante pesado ltc.....	69
Anexo 4.	Subgrupos de atributos obtenidos por cada algoritmo de medición de relevancia de términos para el archivo de indexación mediante pesado tfc.	71
Anexo 5.	Reglas de elección de profesores como jurados de TEG.	73
Anexo 6.	Resultados de las pruebas realizadas al módulo de recomendación de áreas.	78
Anexo 7.	Resultados de las pruebas realizadas al módulo de asignación de jurados.	89

INTRODUCCIÓN

La revolución digital ha hecho que la captura de datos sea sencilla y que su almacenamiento tenga un coste prácticamente nulo. Sin embargo, los datos por sí solos no producen beneficio directo, su verdadero valor radica en la posibilidad de extraer información útil para la toma de decisiones o para comprender el fenómeno que los produjo. Tradicionalmente, este análisis de datos se hacía mediante un proceso manual o semiautomático en el que expertos en el dominio aplicaban técnicas estadísticas para proporcionar resúmenes, generar informes o validar modelos. No obstante, este proceso es irrealizable conforme aumenta la cantidad de datos.

Dada esta situación, hoy en día se está aplicando el análisis inteligente de grandes cantidades de datos, el cual permite procesarlos automáticamente, identificar patrones presentes en los mismos y presentarlos como conocimiento útil, que es lo que se conoce como proceso de minería de datos (MD). Asimismo, puesto que los datos generalmente se encuentran inmersos en grandes cantidades de textos, surge la necesidad de aplicar este proceso de MD a colecciones de documentos no estructurados, a lo que se le denomina minería de textos (MT).

Actualmente, la Escuela de Computación de la UCV realiza la asignación de jurados a Trabajos Especiales de Grado (TEG) siguiendo un proceso subjetivo y discrecional en el que se deciden los jurados más convenientes de acuerdo al área profesional del mismo, lo que requiere una inversión de tiempo considerable e información actualizada sobre los jurados asignados en el semestre. Como solución a este problema, se propone la automatización de este proceso aplicando MT para desarrollar un prototipo de sistema de recomendación que permita clasificar los documentos de acuerdo a las diversas áreas y seleccionar los jurados más pertinentes, según reglas preestablecidas.

A continuación, se describe la estructura general de este documento:

- **Capítulo 1:** contiene una síntesis de la revisión bibliográfica que se llevó a cabo para la realización del presente trabajo. Se da una visión general del proceso de

MD y se presenta una adaptación del mismo a la MT, orientado a la tarea de clasificación de textos. Además, se detallan algunas de las aplicaciones de la MD y se señalan trabajos conseguidos en la literatura basados en técnicas de MT.

- **Capítulo 2:** se expone el problema a resolver, la solución y los objetivos propuestos que se pretenden alcanzar. Igualmente, se detalla el proceso de MT aplicado para estimar el modelo de clasificación y los resultados de su evaluación. También, se señalan los aspectos relacionados con el desarrollo de la aplicación: el módulo de recomendación, de asignación y de mantenimiento. Se da una breve descripción de las tecnologías utilizadas y las consideraciones de diseño, y se muestran los resultados obtenidos.
- Por último, se presentan las conclusiones a las que se llegaron y las recomendaciones para trabajos futuros.

CAPÍTULO 1. MARCO TEÓRICO

1.1 Conceptos básicos de la minería de datos

El proceso global de descubrimiento de conocimiento útil a partir de datos ha adquirido diferentes denominaciones, entre las cuales se encuentran: extracción de conocimiento, descubrimiento de conocimiento en bases de datos o KDD (*Knowledge Discovery in Databases*) y, de manera genérica, minería de datos (MD), aunque ésta última es considerada un paso dentro del mismo, el correspondiente a la extracción de conocimiento. La MD se define como el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y comprensibles a partir de grandes volúmenes de datos (Fayyad, Piatetsky-Shapiro, y Smyth, 1996). El término proceso implica que comprende un conjunto bien definido de pasos, y no trivial significa que involucra alguna búsqueda o inferencia.

Por otro lado, se entiende por patrón una expresión en algún lenguaje que describe a un subgrupo de los datos o un modelo aplicable al subgrupo. La validez de los patrones se refiere a su correcta aplicación a nuevos datos; la novedad se mide por su aporte de información desconocida; la utilidad va ligada a la obtención de conocimiento a partir del mismo aplicable en algún dominio; y la comprensión se relaciona con la facilidad de interpretación y de utilización. Extraer un patrón consiste en adecuar un modelo a los datos, hallar una estructura a partir de éstos, o, en general, hacer cualquier descripción de alto nivel de un grupo de datos, los cuales, por sí solos, constituyen tan solo una colección de hechos (Fayyad et al., 1996).

1.1.1. Tareas de minería de datos

La aplicación de algoritmos de MD puede perseguir dos tipos de tareas:

a. Tareas Predictivas

Son aquellas en las que el sistema halla patrones para predecir valores desconocidos o futuros de otras variables de interés, basado en los datos disponibles. Las principales tareas predictivas son:

a.1. Clasificación

Consiste en el aprendizaje de una función que mapea o clasifica un elemento de datos dentro de una de varias clases predefinidas, representadas por el valor del atributo de clase. Los atributos restantes, que sean significativos para determinar la clase, son utilizados por las técnicas de clasificación para generar modelos que permitan determinar la clase de un ejemplo a partir de los valores de sus atributos significativos. Ejemplos: clasificación de plantas en géneros a partir de sus características morfológicas, análisis de riesgo en préstamos, etc.

a.2. Regresión

Consiste en el aprendizaje de una función que mapea un elemento de datos a una variable de predicción de valor real. Es un caso particular de la tarea de clasificación, cuando el rango de la función es real. Ejemplos: estimar la probabilidad de que un paciente sobreviva dados los resultados de un grupo de pruebas diagnósticas, previsión de ventas, etc.

b. Tareas Descriptivas

Son aquellas en las que el sistema halla patrones humanamente comprensibles que describen los datos para presentarlos al usuario. Las principales tareas descriptivas son:

b.1. Agrupamiento

Busca identificar un conjunto finito de categorías o grupos para describir los datos, basándose en el principio de maximizar la similitud entre elementos de un grupo y minimizar la similitud entre elementos de grupos distintos. Ejemplo: descubrir subpoblaciones homogéneas de consumidores en bases de datos de *marketing*.

b.2. Asociación

Consiste en encontrar un modelo que describa dependencias significativas entre variables. Ejemplo: análisis de la cesta de compra, en el que se pueden encontrar asociaciones como que el 95% de las veces que se compran pañales también se compra leche.

1.2 El proceso de minería de datos

Anand y Brachman (1996) presentan un enfoque práctico y detallado del proceso de MD que enfatiza su naturaleza interactiva, resumiéndolo en los siguientes pasos:

1. **Comprensión del dominio e identificación del objetivo:** se analiza el conocimiento previo sobre el dominio de la aplicación y se especifica el objetivo del proceso desde el punto de vista del usuario.
2. **Selección del conjunto de datos destino** (selección horizontal): consiste en enfocarse en un subgrupo de datos de muestra sobre los cuales va a ser realizado el descubrimiento.
3. **Preprocesamiento de datos:** las operaciones básicas incluyen recolectar la información necesaria para el modelo, eliminar el ruido (limpieza), seleccionar los atributos más significativos (selección vertical), decidir sobre las estrategias para el manejo de campos de datos ausentes, entre otros.
4. **Transformación de los datos:** se pueden realizar transformaciones en los datos en función del objetivo del proceso.
5. **Selección de una tarea particular de MD:** dicha selección se realiza de acuerdo al objetivo del proceso (paso 1).
6. **Selección del modelo e hipótesis:** se decide qué modelos y parámetros podrían ser apropiados y se hace coincidir un algoritmo particular de MD con el criterio general del proceso.
7. **Minería de datos:** consiste en la aplicación de algoritmos de descubrimiento sobre los datos, para la extracción de patrones que permitan descubrir automáticamente conocimiento, generalmente orientado hacia la toma de decisiones. El espacio de patrones a menudo es infinito, por lo que su enumeración involucra algunas formas de búsqueda en dicho espacio.
8. **Interpretación y evaluación de los patrones minados:** consiste en la visualización de los patrones y modelos extraídos o en la visualización de los datos dados los modelos extraídos. Dependiendo de los resultados obtenidos luego de las evaluaciones, existe la posibilidad de regresar a cualquiera de los pasos (del 1 al 7) para más refinamiento.
9. **Aplicación del conocimiento descubierto:** utilizar el conocimiento, ya sea incorporándolo en otro sistema o simplemente documentándolo y reportándolo a

las partes interesadas. Este paso también incluye la revisión y resolución de conflictos potenciales con el conocimiento extraído.

La Figura 1 ilustra el flujo de los pasos básicos del proceso de MD (aquellos que reciben entradas y generan salidas).

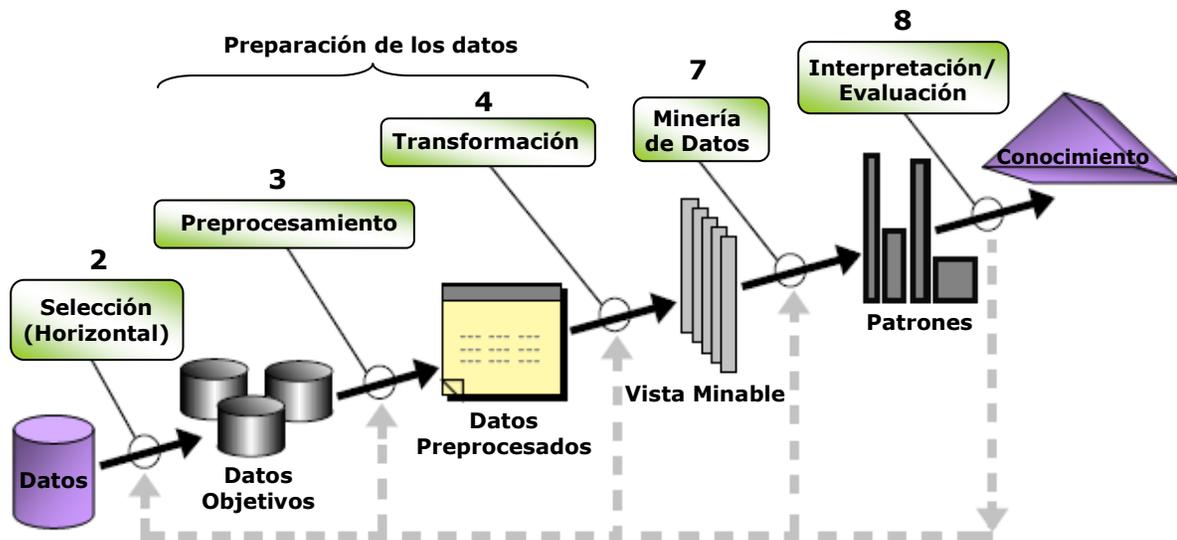


Figura 1. Pasos del proceso de MD.

Nota. Fuente: Adaptado de Anand y Brachman (1996).

Dos aspectos relevantes y significativos del proceso de MD lo constituyen el preprocesamiento y la transformación de los datos, ya que preparan los datos para la minería de datos. A continuación se describen estas tareas:

1.2.1. Preparación de los datos

La preparación de los datos engloba los pasos de preprocesamiento y transformación de los mismos y representa aproximadamente el 70% del proceso de MD, razón por la cual se profundizará en este aspecto. Comprende lo siguiente (Bogado, Dapozo, López, y Porcel, 2007):

- **Limpieza de los datos:** consiste en detectar y corregir errores en los datos, como: campos redundantes u obsoletos, valores inconsistentes, ruido, valores ausentes, valores anómalos (*outliers*). Estos errores pueden tratarse: ignorando o eliminando el dato (los registros o los atributos), creando un nuevo atributo lógico

que indique si el valor de la variable original era redundante/inconsistente/nulo/anómalo, o sustituyéndolo (imputación) por una constante indicada por los expertos, un valor medio, el más parecido, una estimación del valor, un valor generado, entre otros.

- **Selección de atributos:** se determinan los atributos más relevantes en función del objetivo del proceso. Para esto se pueden emplear métodos embebidos, en los que la selección de atributos ocurre como parte del algoritmo de MD, por filtros, donde la selección se realiza previamente a la aplicación del algoritmo mediante algún enfoque independiente de la tarea, o por modelos, en los cuales un algoritmo de MD estima un modelo para la selección.
- **Transformaciones:** cambiar el rango de las variables (normalización o estandarización), su tipo (numerización y discretización) o un atributo o conjunto de atributos en otros.
- **Construcción de nuevos atributos:** construcción de nuevas características a partir de las originales. Es muy dependiente del dominio, por lo que requiere conocimiento sobre el mismo.

1.2.2. Minería de datos

a. Componentes de los algoritmos de MD

Se identifican tres componentes principales en la construcción de algoritmos que implementen las tareas de MD (Fayyad et al., 1996):

- a.1. Modelo de representación:** es el lenguaje empleado para describir los patrones a descubrir. Si la representación es muy limitada ningún entrenamiento producirá un modelo certero de los datos, pero un mayor poder representacional incrementa el peligro de sobreajustar los datos de entrenamiento, reduciendo la exactitud de predicción en datos no examinados.
- a.2. Modelo de evaluación:** los modelos descriptivos pueden ser evaluados en cuanto a su novedad, utilidad y comprensión. Los modelos predictivos son juzgados por la exactitud de predicción empleando métricas como la precisión y el *recall*. La precisión se define como la proporción entre la cantidad de

instancias predichas correctamente para un subgrupo y la cantidad total de instancias predichas para el mismo, mientras que el *recall* es la proporción entre la cantidad de instancias predichas correctamente para un subgrupo y la cantidad total de instancias que pertenecen al mismo en la colección. Existen diversas técnicas para la evaluación de un modelo, entre las que se encuentran las siguientes (De San Pedro y Lasso, 2006):

- Utilizar todo el conjunto de datos para entrenamiento y prueba. Problemas: sobreajuste o subajuste.
- Separar los datos en un conjunto de entrenamiento y un conjunto de prueba. Permite detectar el sobreajuste cuando se obtienen resultados mucho mejores para el conjunto de entrenamiento que para el de prueba, pero los resultados son muy dependientes de la partición. No es favorable si hay escasez de datos.
- Realizar validación cruzada, que consiste en dividir los datos en k particiones conjuntas (*k-fold cross validation*). La elección del valor k dependerá del tamaño y características de la muestra, aunque se suele usar el valor 10. En cada iteración se construirá y evaluará un modelo, utilizando uno de los conjuntos para el entrenamiento y el resto para prueba. El error estimado es el promedio de los errores en cada iteración. Involucra gran costo computacional (se debe inducir el modelo k veces), por lo que no es factible para conjuntos de datos muy grandes.

a.3. La búsqueda: una vez que se fija el modelo de representación y el criterio de evaluación, el problema de MD se reduce simplemente a una tarea de optimización: hallar los parámetros y modelos que optimicen dicho criterio de evaluación.

b. Métodos de minería de datos

Entre los principales métodos de MD destacan los siguientes (Fayyad et al., 1996):

b.1. Regresión no lineal

Son algoritmos que se caracterizan por ajustar combinaciones lineales y no lineales de funciones base (*sigmoides, splines, polinomios*) a combinaciones de las

variables de entrada. Tienen un gran poder representacional, pero pueden ser difíciles de interpretar. Ejemplos: redes neuronales, métodos de spline adaptativo y regresión por búsqueda de proyección (PPR).

b.2. Métodos basados en ejemplos

Utilizan ejemplos representativos del problema a resolver para aproximar un modelo, es decir, las predicciones sobre nuevos ejemplos se derivan a partir de las propiedades de ejemplos similares en el modelo cuya predicción es conocida. Suelen ser poderosos en cuanto a sus propiedades de aproximación, sin embargo pueden ser difíciles de interpretar ya que el modelo está implícito en los datos (no se formula explícitamente).

Entre las técnicas se encuentran los sistemas de razonamiento basado en casos y la técnica de k vecinos más cercanos. En éstos últimos, la clase de cualquier punto nuevo en el espacio de dos dimensiones es la misma que la del punto más cercano en el conjunto de datos de entrenamiento original. Los parámetros a estimar pueden ser el número K de vecinos a utilizar para la predicción y la medida de distancia en sí.

b.3. Árboles de decisión

Los árboles de decisión representan reglas en las que atributos independientes determinan los valores finales. Cada nodo representa una propiedad que puede tomar diversos valores, cada uno de los cuales genera una rama, y los nodos hojas representan las clasificaciones finales. Son útiles en problemas de alta dimensionalidad y pequeño número de valores para cada atributo.

b.4. Gráfico probabilístico y modelos de dependencia

Los modelos gráficos especifican dependencias probabilísticas utilizando una estructura de gráfico. En su forma más simple, el modelo define qué variables son directamente dependientes la una de la otra. Son de particular interés para el proceso de MD, ya que la forma gráfica del modelo lo hace sencillo para la interpretación humana.

1.3 Minería de textos

Cuando se realiza MD, la información se obtiene normalmente de bases de datos, en las cuales ésta se encuentra de manera estructurada. Sin embargo, a pesar de la proliferación de las bases de datos en las últimas décadas, en la actualidad la mayor parte de la información (más de un 80%) se encuentra almacenada en forma de texto. Es por ello que surge la necesidad de poder realizar el proceso de MD sobre grandes cantidades de información textual.

La minería de textos (MT) es el proceso de descubrimiento de patrones interesantes y nuevos conocimientos a partir de una colección de textos, más específicamente, es el proceso encargado del descubrimiento de conocimiento que no existía explícitamente en ningún texto de la colección, pero que surge de relacionar el contenido de varios de ellos. Su objetivo es descubrir tendencias, desviaciones y asociaciones en una gran cantidad de información textual (Montes, 2001).

Como se aprecia en la Figura 2, la MT representa una de las tareas más complejas dentro del procesamiento automático de textos, siendo éste, hoy en día, el área de investigación más importante de la lingüística computacional. Dentro de las tareas de MT, este trabajo se centra en la clasificación de textos.

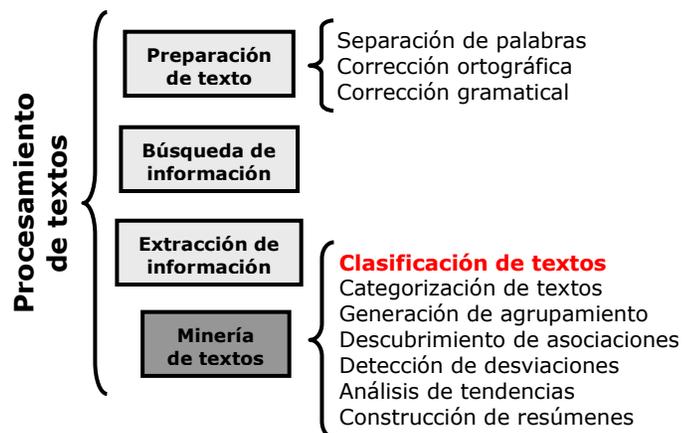


Figura 2. Tareas del procesamiento de textos.

Nota. Fuente: Adaptado de Montes (2001).

Una propuesta para la tarea de clasificación de textos que ha cobrado importancia es la máquina de aprendizaje, la cual plantea la construcción automática de un clasificador para una categoría c_i , mediante la observación de las características de un grupo de documentos que previamente han sido clasificados bajo c_i de forma manual por un experto en el dominio; a partir de dicha observación un proceso inductivo deriva las características que debería tener un documento nuevo para ser clasificado bajo c_i . Con base a esta propuesta, la clasificación de textos se puede definir como la tarea de determinar la asignación de un valor $a_{ij} \in \{0,1\}$ para cada entrada de una matriz de decisión (ver Figura 3), donde $C = \{c_1, \dots, c_k\}$ representa el conjunto de categorías predefinidas, $D = \{d_1, \dots, d_N\}$ el conjunto de documentos a ser clasificados y el valor a_{ij} la decisión de clasificar d_j en la categoría c_i , de manera que si $a_{ij} = 1$ entonces el elemento d_j es clasificado en c_i ; en otro caso, $a_{ij} = 0$.

	d_1	...	d_j	...	d_N
c_1	a_{11}	...	a_{1j}	...	a_{1N}
...
c_i	a_{i1}	...	a_{ij}	...	a_{iN}
...
c_k	a_{k1}	...	a_{kj}	...	a_{kN}

Figura 3. Matriz de decisión para la clasificación de textos.

Nota. Fuente: Adaptado de Sebastiani (1999).

Las categorías sólo son etiquetas simbólicas, no se asume ningún conocimiento adicional sobre su significado como ayuda en el proceso de construcción del clasificador. De igual modo, la asignación de documentos se realiza en base al contenido del documento y no en base a metadatos (por ejemplo, fecha y fuente de publicación).

Una ventaja de esta propuesta es que si se actualiza el grupo original de categorías C o si el sistema se traslada a otro dominio diferente, todo lo que se necesita es la construcción automática inductiva de un nuevo clasificador a partir de un

grupo diferente de documentos clasificados manualmente. En términos de efectividad, hoy en día los clasificadores construidos por medio de técnicas de máquinas de aprendizaje alcanzan niveles de rendimiento bastante aceptables.

1.3.1. Proceso de minería de textos

A continuación, se presenta una síntesis del proceso que se sigue al realizar MT, enfocado en la tarea de clasificación de textos bajo la propuesta de máquinas de aprendizaje, el cual se adapta al proceso de MD, dado que la MT es un caso particular de ésta (Ass y Eikvil, 1999).

a. Preparación de los textos

Un texto es difícil y fácil de manejar a la misma vez; difícil porque los conceptos abstractos que contiene el texto son complejos de representar en una computadora, y fácil porque todos los textos tienen una gran cantidad de datos redundantes. Durante la preparación, los textos, que generalmente son cadenas de caracteres, se transforman a algún tipo de representación estructurada o semi-estructurada que facilite su análisis posterior. Dicha representación depende de la estrategia empleada, así como el tipo de patrones descubiertos viene dado por el tipo de representación. A continuación, se describen las etapas consideradas en la preparación de los textos:

a.1. Preprocesamiento

Generalmente, el preprocesamiento del texto consiste en:

- Remover las palabras frecuentes que no aportan información (Ej.: pronombres, preposiciones, conjunciones, etc.).
- Reducir las palabras a su raíz eliminando prefijos y sufijos (lematización o *stemming*). Esta reducción se lleva a cabo en grupos de palabras que tienen el mismo significado conceptual, como caminar, caminante, caminado y caminando.
- Si los documentos se encuentran codificados mediante algún lenguaje de marcado, remover las etiquetas (HTML u otras).

Como resultado del preprocesamiento se obtiene el conjunto de términos a utilizar: $T = \{t_1, \dots, t_M\}$.

a.2. Indexación

La representación de documentos más utilizada para la indexación, es el llamado *modelo de espacio vectorial*, en el cual los documentos son representados por vectores de términos. La transformación de los textos a dicha representación constituye una etapa adicional en el proceso de MT. Por lo general, la colección de documentos se representa mediante una matriz A (Figura 4) de términos por documento, donde a_{ij} es el peso del término j en el documento i . La matriz A es esparcida (la mayoría de los elementos son 0), ya que normalmente un término no aparece en todos los documentos (Ass y Eikvil, 1999).

	t_1	...	t_j	...	t_M
d_1	a_{11}	...	a_{1j}	...	a_{1M}
...
d_i	a_{i1}	...	a_{ij}	...	a_{iM}
...
d_N	a_{N1}	...	a_{Nj}	...	a_{NM}

Figura 4. Matriz de representación de los documentos mediante indexación de términos.

Nota. Fuente: Elaboración propia.

En la Tabla 1 se formulan diferentes propuestas para determinar el peso a_{ij} , presentadas por Ass y Eikvil (1999), sin embargo la mayoría se basa en dos observaciones empíricas en lo que al texto se refiere:

- Mientras más veces se produzca una palabra en un documento, ésta es más relevante al tema del documento.
- Mientras más veces se produzca una palabra en todos los documentos de la colección, ésta discrimina peor entre documentos.

Tabla 1. Principales propuestas de pesado para la indexación de términos.

Propuesta de pesado		Fórmula
Booleano	$a_{ij} = 1$ Si $f_{ij} > 0$	$a_{ij} = 0$ En otro caso
Frecuencia del término (tf)	$a_{ij} = f_{ij}$	
Frecuencia del término por frecuencia inversa del documento (tf×idf)	$a_{ij} = f_{ij} * \log(N/n_j)$	
Tfc	$a_{ij} = \frac{f_{ij} * \log(N/n_j)}{\sqrt{\sum_{j=1}^M [f_{ij} * \log(N/n_j)]^2}}$	
Ltc	$a_{ij} = \frac{\log(f_{ij}+1.0) * \log(N/n_j)}{\sqrt{\sum_{j=1}^M [\log(f_{ij}+1.0) * \log(N/n_j)]^2}}$	
Entropía	$a_{ij} = \log(f_{ij}+1.0) * \left(1 + \frac{1}{\log(N)} \sum_{i=1}^N [f_{ij}/n_j * \log(f_{ij}/n_j)] \right)$	
Terminología	f_{ij} = frecuencia del término j en el documento i. n_j = número total de veces que el término j aparece en la colección (D). N = cantidad de documentos en la colección (cardinalidad de D). M = cantidad de términos en la colección (cardinalidad de T).	

a.3. Reducción de la dimensionalidad

Esta etapa se corresponde con la selección de atributos en el proceso de MD. En MT la alta dimensionalidad del espacio de términos puede ser problemática, ya que existe una dimensión por cada término que se encuentra al menos una vez en el conjunto de documentos D, aproximándose este número M de términos, por lo general, a cientos de miles. Las técnicas de MD estándar no pueden lidiar con este gran grupo de características ya que el procesamiento es extremadamente costoso en términos computacionales y los resultados se vuelven poco fiables. Debido a esto, en MT se emplean técnicas de reducción de dimensionalidad (RD) que tienen por efecto la reducción de la dimensionalidad del espacio vectorial de M a $M' \ll M$.

La RD se enfoca de dos maneras, aunque esta distinción no tiene impacto en el tipo de técnica escogida (Sebastiani, 1999):

- **RD local:** para cada categoría c_i se seleccionan $M_i' \ll M$ características, mediante las cuales el clasificador operará para la categoría c_i .
- **RD global:** se seleccionan $M' \ll M$ características, mediante las cuales el clasificador operará para todas las categorías $C = \{c_1, \dots, c_k\}$.

De igual forma, se realiza una segunda distinción en base al tipo de características seleccionadas:

- **RD por selección de características**

Dado un $M' \ll M$ fijo, las técnicas para selección de características o reducción del espacio de términos (TSR, por sus siglas en inglés) tienen como objetivo seleccionar, a partir de la colección original de M características, los M' términos que cuando fueron utilizados para la indexación de documentos produjeron la disminución de efectividad más baja, con respecto a la obtenida empleando las representaciones verdaderas. Para la tarea de clasificación se realiza TSR global conservando los $M' \ll M$ términos que obtienen valores más altos de acuerdo a funciones numéricas que miden la "importancia" o relevancia de cada término en los documentos, como las formuladas en la Tabla 2.

Tabla 2. Principales funciones propuestas para medir la relevancia de los términos.

Función	Notación	Fórmula
Frecuencia en documento	n_j	$P(t_j, c_i)$
Ganancia de información	$IG(t_j, c_i)$	$P(t_j, c_i) \cdot \log \frac{P(t_j, c_i)}{P(c_i) \cdot P(t_j)} + P(\bar{t}_j, c_i) \cdot \log \frac{P(\bar{t}_j, c_i)}{P(c_i) \cdot P(\bar{t}_j)}$
Chi-cuadrado	$\chi^2(t_j, c_i)$	$\frac{N \cdot [P(t_j, c_i) \cdot P(\bar{t}_j, \bar{c}_i) - P(t_j, \bar{c}_i) \cdot P(\bar{t}_j, c_i)]^2}{P(t_j) \cdot P(\bar{t}_j) \cdot P(c_i) \cdot P(\bar{c}_i)}$
Coefficiente de correlación	$CC(t_j, c_i)$	$\frac{\sqrt{N} \cdot [P(t_j, c_i) \cdot P(\bar{t}_j, \bar{c}_i) - P(t_j, \bar{c}_i) \cdot P(\bar{t}_j, c_i)]}{\sqrt{P(t_j) \cdot P(\bar{t}_j) \cdot P(c_i) \cdot P(\bar{c}_i)}}$
$t_j =$ término j ; $c_i =$ categoría i ; $N =$ cantidad de documentos en la colección (cardinalidad de D)		

- **RD por extracción de características (re-parametrización)**

Las características seleccionadas no son un subgrupo de las M características originales sino que se sintetiza una colección de M' nuevas características que maximizan la efectividad obtenida, a partir de la colección original, las cuales se obtienen mediante combinaciones o transformaciones de las originales. Resulta útil utilizar características sintetizadas debido a los problemas de polisemia, homonimia y sinonimia dominantes en MT, los cuales pueden ocasionar que los términos no sean dimensiones óptimas para la representación del contenido del documento.

b. Minería de textos

El paso de MT para la tarea de clasificación de textos corresponde al proceso inductivo de construcción de un clasificador para una categoría $c_i \in C$, el cual consiste de las siguientes fases (Sebastiani, 1999):

1. La definición de una función $CSV_i: D \rightarrow [0,1]$ (por sus siglas en inglés, *Categorisation Status Value*) que, dado un documento d_j , retorna un valor de estado de clasificación para el mismo. Esta función toma diferentes significados de acuerdo a los clasificadores.
2. La definición de un umbral τ_i tal que $CSV_i(d_j) \geq \tau_i$ es interpretado como una decisión de clasificar d_j bajo c_i , mientras que $CSV_i(d_j) < \tau_i$ es interpretado como una decisión de no clasificar d_j bajo c_i . El umbral es trivialmente cualquier valor en el intervalo abierto $(0,1)$.

b.1. Algoritmos de clasificación para la MT

Existen diversos algoritmos de clasificación que pueden ser aplicados a la MT, entre los cuales se pueden mencionar los siguientes: redes neuronales (*perceptron*), probabilísticos (redes bayesianas), lineales, generadores de reglas, basados en redes semánticas, árboles, algoritmos genéticos, etc. A continuación, se describen algunos de los principales algoritmos:

- **Simple de Bayes**

El clasificador simple de Bayes utiliza los datos de entrenamiento para estimar parámetros de una distribución de probabilidad, que determina la probabilidad de cada clase dado los valores de las características del documento de una nueva instancia, para lo cual emplea el teorema de Bayes. Se basa en la presunción de independencia condicional en cuanto a la aparición de palabras en documentos, aunque esto generalmente no es cierto ya que en realidad la ocurrencia de un término t' dentro de un documento d_t no es siempre independiente de la ocurrencia de otro término t'' dentro de d_t . Sin embargo, el clasificador simple de Bayes es sorprendentemente

efectivo. Jain y Li (1998) han demostrado que la selección de características es contraproducente para este método.

Los cálculos se resumen mediante la siguiente fórmula:

$$CSV(d_t) \stackrel{def}{=} P(c_i|d_t) \stackrel{est}{=} \prod_{y=1}^M [P(c_i|t_y) \cdot P(t_y|d_t) + P(c_i|\bar{t}_y) \cdot P(\bar{t}_y|d_t)]$$

• K vecinos más cercanos

El método de k vecinos más cercanos o k-NN (por sus siglas en inglés, *k-Nearest Neighbour*), como todos los métodos basados en ejemplos, no construye una representación explícita y declarativa de la categoría de interés, sino que se vale de los criterios de clasificación que los expertos han dado sobre documentos de entrenamiento similares al que se clasifica. Como consecuencia, estos métodos han sido denominados sistemas de aprendizaje lento (*lazy learning*) o aprendizaje basado en memoria (*memory-based learning*), ya que no implican un fase de entrenamiento verdadera.

En k-NN para decidir si d_t debe ser clasificado bajo c_i se observa si los k documentos de entrenamiento más similares a d_t también han sido clasificados bajo c_i ; si la respuesta es positiva para una proporción considerable de éstos, se toma una decisión de clasificación positiva. Matemáticamente, clasificar un documento d_t por medio de k-NN se reduce al cálculo de:

$$CSV_i(d_t) = \sum_{d_j \in TR_k(d_t)} RSV(d_t, d_j) \cdot a_{ij}$$

Donde $RSV(d_t, d_j)$ representa cualquier función de coincidencia entre los documentos d_t y d_j , ya sea probabilística o basada en vectores, como por ejemplo la distancia Euclídea o el coseno entre los dos vectores documento. $TR_k(d_t)$ es la colección de los k documentos de entrenamiento d_j para los cuales $RSV(d_t, d_j)$ es máximo, y los valores a_{ij} se toman de la matriz de decisión correcta de los documentos de entrenamiento (Figura 3).

La construcción de un clasificador k-NN implica determinar el umbral k, generalmente de manera experimental. En la Figura 5 se observa el carácter local que presenta k-NN en cuanto al área de influencia del clasificador, lo que constituye una de las principales ventajas de k-NN e influye en su notable eficiencia, la cual ha sido demostrada en diferentes experimentos encontrados en la literatura (algunos reportados en la Sección 1.4).

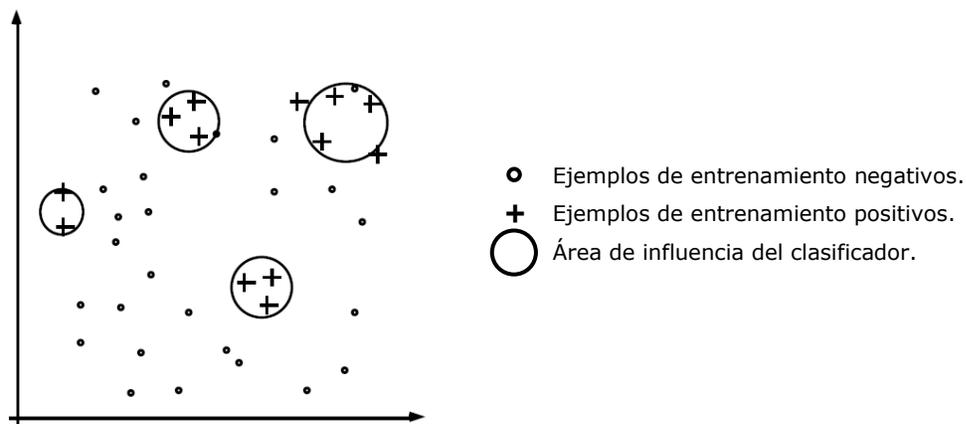


Figura 5. Conducta de clasificación de los clasificadores como k-NN.

Nota. Fuente: Adaptado de Sebastiani (1999).

• Árboles de decisión

En este método el vector documento d_t se compara con un árbol de decisión que se construye a partir de los documentos de entrenamiento, siendo el algoritmo CART uno de los más populares para la realización de esta tarea. CART construye un árbol de decisión binario dividiendo el conjunto de vectores de entrenamiento en cada nodo de acuerdo a una función aplicada a cada elemento (atributo) del vector, por lo tanto, la primera tarea es decidir cuáles elementos hacen la mejor división.

Otros algoritmos de árboles de decisión conocidos son C4.5 (Quinlan, 1993) y CHAID. C4.5 difiere de CART en el hecho de que produce árboles con un número variado de ramas por nodo y en la propuesta de poda. CHAID está restringido a variables categóricas y difiere de CART y C4.5 en que en lugar de primero sobreajustar los datos y luego podar, éste intenta detener el crecimiento del árbol antes de que ocurra el sobreajuste.

• Máquinas de soporte vectorial

Las SVM (por sus siglas en inglés, *support vector machine*), sólo son aplicables a tareas de clasificación binarias, por lo que este método de clasificación de texto debe ser tratado como una serie de problemas dicotómicos de clasificación en los que se realizan divisiones y subdivisiones de las clases, considerando dos cada vez. La clasificación se realiza por medio de la construcción de un hiperplano N-dimensional que separe óptimamente los datos en dos categorías.

Las SVM se encuentran muy relacionadas con las redes neuronales, de hecho, un modelo SVM cuya función núcleo sea sigmoide (con forma de "s") es equivalente a una red neuronal tipo *perceptron* de dos capas. Han demostrado producir buenas prestaciones en la clasificación de textos. Además, integran la reducción de dimensionalidad y la clasificación (Joachims, 1998).

1.4 Aplicaciones y Trabajos Previos

Los campos en los que pueden emplearse técnicas de MD son extremadamente variados, prácticamente en cualquier situación en la que se disponga de un conjunto de datos, como por ejemplo:

- **Análisis de textos:** análisis de documentos digitales, Internet, etc.
- **Ámbitos financieros y de seguros:** detección de fraudes, evasión fiscal, marketing, análisis de clientes, aprobación de préstamos, etc.
- **Producción y distribución:** índices de producción y costes, optimización de la fabricación, control de calidad, prevención de problemas, diagnóstico de fallas, análisis de cestas de la compra, gestión de *stocks* y planificación de transportes.
- **Redes de telecomunicaciones:** patrones de llamadas, gestión de fallos, etc.
- **Social:** data demográfica, tendencias de votación, etc.
- **Dominios científicos:** observaciones astronómicas, datos biológicos, electroquímica, modelos de diagnóstico, gestión de tratamientos, diseño de campañas de prevención y vacunación, biomedicina.
- **Aspectos climatológicos:** predicción de tormentas, etc.
- **Administraciones y organismos públicos:** análisis de políticas de empleo, vivienda, seguridad, educación, transporte, poblacionales, control del tráfico, etc.

La tarea de clasificación de textos de la MT, que a su vez se puede ver como una aplicación de la MD, puede aplicarse a documentos referentes a cualquier ámbito, pero en particular proporciona grandes ventajas en las disciplinas científicas.

Se han realizado importantes avances en el área de medicina, siendo ejemplo de esto el asistente de investigación biológica BioRAT (Buxton, Corney, Jones, y Langdon, 2004), que constituye una poderosa herramienta para encontrar y analizar artículos de investigación, y el clasificador estomatológico de La Red, Peláez y Sánchez (2002). Éste último utiliza una estructura de red con las categorías asociadas a sus patrones mediante compuertas lógicas AND y OR y aplica una función de similitud entre patrones, alcanzando una exactitud superior al 70%.

Un experimento de clasificación con la Colección Reuters-21578, llevado a cabo por Aas y Eikvil (1999), emplea pesado por entropía, RD por extracción de características (método de Lanczos), k vecinos más cercanos y el coseno como medida de similitud para la construcción del clasificador, con el cual obtienen una precisión general de 80%. Fuentes, Montes, Téllez, y Villaseñor (2003) presentan un clasificador automático de desastres naturales en México con el que logran una exactitud aproximada del 90%, utilizando el método de K vecinos más cercano con umbral en la frecuencia e IG como técnica de pesado para la RD.

CAPÍTULO 2. MARCO APLICATIVO

2.1 Planteamiento del problema

Actualmente, el proceso de asignación de jurados a Trabajos Especiales de Grado (TEG) de la Escuela de Computación de la UCV se lleva a cabo siguiendo los siguientes pasos:

1. Los tutores de cada TEG sugieren profesores como jurados para los mismos.
2. Dicha sugerencia es sometida a consideración en el Consejo de Escuela (CE).
3. El CE analiza cada caso según ciertos criterios, principalmente: el área profesional de los jurados sugeridos y las abarcadas por el TEG, la cantidad de veces que han sido asignados como jurado, la disponibilidad de los mismos (que no se encuentren de permiso o que no hayan sido asignados como jurados de otro TEG a presentar en fecha cercana al considerado), entre otros.
4. El CE toma la decisión de mantener la sugerencia o realizar cambios y, finalmente, se realiza la asignación de los jurados.

El proceso descrito presenta ciertas características desfavorables que lo hacen candidato para ser resuelto aplicando MT, entre las cuales se pueden señalar las siguientes:

- Es un proceso subjetivo, puesto que el CE realiza la toma de decisiones en base a su criterio sobre los profesores disponibles que resulten más convenientes como jurados para cierto TEG, de acuerdo al área profesional del mismo.
- El proceso es discrecional, no se encuentra debidamente regulado, por lo que se hace prudencialmente.
- No hay información actualizada sobre los jurados asignados ni sobre la disponibilidad de los profesores.
- La información con la que se cuenta se encuentra en papel, sin una organización que facilite su consulta.

2.2 Propuesta de solución

Como solución al problema planteado, se propone la automatización del proceso de asignación de jurados a TEG a través de la creación de un prototipo de sistema de recomendación que permita, mediante técnicas de minería de texto, realizar la clasificación de los mismos de acuerdo a las diversas áreas profesionales de la Escuela de Computación, para así, asignar como jurados a los profesores pertenecientes al área recomendada que resulten más pertinentes, según ciertos criterios establecidos. Entre los criterios a considerar se encuentran los siguientes:

- Que no se encuentre de permiso.
- Que no exceda cierto umbral en cuanto a la cantidad de asignaciones semestrales.
- Que se encuentre entre los docentes del área que hayan sido jurados de menos TEG para el semestre dado.

En la Figura 6 se muestran los componentes básicos del prototipo de sistema de asignación de jurados a TEG propuesto. Principalmente, se plantea la construcción de un clasificador que permita asignar una o dos áreas (máximo) con las que se encuentra más relacionado un TEG. Una vez clasificado el documento, se procede a la selección de los diferentes tipos de jurados (principales y suplentes) entre los profesores del área o áreas en cuestión, según los criterios mencionados.

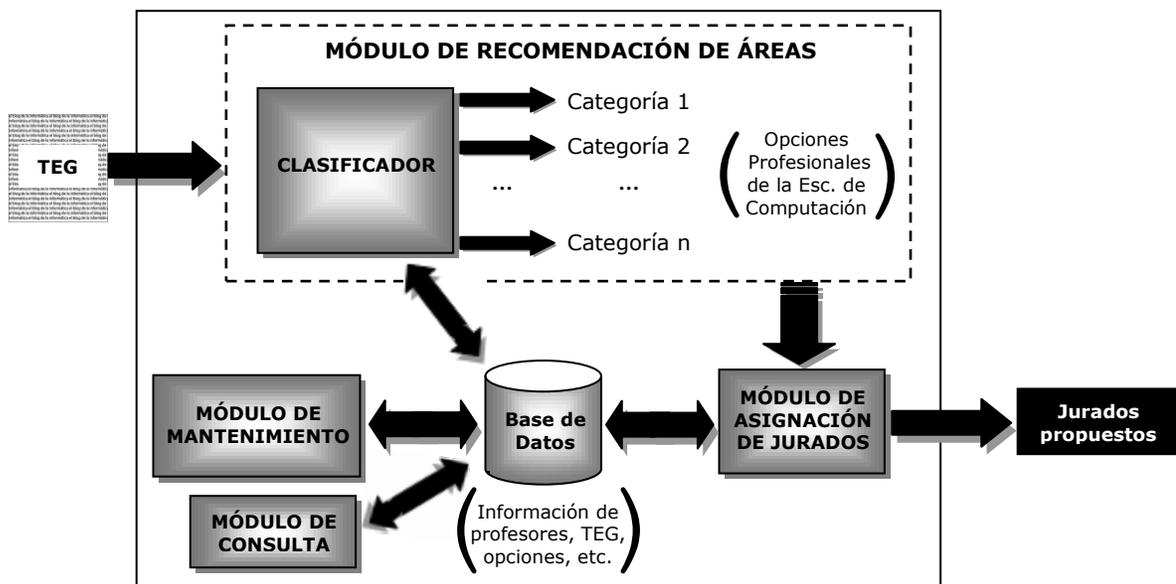


Figura 6. Prototipo de sistema de asignación de jurados a TEG propuesto.

Nota. Fuente: Elaboración propia.

2.3 Objetivos

2.3.1. Objetivo General

Desarrollar un prototipo de sistema basado en minería de textos que asigne de manera automática jurados a TEG.

2.3.2. Objetivos Específicos

- Analizar los requerimientos necesarios para el desarrollo del prototipo de sistema de asignación de jurados.
- Recolectar la información necesaria para la creación del prototipo (documentos de TEG, opciones profesionales, personal docente, entre otros).
- Generar un modelo de clasificación basado en técnicas de minería de textos que recomiende las opciones profesionales relacionadas con los documentos de TEG.
- Construir un histórico de fácil consulta de las asignaciones de jurados realizadas para cada TEG.
- Implementar la solución planteada en una aplicación Web.
- Realizar las pruebas necesarias para comprobar el funcionamiento del sistema desarrollado.

2.4 Proceso de minería de textos

Siguiendo los pasos especificados en la Sección 1.3.1, a continuación se detalla el proceso de minería de textos que se llevó a cabo para generar el modelo de clasificación de textos del prototipo.

2.4.1. Recolección de los datos

Los datos necesarios para realizar la minería de texto son los documentos de TEG. Específicamente, se reunió la siguiente información por documento:

- Título, resumen (o, en su defecto, introducción) y palabras claves, que constituyen la información textual sobre la cual se realizará el proceso de MT.
- Área profesional o mención a la cual pertenece.
- Tutor(es) y jurados asignados (para el histórico de asignaciones).

Adicionalmente, se recolectó la siguiente información requerida para el desarrollo del prototipo:

- Los nombres de las opciones profesionales.
- Los nombres y áreas de especialización de los profesores de la Escuela de Computación.

La lista contentiva de los profesores y sus áreas de especialización se adquirió directamente con ellos y en algunos casos, de estar disponible, de los sitios Web de los Centros de Investigación. La información en digital sobre los TEG se obtuvo de los sistemas CONEST (<https://estudiantes.ciens.ucv.ve/>) y BUSCONEST (<https://busconest.ciens.ucv.ve/>) de la Facultad de Ciencias, la cual corresponde a los TEG presentados por los estudiantes de la Escuela de Computación entre los años 2008 y 2010 (desde que se implantaron dichos sistemas hasta la actualidad), que en total constituyen 99 documentos de todas las menciones. La cantidad de documentos es menor (casi la mitad) que la cantidad de graduandos para el período mencionado, ya que la mayoría corresponde a TEG realizados por dos graduandos. Los TEG de años anteriores al 2008 se adquirieron de la Biblioteca Alonso Gamero en físico, lo que hizo más lento el proceso de recolección al ser necesario digitalizarlos.

La falta de una cantidad importante de documentos de TEG en digital para cada una de las áreas profesionales constituyó una importante limitación en el desarrollo del presente TEG, razón por la cual las áreas a considerar por el clasificador se redujeron a aquellas en las que se concentró la mayoría de los documentos recopilados (Ver Tabla 3). Esta decisión se debió al hecho de que los algoritmos de MT necesitan una cantidad importante de datos para alcanzar un rendimiento aceptable.

Tabla 3. Cantidad de documentos recopilados para las áreas profesionales.

Áreas	Cant. de docs. digitales	Cant. de docs. en físico	Cant. de docs. por área
Aplicaciones de Tecnología Internet (ATI)	24	26	50
Tecnología en Comunicaciones y Redes de Computadoras (Redes)	20	29	49
Bases de Datos (BD)	9	41	50
Inteligencia Artificial (IA)	7	42	49
Total	60	138	N = 198
N = cantidad de documentos de la colección (D)			

Dado que los datos se encontraban contenidos tanto en documentos en físico como en archivos digitales de diferentes formatos (.xls, .csv, .xml), se realizó un proceso de extracción, transformación y carga (*Extraction, Transformation and Loading* o ETL) de los mismos, mediante procedimientos desarrollados en lenguaje Java, con el fin de unificarlos y almacenarlos en una base de datos.

2.4.2. Preparación de los textos

a. Preprocesamiento

Los pasos que se llevaron a cabo para el preprocesamiento de los documentos recopilados se listan a continuación:

- Eliminación de signos de puntuación y demás caracteres especiales. Los acentos también fueron removidos, para facilitar el análisis de los textos.
- Construcción de un diccionario de palabras frecuentes en el español que no aportan información para la tarea de clasificación de textos (Ver Tabla 4).

Tabla 4. Diccionario de palabras no informativas del español.

Preposiciones	"a", "ante", "bajo", "cabe", "con", "contra", "de", "desde", "en", "entre", "hacia", "hasta", "para", "por", "sin", "segun", "so", "sobre", "tras", "mediante", "durante", "excepto", "salvo"
Artículos	"el", "la", "lo", "los", "las", "un", "una", "unos", "unas", "del", "al"
Adjetivos determinativos: demostrativos, posesivos, indefinidos e interrogativos	"este", "esta", "estos", "estas", "ese", "esa", "esos", "esas", "aquel", "aquella", "aquellos", "aquellas", "mio", "mia", "mios", "mias", "tuyo", "tuya", "tuyos", "tuyas", "suyo", "suya", "suyos", "suyas", "nuestro", "nuestra", "nuestros", "nuestras", "vuestro", "vuestra", "vuestros", "vuestras", "mi", "mis", "tu", "tus", "su", "sus", "cuyo", "cuya", "cuyos", "cuyas", "otro", "otra", "otros", "otras", "todo", "toda", "todos", "todas", "cierto", "cierta", "ciertos", "ciertas", "semejante", "semejantes", "tal", "diferente", "diferentes", "diverso", "diversa", "diversos", "diversas", "varios", "varias", "cada", "mismo", "misma", "mismos", "mismas", "tanto", "tanta", "tantos", "tantas", "cualquier", "cualquiera", "cualesquier", "algun", "algunos", "algunas", "ningun", "poco", "poca", "pocos", "pocas", "mucho", "mucha", "demasiado", "demasiada", "demasiados", "demasiadas", "determinado", "determinada", "determinados", "determinadas", "cual", "cuales", "que", "quien", "quienes", "cuanto", "cuantos", "cuanta", "cuantas", "bastante"
Pronombres	"yo", "mi", "me", "conmigo", "nosotros", "nosotras", "nos", "tu", "usted", "ti", "te", "contigo", "ustedes", "vosotros", "vosotras", "os", "el", "ella", "ello", "si", "se", "consigo", "le", "ellos", "ellas", "les", "vos", "esto", "eso", "aquello", "alguien", "nadie", "algo", "nada", "alguno", "algun", "ninguno", "ninguna", "muchos", "muchas", "bastantes"
Adverbios Determinativos: de lugar, tiempo, modo, cantidad, duda, afirmación, negación	"aqui", "alli", "ahi", "aca", "alla", "cerca", "lejos", "fuera", "afuera", "dentro", "adentro", "encima", "debajo", "arriba", "abajo", "delante", "adelante", "alrededor", "detras", "donde", "mientras", "luego", "temprano", "antes", "despues", "pronto", "tarde", "ya", "ahora", "entonces", "hoy", "mañana", "ayer", "nunca", "jamás", "siempre", "todavía", "cuando", "asi", "apenas", "como", "mas", "menos", "tanto", "casi", "muy", "quizá", "quizas", "acaso", "vez", "ciertamente", "tambien", "no", "tampoco", "solo"
Conjunciones	"y", "e", "ni", "o", "u", "ya", "ora", "sea", "fuera", "pero", "sino", "embargo", "obstante", "excepto", "salvo", "porque", "pues", "puesto", "pues", "supuesto", "dado", "luego", "conque", "consiguiente", "manera", "modo", "fin", "aunque", "aun", "siquiera"
Interjección	"ay", "ah", "oh", "huy", "uy", "bah", "hurra", "uf", "ojala", "ea", "puf", "hola", "chao", "caramba"

Estas palabras fueron eliminadas de los documentos aplicando un proceso de comparación con el diccionario.

- Una vez eliminadas las palabras no informativas, se realizó el proceso de lematización (*stemming*) sobre el resto, mediante la aplicación del algoritmo *Porter Stemming* (Porter, 1980) para el español. Se implementó una versión del mismo en lenguaje Java basada en el algoritmo especificado en la página Web de *Snowball* (<http://snowball.tartarus.org/>), un sitio Web que presenta una serie de algoritmos de lematización para diversos idiomas.

Como resultado del preprocesamiento, se obtuvo un total del 3.747 raíces informativas a partir de los documentos recopilados, las cuales constituyen el conjunto de términos a utilizar: $T = \{t_1, \dots, t_M\}$, con $M = 3.747$.

b. Indexación

Se utilizó la representación mediante el modelo de espacio vectorial, calculando el peso a_{ij} del término j en el documento i aplicando tres de las propuestas de pesado sugeridas en la Sección 1.3.1 (a.2), a saber: tfc , lfc y entropía. Las mismas fueron seleccionadas como técnicas de indexación ya que toman en cuenta aspectos como la diferencia de tamaños entre documentos y la reducción del efecto de grandes diferencias en las frecuencias (Ass y Eikvil, 1999), además de la gran efectividad que han registrado en los trabajos previos revisados.

Para el cálculo de los pesos se generó una matriz de frecuencias que contiene las apariciones de cada término en T en los documentos de la colección D (f_{ij} = frecuencia del término j en el documento i) y en total en toda la colección (n_j = número total de veces que el término j aparece en la colección). Finalmente, se obtuvieron tres matrices de documentos como la representada en la Figura 7, una por cada tipo de pesado para la indexación.

	t_1	...	t_j	...	T_{3747}
d_1	a_{11}	...	a_{1j}	...	$a_{1\ 3747}$
...
d_i	a_{i1}	...	a_{ij}	...	$a_{i\ 3747}$
...
d_{198}	$a_{198\ 1}$...	$a_{198\ j}$...	$a_{198\ 3747}$

Figura 7. Matriz de representación de los documentos recopilados.

Nota. Fuente: Elaboración propia.

c. Reducción de dimensionalidad

La gran cantidad de términos obtenidos luego del preprocesamiento ($M = 3.747$) refleja el problema de alta dimensionalidad del espacio de términos. Para la construcción del clasificador se realizó TSR global por selección de características, porque este tipo de RD permite que se consideren solo los términos más relevantes.

c.1. Técnicas para la selección de características

Como técnicas para la selección de características, se consideraron las siguientes funciones que miden la relevancia de los términos en los documentos, tomando en cuenta las sugeridas en la Sección 1.3.1 (a.3): ganancia de información (*IG*), chi-cuadrado (X^2), coeficiente de correlación (*CC*), razón de ganancia (*GR*) y consistencia (*consistency*). Existen diversas herramientas que permiten aplicar estas técnicas con buenos resultados. En este trabajo se utilizó la herramienta Weka (<http://www.cs.waikato.ac.nz/ml/weka/>), versión 3.6.4, ya que es un sistema multiplataforma de amplio uso que se encuentra disponible en Internet de forma gratuita y que, además, es de licencia de software libre.

Cada una de las matrices obtenidas luego de aplicar las tres técnicas de indexación se transformó en un archivo del formato aceptado como entrada por Weka: “.arff”, en el cual cada una de las raíces en T se toman como atributos de tipo numérico. Para más información sobre los archivos “.arff” ver el Anexo 1.

c.2. Definición del número de características

Para definir el número de características a considerar por cada tipo de pesado empleado para la indexación, se utilizó un sistema de votación empleando como criterio de selección final aquellos atributos que fueron votados por cuatro o por los cinco algoritmos. Los atributos seleccionados por cada algoritmo para cada caso de indexación se detallan en los Anexos 2, 3 y 4 y las selecciones finales se presentan en la Tabla 5.

Tabla 5. Selección final de atributos para los tres tipos de indexación aplicados.

Criterio	Ltc	Entropía	Tfc	Ranking
Mayoría	desarroll	web	desarroll	1
	web	siti	web	2
	siti	javascript	aplic	3
	clasificacion	clasificacion	siti	4
	dat	dat	clasificacion	5
	red	red	red	6
	cre	ruby	servici	7
	servici	administr	ruby	8
	ruby	usabil	prototip	9
	prototip	prototip	manej	10
	manej	manej	metod	11
	metod	metod	disposit	12
	control	disposit	protocol	13
	disposit	distribu	human	14
	protocol	protocol	conoc	15
	human	human	segur	16
	conoc	conoc	fundament	17
	profesional	profesional	inteligent	18
	segur	segur	asoci	19
	oos	expert	inalambr	20
	fundament	domini	movil	21
	domini	inteligent	agent	22
	inteligent	inalambr	ontologi	23
	ipv6	regl	profesional	24
	inalambr	artificial	expert	25
	movil	on	domini	26
	agent	bas	ip	27
	interaccion	consistent	802	28
	usabil	ipv6	bluetooth	29
	expert	ip	artificial	30
	ip	802	mineri	31
	802	bluetooth	neuronal	32
	bluetooth	movil	ipv6	33
	descript	agent	iee	34
	regl	descript	regl	35
	artificial	ontologi		36
	ontologi	mineri		37
	mineri	resolucion		38
	neuronal	neuronal		39
	iee	iee		40
	especi	especi		41
M' « M	41 « 3.747	41 « 3.747	35 « 3.747	

2.4.3. Estimación y evaluación del modelo de clasificación

Entre los algoritmos de clasificación descritos en la Sección 1.3.1 (b.1), para generar el modelo de clasificación se seleccionó k vecinos más cercanos. La razón de esta escogencia se debió a que es un método que ha reportado en la literatura un buen rendimiento para la clasificación a pesar de su sencilla implementación, lo cual es una característica favorable ya que al requerir un menor costo computacional los tiempos de respuesta al clasificar son considerablemente menores que los requeridos por otros métodos de mayor complejidad, como por ejemplo las redes neuronales.

Para cada matriz de indexación, se realizó un conjunto de experimentos aplicando el algoritmo k -NN, encontrado en Weka como IBk (dentro de los métodos *lazy*). Los experimentos consistieron en probar el algoritmo configurando sus parámetros a los siguientes valores:

- Valor de k : 1, 3, 5 y 7.
- Medida de distancia: Chebyshev y Euclídea.
- Ponderación de la distancia: $1/\text{distancia}$, $1-\text{distancia}$ o sin ponderación.
- La utilización o no de validación cruzada para seleccionar el mejor valor de k , en cuyo caso el parámetro k proporcionado en realidad indica el máximo número de vecinos a considerar en la búsqueda.

Como método de evaluación se utilizó validación cruzada con 10 particiones (10 *folds cross-validation*). Las Tablas 6, 7 y 8 muestran los resultados obtenidos en los experimentos para cada uno de los tipos de indexación.

Tabla 6. Resultados del clasificador k-NN con la selección por entropía.

K	Distancia	CV	Distancia Ponderada				Clasificaciones			
			NO	1 / distancia	1 - distancia					
1	Chebyshev	T	152	76,7677%	152	76,7677%	152	76,7677%	Correctas	
		F	46	23,2323%	46	23,2323%	46	23,2323%	Incorrectas	
	Euclídea	T	152	76,7677%	152	76,7677%	152	76,7677%	Correctas	
		F	46	23,2323%	46	23,2323%	46	23,2323%	Incorrectas	
		T	160	80,8081%	160	80,8081%	160	80,8081%	Correctas	
		F	38	19,1919%	38	19,1919%	38	19,1919%	Incorrectas	
3	Chebyshev	T	152	76,7677%	152	76,7677%	152	76,7677%	Correctas	
		F	46	23,2323%	46	23,2323%	46	23,2323%	Incorrectas	
	Euclídea	T	145	73,2323%	153	77,2727%	153	77,2727%	Correctas	
		F	53	26,7677%	45	22,7273%	45	22,7273%	Incorrectas	
		T	160	80,8081%	160	80,8081%	160	80,8081%	Correctas	
		F	38	19,1919%	38	19,1919%	38	19,1919%	Incorrectas	
5	Chebyshev	T	152	76,7677%	152	76,7677%	152	76,7677%	Correctas	
		F	46	23,2323%	46	23,2323%	46	23,2323%	Incorrectas	
	Euclídea	T	128	64,6465%	144	72,7273%	140	70,7071%	Correctas	
		F	70	35,3535%	54	27,2727%	58	29,2929%	Incorrectas	
		T	160	80,8081%	159	80,3030%	158	79,7980%	Correctas	
		F	38	19,1919%	39	19,6970%	40	20,2020%	Incorrectas	
7	Chebyshev	T	152	76,7677%	163	82,3232%	161	81,3131%	Correctas	
		F	46	23,2323%	35	17,6768%	37	18,6869%	Incorrectas	
	Euclídea	T	152	76,7677%	152	76,7677%	152	76,7677%	Correctas	
		F	46	23,2323%	46	23,2323%	46	23,2323%	Incorrectas	
		T	116	58,5859%	138	69,6970%	127	64,1414%	Correctas	
		F	82	41,4141%	60	30,3030%	71	35,8586%	Incorrectas	
7	Euclídea	T	160	80,8081%	159	80,3030%	158	79,7980%	Correctas	
		F	38	19,1919%	39	19,6970%	40	20,2020%	Incorrectas	
	7	Euclídea	T	148	74,7475%	159	80,3030%	158	79,7980%	Correctas
			F	50	25,2525%	39	19,6970%	40	20,2020%	Incorrectas

CV = Cross Validate (T=True, F=False); Dist. Pond. = Distancia ponderada

Tabla 7. Resultados del clasificador k-NN con la selección por pesado ltc.

K	Distancia	CV	Distancia Ponderada				Clasificaciones		
			NO	1 / distancia	1 - distancia				
1	Chebyshev	T	159	80,3030%	159	80,3030%	159	80,3030%	Correctas
		F	39	19,6970%	39	19,6970%	39	19,6970%	Incorrectas
	Euclídea	T	159	80,3030%	159	80,3030%	159	80,3030%	Correctas
			F	39	19,6970%	39	19,6970%	39	19,6970%
		F	155	78,2828%	155	78,2828%	155	78,2828%	Correctas
			43	21,7172%	43	21,7172%	43	21,7172%	Incorrectas
3	Chebyshev	T	156	78,7879%	158	79,7980%	158	79,7980%	Correctas
		F	42	21,2121%	40	20,2020%	40	20,2020%	Incorrectas
	Euclídea	T	156	78,7879%	162	81,8182%	161	81,3131%	Correctas
			F	42	21,2121%	36	18,1818%	37	18,6869%
		F	161	81,3131%	161	81,3131%	160	80,8081%	Correctas
			37	18,6869%	37	18,6869%	38	19,1919%	Incorrectas
5	Chebyshev	T	163	82,3232%	162	81,8182%	162	81,8182%	Correctas
		F	35	17,6768%	36	18,1818%	36	18,1818%	Incorrectas
	Euclídea	T	159	80,3030%	157	79,2929%	158	79,7980%	Correctas
			F	39	19,6970%	41	20,7071%	40	20,2020%
		F	146	73,7374%	159	80,3030%	155	78,2828%	Correctas
			52	26,2626%	39	19,6970%	43	21,7172%	Incorrectas
7	Chebyshev	T	161	81,3131%	162	81,8182%	162	81,8182%	Correctas
		F	37	18,6869%	36	18,1818%	36	18,1818%	Incorrectas
	Euclídea	T	157	79,2929%	162	81,8182%	162	81,8182%	Correctas
			F	41	20,7071%	36	18,1818%	36	18,1818%
		F	159	80,3030%	157	79,2929%	158	79,7980%	Correctas
			39	19,6970%	41	20,7071%	40	20,2020%	Incorrectas
CV = Cross Validate (T=True, F=False); Dist. Pond. = Distancia ponderada									

Tabla 8. Resultados del clasificador k-NN con la selección por pesado tfc.

K	Distancia	CV	Distancia Ponderada				Clasificaciones		
			NO	1 / distancia	1 - distancia				
1	Chebyshev	T	152	76,7677%	152	76,7677%	152	76,7677%	Correctas
			46	23,2323%	46	23,2323%	46	23,2323%	Incorrectas
		F	152	76,7677%	152	76,7677%	152	76,7677%	Correctas
			46	23,2323%	46	23,2323%	46	23,2323%	Incorrectas
	Euclídea	T	156	78,7879%	156	78,7879%	156	78,7879%	Correctas
			42	21,2121%	42	21,2121%	42	21,2121%	Incorrectas
		F	156	78,7879%	156	78,7879%	156	78,7879%	Correctas
			42	21,2121%	42	21,2121%	42	21,2121%	Incorrectas
3	Chebyshev	T	152	76,7677%	152	76,7677%	152	76,7677%	Correctas
			46	23,2323%	46	23,2323%	46	23,2323%	Incorrectas
		F	138	69,6970%	140	70,7071%	140	70,7071%	Correctas
			60	30,3030%	58	29,2929%	58	29,2929%	Incorrectas
	Euclídea	T	153	77,2727%	160	80,8081%	160	80,8081%	Correctas
			45	22,7273%	38	19,1919%	38	19,1919%	Incorrectas
		F	155	78,2828%	160	80,8081%	161	81,3131%	Correctas
			43	21,7172%	38	19,1919%	37	18,6869%	Incorrectas
5	Chebyshev	T	152	76,7677%	152	76,7677%	152	76,7677%	Correctas
			46	23,2323%	46	23,2323%	46	23,2323%	Incorrectas
		F	139	70,2020%	141	71,2121%	140	70,7071%	Correctas
			59	29,7980%	57	28,7879%	58	29,2929%	Incorrectas
	Euclídea	T	153	77,2727%	156	78,7879%	153	77,2727%	Correctas
			45	22,7273%	42	21,2121%	45	22,7273%	Incorrectas
		F	158	79,7980%	158	79,7980%	159	80,3030%	Correctas
			40	20,2020%	40	20,2020%	39	19,6970%	Incorrectas
7	Chebyshev	T	152	76,7677%	152	76,7677%	152	76,7677%	Correctas
			46	23,2323%	46	23,2323%	46	23,2323%	Incorrectas
		F	122	61,6162%	131	66,1616%	125	63,1313%	Correctas
			76	38,3838%	67	33,8384%	73	36,8687%	Incorrectas
	Euclídea	T	153	77,2727%	156	78,7879%	153	77,2727%	Correctas
			45	22,7273%	42	21,2121%	45	22,7273%	Incorrectas
		F	150	75,7576%	152	76,7677%	152	76,7677%	Correctas
			48	24,2424%	46	23,2323%	46	23,2323%	Incorrectas

CV = Cross Validate (T=True, F=False); **Dist. Pond.** = Distancia ponderada

Como se observa en los resultados destacados en las Tablas 6, 7 y 8, los mayores porcentajes de clasificaciones correctas fueron **82,32%** para el pesado ltc y por entropía, y 81,31% para el pesado tfc. Como se alcanza la misma exactitud con el pesado ltc y por entropía, se decidió realizar la elección del método de indexación a utilizar en base a los valores de k con los que se alcanzaron dichos resultados, que son: k=3 (ltc) y k=5 (entropía). En el contexto de clasificación de este trabajo, en el que se consideran cuatro áreas profesionales como categorías, resulta más conveniente la elección de k=5, puesto que permite dar prioridad cuando ocurre un empate al hallar el área que clasifica a mayor cantidad de los k vecinos. Para k=5 esto ocurre cuando dos áreas diferentes clasifican a 2 vecinos cada una, y se resuelve en base al promedio de las distancias ponderadas de los dos documentos clasificados bajo

cada una. Con $k=3$ el empate ocurre cuando cada vecino pertenece a un área diferente, por lo que se cuenta con una sola distancia para cada área, lo que imposibilita la aplicación del criterio de promedio de distancias en la resolución del empate.

Finalmente, utilizando los parámetros con los cuales se obtuvo el mejor resultado en los experimentos para la indexación por entropía (Ver Tabla 9), se infirió el modelo de clasificación a utilizar, a saber:

Indexación: mediante pesado por entropía

Algoritmo de clasificación: k vecinos más cercanos, con los siguientes parámetros:

- **K=5**
- **Distancia:** Euclídea
- **Ponderación de la distancia:** $1/\text{distancia}$.

La Figura 8 especifica como se calculan las medidas precisión y *recall* por área, indicadas en la Tabla 9, ya que son las métricas que se utilizan para evaluar la exactitud de predicción del clasificador.

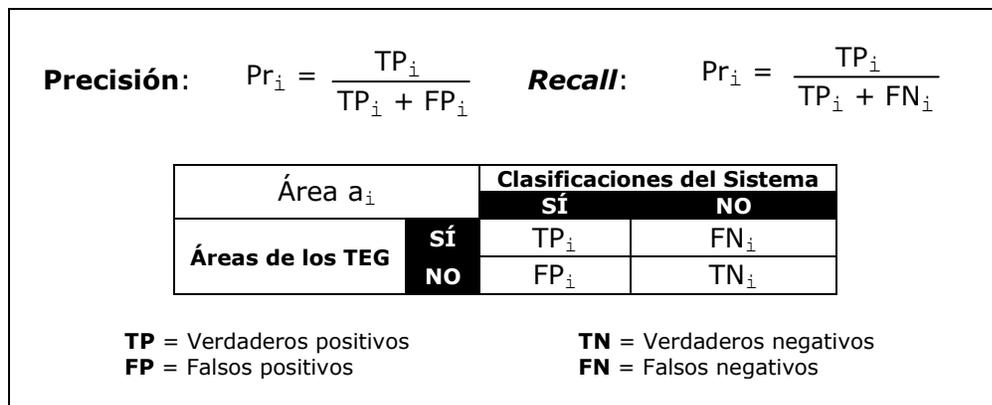


Figura 8. Medidas para la evaluación del clasificador.

Nota. Fuente: Adaptado de Sebastiani (1999).

Tabla 9. Mejor resultado de k-NN para la selección de la indexación con pesado por entropía.

kNN con K=5 y distancia Euclídea ponderada como 1/distancia						
Clasificaciones correctas	163	82,3232%				
Clasificaciones Incorrectas	35	17,6768%				
Nro. Total Instancias	198	100,0000%				
Medidas de Rendimiento	Precisión	Recall				
Clase a = ATI	0,729	0,860				
Clase b = Redes	1,000	0,816				
Clase c = BD	0,695	0,820				
Clase d = IA	0,975	0,796				
Promedio	0,848	0,823				
Matriz de Confusión	Reales	Clasificador				
		a	b	c	d	
		a	43	0	7	0
		b	4	40	5	0
		c	8	0	41	1
d	4	0	6	39		

La precisión indica una proporción entre los verdaderos y falsos positivos de cada área, por lo que una precisión alta refleja que la mayoría de los documentos clasificados bajo dicha área realmente pertenecían a la misma. En cambio, el *recall* indica una proporción entre los verdaderos positivos y falsos negativos de cada área, por lo que un mayor *recall* refleja que la mayoría de los documentos que realmente pertenecen a un área determinada fueron considerados por el clasificador bajo dicha área. La matriz de confusión de la Tabla 9 refleja que todos los documentos clasificados bajo el área de Redes efectivamente pertenecían a dicha área, ya que en la columna b todos los elementos son 0 a excepción de la fila b, debido a esto, es el área para la cual el clasificador es más preciso. Sin embargo, esto no quiere decir que todos los documentos del área de redes se clasificaron bajo esta área, como lo indica su *recall* y la matriz de confusión (la fila b muestra 9 falsos negativos) Asimismo, queda en evidencia que el área para la cual el clasificador reconoce correctamente mayor proporción de documentos es ATI, como se observa en la matriz de confusión, al presentar la menor cantidad de falsos negativos (7), por lo que presenta el mayor *recall*. BD es el área con menor precisión, con 18 clasificaciones incorrectas (suma de los elementos de la columna c, sin incluir la fila c), seguida de ATI, debido a que todos los falsos negativos de ATI fueron clasificados bajo BD y casi todos los falsos negativos de BD (a excepción de uno) fueron clasificados bajo ATI, lo que indica que son las

áreas para las cuales el clasificador presenta mayor confusión. Esto tiene correspondencia con la realidad, ya que las áreas de ATI y BD están muy relacionadas.

2.5 Desarrollo del prototipo

En esta sección se describe detalladamente el desarrollo del prototipo. Con este fin, a continuación, se definen cada uno de los módulos componentes del mismo, los cuales fueron especificados en la Figura 6, las principales tecnologías empleadas y las consideraciones de diseño seguidas para su desarrollo, mediante un modelado de la aplicación basado específicamente en el diagrama de casos de usos.

2.5.1. Módulo de recomendación de áreas

Como se observa en la Figura 9, el módulo de recomendación de áreas recibe, como entrada, los datos correspondientes al título, resumen y palabras claves del TEG a clasificar y genera, como salida, una o dos áreas (máximo) recomendadas como categorías. Para este fin, se siguen los siguientes pasos:

1. Preprocesar el documento introducido, es decir: se eliminan signos de puntuación, palabras no informativas y se realiza *word stemming*. Se consideran únicamente los atributos seleccionados en el modelo, correspondientes a la selección para la indexación por entropía (Tabla 5).
2. Representar el documento según el modelo de espacio vectorial aplicado a los documentos de entrenamiento, mediante la indexación de pesado por entropía. Para esto es necesario utilizar los siguientes parámetros, obtenidos durante la generación del modelo:
 - El valor de N (cantidad de documentos de entrenamiento).
 - La frecuencia de cada atributo en cada documento de entrenamiento (f_{ij}).
 - La frecuencia en documento de cada atributo (n_j).
3. Medir la similitud del documento de entrada con respecto a cada uno de los documentos de entrenamiento, calculando la distancia Euclídea entre éstos ponderada como $1/\text{distancia}$.

4. Hallar los k vecinos más cercanos, que son aquellos cuyos valores de similitud constituyen los k mayores (similitud = 1/distancia).
5. Clasificar el documento de entrada considerando las áreas de los k vecinos, recomendando aquella que los clasifica en su mayoría. Como se señaló en el punto 2.4.3, en caso de empate entre dos áreas se recomiendan ambas, otorgándoles prioridad de acuerdo al promedio de las similitudes de sus pares de documentos.
6. Solicitar confirmación del área o áreas recomendadas al usuario. En caso de no encontrarse conforme, le permite seleccionar directamente la misma. El(Las) área(s) confirmada(s) se envía(n) al módulo de asignación de jurados.

Este proceso de recomendación de una o dos áreas, en el caso especial, para un TEG se resume en la Figura 9.

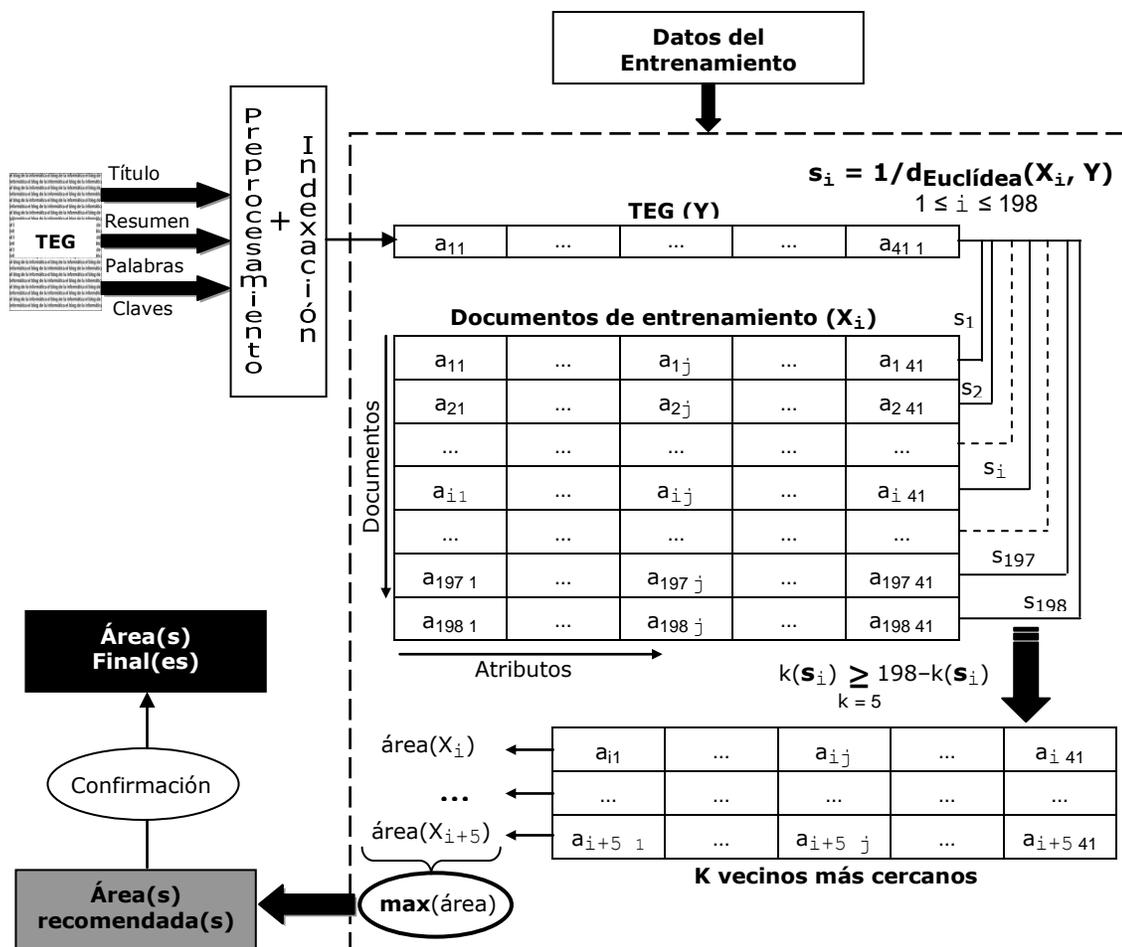


Figura 9. Módulo de recomendación de áreas del prototipo.

Nota. Fuente: Elaboración propia.

2.5.2. Módulo de asignación de jurados

Una vez clasificados los documentos, el módulo de asignación de jurados selecciona, según lo especifican un conjunto de reglas de elección de profesores (Ver Anexo 5), los docentes pertenecientes al área o áreas confirmadas. Este módulo de asignación no se limita a las cuatro áreas consideradas por el clasificador, permitiendo asignar jurados de todas las áreas profesionales de la Escuela de Computación.

A continuación, se explican las reglas de elección que rigen la asignación:

- A cada TEG se le asignan cuatro profesores como jurados, dos como jurados principales y dos como suplentes.
- Los jurados asignados no pueden ser los tutores del TEG.
- Si el TEG tiene una única área asignada, los cuatro jurados se seleccionan dentro de los docentes pertenecientes a dicha área. De no encontrarse disponible ningún profesor del área, se le notifica al usuario el motivo, el cual puede ser que hayan excedido el número máximo de asignación como jurados o que se encuentren de permiso.
- Si el TEG tiene dos áreas asignadas, se seleccionan dos profesores de cada área, uno como principal y uno como suplente. Si no se encuentran disponibles más docentes de un área, se selecciona a un docente de la otra área. Si tampoco se encuentra disponible un docente de la otra área, se notifica el motivo al usuario.
- Para la selección de los jurados principales primero se verifica que el docente no se encuentre de permiso y que no haya superado el número máximo de asignaciones como jurado principal para un semestre lectivo, el cual se fija empíricamente con el valor cuatro, sin embargo puede ser modificado a través del módulo de mantenimiento. Dentro de los profesores disponibles del área que no hayan superado dicho máximo, se seleccionan aquellos que presenten menor cantidad de asignaciones como jurados principales, y en caso de empates, menor cantidad de asignaciones como suplentes. En caso de que también se obtengan

empates en la cantidad de asignaciones como suplentes, se selecciona aleatoriamente.

- Para la selección de los jurados suplentes primero se verifica que el docente no se encuentre de permiso. Entre los profesores disponibles, se seleccionan aquellos que presenten menor cantidad de asignaciones como principales, y en caso de empates, menor cantidad de asignaciones como suplentes. En caso de que también se obtengan empates en la cantidad de asignaciones como suplentes principales, se selecciona aleatoriamente.

2.5.3. Módulo de consulta

El módulo de consulta del prototipo permite realizar búsquedas personalizadas sobre el histórico de asignaciones de profesores como jurados de TEG, configurando los parámetros de consulta a conveniencia del usuario.

Los parámetros de consulta son los siguientes:

- **Período:** permite limitar la consulta de jurados al período lectivo indicado.
- **Profesor:** habilita la consulta de las asignaciones de un profesor particular.
- **Tipo de Jurado:** limita la consulta al tipo de jurado establecido (Principal 1, Principal 2, Suplente 1 o Suplente 2)
- **Estudiante:** restringe la consulta a los jurados de un estudiante determinado.

Mediante la combinación de valores de estos parámetros, se permite al usuario realizar búsquedas específicas o generalizadas, según su preferencia. Como resultado, se muestra la información de los jurados y de los TEG que coincidan con la consulta establecida.

2.5.4. Módulo de mantenimiento

El módulo de mantenimiento del prototipo se encuentra disponible únicamente para el usuario administrador del mismo, ya que permite configurar parámetros y realizar acciones que afectan directamente su desempeño.

Dado que las asignaciones de jurados se hacen semestralmente, el módulo de mantenimiento tienen como utilidad principal la inicialización de las asignaciones semestrales de los docentes una vez finalizado cada semestre. A pesar de esto, dichos valores se preservan con el objetivo de llevar un historial de asignaciones por docente.

Además, el módulo de mantenimiento ofrece la posibilidad de cambiar el parámetro de la cantidad máxima de asignaciones como jurado principal de un profesor para un semestre lectivo, al cual se le asigna el valor cuatro, como se indicó anteriormente, como un estimado de la carga semestral de un profesor. En caso de que para alguna área todos los docentes alcancen el máximo de asignaciones, el sistema lo notifica y ofrece la posibilidad de modificar dicho valor.

También permite al administrador agregar, eliminar o modificar docentes, usuarios, áreas, así como asociar áreas a un docente.

2.5.5. Tecnologías utilizadas

Las principales tecnologías que se emplearon para el desarrollo del prototipo fueron las siguientes:

a. Plataforma Java

La tecnología Java es tanto un lenguaje de programación como una plataforma. El lenguaje de programación Java es un lenguaje de alto nivel orientado a objetos que tiene una sintaxis y estilo particular. Una plataforma Java es un entorno particular en el cual se ejecutan las aplicaciones en lenguaje Java. Existen cuatro plataformas Java, entre las cuales se encuentra la plataforma Java Enterprise Edition o Java EE, que está construida sobre la plataforma Java Estandar Edition o Java SE.

En específico, para la implementación del prototipo se utilizó como *kit* de desarrollo de software o SDK (por sus siglas en inglés, *software development kit*) el Java EE 6 SDK, puesto que proporciona un poderoso conjunto de API's y un entorno para desarrollar y ejecutar aplicaciones Java a gran escala, multi-niveles, escalables, confiables y seguras.

b. JSP y servlets

Las *JavaServer Pages* (JSP) y los servlets son tecnologías Java EE utilizadas a nivel Web en aplicaciones. Los servlets son clases del lenguaje de programación Java que procesan solicitudes dinámicamente y construyen respuestas, generalmente para páginas HTML. Las JSP son documentos basados en texto que son compilados en servlets y que permiten crear páginas Web dinámicas a partir de los parámetros de petición que envíe el navegador Web. Debido a esto, las especificaciones de las JSP van ligadas a una especificación de los servlets.

Específicamente, se utilizaron las especificaciones JSP 2.1 y Servlet 2.5.

c. Apache Tomcat

Tomcat es un contenedor Web o motor de *servlets* (*servlet engine*) que implementa las especificaciones de los *servlets* y JSP (JavaServer Pages), por lo que se puede decir que es un servidor Web que funciona bajo la tecnología JEE de Java. Ofrece una solución para la ejecución de páginas Web dinámicas desarrolladas bajo la tecnología JSP, al ofrecer un entorno donde habitan los JSP y *Servlets* (contenedor). Por defecto se presenta en combinación con el servidor Web Apache.

En este trabajo se utilizó la versión 6.0 de Apache Tomcat, debido a que es la versión implementada para las especificaciones Servlet/JSP 2.5/2.1, con el servidor Web Apache versión 2.0, funcionando en modo *stand-alone* (independiente), es decir, como parte integral del servidor, que es el modo por defecto usado por Tomcat. Principalmente, se selecciona este contenedor Web debido a su interoperabilidad, puesto que funciona en cualquier sistema operativo que disponga de la máquina virtual Java, además de ser software libre y permitir la conexión con diferentes manejadores de bases de datos (MySQL, Oracle, etc.) mediante JDBC (*Java Database Connectivity*).

d. MySQL Server

MySQL es uno de los sistemas de gestión de bases de datos relacionales SQL más populares. Una base de datos es una colección estructurada de datos. Una base de datos relacional almacena datos en tablas separadas en lugar de poner todos los datos en un gran almacén, lo que añade velocidad y flexibilidad. Para añadir, acceder, y procesar los datos almacenados en una base de datos, se necesita un sistema de gestión de base de datos, como MySQL Server.

El servidor de base de datos MySQL es soportado por una gran cantidad de lenguajes y aplicaciones, como el lenguaje Java utilizado en este trabajo, mediante JDBC. Específicamente, se desarrolló con la versión MySQL Server 5.1.36.

e. Eclipse

Eclipse es un entorno de desarrollo integrado o IDE, por sus siglas en inglés (*Integrated development environment*), de código abierto y multiplataforma para desarrollar aplicaciones cliente enriquecidas. El SDK de Eclipse incluye las herramientas de desarrollo de Java, ofreciendo un IDE con un compilador de Java interno. Proporciona al programador *frameworks* muy ricos para el desarrollo de aplicaciones gráficas, definición y manipulación de modelos de software, aplicaciones Web, etc.

Para el desarrollo de este TEG se utilizó la versión 3.5 de Eclipse, conocida como Eclipse Galileo.

2.5.6. Diseño de la interfaz

A continuación, se describen las principales consideraciones de diseño que se tuvieron en cuenta para el desarrollo del prototipo. Se especifica el patrón de diseño seguido y se describen sus características fundamentales. Adicionalmente, se realiza el modelado de la aplicación para comprender el sistema que se desarrolló. Para modelar, analizar y diseñar sistemas orientados a objetos se utiliza el lenguaje de modelado UML (Unified Modeling Language, por sus siglas en inglés). Entre los diagramas de modelado que proporciona UML se encuentran: el diagrama de casos de usos, diagrama de clases, diagrama de estados, diagrama de secuencias, entre otros (Cueva, 1999). Para el modelado de los requerimientos del prototipo se emplea el diagrama de casos de usos. Finalmente, se presentan las principales interfaces del prototipo.

a. Patrón de diseño MVC

El diseño de la aplicación se basó en la arquitectura Modelo-Vista-Controlador (MVC), puesto que permite separar la funcionalidad central del modelo de negocio de la aplicación de la presentación y la lógica de control que utiliza dicha funcionalidad. En la Figura 10 se describen los participantes de la arquitectura MVC.

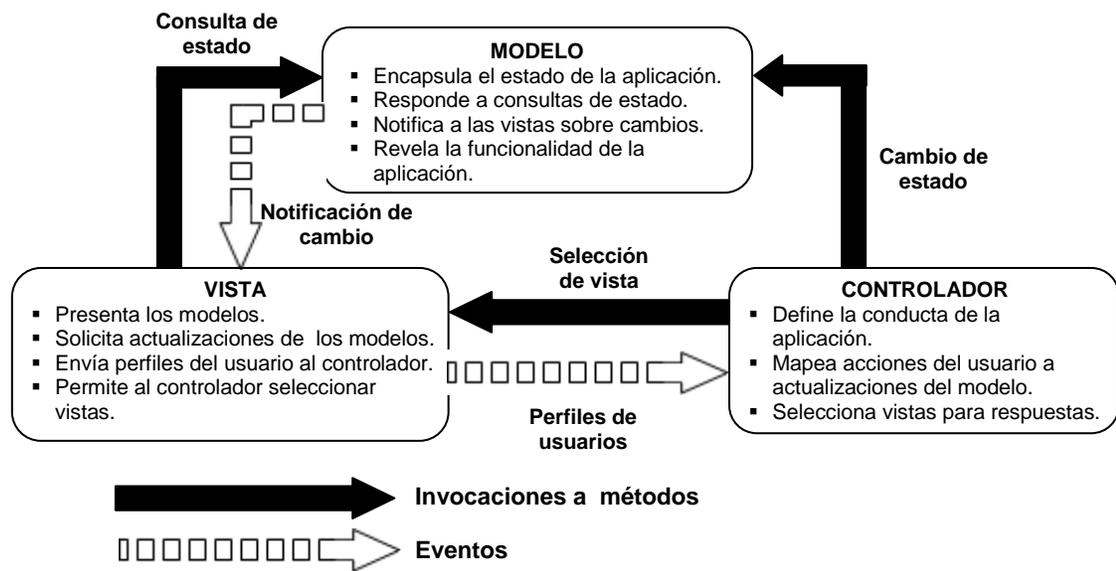


Figura 10. Estructura del patrón de diseño MVC.

Nota. Fuente: Adaptado de Oracle, Java BluePrints, Model-View-Controller (s.f).

Se utiliza el patrón de diseño MVC debido a las grandes ventajas que aporta en el desarrollo de aplicaciones, entre las cuales se mencionan las siguientes:

- Permite la reutilización de componentes del modelo.
- Hace más sencillo agregar múltiples representaciones de los mismos datos.
- Facilita la inserción de nuevos datos, según sea requerido por la aplicación.
- Crea independencia de funcionamiento.
- Hace más fácil el mantenimiento de la aplicación.
- Ofrece maneras más sencillas de probar el funcionamiento de la aplicación.

b. Diagramas de casos de uso

Los casos de usos forman parte del análisis de un sistema, al modelar sus funcionalidades ayuda a describir qué es lo que el sistema debe hacer y cómo el usuario interactúa con él (Gracia, 2003).

A continuación, se presentan los casos de usos modelados para el desarrollo del prototipo, en tres niveles de especificación.

b.1. Casos de uso - Nivel 0

En este caso de uso se modela el prototipo a nivel general, con sus respectivos actores (Ver Figura 11). Los actores se describen en la Tabla 10.

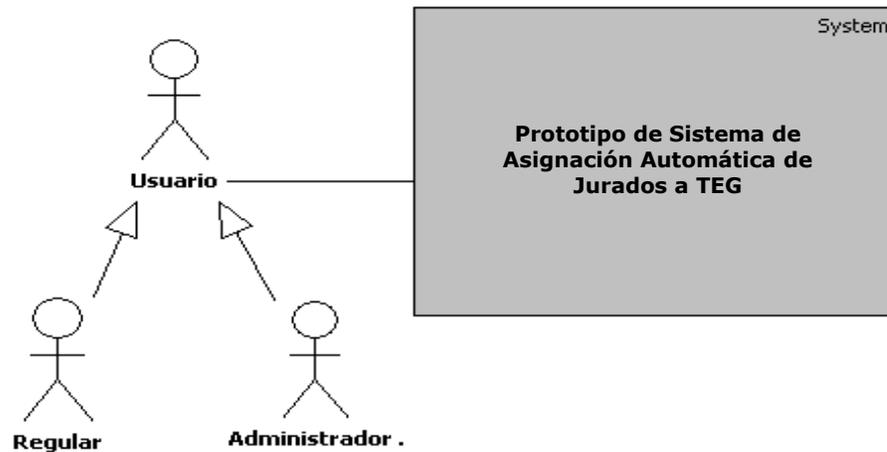


Figura 11. Casos de uso del prototipo - Nivel 0.

Nota. Fuente: Elaboración propia.

Tabla 10. Actores que participan en el prototipo.

Actores	Descripción de los actores
Usuario	Actor que representa a los participantes que interactúan con el prototipo de sistema de asignación de jurados a TEG. Se distinguen dos tipos de usuario: regular y administrador.
Regular	Usuario registrado que utiliza la aplicación, en este caso este usuario corresponde al Jefe de Departamento de la Escuela de Computación.
Administrador	Usuario que tiene pleno acceso al sistema, ya que lo administra y se encarga de su mantenimiento (administrador del sistema).

b.2. Casos de uso - Nivel 1

En este nivel se refleja en términos generales la interacción que tiene cada actor con el prototipo. En la Figura 12, se observa que el usuario regular interactúa con las funcionalidades generales del mismo, a excepción de la funcionalidad administrativa, en cambio, el usuario administrador interactúa con todas las funcionalidades. La descripción de los casos de uso del nivel 1 se expone en la Tabla 11, en la cual se indican los elementos que interactúan en cada caso de uso.

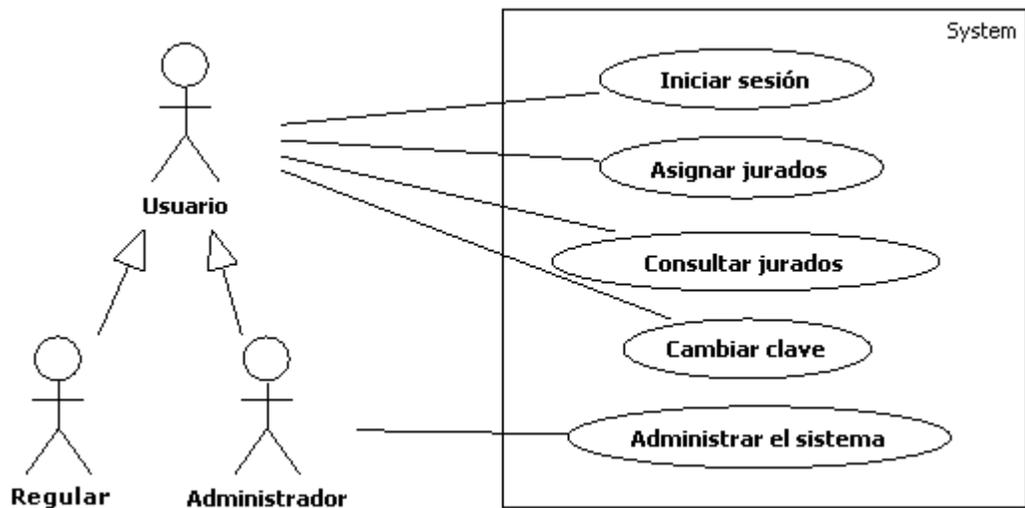


Figura 12. Casos de uso del prototipo - Nivel 1.

Nota. Fuente: Elaboración propia.

Tabla 11. Detalle de los casos de uso del nivel 1.

CASOS DE USO - NIVEL 1	
CASO DE USO 1	
Nombre	Iniciar sesión.
Actor	Usuario (ambos tipos).
Descripción	Permite a los usuarios ingresar al sistema.
Pre-condición	El usuario debe estar registrado en el sistema.
Flujo Básico	<ul style="list-style-type: none"> - El usuario ingresa a la página inicial. - Se solicita su nombre de usuario y clave - Se verifica la validez de los datos. - De tratarse de un usuario y clave válidos, se le da acceso al sistema, de lo contrario, se le notifica la invalidez de los datos introducidos y se le indica la cantidad de intentos restantes. A cada usuario se le permiten 3 intentos inválidos, al cuarto intento inválido se bloquea al usuario.
Post-condición	El usuario puede utilizar las funcionalidades del sistema permitidas según su tipo.
CASO DE USO 2	
Nombre	Asignar jurados.
Actor	Usuario (ambos tipos).
Descripción	Permite al usuario asignar los jurados a un TEG.
Pre-condición	El usuario debe haber iniciado sesión en el sistema y el semestre anterior al actual, el cual se calcula en base a la fecha del sistema, debe encontrarse cerrado.
Flujo Básico	<ul style="list-style-type: none"> - El usuario introduce los datos requeridos del TEG. - El prototipo le recomienda una o dos áreas para la clasificación del mismo. - El usuario confirma las áreas recomendadas o las modifica. - El sistema asigna los jurados al TEG acorde al área o a las áreas finales.

Post-condición	TEG con jurados asignados.
CASO DE USO 3	
Nombre	Consultar jurados.
Actor	Usuario (ambos tipos).
Descripción	Se muestran los jurados asignados a TEG según diferentes criterios de búsqueda.
Pre-condición	El usuario debe haber iniciado sesión y tiene que seleccionar un tipo de búsqueda.
Flujo Básico	<ul style="list-style-type: none"> - El usuario selecciona el tipo de búsqueda de jurados a TEG (por profesor, por estudiante, por semestre, etc.). - El sistema muestra los resultados de acuerdo al tipo de búsqueda seleccionada (de ser válida).
Post-condición	Jurados consultados o ningún resultado (si no hay coincidencias para los parámetros de búsqueda).
CASO DE USO 4	
Nombre	Cambiar clave.
Actor	Usuario (ambos tipos).
Descripción	Se permite al usuario cambiar su propia clave.
Pre-condición	El usuario debe haber iniciado sesión, confirmar la clave actual y colocar una nueva clave válida (de 6 a 12 caracteres alfanuméricos, solo se permiten los siguientes caracteres especiales: `', `/', `-', `_', `=')).
Flujo Básico	<ul style="list-style-type: none"> - El usuario confirma su clave actual e introduce dos veces la nueva clave. - Se validan los datos. - De ser validos, se cambia la clave, de lo contrario, se notifica al usuario el fallo en el cambio.
Post-condición	Clave de usuario cambiada.
CASO DE USO 5	
Nombre	Administrar el sistema.
Actor	Administrador.
Descripción	Se permite al usuario administrador acceder a funcionalidades que afectan el funcionamiento del sistema.
Pre-condición	El usuario debe haber iniciado sesión como usuario administrador.
Flujo Básico	<ul style="list-style-type: none"> - El administrador inicia sesión. - El prototipo le presenta las funcionalidades extras a las que tiene acceso como usuario administrador.
Post-condición	El sistema es modificado para su mantenimiento por parte del administrador.

b.3. Casos de uso - Nivel 2

En este nivel se especifican las funcionalidades del prototipo ofrecidas a los actores del mismo, las cuales se especifican en la Figura 13. Dichas funcionalidades se describen en la Tabla 12.

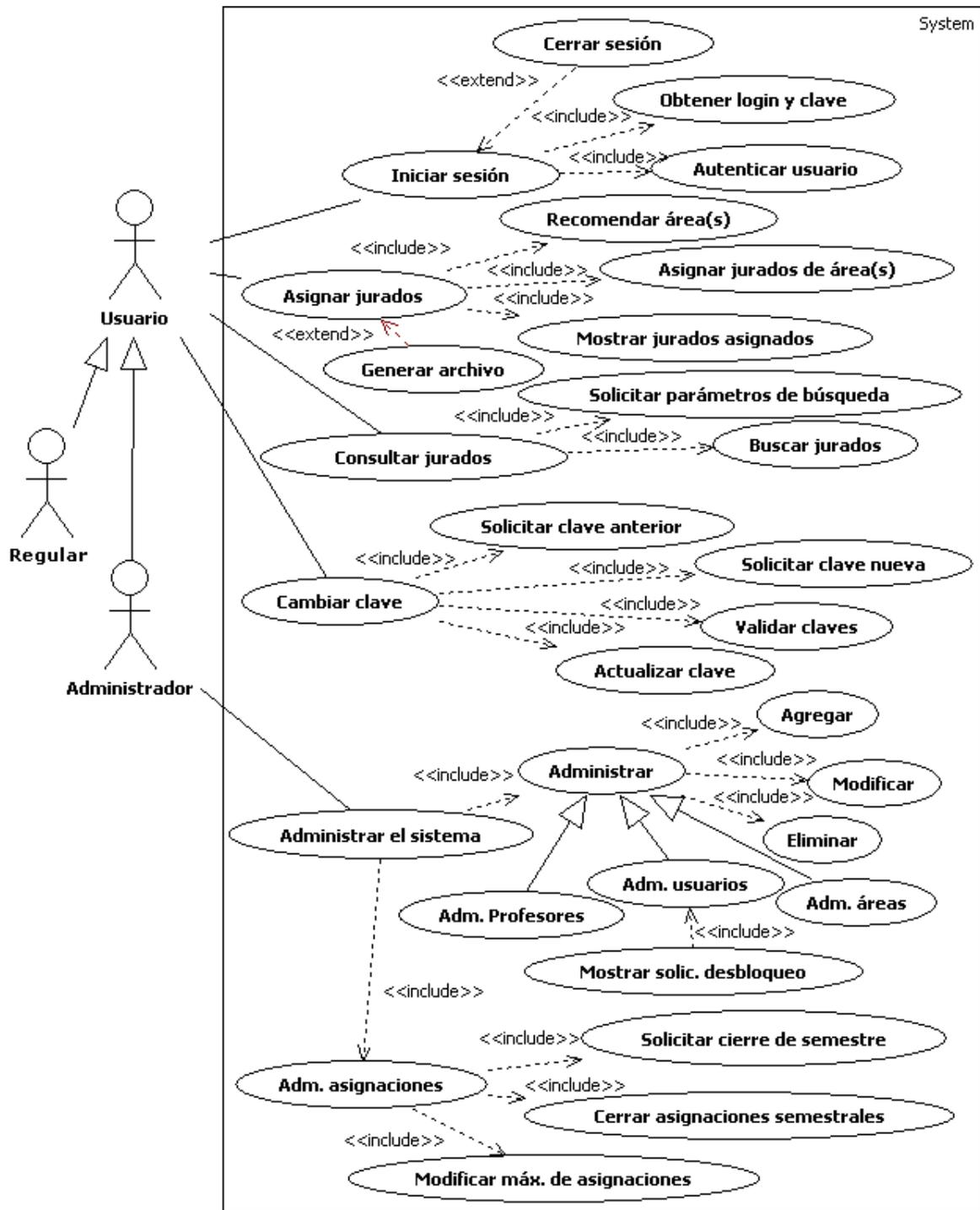


Figura 13. Casos de uso del prototipo - Nivel 2.

Nota. Fuente: Elaboración propia.

Tabla 12. Detalle de los casos de uso del nivel 2.

CASOS DE USO - NIVEL 2	
CASO DE USO 1	
Nombre	Obtener <i>login</i> y clave.
Actor	Usuario (ambos tipos).
Descripción	Se obtienen las credenciales de acceso al sistema.
Pre-condición	N.A.
Flujo Básico	Se solicita el nombre de usuario y la clave, para que una vez ingresados puedan ser validados para poder realizar la autenticación y así tener acceso al sistema.
Post-condición	<i>Login</i> y clave ingresados al sistema para su verificación.
CASO DE USO 2	
Nombre	Autenticar usuario.
Actor	Usuario (ambos tipos).
Descripción	Se validan y verifican las credenciales de acceso.
Pre-condición	El usuario tiene que haber ingresado su <i>login</i> y clave.
Flujo Básico	Primero se verifica que los datos sean válidos (no nulos, alfanuméricos y que no superen el tamaño indicado) y, de serlo, se verifica que exista una coincidencia entre el <i>login</i> y la clave dada en la base de datos. De ser incorrectos los datos, se verifica la cantidad de intentos inválidos que han sido efectuados por el usuario, si esa cantidad supera los 3 intentos inválidos el usuario es bloqueado.
Post-condición	Se permite el acceso al sistema o se regresa a la página de inicio de sesión y en caso de tratarse del último intento inválido el usuario queda bloqueado.
CASO DE USO 3	
Nombre	Cerrar sesión.
Actor	Usuario (ambos tipos).
Descripción	Se cierra la sesión del usuario.
Pre-condición	Sesión iniciada.
Flujo Básico	El usuario solicita el cierre de su sesión, la misma se cierra y se cierra y se regresa a la pantalla de inicio de sesión.
Post-condición	Sesión cerrada.
CASO DE USO 4	
Nombre	Recomendar área(s).
Actor	Usuario (ambos tipos).
Descripción	Se recomienda una o dos áreas bajo las cuales el TEG puede ser clasificado.
Pre-condición	Haber ingresado los datos requeridos del TEG (título, resumen y palabras clave)
Flujo Básico	El usuario ingresa los datos del TEG, el sistema los procesa y aplica el modelo definido por el clasificador del sistema, el cual produce como salida una o dos áreas que son recomendadas al usuario, y se solicita su confirmación.
Post-condición	Área(s) recomendada(s) al usuario.
CASO DE USO 5	
Nombre	Asignar jurado de área(s).
Actor	Usuario (ambos tipos).
Descripción	Se asignan profesores del área(s) como jurados al TEG
Pre-condición	El TEG debe haber sido clasificado y confirmadas el área o áreas a considerar.
Flujo Básico	Una vez que ha sido confirmada o modificada el o las áreas

	recomendadas, se seleccionan los profesores del área o áreas finales de acuerdo a los criterios de elección preestablecidos como jurados (principales y suplentes) del TEG.
Post-condición	Jurado asignado al TEG.
CASO DE USO 6	
Nombre	Mostrar jurados asignados.
Actor	Usuario (ambos tipos).
Descripción	Se muestran por pantalla los jurados asignados.
Pre-condición	Realizar la asignación de jurados.
Flujo Básico	Como salida de la asignación se muestran por pantalla los profesores seleccionados como jurados del TEG.
Post-condición	Jurados mostrados por pantalla.
CASO DE USO 7	
Nombre	Generar archivo.
Actor	Usuario (ambos tipos).
Descripción	Se genera un archivo con la información de la asignación de jurados al TEG.
Pre-condición	Haberse asignado y mostrado los jurados del TEG.
Flujo Básico	Una vez mostrados los jurados asignados se realiza un archivo con la información de los mismos y del TEG.
Post-condición	Archivo contenedor de los jurados asignados al TEG generado.
CASO DE USO 8	
Nombre	Solicitar parámetros de búsqueda.
Actor	Usuario (ambos tipos).
Descripción	Se solicitan los parámetros necesarios para la búsqueda de los jurados a consultar.
Pre-condición	N.A.
Flujo Básico	Se solicita al usuario la información necesaria para consultar las asignaciones de jurados, como: el semestre, profesor, estudiante y el tipo de jurados
Post-condición	Parámetros de búsqueda configurados.
CASO DE USO 9	
Nombre	Buscar jurados.
Actor	Usuario (ambos tipos).
Descripción	Se consultan los jurados de acuerdo a los parámetros de búsqueda establecidos.
Pre-condición	Haber establecido los parámetros de búsqueda.
Flujo Básico	En base a los parámetros de búsqueda seleccionados por el usuario, el sistema realiza la consulta a la BD y muestra el resultado de la misma al usuario.
Post-condición	Asignaciones de jurados a TEG mostradas o ningún resultado (si no hay coincidencias para los parámetros de búsqueda).
CASO DE USO 10	
Nombre	Solicitar clave anterior.
Actor	Usuario (ambos tipos).
Descripción	Solicita la clave actual del usuario.
Pre-condición	El usuario debe estar registrado y haber iniciado sesión.
Flujo Básico	Se pide al usuario que ingrese la clave actual para verificarla antes de permitirle cambiar la misma.
Post-condición	Clave actual ingresada.
CASO DE USO 11	
Nombre	Solicitar clave nueva.
Actor	Usuario (ambos tipos).

Descripción	Solicita la clave actual del usuario.
Pre-condición	El usuario debe estar registrado y haber iniciado sesión.
Flujo Básico	Se pide al usuario que ingrese la clave nueva para validarla antes de actualizar la misma.
Post-condición	Clave nueva ingresada.
CASO DE USO 12	
Nombre	Validar claves.
Actor	Usuario (ambos tipos).
Descripción	Verificación y validación de las claves.
Pre-condición	Deben haberse ingresado la clave actual y la nueva.
Flujo Básico	Primero se verifica que la clave actual sea correcta, de ser así se valida la nueva clave.
Post-condición	Claves verificadas y validadas.
CASO DE USO 13	
Nombre	Actualizar claves.
Actor	Usuario (ambos tipos).
Descripción	Actualización de la clave del usuario.
Pre-condición	Clave actual correcta y clave nueva válida.
Flujo Básico	Una vez verificadas y validadas las claves, se procede a actualizar la clave del usuario. En caso contrario, se notifica al usuario sobre el error.
Post-condición	Clave cambiada o notificación de error.
CASO DE USO 14	
Nombre	Administrar.
Actor	Administrador.
Descripción	Administración de la información de profesores, usuarios y áreas.
Pre-condición	Ser usuario administrador.
Flujo Básico	Se ofrece la opción de administración de profesores, usuarios y áreas al administrador del sistema, la cual se resumen en la posibilidad de agregar, modificar o eliminar información relacionada con éstos. En cuanto a los usuarios, además se verifica si hay solicitudes de desbloqueo pendientes y, de ser así, permite realizar esta acción.
Post-condición	Información de profesores, usuarios y áreas actualizada.
CASO DE USO 15	
Nombre	Administrar asignaciones.
Actor	Administrador.
Descripción	Modificación de información relacionada con las asignaciones de jurados a TEG.
Pre-condición	Ser usuario administrador.
Flujo Básico	Se presenta al administrador las opciones relacionadas con las asignaciones de jurados a TEG, que son: la solicitud automática de cierre de semestre (reseteo de asignaciones de jurados) de acuerdo a la fecha del sistema, la ejecución del cierre, y la modificación de un parámetro relacionado con los criterios de selección de jurados dentro de un área, que es el máximo de asignaciones permitidas para un semestre (para todos los profesores).
Post-condición	Opciones de asignaciones de jurados actualizadas.

c. Interfaces principales

A continuación, las Figuras 14 a la 18 presentan las interfaces correspondientes a las funcionalidades principales del prototipo de asignación de jurados, relacionadas con los casos de usos asignar jurados y consultar jurados.

The screenshot shows the SAJ TEG web application interface. At the top, there is a navigation bar with the SAJ TEG logo and the title 'SISTEMA DE ASIGNACIÓN DE JURADOS A TEG'. Below the navigation bar, there is a user menu with options like 'Inicio', 'Contáctenos', and 'Créditos'. The main content area is titled 'Datos del TEG' and contains a form for assigning jurors to a student's thesis. The form includes fields for 'Tutor 1', 'Tutor 2', and 'Tutor 3', a 'Cantidad de Autores' selector, and a 'Datos del Estudiante 1' section with fields for 'Cédula', 'Hombres', 'Apellidos', and 'Email'. A 'Resumen' section contains a text area with a description of the thesis topic: 'MINERÍA DE TEXTOS PARA LA ASIGNACIÓN AUTOMÁTICA DE JURADOS A TRABAJOS ESPECIALES DE GRADO'. Below the text area, there is a 'Palabras Claves' field with the text 'Minería de textos, clasificación, Porter Stemming, k vecinos, redes neuronales, Weka, JavaServer Pages (JSP)'. At the bottom of the form, there are three buttons: 'Borrar', 'Siguiente', and 'Cancelar'. The footer of the page contains the text: 'Universidad Central de Venezuela | Fac. de Ciencias | Esc. de Computación | SAJ - Sistema de Asignación de Jurados a TEG | Optimizado para los navegadores Explorer 8 y Firefox 5 | Resolución recomendada: 1024x768'.

Figura 14. Interfaz del módulo de recomendación de áreas – Paso 1.

Nota. Fuente: Elaboración propia.



Figura 15. Interfaz del módulo de recomendación de áreas – Paso 2.

Nota. Fuente: Elaboración propia.



Figura 16. Interfaz del módulo de asignación de jurados – Paso 3.

Nota. Fuente: Elaboración propia.

The screenshot displays the SAJ TEG web application interface. At the top, there is a header with the SAJ TEG logo and a navigation bar containing links for Inicio, Contáctenos, and Créditos. The current page is titled 'Consultar Jurados - Paso 1'. On the left, there is a 'Menú de Opciones' and an 'Administración' section. The main content area features a form titled 'PARÁMETROS DE LA CONSULTA' with the following fields:

- Período: Seleccione un período
- Profesor: Seleccione un Profesor
- Tipo de Jurado: Seleccione un tipo de jurado
- Estudiante: Seleccione un estudiante

The 'Estudiante' dropdown menu is open, showing a list of names: Adriana Cova, César Cortés, Dayana R. Torres N., Edsel Hernández, and Vanessa C. Torres N. The footer contains the text: 'Universidad Central de Venezuela | Fac. de Ciencias | Esc. de Computación | SAJ - Sistema de Asignación de Jurados a TEG | Optimizado para los navegadores IE Explorer 6 y Firefox 5 | Resolución recomendada: 1024x768'.

Figura 17. Interfaz del módulo de consulta – Paso 1.

Nota. Fuente: Elaboración propia.

SISTEMA DE ASIGNACIÓN DE JURADOS A TEG

Inicio | Contáctenos | Créditos

Caracas, 01 de mayo de 2011, 09:26 am

Usuario: admin | Cerrar Sesión

Inicio > Consultar Jurados - Paso 1 > Consultar Jurados - Paso 2

Menú de Opciones

- Asignar Jurados
- Consultar Jurados
- Cantidad Asignaciones
- Cambiar Clave
- Cerrar Sesión

Administración

- Configuración General
- Opciones Profesores
- Opciones Profesor-Area
- Opciones Usuarios
- Opciones Áreas

Datos del Jurado

Fecha Asignacion de Jurado: 03-06-2011

C.I. PROFESOR	NOMBRE PROFESOR	TIPO JURADO
11035937	Marcel José Castro González	Principal 1
5073618	Nora Elena Montaña Fermin	Principal 2
9663887	Antonio Mauricio Silva Sprock	Suplente 1
10334608	Iván José Flores Vitelli	Suplente 2

Datos de Autores

Cédula Autor 1: 187089157 Nombre Autor 1: Dayana R. Torres N.
 Email Autor 1: dayana.torres@ciens.ucv.ve

Datos de Tutores

Tutor 1: Esmeralda Ramos Guanchez C.I. 4681866
 Tutor 2: Haydemar María Nuñez Castro C.I. 5538772

Datos de la tesis

Título de la Tesis: Minería de textos para la asignación automática de jurados a Trabajos Especiales de Grado

Resumen de la Tesis: El presente trabajo tiene como objetivo el desarrollo de un sistema automatizado basado en técnicas de minería de textos para la asignación de jurados a Trabajos Especiales de Grado (TEG), a través de la clasificación de los mismos de acuerdo a las áreas profesionales a las cuales pertenece el personal docente de la Escuela de Computación. Para la creación del clasificador se aplica el proceso de minería de textos sobre una recopilación de documentos de TEG, cuyas

Palabras Claves de la Tesis: Minería de textos, clasificación, Porter Stemming, k vecinos, redes neuronales, Weka, JavaServer Pages (JSP).

Generar Archivo

Universidad Central de Venezuela | Fac. de Ciencias | Esc. de Computación | SAJ - Sistema de Asignación de Jurados a TEG
 Optimizado para los navegadores: Explorer 8 y Firefox 5 | Resolución recomendada: 1024x768

Figura 18. Interfaz del módulo de consulta – Paso 2.

Nota. Fuente: Elaboración propia.

2.5.7. Pruebas y resultados

Dado que el prototipo de asignación de jurados a Trabajos Especiales de Grado, se compone de dos módulos principales: el módulo de recomendación de áreas y el módulo de asignación de jurados, las pruebas realizadas al prototipo se centran en la el cálculo del error de clasificación presentado por el primero, y en verificar que las asignaciones de jurados realizadas por el segundo se apeguen a las reglas de elección establecidas.

Para verificar y validar los resultados que ofrece el módulo de recomendación de áreas, se realizaron diversas pruebas que consistieron en clasificar 23 documentos de TEG de las áreas consideradas, que no fueron utilizados durante la generación del

modelo de clasificación. La distribución de los documentos por áreas se observa en la Tabla 13. El conjunto de prueba completo se detalla en el Anexo 6.

Tabla 13. Distribución de los documentos de prueba por áreas.

Áreas	Cantidad de Documentos
ATI	9
Redes	6
BD	1
IA	2
ATI y BD	5
Total	23

Los resultados de la evaluación del módulo de recomendación de áreas del prototipo se resumen en la Tabla 14, en la cual se comparan las áreas de los documentos de TEG con las recomendadas por el prototipo. Se consideran correctas las clasificaciones que consideran el área del TEG o al menos una de sus áreas, en el caso de pertenecer a más de una (ATI y BD).

Tabla 14. Resultados de las pruebas realizadas al prototipo.

Nro. TEG	Área del TEG	Área Recomendada	Tipo Clasificación
1	ATI	ATI	Correcta
2	REDES	REDES y BD	Correcta
3	ATI	ATI	Correcta
4	REDES	REDES	Correcta
5	ATI y BD	BD y ATI	Correcta
6	REDES	REDES	Correcta
7	ATI y BD	ATI	Correcta
8	ATI	ATI	Correcta
9	REDES	REDES e IA	Correcta
10	REDES	REDES	Correcta
11	ATI	ATI	Correcta
12	ATI	ATI y BD	Correcta
13	ATI	ATI	Correcta
14	IA	REDES	Incorrecta
15	REDES	REDES	Correcta
16	ATI	BD y ATI	Correcta
17	BD	BD	Correcta
18	BD y ATI	ATI y BD	Correcta
19	ATI	ATI	Correcta
20	ATI y BD	BD y ATI	Correcta
21	ATI	ATI	Correcta
22	ATI y BD	BD	Correcta
23	IA	BD e IA	Correcta

En las pruebas realizadas se alcanza un total de documentos clasificados correctamente del 95,65%, mayor que el mejor resultado obtenido durante la estimación del modelo, de 82,32%, en la cual solo se consideraba un área por documento (Tabla 15). También se evidencia que IA es el área para la cual el clasificador reporta menor precisión, al ser la única área que presenta una clasificación incorrecta. Se observa un buen rendimiento por parte del clasificador, puesto que la mayoría de las áreas recomendadas incluyen al menos a un área a la cual el TEG pertenece.

Tabla 15. Comparación de los resultados obtenidos al estimar el modelo y los alcanzados por el prototipo.

Mejor resultado obtenido durante la estimación del modelo de clasificación			Pruebas realizadas al Prototipo	
Clasificaciones correctas	163	82,323%	22	95,652 %
Clasificaciones incorrectas	35	17,676%	1	4,347 %
Nro. Total Instancias	198	100%	23	100%

A continuación, a manera de ejemplo, se presenta una de las pruebas realizadas al módulo de asignación de jurados, las cuales consistieron en inicializar en cero la cantidad de asignaciones de un grupo de profesores de cada área y asignar los jurados de cada documento del conjunto de prueba. Las pruebas completas se presentan en el Anexo 7. Este ejemplo corresponde al documento de prueba Nro. 12.

Tabla 16. Estado de los profesores de las áreas recomendadas para el TEG de prueba Nro. 12.

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	1	Andrés Sanoja	ATI
2	1	Eleonora Acosta	
2	2	Yusneyi Carballo	
2	2	Andrés Castro	
1	4	Jossie Zambrano	
2	1	Sergio Rivas	
0	0	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	1	Antonio Silva	
0	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	

Tabla 17. Jurados asignados al TEG de prueba Nro. 12.

Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI	ATI y BD	Yusneyi Carballo	Prof. Jossie Iranidis Zambrano	Profa. Rossana Díaz Salguero	Prof. Andrés Fernando Sanoja Vargas	Prof. Wilfredo Andres Rangel Azuaje

Haciendo un seguimiento de la cantidad de asignaciones de los docentes, tanto como jurados principales como suplentes, se corroboró que los jurados seleccionados son los que corresponden según las reglas de elección establecidas (Ver Anexo 5).

CONCLUSIONES Y RECOMENDACIONES

Se desarrolló un prototipo que permite la asignación automática de jurados a TEG, recomendando una o dos áreas (máximo) bajo las cuales el mismo puede ser clasificado a través de la aplicación de técnicas de minería de textos (MT), lográndose el objetivo general propuesto en este TEG.

Según los resultados alcanzados por el clasificador del prototipo, de un 95% de instancias clasificadas correctamente, también se concluye que resulta más efectiva la consideración de más de un área a aplicar MT para clasificar automáticamente un TEG, puesto que en muchos casos los documentos no tienen relación con una sola área. Esto se ve reflejado en el incremento logrado en cuanto al porcentaje de clasificaciones correctas, a comparación con el mejor resultado obtenido en las pruebas realizadas con la herramienta Weka para la estimación del modelo de clasificación aplicado, de un 82%, en las cuales los clasificadores utilizados asignaban una sola área a cada documento. Como consecuencia de esta disminución del error de clasificación, se evidencia que el clasificador logra una mayor precisión, así como el hecho de que casi la totalidad de los documentos que debían ser considerados bajo un área efectivamente fueron recomendados para dicha área, indica un aumento del *recall* global del sistema.

Además, queda demostrada la efectividad de la representación de documentos por medio del modelo de espacio vectorial mediante la aplicación de técnicas de indexación más sofisticadas que la frecuencia cruda de la palabra o el pesado booleano, como la entropía. Asimismo, se evidencian las ventajas derivadas de la reducción de la dimensionalidad del espacio de características a través de técnicas como InfoGain, X^2 y CC, que en este trabajo permitieron que de un total de 3.747 raíces de palabras informativas se creara el modelo utilizando una selección de tan solo 41 de éstas, con las cuales se obtuvieron mejores prestaciones para el clasificador que considerando el conjunto completo de raíces. Principalmente, los resultados de este trabajo confirman el buen rendimiento reportado en la literatura del método K vecinos más cercanos para la clasificación de textos, a pesar de ser un algoritmo sencillo de implementar y de bajo costo computacional.

El histórico de asignaciones pone a la disposición información en digital, actualizada y de fácil consulta sobre los profesores asignados como jurados de los TEG, su disponibilidad y los autores de los TEG, mediante un sistema de búsqueda.

La principal ventaja del módulo de asignación de jurados reside en que permite que la toma de decisiones con respecto a la elección de los profesores como jurados pase de ser un proceso subjetivo y discrecional analizado en un Consejo, a ser un proceso automatizado que garantiza la escogencia de los mismos de acuerdo a los criterios establecidos, que busca garantizar una elección equitativa con igualdad de condiciones.

Por otro lado, aunque el prototipo ofrece la opción de seleccionar el (las) área(s) de un TEG según la preferencia del usuario, obviando la recomendación dada por el clasificador, con el fin de permitir la asignación de jurados a TEG pertenecientes a todas las áreas de la Escuela de Computación, se recomienda como trabajo a futuro la inclusión de las áreas no consideradas por el mismo. Este TEG detalla el proceso a seguir al realizar MT y especifica un modelo de clasificación que ha demostrado ser efectivo, por lo que la inclusión de las áreas restantes implicaría básicamente la recolección de una cantidad considerable de documentos de dichas áreas para sumarlos al conjunto de documentos de entrenamiento utilizado. La recolección de documentos de TEG constituyó la principal limitación en este trabajo, puesto que la Escuela de Computación no contaba con un sistema de gestión de TEG con información en digital de los mismos, por lo que la mayoría de los documentos recopilados fueron obtenidos en físico de la Biblioteca Alonso Gamero, lo que significó un lento y tedioso trabajo de recopilación. Debido a que el prototipo desarrollado también lleva un histórico de las asignaciones de jurados a TEG con la información del mismo, su uso facilita la recolección futura de información en digital relacionada con los TEG de las diferentes áreas, que podría contribuir con la ampliación del mismo.

De igual modo, se propone la estimación del modelo empleando otros algoritmos de clasificación, para la comparación de los resultados obtenidos. Específicamente, se recomiendan los algoritmos Simple de Bayes y máquinas de

soporte vectorial, ya que arrojaron buenos resultados (superiores al 80% de precisión predictiva) en pruebas realizadas durante la generación del modelo.

REFERENCIAS

Aas, K y Eikvil, L. (1999). *Text Categorisation: A Survey*. Extraído el 19 de Mayo de 2011 desde http://www.cis.uni-muenchen.de/kurse/pmaier/ML_05/material/aas99text.pdf

Anand, T. y Brachman, R. (1996). The Process of Knowledge Discovery in Databases: A Human-Centered Approach. En Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., y Uthurusamy, R. (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 37–58). USA: AAAI Press.

Apache, Tomcat. (s.f.). Extraído el 21 de Febrero de 2011 desde <http://tomcat.apache.org/tomcat-6.0-doc/index.html>

Bogado, V., Dapozo, G., López, M., Porcel, E. (2007). *Técnicas de preprocesamiento para mejorar la calidad de los datos en un estudio de caracterización de ingresantes universitarios*. Extraído el 16 de Octubre de 2010 desde <http://www.ing.unp.edu.ar/wicc2007/trabajos/ISBD/111.pdf>

Buxton, B., Corney, D., Jones, D., y Langdon, W. (2004). BioRAT: extracting biological information from full-length papers. *Bioinformatics*, 20(17), 3206–3213. Extraído el 23 de Julio de 2010 desde <http://bioinformatics.oxfordjournals.org/content/20/17/3206.full.pdf>

Cueva, J. (1999). *Introducción a UML*. Extraído el 12 de marzo de 2011 desde <http://gidis.ing.unlpam.edu.ar/downloads/pdfs/IntroduccionUML.PDF>

De San Pedro, M. y Lasso, M. (2006). *Aprendizaje Automático y Minería de Datos. Teoría 5: Evaluación de modelos*. Extraído el 17 de Marzo de 2011 desde <http://www.dirinfo.unsl.edu.ar/~aamd/Teorias/teo5md4.pdf>

Fayyad, U., Piatetsky-Shapiro, G. y Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37-54. Extraído el 13 de Noviembre de

2010 desde <http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>

Fuentes, O., Montes, M., Téllez, A., y Villaseñor, L. (2003). *Clasificación Automática de Textos de Desastres Naturales en México*. Trabajo presentado en el 10º Congreso Internacional de Investigación en Ciencias Computacionales, 263-269. Octubre, México.

García, D. (s.f.). *Manual de Weka*. Extraído el 12 de Mayo de 2011 desde <http://www.metaemotion.com/diego.garcia.morate/download/weka.pdf>

Gracia, J. (2003). *UML: Casos de Uso. Use case*. Extraído el 11 de Mayo de 2011 desde <http://www.ingenierosoftware.com/analisisydiseno/casosdeuso.php>

Jain, A. y Li, Y. (1998). *The Computer Journal*. Extraído el 16 de Enero de 2011 desde <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.100.7400>

Joachims, T. (1998). *Text categorization with support vector machines: Learning with many relevant features*. Proc. 10th European Conference on Machine Learning (ECML), Springer Verlag. Extraído el 3 de Septiembre de 2010 desde <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.117.4939&rep=rep1&type=pdf>

La Red, D., Peláez, J. y Sánchez, P. (2002). Un Clasificador de Texto Por Aprendizaje. *Revista Iberoamericana de Inteligencia Artificial*, 6(15), 53-59. Extraído el 20 de Julio de 2010 desde <http://exa.unne.edu.ar/depar/areas/informatica/SistemasOperativos/ClasTextoApre.pdf>

Langley, P. y Simon, H. (1995). Applications of Machine Learning and Rule Induction. *Communications of the ACM*, 38, 55-64. Obtenido el 4 de Septiembre de 2010 desde la base de datos de Portal ACM.

Montes, M. (2001) *Minería de texto: Un nuevo reto computacional*. Extraído el 7 de Marzo de 2011 desde <http://ccc.inaoep.mx/~mmontesg/publicaciones/2001/MineriaTexto-md01.pdf>

MySQL, Panorámica del sistema de gestión de base de datos MySQL. (s.f.). Extraído el 13 de Diciembre de 2011 desde <http://dev.mysql.com/doc/refman/5.0/es/what-is.html>

Oracle, Java BluePrints, Model-View-Controller. (s.f.). Extraído el 15 de marzo de 2011 desde <http://java.sun.com/blueprints/patterns/MVC-detailed.html>

Oracle, Java SE 6 Documentation. (s.f.). Extraído el 13 de Septiembre de 2010 desde <http://download.oracle.com/javase/6/docs/index.html>

Package weka classifiers functions. (s.f.). Extraído el 10 de Febrero de 2011 desde <http://weka.sourceforge.net/doc/weka/classifiers/functions/package-summary.html>

Pedersen, J., Weigend A. y Wiener, E. (1995). *A neural network approach to topic spotting*. Proceedings of SDAIR-95, 4th Annual Symposium on Document Analysis and Information Retrieval, Las Vegas, 317-332. Extraído el 21 de Marzo de 2010 desde <http://www.jopedersen.com/Publications/wiener95neural.pdf>

Porter, M. (1980). *An Algorithm for Suffix Stripping*. Extraído el 7 de Enero de 2011 desde <http://www.cs.odu.edu/~jbollen/IR04/readings/readings5.pdf>

Quinlan, J. (1993). *C4.5: Programming for machine learning*. Extraído el 29 de Marzo de 2011 desde <http://www.springerlink.com/content/v986m1562062hk51/>

Sebastiani, F. (1999). *A Tutorial on Automated Text Categorisation*. Extraído el 16 de Marzo de 2010 desde <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.122.670>

The Stemming Algorithm. (s.f.). Extraído el 3 de Febrero de 2011 desde <http://snowball.tartarus.org/algorithms/spanish/stemmer.html>

Wikipedia, Eclipse (software). (s.f.). Extraído el 11 de Enero de 2011 desde http://es.wikipedia.org/wiki/Eclipse_%28software%29

Wikipedia, JAVA SE. (s.f.). Extraído el 16 de Octubre de 2010 desde http://es.wikipedia.org/wiki/Java_SE

ANEXOS

Anexo 1. Descripción del formato de archivos “.arff” (García, D.).

Nativamente Weka trabaja con un formato denominado arff, acrónimo de Attribute-Relation File Format. Este formato está compuesto por una estructura claramente diferenciada en tres partes:

1. **Cabecera.** Se define el nombre de la relación. Su formato es el siguiente:

```
@relation <nombre-de-la-relación>
```

Donde <nombre-de-la-relación> es de tipo String*. Si dicho nombre contiene algún espacio será necesario expresarlo entrecomillado.

2. **Declaraciones de atributos.** En esta sección se declaran los atributos que compondrán nuestro archivo junto a su tipo. La sintaxis es la siguiente:

```
@attribute <nombre-del-atributo> <tipo>
```

Donde <nombre-del-atributo> es de tipo String teniendo las mismas restricciones

que el caso anterior. Weka acepta diversos tipos, estos son:

a) **NUMERIC** Expresa números reales**.

b) **INTEGER** Expresa números enteros.

c) **DATE** Expresa fechas, para ello este tipo debe ir precedido de una etiqueta de formato entrecomillada. La etiqueta de formato está compuesta por caracteres separadores (guiones y/o espacios) y unidades de tiempo:

- dd Día.
- MM Mes.
- yyyy Año.
- HH Horas.
- mm Minutos.
- ss Segundos.

d) **STRING** Expresa cadenas de texto, con las restricciones del tipo String comentadas anteriormente.

e) **ENUMERADO** El identificador de este tipo consiste en expresar entre llaves y separados por comas los posibles valores (caracteres o cadenas de caracteres) que puede tomar el atributo. Por ejemplo, si tenemos un atributo que indica el tiempo podría definirse:

```
@attribute tiempo {soleado,lluvioso,nublado}
```

3. **Sección de datos.** Declaramos los datos que componen la relación separando entre comas los atributos y con saltos de línea las relaciones.

```
@data
```

```
4,3.2
```

*Entendiendo como tipo String el ofrecido por Java.

**Debido a que Weka es un programa Anglosajón la separación de la parte decimal y entera de los números reales se realiza mediante un punto en vez de una coma.

Aunque éste es el modo "completo" es posible definir los datos de una forma abreviada (sparse data). Si tenemos una muestra en la que hay muchos datos que sean 0 podemos expresar los datos prescindiendo de los elementos que son nulos, rodeando cada una de las filas entre llaves y situando delante de cada uno de los datos el número de atributo***.

Un ejemplo de esto es el siguiente

```
@data
{1 4, 3 3}
```

En este caso hemos prescindido de los atributos 0 y 2 (como mínimo) y asignamos al atributo 1 el valor 4 y al atributo 3 el valor 3.

En el caso de que algún dato sea desconocido se expresará con un símbolo de cerrar interrogación ("?").

Es posible añadir comentarios con el símbolo "%", que indicará que desde ese símbolo hasta el final de la línea es todo un comentario. Los comentarios pueden situarse en cualquier lugar del fichero.

Un ejemplo de un archivo de prueba.

```
prueba.arff
1 % Archivo de prueba para Weka.
2 @relation prueba
3
4 @attribute nombre STRING
5 @attribute ojo_izquierdo {Bien,Mal}
6 @attribute dimension NUMERIC
7 @attribute fecha_analisis DATE "dd-MM-yyyy HH:mm"
8
9 @data
10 Antonio,Bien,38.43,"12-04-2003 12:23"
11 'Maria Jose',?,34.53,"14-05-2003 13:45"
12 Juan,Bien,43,"01-01-2004 08:04"
13 Maria,?,?, "03-04-2003 11:03"
```

***La numeración de atributos comienza desde el 0, por lo que primero es el 0 y no el 1.

Anexo 2. Subgrupos de atributos obtenidos por cada algoritmo de medición de relevancia de términos para el archivo de indexación mediante pesado por entropía.

Algoritmo	InfoGain (AttributeEval)	GainRatio (AttributeEval)	ChiSquared (AttributeEval)	Consistency (SubsetEval)	Cfs (SubsetEval)	Ranking de Atributos
Método de Búsqueda	Ranker	Ranker	Ranker	BestFirst	BestFirst	
Método de Evaluación	Cross- Validation (10 particiones)					
Pesado Por entropía	inalambr	inalambr	inalambr	web	web	1
	web	clasificacion	clasificacion	clasificacion	siti	2
	clasificacion	802	disposit	manej	clasificacion	3
	disposit	human	web	human	dat	4
	human	artificial	human	inalambr	red	5
	red	iee	802	dat	ruby	6
	802	sem	red	ruby	bas	7
	artificial	bluetooth	artificial	administr	manej	8
	agent	neuronal	inteligent	disposit	metod	9
	manej	siti	agent	javascript	disposit	10
	inteligent	wlan	ruby	protocol	protocol	11
	ruby	11	siti	configuracion	human	12
	siti	agent	manej	segur	conoc	13
	bas	distribu	movil	red	profesional	14
	domini	ruby	iee	agil	domini	15
	movil	especi	conoc	prototip	802	16
	iee	cabl	rails	xp	inalambr	17
	dat	ontologi	expert	siti	agent	18
	rails	network	sem	cre	artificial	19
	expert	mineri	dat	tecnologi	distribu	20
	conoc	morfologi	protocol	proyect	expert	21
	sem	inteligent	domini	recuper	inteligent	22
	on	disposit	bluetooth	especi	bluetooth	23
	protocol	seminal	bas	distribu	neuronal	24
	bluetooth	analisis	neuronal	conoc	segur	25
	neuronal	resolucion	on	profesional	movil	26
	metod	liqu	profesional	expert	prototip	27
	profesional	fabric	conect	inteligent	mineri	28
	prototip	descript	segur	reforz	javascript	29
	segur	rails	11	desarroll	tecnologi	30
	conect	pda	wlan	tecnic	administr	31
	administr	movil	distribu	objet	recuper	32
	distribu	conoc	metod	usabil	descript	33
	11	ip	regl	codig	ontologi	34
	wlan	regl	cabl	metod	xp	35
	regl	on	prototip	apoy	academ	36
	usabil	representacion	morfologi	iee	proyect	37
	cabl	expert	mineri	regl	codig	38
	especi	domini	ontologi	aplic	configuracion	39
	mineri	javascript	especi	hoc	laboratori	40
	morfologi	web	network	object	desarroll	41
	network	ataqu	ip	extrem	cre	42
	ontologi	consistent	fabric	cienci	on	43
	ip	diagnost	usabil	academ	objet	44
	diagnost	ipv6	analisis	conest	agil	45
	seminal	usabil	seminal	bibliotec	conjunt	46

analís	conect	liqu	característ	generación	47
fabric	profesional	resolución	postgr	organiz	48
liqu	bas	descript	distribuid	ipv6	49
resolución	sustent	diagnost	control	manipul	50
representación	protocol	administr	are	regl	51
javascript	red	representación	eficient	especial	52
descript	segur	javascript	estánd	cienci	53
interacción	metod	pda	visualización	usabil	54
ataqu	manej	ataqu	mensaj	reingenieri	55
consistent	dat	consistent	organiz	seguimient	56
pda	prototip	interacción	gui	mysql	57
sustent	administr	ipv6	empres	are	58
ipv6	investig	sustent	fundament	adicional	59
investig	caract	investig	domini	automatización	60
divulgación	indol	indol	autenticación	mensaj	61
indol	destin	vuelte	explicación	empres	62
vuelte	divulgación	interrelación	ido	band	63
destin	notación	ideal	artificial	oos	64
interrelación	egres	oos		consistent	65
oos	jubil	divulgación		decisión	66
ven	cultural	caract		fundament	67
ideal	científ	destin		dirección	68
caract	vincul	científ		asoci	69
científ	deport	distant		autenticación	70
vincul	interrelación	transf		ip	71
deport	vuelte	hipermedi		comienz	72
cultural	ideal	ven		físic	73
median	hipermedi	geograf		infraestructur	74
atención	transf	atención		comunic	75
hipermedi	wid	transparent		resolución	76

Anexo 3. Subgrupos de atributos obtenidos por cada algoritmo de medición de relevancia de términos para el archivo de indexación mediante pesado ltc.

Algoritmo	InfoGain (AttributeEval)	GainRatio (AttributeEval)	ChiSquared (AttributeEval)	Consistency (SubsetEval)	Cfs (SubsetEval)	Ranking de Atributos
Método de Búsqueda	Ranker	Ranker	Ranker	BestFirst	BestFirst	
Método de Evaluación	Cross-Validation (10 particiones)	Cross-Validation (10 particiones)	Cross-Validation (10 particiones)	Cross-Validation (10 particiones)	Cross-Validation (10 particiones)	
Pesado ltc	inalambr	inalambr	inalambr	web	desarroll	1
	web	clasificacion	disposit	cre	web	2
	clasificacion	802	clasificacion	servici	interaccion	3
	disposit	artificial	web	manej	clasificacion	4
	manej	iee	manej	segur	dat	5
	red	interaccion	red	desarroll	red	6
	human	sem	802	control	cre	7
	802	agent	agent	protocol	servici	8
	agent	neuronal	inteligent	inalambr	ruby	9
	desarroll	bluetooth	artificial	human	manej	10
	artificial	expert	human	dat	metod	11
	inteligent	11	ruby	conoc	disposit	12
	ruby	disposit	expert	red	protocol	13
	expert	wlan	movil	ruby	human	14
	movil	ip	desarroll	administr	expert	15
	siti	siti	siti	distribuid	inalambr	16
	protocol	conexion	protocol	disposit	agent	17
	iee	ontologi	iee	iee	artificial	18
	servici	ruby	interaccion	aplic	siti	19
	interaccion	mineri	rails	oracl	control	20
	dat	cabl	dat	fundament	conoc	21
	rails	morfologi	sem	clasificacion	profesional	22
	conoc	especi	bluetooth	configuracion	segur	23
	on	fundament	neuronal	central	fundament	24
	metod	inteligent	on	tecnologi	802	25
	control	oos	conoc	prototip	movil	26
	sem	fabric	segur	profesional	neuronal	27
	segur	analisis	11	apoy	distribu	28
	neuronal	rails	profesional	domini	aplic	29
	bluetooth	transmision	conect	direccion	prototip	30
	prototip	descript	metod	movil	inteligent	31
	profesional	human	domini	aplicacion	bluetooth	32
	domini	movil	cre	siti	laboratori	33
	cre	regl	wlan	proces	administr	34
	11	on	ip	gener	recuper	35
	conect	web	servici	extraccion	tecnologi	36
	wlan	representacion	regl	objet	consistent	37
	ip	metod	prototip	academ	central	38
	regl	manej	control	reingenieri	mysql	39
	usabil	usabil	conexion	proyect	configuracion	40
	cabl	protocol	cabl	conest	oos	41
	conexion	conect	mineri	manipulacion	asoci	42
	morfologi	profesional	morfologi	metod	ontologi	43
	mineri	ipv6	ontologi	conjunt	analisis	44
	ontologi	domini	especi	estand	organiz	45
	especi	red	fundament	apo	plant	46
fundament	segur	oos	mensaj	domini	47	

	oos	conoc	transmision	organiz	direccion	48
	transmision	cre	fabric	oos	r	49
	fabric	desarroll	analís	inteligent	ip	50
	analís	dat	usabil	recuper	manipul	51
	representacion	servici	descript	ipv6	descript	52
	descript	prototip	representacion	agent	proces	53
	ven	control	ipv6	resolucion	gener	54
	ipv6	agrup	atencion	especi	javascript	55
	median	clas	ven		solucion	56
	transparent	egres	transparent		xp	57
	interrelacion	deport	ideal		objet	58
	cultural	cultural	agrup		academ	59
	ideal	remot	vuelt		agil	60
	atencion	ven	caract		usabil	61
	vuelt	median	interrelacion		bas	62
	cientif	divulgacion	median		smbd	63
	agrup	transparent	cientif		oracl	64
	caract	jubil	cultural		distribuid	65
	egres	cientif	jubil		recurs	66
	deport	caract	deport		are	67
	indol	ideal	geograf		eficient	68
	divulgacion	interrelacion	egres		fue	69
	destin	atencion	destin		conjunt	70
	jubil	vuelt	indol		generacion	71
	geograf	destin	dem		apo	72
	plant	indol	distant		mensaj	73
	hipermedi	distant	divulgacion		estructur	74
	distant	usenet	hipermedi		ipv6	75
	clas	geograf	vincul		regl	76
	cooper	transf	transf		mineri	77

Anexo 4. Subgrupos de atributos obtenidos por cada algoritmo de medición de relevancia de términos para el archivo de indexación mediante pesado tfc.

Algoritmo	InfoGain (AttributeEval)	GainRatio (AttributeEval)	ChiSquared (AttributeEval)	Consistency (SubsetEval)	Cfs (SubsetEval)	Ranking de Atributos
Método de Búsqueda	Ranker	Ranker	Ranker	BestFirst	BestFirst	
Método de Evaluación	Cross-Validation (10 particiones)					
Pesado tfc	inalambr	inalambr	inalambr	web	web	1
	web	clasificacion	disposit	servici	siti	2
	disposit	802	web	ruby	clasificacion	3
	clasificacion	artificial	clasificacion	manej	red	4
	manej	siti	manej	inalambr	ruby	5
	red	iee	red	clasificacion	manej	6
	human	sem	human	objet	metod	7
	802	disposit	802	protocol	disposit	8
	agent	agent	agent	human	human	9
	artificial	bluetooth	intelligent	segur	expert	10
	intelligent	neuronal	artificial	oracl	inalambr	11
	ruby	11	siti	fundament	agent	12
	siti	wlan	ruby	desarroll	desarroll	13
	desarroll	expert	expert	administr	dat	14
	expert	conexion	movil	disposit	servici	15
	movil	ruby	iee	aplic	protocol	16
	iee	ontologi	protocol	dat	conoc	17
	rails	morfologi	desarroll	prototip	artificial	18
	conoc	mineri	rails	control	objet	19
	protocol	cabl	sem	intelligent	distribu	20
	metod	especi	segur	iee	profesional	21
	servici	fundament	bluetooth	red	fundament	22
	on	intelligent	neuronal	distribuid	intelligent	23
	sem	analisis	on	ontologi	neuronal	24
	segur	fabric	conoc	interaccion	interaccion	25
	bluetooth	rails	metod	programacion	administr	26
	neuronal	descript	11	configuracion	segur	27
	prototip	r	profesional	asoci	asoci	28
	aplic	transmision	conect	implementacion	802	29
	profesional	ip	domini	tecnologi	aplic	30
	domini	human	wlan	are	movil	31
	11	movil	aplic	interpret	tecnologi	32
	conect	manej	regl	conoc	prototip	33
	wlan	regl	servici	direccion	consistent	34
	regl	web	prototip	ipv6	ontologi	35
	morfologi	on	cabl	movil	bluetooth	36
	usabil	representacion	morfologi	agent	agil	37
	cabl	oos	mineri	aplicacion	programacion	38
	mineri	aplic	conexion	tecnic	mysql	39
	conexion	conect	especi	siti	control	40
	ontologi	segur	ontologi	proces	are	41
	fundament	profesional	fundament	javascript	estructur	42
	especi	protocol	ip	neces	domini	43
	ip	red	transmision	academ	particul	44
	transmision	ipv6	fabric	agil	direccion	45
	fabric	usabil	analisis	mysql	recuper	46
	analisis	domini	usabil	metod	informacion	47

	representacion	conoc	descript	eficient	implementacion	48
	descript	metod	representacion	organiz	javascript	49
	r	prototip	r	gui	bas	50
	oos	servici	oos	prest	manipulacion	51
	asoci	desarroll	asoci	recuper	oracl	52
	vuelt	divulgacion	ipv6	diagnost	eficient	53
	ideal	caract	indol	regl	utilizacion	54
	median	destin	vuelt		fue	55
	indol	indol	caract		configuracion	56
	ven	asoci	divulgacion		generacion	57
	destin	vuelt	cientif		apo	58
	ipv6	egres	deport		razon	59
	interelacion	cientif	destin		mensaj	60
	caract	deport	cultural		aument	61
	atencion	cultural	dem		empres	62
	transparent	imag	egres		ip	63
	cientif	cooper	ven		laboratori	64
	divulgacion	dispuest	ideal		manipul	65
	cultural	dem	interelacion		mineri	66

Anexo 5. Reglas de elección de profesores como jurados de TEG.

```
cant_jurados_principales_asignados_a_TEG = 2;  
cant_jurados_suplentes_asignados_a_TEG = 2;  
cant_jurados_asignados_a_TEG = cant_jurados_principales_asignados_a_TEG +  
cant_jurados_suplentes_asignados_a_TEG;  
jurados_asignados_a_TEG = {principal1, principal2, suplente1, suplente2};
```

Si ((cant_areas_TEG == 1) AND (area1_TEG == a1))

ENTONCES

```
principal1_area = a1;  
principal2_area = a1;  
suplente1_area = a1;  
suplente2_area = a1;
```

FSI

Si ((cant_areas_TEG == 2) AND (area1_TEG == a1) AND (area2_TEG == a2))

ENTONCES

```
principal1_area = a1;  
principal2_area = a2;  
suplente1_area = a1;  
suplente2_area = a2;
```

FSI

SI (tipo_jurado_a_elegir == "principal1")

ENTONCES

SI ((profesor_ci != tutor1_ci) AND (profesor_ci != tutor2_ci) AND
(profesor_area == principal1_area) AND (profesor_disponible=="si") AND
(profesor_cant_asig_semestral < max_asig_semestral) AND
profesor_cant_asig_princ <= menor_cant_asig_princ_profs_de_principal1_area)

ENTONCES

SI (cant_profs_cumplen_condicion > 1)

ENTONCES

SI (profesor_cant_asig_sup <=
menor_cant_asig_sup_profs_de_principal1_area)

ENTONCES

SI (cant_profs_cumplen_condicion > 1)

ENTONCES

principal1 = selección_aleatoria_profesor();

SINO

principal1 = profesor;

FSI

SINO

principal1 = profesor;

FSI

FSI

FSI

FSI

SI (tipo_jurado_a_elegir == "principal2")

ENTONCES

SI ((profesor_ci != tutor1_ci) AND (profesor_ci != tutor2_ci) AND
 (profesor_area == principal2_area) AND (profesor_disponible=="si") AND
 (profesor_cant_asig_semestral < max_asig_semestral) AND
 profesor_cant_asig_princ <= menor_cant_asig_princ_profs_de_principal1_area)

ENTONCES

SI (cant_profs_cumplen_condicion > 1)

ENTONCES

SI (profesor_cant_asig_sup <= menor_cant_asig_sup_profs_de_principal2_area)

ENTONCES

SI (cant_profs_cumplen_condicion > 1)

ENTONCES

principal2 = selección_aleatoria_profesor();

SINO

principal2 = profesor;

FSI

SINO

principal2 = profesor;

FSI

FSI

FSI

FSI

SI (tipo_jurado_a_elegir == "suplente1")

ENTONCES

SI ((profesor_ci != tutor1_ci) AND (profesor_ci != tutor2_ci) AND
(profesor_area == suplente1_area) AND (profesor_disponible=="si") AND
profesor_cant_asig_princ <= menor_cant_asig_princ_profs_de_suplente1_area)

ENTONCES

SI (cant_profs_cumplen_condicion > 1)

ENTONCES

SI (profesor_cant_asig_sup <=
menor_cant_asig_sup_profs_de_suplente1_area)

ENTONCES

SI (cant_profs_cumplen_condicion > 1)

ENTONCES

suplente1 = selección_aleatoria_profesor();

SINO

suplente1 = profesor;

FSI

SINO

suplente1 = profesor;

FSI

FSI

FSI

FSI

SI (tipo_jurado_a_elegir == "suplente2")

ENTONCES

SI ((profesor_ci != tutor1_ci) AND (profesor_ci != tutor2_ci) AND
(profesor_area == suplente2_area) AND (profesor_disponible=="si") AND
profesor_cant_asig_princ <= menor_cant_asig_princ_profs_de_suplente2_area)

ENTONCES

SI (cant_profs_cumplen_condicion > 1)

ENTONCES

SI (profesor_cant_asig_sup <=
menor_cant_asig_sup_profs_de_suplente2_area)

ENTONCES

SI (cant_profs_cumplen_condicion > 1)

ENTONCES

suplente2 = selección_aleatoria_profesor();

SINO

suplente2 = profesor;

FSI

SINO

suplente2 = profesor;

FSI

FSI

FSI

FSI

Anexo 6. Resultados de las pruebas realizadas al módulo de recomendación de áreas.

Nro.	Área(s) TEG	ATI	Área(s) Recomendada(s)	ATI
1	Título	MI ESCUELA EN INTERNET: UNA HERRAMIENTA PARA LA GENERACIÓN DE SITIOS WEB ESCOLARES		
	Resumen	EL OBJETIVO PRINCIPAL DE ESTE TRABAJO ESPECIAL DE GRADO ES EL DESARROLLO DE UNA HERRAMIENTA QUE PERMITA GENERAR SITIOS WEB ESCOLARES, DE ESTA MANERA LAS ESCUELAS PODRÍAN TENER PRESENCIA EN INTERNET Y MOSTRAR TODA LA INFORMACIÓN QUE NECESITEN DAR A CONOCER A LOS USUARIOS POTENCIALES O INTERESADOS. A TRAVÉS DE ESTA HERRAMIENTA GENERADORA DE SITIOS WEB ESCOLARES SE PODRÍAN CREAR Y MODIFICAR SITIOS WEB PARA ESCUELAS QUE ASÍ LO REQUIERAN, SIN NECESIDAD DE CONOCIMIENTOS AVANZADOS DE LAS TECNOLOGÍAS EMPLEADAS.		
	Pals. Claves	GENERADOR DE SITIOS WEB ESCOLARES, SITIOS WEB ESCOLARES, GENERADOR, SITIOS ESCOLARES, ESCUELAS, PÁGINAS WEB DE ESCUELAS, SITIOS WEB DE ESCUELAS.		
	Área(s) TEG	1=Redes	Área(s) Recomendada(s)	1=Redes; 2=BD
2	Título	DISEÑO Y DESARROLLO DE PLATAFORMA INTEGRADA GSM-SMPP PARA LA ADMINISTRACIÓN DE MENSAJES DE TEXTO SMS Y PUBLICACIÓN WEB.		
	Resumen	EL OBJETIVO DEL PRESENTE TRABAJO ESPECIAL DE GRADO CONSISTE EN LA IMPLEMENTACIÓN DE UNA PLATAFORMA CAPAZ DE PROVEER, A EMPRESAS QUE PRESTAN SERVICIOS DE VALOR AGREGADO MEDIANTE LA MENSAJERÍA CORTA DE TEXTO, EL CONTROL DE TODOS SUS ENVÍOS, ASÍ COMO LAS RESPECTIVAS RESPUESTAS DE LOS MISMOS, ADEMÁS DE SER CONFIGURABLE PARA CUALQUIER TIPO DE OPERADORA, NO SÓLO A NIVEL NACIONAL SINO TAMBIÉN INTERNACIONAL. ESTE DESARROLLO FUE MOTIVADO POR LA NECESIDAD DE LA EMPRESA A LAS COMUNICACIONES, QUIÉN PRESTA ESTE TIPO DE SERVICIOS, CON EL FIN DE AMPLIAR SU PLATAFORMA Y PODER BRINDAR UN MEJORAR SERVICIO A SUS CLIENTES; HACIENDO USO DE LAS ESPECIFICACIONES DEL PROTOCOLO ESTÁNDAR DE CLASE MUNDIAL SMPP (SHORT MESSAGE PEER-TO-PEER), LA IMPLEMENTACIÓN FUE LLEVADA A CABO EN LENGUAJE JAVA Y COMO MANEJADOR DE BASE DE DATOS MYSQL.		
	Pals. Claves	SMPP, SHORT MESSAGE PEER TO PEER, SMS, SHORT MESSAGING SYSTEM, MENSAJERÍA DE TEXTO.		
	Área(s) TEG	ATI	Área(s) Recomendada(s)	ATI
3	Título	HERRAMIENTA DE APOYO PARA LA FORMULACIÓN DE PROYECTOS SOCIO-PRODUCTIVOS COMUNALES		
	Resumen	A TRAVÉS DE ESTE TRABAJO ESPECIAL DE GRADO SE CONSTRUYE UN PRODUCTO DE SOFTWARE QUE APOYA LA FORMULACIÓN DE PROYECTOS SOCIO-PRODUCTIVOS COMUNALES. ESTA HERRAMIENTA PERMITE A LAS USUARIAS Y USUARIOS, LUEGO DE LA FASE DE ESTUDIO Y DIAGNÓSTICO DE LA REALIDAD COMUNAL, PODER CONVERTIR SUS IDEAS-PROYECTOS, RELACIONADAS CON ACTIVIDADES PRODUCTIVAS, EN UN PROYECTO SOCIO-PRODUCTIVO EN SÍ. ESTO A TRAVÉS DE LA APLICACIÓN DE UNA METODOLOGÍA PARA LA FORMULACIÓN DE PROYECTOS Y LA GENERACIÓN DEL MISMO EN FORMATO DIGITAL. SE TOMÓ COMO PROPUESTA PILOTO LA METODOLOGÍA DISEÑADA POR EL BANCO DE DESARROLLO DE LA MUJER, PARA FINANCIAR PROYECTOS SOCIO-PRODUCTIVOS A MUJERES DE NUESTRO PUEBLO, PROVEYÉNDOLAS, DE ESTA MANERA, DE UN INSTRUMENTO PARA EL MANEJO DE DICHA METODOLOGÍA. LA HERRAMIENTA DESARROLLADA ES UNA APLICACIÓN WEB A LA QUE PUEDEN ACCEDER LAS USUARIAS A FIN DE PRESENTAR SUS PROYECTOS ANTE LA INSTITUCIÓN. ASÍ COMO TAMBIÉN PUEDE SER UTILIZADA POR CUALQUIER INTERNAUTA QUE DESEE ORIENTARLA HACIA OTROS DESTINOS, YA SEA PARA FINES SÓLO DE PROYECTOS COMO PARA LA CONSECUCCIÓN DE FINANCIAMIENTOS A TRAVÉS DE OTROS ENTES CON COMPETENCIA EN LA MATERIA.		
	Pals. Claves	PROYECTOS SOCIO-PRODUCTIVOS, GOBIERNO ELECTRÓNICO, HERRAMIENTA		

		WEB.		
Área(s) TEG	Redes	Área(s) Recomendada(s)	Redes	
4	Título	UNA HERRAMIENTA BASADA EN TECNOLOGÍA PALM PILOT PARA EL SISTEMA INDUSTRIAL DE CRÉDITO: ANÁLISIS DE FACTIBILIDAD Y DISEÑO DE UNA INTERFAZ DE COMUNICACIÓN.		
	Resumen	<p>EL BANCO DE LA GENTE EMPRENDEDORA TIENE PLANTEADO DENTRO DE SUS ESTRATEGIAS ORGANIZACIONAL LA IMPLANTACIÓN DE UN SISTEMA QUE APOYE LA GESTIÓN DE LA ASISTENTE ADMINISTRATIVO Y EL ASESOR DE NEGOCIO, AUTOMATIZANDO EL PROCESO DE TRANSCRIPCIÓN DE INFORMACIÓN DE LAS SOLICITUDES DE CRÉDITO DE LOS CLIENTES Y MEJORANDO LOS PROCESOS DE OBTENCIÓN DE INFORMACIÓN DEL ESTATUS DE LA CARTERA DE CRÉDITO PARA EL ASESOR DE NEGOCIO; EL MEJORAMIENTO DE AMBOS PROCESOS AUMENTARÍA LA PRODUCTIVIDAD DE AMBAS ÁREAS, LO QUE SE TRADUCIRÍA EN UN OFRECIMIENTO ADECUADO DEL SERVICIO QUE SE LE PRESTA A LOS CLIENTES Y EL CUMPLIMIENTO DEL ESLOGAN DE LA ORGANIZACIÓN CRÉDITOS RÁPIDOS Y SIN TANTO PAPELEO. OTRA VENTAJA PARA EL DESARROLLO DE UN SISTEMA DE ESTA ENVERGADURA ES LO RELACIONADO AL AHORRO DE RECURSOS MATERIALES TALES COMO PAPEL, TINTA, CONSUMIBLES, ESPACIO FÍSICO DE EQUIPOS DESTINADOS PARA ASESORES DE NEGOCIO, ETC. POR LO ANTES EXPUESTO QUE SE ESTÁ DESARROLLANDO UN SISTEMA BASADO EN TECNOLOGÍA PALM PILOT (DISPOSITIVOS MÓVILES) PARA LA CAPTURA DE INFORMACIÓN EN CAMPO DE LAS SOLICITUDES DE CRÉDITO DE CLIENTES, QUE A TRAVÉS DE SU RESPECTIVA INTERFAZ PODRÁ LLEVAR DICHA INFORMACIÓN AL CORE PRINCIPAL DE LA ORGANIZACIÓN ELIMINANDO LAS TRANSCRIPCIONES INNECESARIAS, EL GASTO EXCESIVO DE PAPEL, USO DE FORMULARIOS, CONSUMIBLES, ETC. Y ELIMINARA EN GRAN MEDIDA EL USO Y TIEMPO REQUERIDO DE LA FUERZA DE VENTA EN UN COMPUTADOR PERSONAL, YA QUE LA MAYORÍA DE LA INFORMACIÓN REQUERIDA ESTARÁ DISPONIBLE EN SU DISPOSITIVO DE BOLSILLO (PALM PILOT) A TRAVÉS DE LAS DISTINTAS SINCRONIZACIONES CON LA INTERFAZ, MEJORANDO ASÍ EFICIENTEMENTE EL SERVICIO QUE SE LE PRESTA A LOS CLIENTES DE ESTA PRESTIGIOSA ORGANIZACIÓN FINANCIERA.</p>		
	Pals. Claves	ESTRATEGIA ORGANIZACIONAL, PROCESOS, INTERFASE, DISPOSITIVOS MÓVILES, AHORRO, PRODUCTIVIDAD Y SERVICIO.		
Área(s) TEG	1=ATI; 2=BD	Área(s) Recomendada(s)	1=BD; 2=ATI	
5	Título	PROTOTIPO DEL SISTEMA DE GESTIÓN DE PREPARADORES DE LA ESCUELA DE COMPUTACIÓN, MÓDULOS GESTIÓN DE RETIROS Y ADMINISTRATIVO.		
	Resumen	<p>EL DESARROLLO DEL PRESENTE TRABAJO, CONSTITUYE LA CONSTRUCCIÓN DE UN PROTOTIPO DEL SISTEMA DE GESTIÓN DE PREPARADORES DE LA ESCUELA DE COMPUTACIÓN (SGP--EC), MÓDULOS ADMINISTRATIVO Y GESTIÓN DE RETIROS PARA EL MEJORAMIENTO DE LOS PROCESOS DE DICHA GESTIÓN, SIGUIENDO LAS FASES DE LA METODOLOGÍA OPENUP PRODUCIENDO EN LA PRIMERA FASE UN FLUJO DE TRABAJO (WORKFLOW) QUE REPRESENTA LAS ACTIVIDADES CORRESPONDIENTES A LOS USUARIOS ACTUALES. TAMBIÉN SE LLEVÓ A CABO LA DEFINICIÓN Y ESPECIFICACIÓN DE LOS ELEMENTOS NECESARIOS PARA LA CREACIÓN DE LOS SERVICIOS WEB QUE PERMITEN EL SUMINISTRO DE LOS DATOS PARA LOS MÓDULOS DE GESTIÓN DE RETIROS Y ADMINISTRATIVO EMPLEANDO LA HERRAMIENTA TECNOLÓGICA RUBY ON RAILS. DURANTE LA ÚLTIMA FASE SE REALIZÓ LA SIMULACIÓN DE LA EJECUCIÓN DE LAS FUNCIONALIDADES. ALCANZANDO EL OBJETIVO PRINCIPAL DE ESTE TRABAJO ESPECIAL DE GRADO SE CONSIGUIÓ COMO RESULTADO EL DESARROLLO EL MÓDULO DE GESTIÓN DE RETIROS, DEL SISTEMA DE GESTIÓN DE PREPARADORES DE LA ESCUELA DE COMPUTACIÓN, EN CONJUNTO CON EL MÓDULO ADMINISTRATIVO DEL SISTEMA.</p>		
	Pals. Claves	GESTIÓN DE PREPARADORES, GESTIÓN POR PROCESOS, WORKFLOW, SERVICIOS WEB, OPENUP, RUBY ON RAILS.		
Área(s) TEG	Redes	Área(s) Recomendada(s)	Redes	
6	Título	IMPLEMENTACIÓN DE UN DISPOSITIVO INTEGRADO DE SEGURIDAD UTM Y EVALUACIÓN DE LA FACTIBILIDAD DE SU CONFIGURACIÓN A TRAVÉS DE LA METODOLOGÍA DE INTRUSIÓN LPT		
	Resumen	EL PRESENTE TRABAJO ESPECIAL DE GRADO CONSISTE EN EVALUAR LA FACTIBILIDAD DE LA CONFIGURACIÓN DE UN UTM (UNIFIED THREAT		

		<p>MANAGEMENT) FIREBOX 750E WATCHGUARD, EL CUAL PERMITE LA ADMINISTRACIÓN DE LA RED Y PROPORCIONA SEGURIDAD A LA MISMA, A TRAVÉS DE LA METODOLOGÍA DE INTRUSIÓN LPT (LICENSED PENETRATION TESTER). EL FUERTE DE LOS DISPOSITIVOS UTM ES LA INTEGRACIÓN DE LOS DIFERENTES SERVICIOS DE SEGURIDAD EN UN SOLO DISPOSITIVO, MEDIANTE EL CUAL SE PROVEE DE MANERA CENTRALIZADA LA SEGURIDAD Y ADMINISTRACIÓN DE LA RED, FACILITANDO LA LABOR DEL ADMINISTRADOR DE RED Y DEL ANALISTA DE SEGURIDAD. ES IMPORTANTE ACOTAR QUE LA INSTALACIÓN Y CONFIGURACIÓN DESCRITAS EN EL T.E.G. SON GENÉRICAS PARA CUALQUIER DISPOSITIVO UTM DE LA MARCA WATCHGUARD. EL TRABAJO SE INICIA CON UN ESTUDIO TEÓRICO DE LOS ASPECTOS MÁS RELEVANTES QUE SE DEBEN TOMAR EN CONSIDERACIÓN EN LO QUE A SEGURIDAD INFORMÁTICA SE REFIERE Y UN ANÁLISIS DE LAS TÉCNICAS MÁS COMUNES UTILIZADAS POR LOS HACKERS. PARA LA INSTALACIÓN DEL DISPOSITIVO SE UTILIZAN LAS RECOMENDACIONES DEL FABRICANTE Y PARA SU CONFIGURACIÓN SE TOMA EN CUENTA LA POLÍTICA DE USO DE INTERNET DE LA INSTITUCIÓN Y SUS REQUERIMIENTOS EN CUANTO AL SERVICIO. SE APLICA LA METODOLOGÍA LPT AL DISPOSITIVO PARA VERIFICAR LA FACTIBILIDAD DE LA CONFIGURACIÓN DEL UTM. LOS DISPOSITIVOS UTM BRINDAN PROTECCIÓN DE FIREWALL CONVENCIONAL, PROTECCIÓN A ANTISPAM - ANTIPHISHING - FILTRO DE CONTENIDOS - ANTIVIRUS Y DETECCIÓN/PREVENCIÓN DE INTRUSOS IDS (INTRUSION DETECTION SYSTEM) / IPS (INTRUSION PREVENTION SYSTEM)). ESTOS DISPOSITIVOS TAMBIÉN PERMITEN LA CREACIÓN DE REDES VPN (VIRTUAL PRIVATE NETWORK), UTILIZAR PROTOCOLOS DE RUTEO, TRADUCCIONES DE DIRECCIONES Y REGISTRO DE LOGS, ENTRE OTROS. LA METODOLOGÍA PARA PRUEBAS DE INTRUSIÓN LPT ES PROPIA DE LA CASA DE CONSULTORES DE SEGURIDAD EC COUNCIL, LA CUAL DEFINE UNA SERIE DE PASOS PARA PROBAR LA SEGURIDAD DE DIFERENTES EQUIPOS, APLICACIONES Y ENTORNOS DE TRABAJO.</p>
	Pals. Claves	UTM, SEGURIDAD INFORMÁTICA, LPT, HACKERS, SERVICIOS DE SEGURIDAD INFORMÁTICA.
	Área(s) TEG	1=ATI; 2=BD
	Área(s) Recomendada(s)	ATI
7	Título	DESARROLLO DEL SISTEMA DE GESTIÓN ACADÉMICA CONEST 2.0: MÓDULO DE REPORTES, CONSTANCIAS Y TAREAS EN LOTE
	Resumen	MÓDULO DE REPORTES, CONSTANCIAS Y TAREAS EN LOTE DEL SISTEMA CONEST 2.0, BASADAS EN LAS NUEVAS TECNOLOGÍAS RUBY 1.9, RAILS 2.2 Y MYSQL 5.0. LAS FUNCIONALIDADES OFRECEN UNA MAYOR USABILIDAD Y RENDIMIENTO QUE LAS PRESENTES EN EL SISTEMA CONEST Y PERMITE EL AHORRO DE RECURSOS EN LA IMPRESIÓN FÍSICA DE DOCUMENTOS.
	Pals. Claves	CONEST, CONEST 2.0, RUBY, RAILS, MYSQL, BACKGROUNDDB, REPORTES, CONSTANCIAS, TAREAS EN LOTE.
	Área(s) TEG	ATI
	Área(s) Recomendada(s)	ATI
8	Título	PROTOTIPO DEL SISTEMA DE GESTIÓN DE PREPARADORES DE LA ESCUELA DE COMPUTACIÓN, MÓDULOS GESTIÓN DE CONCURSOS Y NOMBRAMIENTOS.
	Resumen	EL PROPÓSITO DE ESTE TRABAJO ESPECIAL DE GRADO SE ORIENTÓ EN LA ELABORACIÓN DE UN PROTOTIPO DEL SISTEMA DE GESTIÓN DE PREPARADORES DE LA ESCUELA DE COMPUTACIÓN, ENFOCADO ESPECÍFICAMENTE EN EL DESARROLLO DE LOS MÓDULOS DE GESTIÓN DE CONCURSOS Y NOMBRAMIENTOS, TOMANDO COMO BASE EL MODELO WORKFLOW PROPUESTO PARA ESTOS PROCESOS. PARA LLEVAR A CABO LA IMPLEMENTACIÓN DE ESTE SISTEMA SE USO OPENUP COMO METODOLOGÍA DE DESARROLLO DE SOFTWARE Y LA HERRAMIENTA DE RUBY ON RAILS COMO PLATAFORMA DE CONSTRUCCIÓN. EL RESULTADO OBTENIDO FUE LA CREACIÓN DE UN PROTOTIPO ENCARGADO DE CONTROLAR LA GESTIÓN DE LOS PREPARADORES, ESPECÍFICAMENTE EN LO CONCERNIENTE A LOS MÓDULOS GESTIÓN DE CONCURSO Y GESTIÓN DE NOMBRAMIENTO. ESTE PROTOTIPO DEPENDE DE LA INTERACCIÓN CON LA BASE DE DATOS A TRAVÉS DE SERVICIOS WEB. DADO QUE EL DESARROLLO DE LA BASE DE DATOS DEL SISTEMA CORRESPONDE A UN TRABAJO A FUTURO, EL PROTOTIPO FUE DOTADO DE UNA SOLUCIÓN PRELIMINAR, A TRAVÉS DE ARCHIVOS XML QUE SIMULAN LA CAPA DE DATOS, CON LA CUAL SE MUESTRAN LOS RESULTADOS A TRAVÉS DE LA APLICACIÓN, MIENTRAS SE ALCANZA EN OTRA ETAPA LA IMPLANTACIÓN DEL ACCESO A LOS DATOS MEDIANTE LOS SERVICIOS WEB.

	Pals. Claves	GESTIÓN DE PREPARADORES, GESTIÓN POR PROCESOS, WORKFLOW, SERVICIOS WEB, OPENU, RUBY ON RAILS		
	Área(s) TEG	Redes	Área(s) Recomendada(s)	1=Redes; 2=IA
9	Título	DISEÑO E IMPLEMENTACIÓN DE UN SERVICIO DE E-MAIL AUNADO A TÉCNICAS ANTISPAM Y BASADO EN SOFTWARE DE DOMINIO PÚBLICO		
	Resumen	<p>EL PRESENTE TRABAJO ESPECIAL DE GRADO ESTÁ ENFOCADO HACIA LA IMPLEMENTACIÓN DE UN SISTEMA DE CORREO ELECTRÓNICO CONJUGADO CON TÉCNICAS Y HERRAMIENTAS PARA EL CONTROL DE SPAM BASADO EN SOFTWARE DE DOMINIO PÚBLICO. EL OBJETIVO DEL TRABAJO ES PROPONER UNA SOLUCIÓN QUE PERMITA A CUALQUIER PERSONA U ORGANISMO IMPLEMENTAR UN SISTEMA DE CORREO ELECTRÓNICO CON UN EFICAZ MANEJO DE MENSAJES NO DESEADOS. EL PROYECTO ESTÁ COMPUESTO POR UN PROCESO DE INVESTIGACIÓN, INSTALACIÓN Y CONFIGURACIÓN DE LOS COMPONENTES NECESARIOS PARA IMPLEMENTAR UNA PLATAFORMA DE CORREO ELECTRÓNICO CON CONTROL DE SPAM. LA SOLUCIÓN ANTISPAM SE BASA EN SPAMASSASSIN, EL CUAL ES UN SOFTWARE DE CÓDIGO ABIERTO QUE IMPLEMENTA FUNCIONALIDADES NECESARIAS PARA EL CONTROL DEL SPAM. COMO CASO DE ESTUDIO SE SELECCIONÓ A LA ASAMBLEA NACIONAL, EN LA CUAL SE EFECTUÓ REINGENIERÍA TOTAL DE LA PLATAFORMA DE CORREO, QUEDANDO ESTABLECIDO ZIMBRA MESSAGING AND COLLABORATION SUITE 5.0 COMO PLATAFORMA DE MENSAJERÍA Y COLABORACIÓN. ESTA SUITE USA COMO BACKGROUND EL SISTEMA OPERATIVO DEBIAN ETCH RELEASE 4. TODOS ESTOS SISTEMAS SE INTEGRARON PARA PROPORCIONAR UN SERVICIO DE CORREO LIBRE DE SPAM EN CONCORDANCIA CON LOS LINEAMIENTOS DEL DECRETO N° 3.390. CON LA IMPLEMENTACIÓN DE LA PLATAFORMA DE CORREO ELECTRÓNICO CON CONTROL DE SPAM, SE LOGRÓ OBTENER UNA SOLUCIÓN QUE PROPORCIONA BENEFICIOS ADICIONALES (ANTIVIRUS, FIREWALL, ETC.) QUE BRINDA MÁS Y MEJORES OPCIONES A LOS USUARIOS Y ADMINISTRADORES. DE IGUAL FORMA SE OBTUVO UNA PLATAFORMA FÁCILMENTE ESCALABLE EN CUANTO A SERVICIOS Y FUNCIONALIDADES, ADEMÁS DE SER EXTENSIBLE, YA QUE ESTA SOLUCIÓN SE PODRÍA IMPLEMENTAR EN OTROS ENTES PÚBLICOS Y PRIVADOS.</p>		
	Pals. Claves	SPAM E-MAIL ZIMBRA POSTFIX SPAMASSASSIN MTA CORREO ELECTRONICO		
	Área(s) TEG	Redes	Área(s) Recomendada(s)	Redes
10	Título	ANÁLISIS E IMPLEMENTACIÓN DE UN ESQUEMA BÁSICO DE SEGURIDAD USANDO SOFTWARE LIBRE		
	Resumen	<p>ACTUALMENTE LAS REDES DE ÁREA LOCAL SON TAN HABITUALES QUE CASI NO HAY OFICINA QUE NO LA POSEA; INCLUSO GRAN PARTE DE HOGARES LA TIENEN. LA MAYORÍA DE ESTAS REDES ESTÁN CONECTADAS A INTERNET, HECHO QUE LAS HACE VULNERABLES A LOS PELIGROS LATENTES EN EL INTERNET, POR ESE MOTIVO ES NECESARIO PROPORCIONAR A ESTAS REDES ALGÚN NIVEL DE SEGURIDAD. EN ESTE TRABAJO SE ANALIZARON EN FORMA DETALLADA CADA UNA DE LAS RECOMENDACIONES PROPUESTAS POR EL DOCUMENTO ISO/IEC 27002; CON EL OBJETO DE BUSCAR UN ESQUEMA BÁSICO DE SEGURIDAD PARA LAS REDES DE ÁREA LOCAL Y PROPONER UN CONJUNTO MÍNIMO DE RECOMENDACIONES AL RESPECTO, ASÍ COMO TAMBIÉN SU IMPLEMENTACIÓN USANDO SOFTWARE LIBRE. EL ESQUEMA BÁSICO DE SEGURIDAD ENCONTRADO ESTÁ COMPUESTO POR UN CONJUNTO MÍNIMO DE RECOMENDACIONES (CLÁUSULAS) CUYO CUMPLIMIENTO PROPORCIONA UN NIVEL DE SEGURIDAD ACEPTABLE EN REDES NO CRÍTICAS; ES DECIR EN REDES DE ORGANIZACIONES CON POCOS ENEMIGOS POTENCIALES, REDES QUE NO PERTENECEN A GRUPOS DE ESPIONAJE, NI AGENCIAS DE INVESTIGACIÓN, NI CENTROS DE SEGURIDAD GUBERNAMENTAL, NI BANCOS, NI OTRAS ORGANIZACIONES DEL ESTILO. ADICIONALMENTE, SE RECOMIENDA ABARATAR COSTOS DE IMPLEMENTACIÓN, INCREMENTAR LA INDEPENDENCIA Y MANTENERSE ACTUALIZADO. PARA LOGRARLO SE PROPONE USAR SOFTWARE LIBRE, EL CUAL CON COSTOS BAJOS Y ALTA CONFIABILIDAD PERMITE IMPLEMENTAR EL CONJUNTO MÍNIMO DE RECOMENDACIONES EXTRAÍDO DEL DOCUMENTO ISO/IEC 27002. LAS RECOMENDACIONES DEL ESQUEMA BÁSICO PROPUESTO SE IMPLEMENTARON EN UN CASO DE ESTUDIO, USANDO HERRAMIENTAS DE SOFTWARE LIBRE, DEMOSTRANDO ASÍ LA EFECTIVIDAD DE LA PROPUESTA.</p>		
	Pals. Claves	INTERNET, INTRANET, SEGURIDAD, SOFTWARE LIBRE, ISO//IEC 2277000022.		

	Área(s) TEG	ATI (y Sist. Inf. Pero no se considera esa área)	Área(s) Recomendada(s)	ATI
11	Título	DESARROLLO DE UN SISTEMA DE INFORMACIÓN COLABORATIVO PARA EL CONTROL DE VISITAS MÉDICAS		
	Resumen	EN ESTE TRABAJO ESPECIAL DE GRADO SE PRESENTA EL DESARROLLO DE UN SISTEMA DE INFORMACIÓN COLABORATIVO PARA FACILITAR EL CONTROL Y LA ADMINISTRACIÓN DE LA INFORMACIÓN DE ORGANIZACIONES DE MERCADEO DE PRODUCTOS FARMACÉUTICOS. PARA EL DESARROLLO DEL MISMO SE UTILIZAN DIVERSAS TECNOLOGÍAS DE PROGRAMACIÓN, ORGANIZADAS EN EL MODELO MULTI-CAPA, ENTRE LAS QUE PODEMOS NOMBRAR: PHP HYPERTEXT PROCESSOR Y SU LIBRERÍA DE FUNCIONALIDADES ZEND FRAMEWORK PARA LA CAPA LÓGICA DE NEGOCIO, BASE DE DATOS MYSQL PARA LA CAPA DE PERSISTENCIA DE DATOS Y JAVASCRIPT Y SU LIBRERÍA DE FUNCIONALIDADES JQUERY. LA ARQUITECTURA QUE SE UTILIZA PARA CONSTRUIR EL SISTEMA DE INFORMACIÓN ES LA LLAMADA MODELO VISTA CONTROLADOR. TODO EL DESARROLLO SE APOYA EN LA METODOLOGÍA DE DESARROLLO SCRUM TOMANDO LAS HISTORIAS DE USUARIO DE LA PROGRAMACIÓN EXTREMA (XP). CON EL DESARROLLO DE ESTE SISTEMA, SE LOGRA LA DISMINUCIÓN DE LOS TIEMPOS DE RESPUESTA EN LAS OPERACIONES DE LA EMPRESA, ASÍ COMO LA AUTOMATIZACIÓN DE DIVERSAS TAREAS, EL CONTROL DE LAS ACTIVIDADES Y CENTRALIZACIÓN DE LA INFORMACIÓN.		
	Pals. Claves	SISTEMAS DE INFORMACIÓN, APLICACIONES DE TRABAJO COLABORATIVO, CONTROL DE VISITAS, CONTROLES DE EVENTOS, LENGUAJES DE PROGRAMACIÓN, SERVIDOR WEB, BASE DE DATOS, PROGRAMACIÓN EXTREMA, SCRUM, MYSQL, PHP, ZEND FRAMEWORK, JQUERY, JAVASCRIPT, AJAX, CSS, HTML, WEB 2.0.		
	Área(s) TEG	ATI	Área(s) Recomendada(s)	1=ATI; 2=BD
12	Título	SISTEMA DE GESTIÓN DE PUNTOS DE COMPROMISO Y PUNTOS DE ACEPTACIÓN PARA LOS BANCOS AFILIADOS A LA CORPORACIÓN SUICHE 77B		
	Resumen	EL PRESENTE TRABAJO EXPONE FUNDAMENTOS CONCEPTUALES, METODOLÓGICOS Y DE IMPLEMENTACIÓN QUE RESUMEN LA INVESTIGACIÓN DEL TRABAJO ESPECIAL DE GRADO REALIZADO CON EL OBJETIVO DE DESARROLLAR UNA APLICACIÓN WEB PARA LA CORPORACIÓN SUICHE 7B C.A. QUE PERMITA EL INTERCAMBIO ENTRE LAS DIFERENTES INSTITUCIONES BANCARIAS DE INFORMACIÓN ASOCIADA A POSIBLES FRAUDES EN OPERACIONES CON TARJETAS DE CRÉDITO Y DÉBITO. LA APLICACIÓN EN CUESTIÓN PERMITIRÁ LA DETECCIÓN DE LOS PUNTOS DE COMPROMISO (PDC) Y PUNTOS DE ACEPTACIÓN (PDA) UTILIZADOS EN ACCIONES DE FRAUDE, COMO PARTE DE UNA NUEVA ESTRATEGIA DE COOPERACIÓN INTERBANCARIA QUE AYUDE A COMBATIR LA COPIA Y USO ILEGAL DE TARJETAS DE DÉBITO Y CRÉDITO A LOS USUARIOS DE LAS INSTITUCIONES BANCARIAS DEL PAÍS. PARA ALCANZAR ESTOS OBJETIVOS SE APLICÓ UNA METODOLOGÍA BASADO EN LA ADAPTACIÓN DE XP (PROGRAMACIÓN EXTREMA), DETALLANDO LAS FASES DE PLANIFICACIÓN, DISEÑO, CODIFICACIÓN Y PRUEBAS PARA CADA UNA DE LAS ITERACIONES QUE SE DIVIDIÓ EL DESARROLLO DE LA APLICACIÓN WEB. SE OBTUVO COMO RESULTADO UNA APLICACIÓN WEB PARA EL MANEJO DEL PERSONAL DE LAS DIFERENTES COMISIONES DE FRAUDE DE LAS INSTITUCIONES BANCARIAS ASÍ COMO PARA EL PERSONAL DE LA GERENCIA DE OPERACIONES DE LA CORPORACIÓN SUICHE 7B, CONVIRTIÉNDOSE ASÍ, EN LA PRIMERA APLICACIÓN EN EL PAÍS PARA TAL FIN.		
	Pals. Claves	SISTEMA DE GESTIÓN DE PUNTOS DE COMPROMISO Y PUNTOS DE ACEPTACIÓN (SIPCA), PUNTO DE COMPROMISO (PDC), PUNTO DE ACEPTACIÓN (PDA), CLONACIÓN, CAJERO AUTOMÁTICO (ATM), PUNTO DE VENTA (POS), SCORE.		
	Área(s) TEG	ATI	Área(s) Recomendada(s)	ATI
13	Título	PROTOTIPO PARA LA GESTIÓN Y CLASIFICACIÓN DE COLECCIONES DE DOCUMENTOS DIGITALES		
	Resumen	EL OBJETIVO DEL PRESENTE TRABAJO ESPECIAL DE GRADO CONSISTE EN EL DESARROLLO DE UNA APLICACIÓN WEB QUE PERMITA APOYAR LOS PROCESOS DE DIGITALIZACIÓN DE DOCUMENTOS, CLASIFICACIÓN Y		

		VISUALIZACIÓN DE DOCUMENTOS DIGITALES BASADOS EN UNA TAXONOMÍA. DENTRO DE LAS NECESIDADES CONTEMPLADAS EN LA APLICACIÓN SE TIENE LO SIGUIENTE: CONTAR CON UNA APLICACIÓN GENERALIZADA QUE LE PERMITA A LOS USUARIOS DE FORMA RÁPIDA, EFICIENTE Y FÁCIL DE USAR REALIZAR UNA ESTRUCTURA JERÁRQUICA DE LOS DOCUMENTOS DIGITALIZADOS Y ASÍ MISMO SE PUEDA MANEJAR LA DATA DE LOS MISMOS DE UNA FORMA EFICIENTE. PARA CUMPLIR CON LOS OBJETIVOS DE LA APLICACIÓN SE UTILIZÓ LA METODOLOGÍA ÁGIL XP (PROGRAMACIÓN EXTREMA), LA CUAL SE BASA FUNDAMENTALMENTE, EN LA CONSTRUCCIÓN PROGRESIVA DEL SOFTWARE SIN HACER ÉNFASIS EN ETAPAS DE DISEÑOS.		
	Pals. Claves	DIGITALIZACIÓN, DOCUMENTOS DIGITALES, CLASIFICACIÓN, TAXONOMÍAS		
	Área(s) TEG	IA	Área(s) Recomendada(s)	REDES
14	Título	MINERÍA DE TEXTOS PARA LA ASIGNACIÓN AUTOMÁTICA DE JURADOS A TRABAJOS ESPECIALES DE GRADO		
	Resumen	ESTE TRABAJO TIENE COMO OBJETIVO EL DESARROLLO DE UN PROTOTIPO BASADO EN MINERÍA DE TEXTOS PARA LA ASIGNACIÓN AUTOMÁTICA DE JURADOS A TRABAJOS ESPECIALES DE GRADO (TEG), A TRAVÉS DE LA CLASIFICACIÓN DEL MISMO DENTRO DE LAS ÁREAS PROFESIONALES DE LA ESCUELA DE COMPUTACIÓN. PARA LA CREACIÓN DEL CLASIFICADOR SE APLICÓ EL PROCESO DE MINERÍA DE TEXTOS SOBRE UNA RECOPIACIÓN DE DOCUMENTOS DE TEG, LO CUALES FUERON PREPROCESADOS, ELIMINANDO SIGNOS DE PUNTUACIÓN, PALABRAS NO INFORMATIVAS Y APLICANDO WORD STEMMING, Y REPRESENTADOS EMPLEANDO EL MODELO DE ESPACIO VECTORIAL CON DIFERENTES TÉCNICAS DE INDEXACIÓN. ASIMISMO, SE PROBARON DIFERENTES MEDIDAS DE RELEVANCIA PARA LA REDUCCIÓN DE LA DIMENSIONALIDAD DEL ESPACIO DE ATRIBUTOS. PARA LA ESTIMACIÓN DEL MODELO DE CLASIFICACIÓN SE REALIZARON PRUEBAS APLICANDO EL ALGORITMO K VECINOS MÁS CERCANOS CON DIFERENTES PARÁMETROS, ALCANZANDO UN 82% DE CLASIFICACIONES CORRECTAS. A TRAVÉS DE LA RECOMENDACIÓN DE HASTA DOS ÁREAS, SE LOGRÓ INCREMENTAR EL PORCENTAJE DE CLASIFICACIONES CORRECTAS A 95%. ESTE MÓDULO DE RECOMENDACIÓN DE ÁREAS SE INTEGRÓ CON EL MÓDULO DE ASIGNACIÓN, CON EL FIN DE REALIZAR LA ASIGNACIÓN DE LOS JURADOS, CONSIDERANDO A LOS PROFESORES PERTENECIENTES AL O LAS ÁREAS RECOMENDADAS, EN BASE A CRITERIOS DE ELECCIÓN PREVIAMENTE DEFINIDOS. EN CONCLUSIÓN, FUE POSIBLE OBTENER UN BUEN RENDIMIENTO DE CLASIFICACIÓN DE DOCUMENTOS DE TEG APLICANDO MINERÍA DE TEXTOS Y SE LOGRÓ AUTOMATIZAR EL PROCESO DE ASIGNACIÓN DE JURADOS A TEG, LO QUE A SU VEZ PERMITE LLEVAR UN REGISTRO HISTÓRICO DE LAS ASIGNACIONES, CONTRIBUYENDO ADEMÁS CON LA AUTOMATIZACIÓN DE PROCESOS DE LA ESCUELA DE COMPUTACIÓN.		
	Pals. Claves	MINERÍA DE TEXTOS, MODELOS DE CLASIFICACIÓN, K VECINOS MÁS CERCANOS, ALGORITMO DE PORTER <i>STEMMING</i> , WEKA.		
	Área(s) TEG	Redes	Área(s) Recomendada(s)	Redes
15	Título	ANÁLISIS E IMPLANTACIÓN DE UN SGSI ISO/IEC 27001:2005, PARA UNA EMPRESA DEL MERCADO DE CAPITALES		
	Resumen	LA ORGANIZACIÓN DE ESTÁNDARES INTERNACIONALES (ISO) PUBLICÓ (OCT., 2005) UNA NORMA QUE PERMITE ESTABLECER UN SISTEMA DE GESTIÓN DE SEGURIDAD DE LA INFORMACIÓN (SGSI), CONOCIDA COMO LA NORMA ISO/IEC 27001:2005. ESTA PERMITE MANEJAR LA SEGURIDAD DE LA INFORMACIÓN DENTRO DE LAS ORGANIZACIONES SIN IMPORTAR SU TAMAÑO, CAPITAL ECONÓMICO O DEDICACIÓN, EVITANDO DE ESTA MANERA EL USO DELIBERADO O EXCESIVO DE TECNOLOGÍA. ADEMÁS HA SIDO DE GRAN ACEPTACIÓN Y HA VENIDO TOMANDO FUERZAS YA QUE ES CERTIFICABLE (A DIFERENCIA DE OTROS ESTÁNDARES O NORMAS PUBLICADOS), OBSERVÁNDOSE DÍA A DÍA UN CRECIMIENTO EN LA CANTIDAD DE EMPRESAS CERTIFICADAS. POR ELLO EL PRESENTE TRABAJO ESPECIAL DE GRADO TIENE COMO PROPÓSITO PRESENTAR UN ANÁLISIS E IMPLANTACIÓN DE UN SGSI BAJO LA NORMA ISO/IEC 27001:2005, CON LA FINALIDAD DE PROVEER UN AMBIENTE DE SEGURIDAD DE LA INFORMACIÓN AL PROYECTO DE INTERNET BANKING EN UNA EMPRESA DEL MERCADO DE CAPITALES (MERCAP). PARA ESTO SE REALIZA PRIMERO UN ANÁLISIS DE LA NORMA ISO/IEC 27001:2005 PARA FACILITAR SU ENTENDIMIENTO, ESTRUCTURACIÓN		

		Y OBLIGACIONES PARA DAR CONFORMIDAD. DESPUÉS SE HACE UN ESTUDIO DEL ESTADO DE LA TOPOLOGÍA DE RED Y DE LAS APLICACIONES PARA SU ADECUACIÓN AL PROYECTO DE INTERNET BANKING, PROPONIÉNDOSE LUEGO UN DISEÑO DE RED E INTERACCIÓN DE APLICACIONES. SEGUIDAMENTE, SE PROPORCIONA UNA METODOLOGÍA DE IMPLANTACIÓN DE LA NORMA Y FINALMENTE SE IMPLANTA UN SISTEMA DE GESTIÓN DE SEGURIDAD DE LA INFORMACIÓN BAJO LA NORMA ISO/IEC 27001:2005. PARA FINALIZAR SE MUESTRAN LAS CONCLUSIONES, LIMITACIONES, TRABAJOS FUTUROS Y DOCUMENTACIÓN GENERADA PARA DAR CONTINUIDAD A LA IMPLANTACIÓN DEL SGSI EN MERCAP.
	Pals. Claves	
	Área(s) TEG	ATI Área(s) Recomendada(s) 1=BD; 2=ATI
16	Título	DISEÑO Y DESARROLLO DE UN MÓDULO DE COMUNICACIÓN ENTRE EL FRAMEWORK RUBY ON RAILS Y LA CENTRAL TELEFÓNICA POR SOFTWARE ASTERISK, PARA EL MERCADEO DE PRODUCTOS Y SERVICIOS EMPRESARIALES
	Resumen	EN EL PRESENTE TRABAJO ESPECIAL DE GRADO SE DISEÑÓ E IMPLEMENTÓ UN MÓDULO DE CONEXIÓN ENTRE EL FRAMEWORK WEB RUBY ON RAILS Y UNA CENTRAL TELEFÓNICA POR SOFTWARE LLAMADA ASTERISK, COMO PARTE DE UNA APLICACIÓN WEB QUE INTEGRÓ ELEMENTOS DE LA TECNOLOGÍA VOZ SOBRE IP (VOIP). DICHO MÓDULO SE IMPLEMENTÓ CON EL APOYO UNA HERRAMIENTA QUE FACILITA EL INTERCAMBIO DE INFORMACIÓN A BAJO NIVEL ENTRE DICHAS TECNOLOGÍAS, LLAMADA TELEGRAPH. LA APLICACIÓN AUTOMATIZÓ EL PROCESO DE OFERTA DE PRODUCTOS Y SERVICIOS EMPRESARIALES, AGILIZANDO EL PROCESO DE CONTACTO VÍA TELEFÓNICA A LOS POTENCIALES CLIENTES. DURANTE EL DESARROLLO SE UTILIZÓ UNA INTERFAZ DE ASTERISK LLAMADA AMI, SOBRE LA CUAL SE ENVÍAN Y RECIBEN COMANDOS DESDE LA APLICACIÓN PARA LOGRAR LOS OBJETIVOS PRINCIPALES DEL PROYECTO, LOS CUALES SE LISTAN A CONTINUACIÓN: 1. REALIZACIÓN DE LLAMADAS DESDE LA APLICACIÓN, ES DECIR QUE EL SOFTWARE MARQUE AUTOMÁTICAMENTE UN NÚMERO RECUPERADO DE UN REPOSITORIO DE DATOS. 2. MONITOREO DE LLAMADAS, CONSISTE EN SUPERVISAR LAS LLAMADAS REALIZADAS. 3. TRANSFERENCIA DE LLAMADAS, SE TRANSFIERE UNA LLAMADA EN CURSO DE UNA EXTENSIÓN A OTRA. 4. ESTATUS DE ASTERISK, ES DECIR CONOCER EL ESTATUS DE LAS EXTENSIONES Y LLAMADAS REALIZADAS EN ASTERISK.
	Pals. Claves	APLICACIONES CON TECNOLOGÍA INTERNET, RUBY ON RAILS (ROR), PROGRAMACIÓN EXTREMA (XP), ASTERISK, TELEGRAPH, VOZ SOBRE IP (VOIP)
	Área(s) TEG	BD Área(s) Recomendada(s) BD
17	Título	DESARROLLO DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA UN DEPARTAMENTO DE FINANZAS DENTRO DEL SECTOR PETROQUÍMICO
	Resumen	EL PRESENTE TRABAJO ESPECIAL DE GRADO (TEG) CONSISTE EN EL DESARROLLO DE UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS (BI, BUSINESS INTELLIGENCE POR SUS SIGLAS EN INGLÉS) PARA UN DEPARTAMENTO DE FINANZAS DEL SECTOR PETROQUÍMICO, EL CUAL MANEJA UNA GRAN CANTIDAD DE INFORMACIÓN Y ESTA ES UTILIZADA PARA REALIZAR LOS ANÁLISIS, REPORTES E INFORMES NECESARIOS. EN LA ACTUALIDAD ESTAS TAREAS SON REALIZADAS DE MANERA MANUAL APOYÁNDOSE EN HOJAS DE CÁLCULOS, LO QUE PUEDE TRAER COMO CONSECUENCIA ERRORES DE MANIPULACIÓN EN LOS DATOS Y UNA GRAN PÉRDIDA DE TIEMPO EN TODO ESTE PROCESO. NUESTRA SOLUCIÓN DE INTELIGENCIA DE NEGOCIO DISMINUYE LOS TIEMPOS EMPLEADOS PARA LA EXTRACCIÓN Y CARGA DE LOS DATOS ASÍ COMO PARA LA GENERACIÓN DE REPORTES, INFORMES, GRÁFICA, ENTRE OTROS Y ASEGURAN LA ENTREGA DE INFORMACIÓN EXACTA, ÚTIL Y NECESARIA PARA APOYAR LA TOMA DE DECISIONES. EL OBJETIVO DE ESTE TEG ES APROVECHAR LAS VENTAJAS QUE NOS OFRECE UNA PLATAFORMA BI LA CUAL ES UNA ARQUITECTURA Y COLECCIÓN DE OPERACIONES INTEGRADAS Y BASES DE DATOS QUE PROPORCIONAN A LA COMUNIDAD EMPRESARIAL EL ACCESO FÁCIL A LOS DATOS; ES DECIR SE APROVECHA LOS PROCESO ANALÍTICO (OLAP) EL CUAL NOS PERMITE UN MEJOR RENDIMIENTO PARA ACCEDER A LOS DATOS ULTIMDIMENSIONALES PROPORCIONANDO UNA BASE TÉCNICA PARA LOS ANÁLISIS Y CÁLCULOS REQUERIDOS, AYUDANDO A QUE SE TOMEN

		DECISIONES DE MANERA RÁPIDA Y PRECISA. PARA ALCANZAR DICHO OBJETIVO SE UTILIZÓ UNA METODOLOGÍA DE DESARROLLO DE SOLUCIONES DE INTELIGENCIA DE NEGOCIO LLAMADA \THE BUSINESS DIMENSIONAL LIFECYCLE\PROPUESTA POR RALPH KIMBALL. Y COMO PLATAFORMA DE DESARROLLO SE UTILIZÓ UNA HERRAMIENTA DE CÓDIGO ABIERTO (OPEN SOURCE) LLAMADA PENTHAHO BI LA CUAL ES UNA PLATAFORMA QUE OFRECE FUNCIONALIDADES COMO LOS SON LA INTEGRACIÓN DE DATOS (ETL), REPORTES, ANÁLISIS, HERRAMIENTAS OLAP, MINERÍA DE DATOS, ENTRE OTRAS.
	Pals. Claves	INTELIGENCIA DE NEGOCIO, DATAWAREHOUSE, PENTHAHO, ANÁLISIS, OLAP, INTEGRACIÓN DE DATOS, METODOLOGÍA DE RALPH KIMBALL.
	Área(s) TEG	1=BD; 2=ATI Área(s) Recomendada(s) 1=ATI; 2=BD
18	Título	DESARROLLO DE UN SISTEMA DE INFORMACIÓN DE PROCESAMIENTO DE TRANSACCIONES (TPS) BASADO EN LA METODOLOGÍA DE DESARROLLO AGILE UNIFIED PROCESS (AUP) PARA LA CAPTURA DE DATOS DE LA ENCUESTA DE ESTABLECIMIENTOS TURÍSTICOS DE ALOJAMIENTO COLECTIVO.
	Resumen	EL DEPARTAMENTO DE ESTADÍSTICAS ECONÓMICAS DEL INSTITUTO NACIONAL DE ESTADÍSTICAS (INE) PROPONE UN NUEVO PROYECTO DENOMINADO ENCUESTA DE ESTABLECIMIENTO DE TURISMO DE ALOJAMIENTO COLECTIVO, EL CUAL EN SU PRIMERA FASE PLANEA LA RECOLECCIÓN DE LOS DATOS DE LA ENCUESTA FÍSICA, Y EN SU SEGUNDA FASE LA CREACIÓN DE TABULADORES O REPORTES SUSTENTADOS EN LOS DATOS OBTENIDOS EN LA PRIMERA FASE. DEBIDO A QUE LA MUESTRA DE ESTABLECIMIENTOS A NIVEL NACIONAL ES BASTANTE AMPLIA, NO RESULTA CONVENIENTE PARA EL DEPARTAMENTO TENER LOS DATOS DE LA ENCUESTA EN FÍSICO, ADEMÁS ESTO DIFICULTARÍA LA EJECUCIÓN DE LA SEGUNDA FASE DEL PROYECTO. POR ELLO SURGE EN EL PRESENTE TRABAJO ESPECIAL DE GRADO (TEG) LA NECESIDAD DE DESARROLLAR UN SISTEMA QUE PERMITA LA CAPTURA DE LOS DATOS DE LA ENCUESTA Y LOS ALMACENE EN UNA BASE DE DATOS CENTRALIZADA. EL SISTEMA DEBE SER WEB, YA QUE PERMITE A LOS TRANSCRIPTORES O INFORMANTES DE LOS DISTINTOS ESTABLECIMIENTOS LLENAR LA ENCUESTA DIGITAL DESDE CUALQUIER LUGAR PARA ALCANZAR DICHO OBJETIVO SE UTILIZÓ LA METODOLOGÍA DE DESARROLLO DEL PROCESO ÁGIL UNIFICADO (AUP, POR SUS SIGLAS EN INGLÉS), LA CUAL OFRECE LA FLEXIBILIDAD NECESARIA PARA EL DESARROLLO Y LA DOCUMENTACIÓN DEL SISTEMA. EL PATRÓN DE DISEÑO MODELO- VISTA- CONTROLADOR (MVC), EL CUAL PERMITE EL DISEÑO MODULAR DEL SISTEMA, DISMINUYENDO ASÍ LAS COMPLICACIONES ANTE LOS CAMBIOS, Y SE EMPLEARON LOS LENGUAJE DE PROGRAMACIÓN WEB PHP Y JAVASCRIPT LOS CUALES FACILITAN CREAR LAS FUNCIONALIDADES DEL SISTEMA Y EL MANEJADOR DE BASE DE DATOS POSTGRESQL PARA ALMACENAR LOS DATOS DE LA ENCUESTA
	Pals. Claves	TURISMO, ESTABLECIMIENTOS TURÍSTICOS, TURISMO DE OFERTA, ALOJAMIENTO COLECTIVO, AUP, MVC, JAVASCRIPT, PHP, POSTGRESQL.
	Área(s) TEG	ATI Área(s) Recomendada(s) ATI
19	Título	DESARROLLO DE UNA APLICACIÓN PARA PROCESAR Y ORGANIZAR OFICIOS DIGITALIZADOS DE MODIFICACIÓN Y APERTURA DE MATERIAS DE LA DIVISIÓN DE CONTROL DE ESTUDIOS DE LA FACULTAD DE CIENCIAS DE LA UNIVERSIDAD CENTRAL DE VENEZUELA.
	Resumen	EL OBJETIVO DEL PRESENTE TRABAJO ESPECIAL DE GRADO CONSISTE EN EL DESARROLLO DE UN SISTEMA QUE AUTOMATICE EL PROCESO DE DIGITALIZACIÓN DE OFICIOS DE APERTURAS Y MODIFICACIONES DE MATERIAS DE LA FACULTAD DE CIENCIAS DE LA UNIVERSIDAD CENTRAL DE VENEZUELA, QUE HAN SIDO ALMACENADOS EN LAS INSTALACIONES DE LA DIVISIÓN DE CONTROL DE ESTUDIOS, ESPECÍFICAMENTE UN ARCHIVO UBICADO EN LA OFICINA DE LA DIRECTORA DE DICHA DIVISIÓN. EL SISTEMA ESTÁ DIVIDIDO EN DOS PARTES, DONDE SE TIENE UNA APLICACIÓN LOCAL ENCARGADA DE EMPAQUETAR Y ENVIAR LAS IMÁGENES AL SERVIDOR, Y EN SEGUNDO LUGAR SE TIENE UNA APLICACIÓN WEB ENCARGADA DEL PROCESAMIENTO DE LOS OFICIOS. PARA CUMPLIR CON LOS OBJETIVOS DE LA APLICACIÓN SE UTILIZÓ LA METODOLOGÍA ÁGIL XP (PROGRAMACIÓN EXTREMA), LA CUAL SE BASA FUNDAMENTALMENTE, EN LA CONSTRUCCIÓN PROGRESIVA DEL SOFTWARE SIN HACER ÉNFASIS EN ETAPAS DE DISEÑOS.

	Pals. Claves	PROCESAMIENTO DIGITAL DE IMÁGENES, OPTICAL CHARACTER RECOGNITION (OCR), APLICACIONES WEB, FACULTAD DE CIENCIAS UCV.		
	Área(s) TEG	1=ATI; 2=BD	Área(s) Recomendada(s)	1=BD; 2=ATI
20	Título	DESARROLLO DE UNA APLICACION WEB DE PROGRAMA DE INCENTIVOS BAJO LA ARQUITECTURA CLIENTE-SERVIDOR UTILIZANDO LA METODOLOGIA DE DESARROLLO RUP-SCRUM		
	Resumen	<p>EL OBJETIVO DE ESTE TRABAJO ESPECIAL DE GRADO ES CONSTRUIR UNA APLICACION WEB QUE SIRVA COMO BASE PARA EL DESARROLLO DE PROGRAMAS DE INCENTIVOS, A TRAVES DE LA FACIL MODIFICACION DE LA INTERFAZ SEGUN LA IMAGEN DE LA EMPRESA ASI COMO LAS REGLAS DE NEGOCIO QUE DEBEN APLICAR, UTILIZANDO UNA METODOLOGIA DE DESARROLLO HIBRIDA ENTRE RUP Y SCRUM, DONDE SE TOMEN LOS ASPECTOS DE CADA UNA QUE PERMITAN DESARROLLAR UN BUEN PRODUCTO CON TIEMPOS DE ESPERA CORTOS. UN PROGRAMA DE INCENTIVOS PRETENDE INCENTIVAR LA FUERZA DE TRABAJO DE UNA EMPRESA EN LA PROMOCION Y VENTA DE LOS PRODUCTOS QUE OFRECE DICHA EMPRESA. EL DESARROLLO DE LA APLICACION WEB SE LLEVO A CABO UTILIZANDO COMO METODO DE DESARROLLO UNA COMBINACION ENTRE RUP Y SCRUM. EL TRABAJO SE DIVIDIO EN CUATRO (4) FASES SEGUN LA METODOLOGIA RUP (A SABER): INICIO, ELABORACION, CONSTRUCCION Y TRANSICION. EN LA FASE DE INICIO SE ANALIZO EL PROBLEMA, SE DETERMINARON LAS PRINCIPALES FUNCIONALIDADES DEL SISTEMA Y SE REALIZARON LAS PRIMERAS VERSIONES DEL MODELO CONCEPTUAL, DEL MODELO DE DATOS Y DE LA INTERFAZ DEL SISTEMA TOMANDO EN CUENTA LA NAVEGABILIDAD DENTRO DE LA APLICACION WEB. EN LA FASE DE ELABORACION SE PROFUNDIZO EL ANALISIS Y SE DETERMINARON EL RESTO DE LAS FUNCIONALIDADES DEL SISTEMA. LA FASE DE CONSTRUCCION SE DESARROLLO DE ACUERDO A LA METODOLOGIA SCRUM REALIZANDO UN LISTADO DE LAS FUNCIONALIDADES DEL SISTEMA AGRUPADAS POR MODULOS, SEGUN SU AFINIDAD, E IMPLEMENTANDOLAS SEMANALMENTE DE MODO QUE AL FINAL DE CADA SEMANA SE OBTUVO UN INCREMENTO FUNCIONAL DE LA APLICACION. FINALMENTE EN LA FASE DE TRANSICION SE INSTALO EL PRODUCTO Y SE REALIZARON PRUEBAS COLECTIVAS MEDIANTE UNA MATRIZ DE PRUEBAS. ESTA COMBINACION DE METODOLOGIAS DE DESARROLLO DE SOFTWARE PERMITIO FORMALIZAR LA DOCUMENTACION DEL PROYECTO, A TRAVES DE LOS DIFERENTES DIAGRAMAS REALIZADOS, Y MINIMIZAR LOS TIEMPOS DE ENTREGA, AL DIVIDIR EL TRABAJO DE TAL FORMA QUE SEMANALMENTE SE OBTUVIERA UN INCREMENTO FUNCIONAL DEL SISTEMA. A LA VEZ QUE SE LOGRO UN PRODUCTO DE CALIDAD QUE CUMPLIO CON LAS EXPECTATIVAS DEL USUARIO FINAL.</p>		
	Pals. Claves	PROGRAMA DE INCENTIVOS, CARRITO DE COMPRA, METODOLOGÍA, LEALTAD, MULTIMONEDA, TIENDA VIRTUAL DE RECOMPENSAS		
Área(s) TEG	ATI	Área(s) Recomendada(s)	ATI	
21	Título	DESARROLLO DE UN PORTAL WEB PARA APOYAR EL APRENDIZAJE CONTINUO TECNOLÓGICO.		
	Resumen	<p>ESTE TRABAJO (ENMARCADO DENTRO DEL PROYECTO FONACIT NRO. 2005000166) MUESTRA ELDESARROLLO DE UN PORTAL QUE PROPORCIONA ELEMENTOS COMUNICACIONALES PARA PROMOVER ELAPRENDIZAJE COLABORATIVO Y ORGANIZACIONAL. SE ESTABLECE UNA INFRAESTRUCTURA QUE BRINDA ELSOPORTE NECESARIO PARA QUE LAS PERSONAS ALFABETIZADAS TECNOLÓGICAMENTE Y SUS FACILITADORES, PUEDAN DISTRIBUIR CONOCIMIENTO TECNOLÓGICO BIEN SEA, PARA REFORZAR LO APRENDIDO O PARAACTUALIZARSE, DANDO CONTINUIDAD AL PROCESO DE ENSEÑANZA-APRENDIZAJE. LA INFRAESTRUCTURA PERMITE QUE LOS USUARIOS PUEDAN INCORPORAR Y DISTRIBUIR CONOCIMIENTO ENDIFERENTES MODALIDADES DE COMUNICACIÓN (FOROS, NOTICIAS Y ARTÍCULOS), ASÍ COMO PROVEEMECANISMOS PARA LA INCORPORACIÓN DE NUEVOS COMPONENTES COMUNICACIONALES QUE FACILITENQUE MEJOREN LA COMUNICACIÓN ENTRE LOS USUARIOS. LA SOLUCIÓN PROPUESTA INCLUYE COMO PASO INICIAL EL MODELADO DE NEGOCIO, PARA LO CUAL SEUTILIZA LA NOTACIÓN BPM (BUSINESS PROCESS MODEL), IDENTIFICANDO LOS PROCESOS PRINCIPALESQUE DEBEN SER IMPLANTADOS A TRAVÉS DEL PORTAL. EN EL DESARROLLO DEL PORTAL SE UTILIZA LAHERRAMIENTA</p>		

		<p>JOOMLA (SOLUCIÓN DE SOFTWARE LIBRE), ESTA PROVEE FUNCIONALES PARA LA CONSTRUCCIÓN RÁPIDA DE PORTALES, DEBIDO A ESTO, SE PROPONE UN MÉTODO DE DESARROLLO BASADO EN EL MODELADO ÁGIL (MA), ESPECIFICANDO COMO ARTEFACTOS PRINCIPALES: EL MODELO DE CASOS DE USO Y ACTIVIDADES (UTILIZANDO LA NOTACIÓN UML - UNIFIED MODELING LANGUAGE), PATRONES DE INTERACCIÓN PROPUESTOS POR WELIE (SOLUCIONES A PROBLEMAS ESPECÍFICOS DE INTERACCIÓN) Y COMO ARQUITECTURA DE SOFTWARE LA PROPUESTA POR JOOMLA. LA EXPERIENCIA LOGRA ESTABLECER DOS CONEXIONES IMPORTANTES, LA PRIMERA, TIENE QUE VER CON LAS TEORÍAS DE APRENDIZAJES Y CÓMO ÉSTAS PUEDEN SER IMPLANTADAS A TRAVÉS DE UNA INFRAESTRUCTURA TECNOLÓGICA; LA SEGUNDA, ESTABLECE LA RELACIÓN QUE EXISTE ENTRE UNA HERRAMIENTA PARA LA CONSTRUCCIÓN RÁPIDA DE PORTALES Y EL PROCESO DE DESARROLLO ÁGIL.</p>		
	Pals. Claves	<p>APRENDIZAJE COLABORATIVO, APRENDIZAJE ORGANIZACIONAL, ALFABETIZACIÓN TECNOLÓGICA, APRENDIZAJE CONTINUO TECNOLÓGICO, MODELADO ÁGIL, PROCESO DE DESARROLLO.</p>		
	Área(s) TEG	1=ATI; 2=BD	Área(s) Recomendada(s)	BD
22	Título	<p>DESARROLLO DE UNA APLICACIÓN WEB PARA EL CENTRO DE CÓMPUTO DE UNA PROCESADORA DE MEDIOS DE PAGO QUE AUTOMATICE LA GESTIÓN DE INCIDENCIAS.</p>		
	Resumen	<p>EL PRESENTE TRABAJO ESPECIAL DE GRADO EXPONE EL DESARROLLO DE UNA APLICACIÓN WEB CAPAZ DE GESTIONAR EFICIENTEMENTE, LAS INCIDENCIAS OCURRIDAS EN EL CENTRO DE CÓMPUTO. LA APLICACIÓN DESARROLLADA DEBERÁ FACILITAR, EL REGISTRO ACTUALIZACIÓN, CONSULTA Y SEGUIMIENTO DE LAS INCIDENCIAS. ADICIONALMENTE, DEBERÁ MEJORAR EL RENDIMIENTO, PRODUCTIVIDAD Y CAPACIDAD DE RESPUESTA AL PERSONAL DEL CENTRO DE CÓMPUTO. LA APLICACIÓN SE DESARROLLÓ SIGUIENDO UNA METODOLOGÍA DENOMINADA PROCESO UNIFICADO, LA CUAL UTILIZA COMO UNA DE SUS HERRAMIENTAS PRINCIPALES LOS DIAGRAMAS DE CASOS DE USO PARA REPRESENTAR LAS FUNCIONALIDADES Y REQUERIMIENTOS DE LA APLICACIÓN, SE CARACTERIZA POR SER CAPAZ DE ADAPTARSE A CAMBIOS DE REQUISITOS EN CUALQUIER PUNTO DEL CICLO DE DESARROLLO DE LA APLICACIÓN Y ESTIMULA EL CONTINUO CONTACTO QUE DEBE MANTENERSE ENTRE EL DESARROLLADOR Y EL CLIENTE. PARA LA IMPLEMENTACIÓN SE UTILIZA COMO HERRAMIENTAS DE DESARROLLO, PHP (HYPERTEXT PRE-PROCESSOR), SISTEMA MANEJADOR DE BASES DE DATOS MYSQL Y JAVASCRIPT, ESTAS HERRAMIENTAS SE CARACTERIZAN POR SER SOFTWARE LIBRE Y JUNTAS PROVEEN LAS FUNCIONALIDADES NECESARIAS PARA LA CONSTRUCCIÓN DE LA APLICACIÓN. EL RESULTADO OBTENIDO ES UNA APLICACIÓN SENCILLA DE MANEJAR, QUE LOGRA OPTIMIZAR Y FACILITAR EL PROCESO DE GESTIÓN DE INCIDENCIAS, A TRAVÉS DE UNA INFRAESTRUCTURA TECNOLÓGICA QUE ABRE LAS PUERTAS HACIA UN CRECIMIENTO COMPETITIVO Y PERMANENTE.</p>		
	Pals. Claves	<p>MEDIOS DE PAGO, PROCESADORA DE MEDIOS DE PAGO, CENTRO DE CÓMPUTO, INCIDENCIAS, SISTEMA DE REGISTRO DE INCIDENCIAS.</p>		
	Área(s) TEG	IA	Área(s) Recomendada(s)	1=BD; 2=IA
23	Título	<p>PROTOTIPO DE HERRAMIENTA PARA EL DESARROLLO DE APLICACIONES BASADAS EN LÓGICA DIFUSA</p>		
	Resumen	<p>LA LÓGICA DIFUSA BRINDA UNA BASE MATEMÁTICA PARA PODER MODELAR CONCEPTOS IMPRECISOS QUE SE UTILIZAN EN EL LENGUAJE DIARIO Y QUE NO PUEDEN SER DEFINIDOS DE MANERA EXACTA. ESTO PERMITE RESOLVER UN GRAN NÚMERO DE PROBLEMAS QUE NO PODRÍAN CONSEGUIR UNA SOLUCIÓN DE MANERA SATISFACTORIA DADA LA NATURALEZA DEL DOMINIO DONDE SE PRESENTAN Y LOS PARÁMETROS QUE LOS RODEAN. DESDE 1965, AÑO EN EL CUAL SE PUBLICÓ EL ARTÍCULO "FUZZY SETS" (ZADEH, 1965), EL USO DE LA LÓGICA DIFUSA COMENZÓ A POPULARIZARSE Y, ACTUALMENTE, LA IMPLEMENTACIÓN EXITOSA DE HERRAMIENTAS BASADAS EN LÓGICA DIFUSA HA TRAÍDO COMO CONSECUENCIA UN INCREMENTO DE SU UTILIZACIÓN EN APLICACIONES PARA SISTEMAS DE CONTROL INDUSTRIALES, RECONOCIMIENTO DE PATRONES, SISTEMAS EXPERTOS, ELECTRODOMÉSTICOS, MINERÍA DE DATOS, ETC. PARA EL DISEÑO DE</p>		

		SISTEMAS BASADOS EN LÓGICA DIFUSA EXISTE UNA AMPLIA GAMA DE HERRAMIENTAS DISPONIBLES, SIN EMBARGO LA MAYORÍA NO SON DE LIBRE DISTRIBUCIÓN, O SU USABILIDAD ES LIMITADA. DEBIDO A ESTO, EL OBJETIVO PRINCIPAL DE ESTE TRABAJO ES DESARROLLAR UN PROTOTIPO DE HERRAMIENTA PARA EL DESARROLLO DE SISTEMAS BASADOS EN LOGICA DIFUSA, EL CUAL PERMITA A LOS USUARIOS CONTAR CON UN AMBIENTE DE DESARROLLO POTENTE Y FACIL DE UTILIZAR, Y QUE ADICIONALMENTE SEA DE DISTRIBUCIÓN LIBRE.
	Pals. Claves	LÓGICA DIFUSA, CONOCIMIENTO EXPERTO, HERRAMIENTA DE DESARROLLO

Anexo 7. Resultados de las pruebas realizadas al módulo de asignación de jurados.

PRUEBA DOCUMENTO 1						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI	ATI	Andrés Sanoja y Eleonora Acosta	Prof. Andrés Germán Castro Orta	Prof. Sergio José Rivas Atanacio	Profa. Jossie Iranidis Zambrano	Profa. Yusneyi Yasmira Carballo Barrera

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
0	0	Andrés Sanoja	ATI
0	0	Eleonora Acosta	
0	1	Yusneyi Carballo	
1	0	Andrés Castro	
0	1	Jossie Zambrano	
1	0	Sergio Rivas	
0	0	Daniel Villavicencio	REDES
0	0	Robinson Rivas	
0	0	Karima Velásquez	
0	0	David Pérez	
0	0	María E. Villapol	
0	0	Rafael Angulo	
0	0	Wuifredo Rangel	BD
0	0	Concettina Di Vasta	
0	0	Antonio Silva	
0	0	Rossana Díaz	
0	0	Paola Saputelli	
0	0	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
0	0	Iván Flores	
0	0	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 2						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
REDES	REDES y BD	Daniel Villavicencio y Robinson Rivas	Prof. David Alejandro Pérez Abreu	Profa. Paola Saputelli	Profa. Karima Daniela Velásquez Castro	Profa. Concettina Di Vasta Valente

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
0	0	Andrés Sanoja	ATI
0	0	Eleonora Acosta	
0	1	Yusneyi Carballo	
1	0	Andrés Castro	
0	1	Jossie Zambrano	
1	0	Sergio Rivas	

0	0	Daniel Villavicencio	REDES
0	0	Robinson Rivas	
0	1	Karima Velásquez	
1	0	David Pérez	
0	0	María E. Villapol	
0	0	Rafael Angulo	
0	0	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	0	Antonio Silva	
0	0	Rossana Díaz	
1	0	Paola Saputelli	
0	0	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
0	0	Iván Flores	
0	0	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 3						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI	ATI	Yusneyi Carballo y Andrés Castro	Profa. Alecia Eleonora Acosta Freites	Prof. Andrés Fernando Sanoja Vargas	Prof. Jossie Iranidis Zambrano	Prof. Sergio José Rivas Atanacio

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
1	0	Andrés Sanoja	ATI
1	0	Eleonora Acosta	
0	1	Yusneyi Carballo	
1	0	Andrés Castro	
0	2	Jossie Zambrano	
1	1	Sergio Rivas	
0	0	Daniel Villavicencio	REDES
0	0	Robinson Rivas	
0	1	Karima Velásquez	
1	0	David Pérez	
0	0	María E. Villapol	
0	0	Rafael Angulo	
0	0	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	0	Antonio Silva	
0	0	Rossana Díaz	
1	0	Paola Saputelli	
0	0	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
0	0	Iván Flores	
0	0	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 4						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
REDES	REDES	Karima Velásquez y David Pérez	Prof. Rafael Alejandro Angulo Ronzulli	Prof. Daniel Villavicencio	Prof. Robinson Samuel Rivas Suárez	Profa. María Elena Villapol

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
1	0	Andrés Sanoja	ATI
1	0	Eleonora Acosta	
0	1	Yusneyi Carballo	
1	0	Andrés Castro	
0	2	Jossie Zambrano	
1	1	Sergio Rivas	
1	0	Daniel Villavicencio	
0	1	Robinson Rivas	
0	1	Karima Velásquez	
1	0	David Pérez	
0	1	María E. Villapol	
1	0	Rafael Angulo	BD
0	0	Wuifredo Rangel	
0	1	Concettina Di Vasta	
0	0	Antonio Silva	
0	0	Rossana Díaz	
1	0	Paola Saputelli	
0	0	Mercy Ospina	
0	0	Marcel Castro	
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
0	0	Iván Flores	
0	0	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 5						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI y BD	BD y ATI	Jossie Zambrano y Wuifredo Rangel	Profa. Mercy H Ospina Torres	Profa. Yusneyi Yasmira Carballo Barrera	Prof. Antonio Mauricio Silva Sprock	Prof. Andrés Germán Castro Orta

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
1	0	Andrés Sanoja	ATI
1	0	Eleonora Acosta	
1	1	Yusneyi Carballo	
1	1	Andrés Castro	
0	2	Jossie Zambrano	
1	1	Sergio Rivas	
1	0	Daniel Villavicencio	
0	1	Robinson Rivas	
0	1	Karima Velásquez	
1	0	David Pérez	
0	1	María E. Villapol	
1	0	Rafael Angulo	

0	0	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	1	Antonio Silva	
0	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
0	0	Iván Flores	
0	0	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 6						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
REDES	REDES	María E. Villapol y Rafael Angulo	Prof. Robinson Samuel Rivas Suárez	Profa. Karima Daniela Velásquez Castro	Prof. Daniel Villavicencio	Prof. David Alejandro Pérez Abreu

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
1	0	Andrés Sanoja	ATI
1	0	Eleonora Acosta	
1	1	Yusneyi Carballo	
1	1	Andrés Castro	
0	2	Jossie Zambrano	
1	1	Sergio Rivas	
1	1	Daniel Villavicencio	REDES
1	1	Robinson Rivas	
1	1	Karima Velásquez	
1	1	David Pérez	
0	1	María E. Villapol	
1	0	Rafael Angulo	
0	0	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	1	Antonio Silva	
0	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
0	0	Iván Flores	
0	0	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 7						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI y BD	ATI	Sergio Rivas y Concettina Di Vasta	Prof. Jossie Iranidis Zambrano	Prof. Andrés Fernando Sanoja Vargas	Profa. Alecia Eleonora Acosta Freites	Prof. Andrés Germán Castro Orta

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	0	Andrés Sanoja	ATI
1	1	Eleonora Acosta	
1	1	Yusneyi Carballo	
1	2	Andrés Castro	
1	2	Jossie Zambrano	
1	1	Sergio Rivas	
1	1	Daniel Villavicencio	REDES
1	1	Robinson Rivas	
1	1	Karima Velásquez	
1	1	David Pérez	
0	1	María E. Villapol	
1	0	Rafael Angulo	
0	0	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	1	Antonio Silva	
0	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
0	0	Iván Flores	
0	0	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 8						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI	ATI	Andrés Sanoja	Prof. Sergio José Rivas Atanacio	Profa. Alecia Eleonora Acosta Freites	Profa. Yusneyi Yasmira Carballo Barrera	Prof. Jossie Iranidis Zambrano

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	0	Andrés Sanoja	ATI
2	1	Eleonora Acosta	
1	2	Yusneyi Carballo	
1	2	Andrés Castro	
1	3	Jossie Zambrano	
2	1	Sergio Rivas	
1	1	Daniel Villavicencio	REDES
1	1	Robinson Rivas	
1	1	Karima Velásquez	
1	1	David Pérez	
0	1	María E. Villapol	
1	0	Rafael Angulo	
0	0	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	1	Antonio Silva	
0	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	

0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
0	0	Iván Flores	
0	0	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 9						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
REDES	REDES e IA	Daniel Villavicencio	Prof. María Elena Villapol	Prof. Iván José Flores Vitelli	Prof. Rafael Alejandro Angulo Ronzulli	Prof. Nora Elena Montaña Fermín

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	0	Andrés Sanoja	ATI
2	1	Eleonora Acosta	
1	2	Yusneyi Carballo	
1	2	Andrés Castro	
1	3	Jossie Zambrano	
2	1	Sergio Rivas	
1	1	Daniel Villavicencio	REDES
1	1	Robinson Rivas	
1	1	Karima Velásquez	
1	1	David Pérez	
1	1	María E. Villapol	
1	1	Rafael Angulo	BD
0	0	Wuilfredo Rangel	
0	1	Concettina Di Vasta	
0	1	Antonio Silva	
0	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	IA
0	0	Marcel Castro	
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 10						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
REDES	REDES	Robinson Rivas	Prof. Rafael Alejandro Angulo Ronzulli	Prof. David Alejandro Pérez Abreu	Prof. María Elena Villapol	Prof. Daniel Villavicencio

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	0	Andrés Sanoja	ATI
2	1	Eleonora Acosta	
1	2	Yusneyi Carballo	
1	2	Andrés Castro	
1	3	Jossie Zambrano	
2	1	Sergio Rivas	

1	2	Daniel Villavicencio	REDES
1	1	Robinson Rivas	
1	1	Karima Velásquez	
2	1	David Pérez	
1	2	María E. Villapol	
2	1	Rafael Angulo	
0	0	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	1	Antonio Silva	
0	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 11						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI	ATI	Eleonora Acosta	Prof. Andrés Germán Castro Orta	Profa. Yusneyi Yasmira Carballo Barrera	Prof. Jossie Iranidis Zambrano	Prof. Andrés Fernando Sanoja Vargas

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	1	Andrés Sanoja	ATI
2	1	Eleonora Acosta	
2	2	Yusneyi Carballo	
2	2	Andrés Castro	
1	4	Jossie Zambrano	
2	1	Sergio Rivas	
1	2	Daniel Villavicencio	REDES
1	1	Robinson Rivas	
1	1	Karima Velásquez	
2	1	David Pérez	
1	2	María E. Villapol	
2	1	Rafael Angulo	
0	0	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	1	Antonio Silva	
0	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 12						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI	ATI y BD	Yusneyi Carballo	Prof. Jossie Iranidis Zambrano	Profa. Rossana Díaz Salguero	Prof. Andrés Fernando Sanoja Vargas	Prof. Wilfredo Andres Rangel Azuaje

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	2	Andrés Sanoja	ATI
2	1	Eleonora Acosta	
2	2	Yusneyi Carballo	
2	2	Andrés Castro	
2	4	Jossie Zambrano	
2	1	Sergio Rivas	
1	2	Daniel Villavicencio	
1	1	Robinson Rivas	REDES
1	1	Karima Velásquez	
2	1	David Pérez	
1	2	María E. Villapol	
2	1	Rafael Angulo	
0	1	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	1	Antonio Silva	
1	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	IA
0	0	Marcel Castro	
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 13						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI	ATI	Andrés Castro	Profa. Alecia Eleonora Acosta Freites	Prof. Sergio José Rivas Atanacio	Prof. Andrés Fernando Sanoja Vargas	Profa. Yusneyi Yasmira Carballo Barrera

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	3	Andrés Sanoja	ATI
3	1	Eleonora Acosta	
2	3	Yusneyi Carballo	
2	2	Andrés Castro	
2	4	Jossie Zambrano	
3	1	Sergio Rivas	
1	2	Daniel Villavicencio	
1	1	Robinson Rivas	REDES
1	1	Karima Velásquez	
2	1	David Pérez	
1	2	María E. Villapol	
2	1	Rafael Angulo	

0	1	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	1	Antonio Silva	
1	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 14						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
IA	REDES	Marcel Castro	Profa. Karima Daniela Velásquez Castro	Prof. Robinson Samuel Rivas Suárez	Prof. Daniel Villavicencio	Profa. María Elena Villapol

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	3	Andrés Sanoja	ATI
3	1	Eleonora Acosta	
2	3	Yusneyi Carballo	
2	2	Andrés Castro	
2	4	Jossie Zambrano	
3	1	Sergio Rivas	
1	3	Daniel Villavicencio	REDES
2	1	Robinson Rivas	
2	1	Karima Velásquez	
2	1	David Pérez	
1	3	María E. Villapol	
2	1	Rafael Angulo	
0	1	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	1	Antonio Silva	
1	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 15						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
REDES	REDES	Karima Velásquez	Profa. María Elena Villapol	Prof. Daniel Villavicencio	Prof. Rafael Alejandro Angulo Ronzulli	Prof. David Alejandro Pérez Abreu

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	3	Andrés Sanoja	ATI
3	1	Eleonora Acosta	
2	3	Yusneyi Carballo	
2	2	Andrés Castro	
2	4	Jossie Zambrano	
3	1	Sergio Rivas	
2	3	Daniel Villavicencio	REDES
2	1	Robinson Rivas	
2	1	Karima Velásquez	
2	2	David Pérez	
2	3	María E. Villapol	
2	2	Rafael Angulo	
0	1	Wuifredo Rangel	BD
0	1	Concettina Di Vasta	
0	1	Antonio Silva	
1	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 16						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI	BD y ATI	Jossie Zambrano	Profa. Concettina Di Vasta Valente	Prof. Andrés Germán Castro Orta	Prof. Antonio Mauricio Silva Sprock	Prof. Andrés Fernando Sanoja Vargas

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	4	Andrés Sanoja	ATI
3	1	Eleonora Acosta	
2	3	Yusneyi Carballo	
3	2	Andrés Castro	
2	4	Jossie Zambrano	
3	1	Sergio Rivas	
2	3	Daniel Villavicencio	REDES
2	1	Robinson Rivas	
2	1	Karima Velásquez	
2	2	David Pérez	
2	3	María E. Villapol	
2	2	Rafael Angulo	
0	1	Wuifredo Rangel	BD
1	1	Concettina Di Vasta	
0	2	Antonio Silva	
1	0	Rossana Díaz	
1	0	Paola Saputelli	
1	0	Mercy Ospina	

0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 17						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
BD	BD	Antonio Silva	Prof. Wilfredo Andres Rangel Azuaje	Profa. Paola Saputelli	Profa. Mercy H Ospina Torres	Profa. Rossana Díaz Salguero

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	4	Andrés Sanoja	ATI
3	1	Eleonora Acosta	
2	3	Yusneyi Carballo	
3	2	Andrés Castro	
2	4	Jossie Zambrano	
3	1	Sergio Rivas	
2	3	Daniel Villavicencio	REDES
2	1	Robinson Rivas	
2	1	Karima Velásquez	
2	2	David Pérez	
2	3	María E. Villapol	
2	2	Rafael Angulo	
1	1	Wilfredo Rangel	BD
1	1	Concettina Di Vasta	
0	2	Antonio Silva	
1	1	Rossana Díaz	
2	0	Paola Saputelli	
1	1	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 18						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
BD y ATI	ATI y BD	Rossana Díaz y Sergio Rivas	Profa. Yusneyi Yasmira Carballo Barrera	Prof. Antonio Mauricio Silva Sprock	Prof. Jossie Iranidis Zambrano	Profa. Concettina Di Vasta Valente

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	4	Andrés Sanoja	ATI
3	1	Eleonora Acosta	
3	3	Yusneyi Carballo	
3	2	Andrés Castro	
2	5	Jossie Zambrano	
3	1	Sergio Rivas	
2	3	Daniel Villavicencio	REDES
2	1	Robinson Rivas	
2	1	Karima Velásquez	
2	2	David Pérez	
2	3	María E. Villapol	
2	2	Rafael Angulo	
1	1	Wuifredo Rangel	BD
1	2	Concettina Di Vasta	
1	2	Antonio Silva	
1	1	Rossana Díaz	
2	0	Paola Saputelli	
1	1	Mercy Ospina	
0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 19						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI	ATI	Andrés Sanoja	Prof. Jossie Iranidis Zambrano	Profa. Alecia Eleonora Acosta Freites	Prof. Sergio José Rivas Atanacio	Prof. Andrés Germán Castro Orta

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
2	4	Andrés Sanoja	ATI
4	1	Eleonora Acosta	
3	3	Yusneyi Carballo	
3	3	Andrés Castro	
3	5	Jossie Zambrano	
3	2	Sergio Rivas	
2	3	Daniel Villavicencio	REDES
2	1	Robinson Rivas	
2	1	Karima Velásquez	
2	2	David Pérez	
2	3	María E. Villapol	
2	2	Rafael Angulo	
1	1	Wuifredo Rangel	BD
1	2	Concettina Di Vasta	
1	2	Antonio Silva	
1	1	Rossana Díaz	
2	0	Paola Saputelli	
1	1	Mercy Ospina	

0	0	Marcel Castro	IA
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 20						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI y BD	BD y ATI	Eleonora Acosta y Paola Saputelli	Prof. Wilfredo Andres Rangel Azuaje	Prof. Andrés Fernando Sanoja Vargas	Profa. Mercy H Ospina Torres	Prof. Sergio José Rivas Atanacio

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
3	4	Andrés Sanoja	ATI
4	1	Eleonora Acosta	
3	3	Yusneyi Carballo	
3	3	Andrés Castro	
3	5	Jossie Zambrano	
3	3	Sergio Rivas	
2	3	Daniel Villavicencio	REDES
2	1	Robinson Rivas	
2	1	Karima Velásquez	
2	2	David Pérez	
2	3	María E. Villapol	
2	2	Rafael Angulo	
2	1	Wuilfredo Rangel	BD
1	2	Concettina Di Vasta	
1	2	Antonio Silva	
1	1	Rossana Díaz	
2	0	Paola Saputelli	
1	2	Mercy Ospina	IA
0	0	Marcel Castro	
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 21						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI	ATI	Yusneyi Carballo	Prof. Andrés Germán Castro Orta	Prof. Sergio José Rivas Atanacio	Prof. Andrés Fernando Sanoja Vargas	Prof. Jossie Iranidis Zambrano

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
3	5	Andrés Sanoja	ATI
4	1	Eleonora Acosta	
3	3	Yusneyi Carballo	
4	3	Andrés Castro	
3	6	Jossie Zambrano	
4	3	Sergio Rivas	

2	3	Daniel Villavicencio	REDES
2	1	Robinson Rivas	
2	1	Karima Velásquez	
2	2	David Pérez	
2	3	María E. Villapol	
2	2	Rafael Angulo	
2	1	Wuifredo Rangel	BD
1	2	Concettina Di Vasta	
1	2	Antonio Silva	
1	1	Rossana Díaz	
2	0	Paola Saputelli	
1	2	Mercy Ospina	IA
0	0	Marcel Castro	
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 22						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
ATI y BD	BD	Andrés Castro y Mercy Ospina	Profa. Rossana Díaz Salguero	Prof. Antonio Mauricio Silva Sprock	Profa. Concettina Di Vasta Valente	Profa. Paola Saputelli

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
3	5	Andrés Sanoja	ATI
4	1	Eleonora Acosta	
3	3	Yusneyi Carballo	
4	3	Andrés Castro	
3	6	Jossie Zambrano	
4	3	Sergio Rivas	
2	3	Daniel Villavicencio	REDES
2	1	Robinson Rivas	
2	1	Karima Velásquez	
2	2	David Pérez	
2	3	María E. Villapol	
2	2	Rafael Angulo	
2	1	Wuifredo Rangel	BD
1	3	Concettina Di Vasta	
2	2	Antonio Silva	
2	1	Rossana Díaz	
2	1	Paola Saputelli	
1	2	Mercy Ospina	IA
0	0	Marcel Castro	
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
0	0	Haydemar Núñez	

PRUEBA DOCUMENTO 23						
Área del TEG	Área Recomendada	Tutor(es)	Principal1	Principal2	Suplente1	Suplente2
IA	BD e IA	Esmeralda Ramos	Profa. Mercy H Ospina Torres	Profa. Haydemar María Nuñez Castro	Profa. Concettina Di Vasta Valente	Prof. Marcel José Castro González

Cant. Asig. Princ.	Cant. Asig. Sup.	Profesor	Área
3	5	Andrés Sanoja	ATI
4	1	Eleonora Acosta	
3	3	Yusneyi Carballo	
4	3	Andrés Castro	
3	6	Jossie Zambrano	
4	3	Sergio Rivas	
2	3	Daniel Villavicencio	
2	1	Robinson Rivas	
2	1	Karima Velásquez	
2	2	David Pérez	
2	3	María E. Villapol	
2	2	Rafael Angulo	
2	1	Wuilfredo Rangel	BD
1	4	Concettina Di Vasta	
2	2	Antonio Silva	
2	1	Rossana Díaz	
2	1	Paola Saputelli	
2	2	Mercy Ospina	
0	1	Marcel Castro	
0	0	Esmeralda Ramos	
0	0	Antonio Silva	
1	0	Iván Flores	
0	1	Nora Montaña	
1	0	Haydemar Núñez	