

# UNIVERSIDAD DE SONORA DIVISIÓN DE INGENIERÍA



## POSGRADO EN INGENIERÍA INDUSTRIAL MAESTRÍA EN INGENIERÍA EN SISTEMAS Y TECNOLOGÍA

CATEGORIZACIÓN Y RECUPERACIÓN DE CONOCIMIENTO  
ORGANIZACIONAL EN EL DEPARTAMENTO DE INGENIERÍA  
INDUSTRIAL

# T E S I S

PRESENTADA POR

**MIGUEL ÁNGEL ROMERO OCHOA**

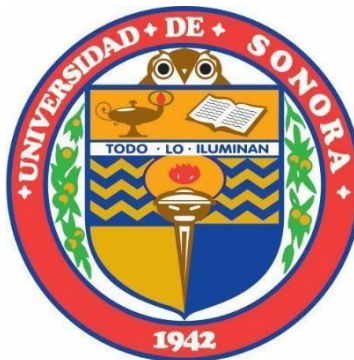
Desarrollada para cumplir con uno de los  
requerimientos parciales para obtener  
el grado de Maestro en Ingeniería

DIRECTOR DE TESIS  
DR. MARIO BARCELÓ VALENZUELA

HERMOSILLO, SONORA, MÉXICO.

SEPTIEMBRE 2014

# Repositorio Institucional UNISON



**"El saber de mis hijos  
hará mi grandeza"**



Excepto si se señala otra cosa, la licencia del ítem se describe como openAccess

## **RESUMEN**

En el presente documento se propone un sistema de información vía Web para la clasificación y recuperación de fuentes de información (conocimiento) con que cuenta el departamento de estudio: texto, imagen, audio o vídeo. Para ello se llevó a cabo una revisión del estado del arte donde se abordan diferentes temas relacionados a la clasificación: técnicas de clasificación, algoritmos de clasificación supervisados, algoritmos de clasificación no supervisados, algoritmos de clasificación semi supervisados y clasificación de las distintas fuentes de conocimiento.

La metodología propuesta, consta de cuatro etapas: Identificación de procesos principales, Describir información de procesos principales, Clasificar la fuente de conocimiento y Recuperación de la fuente de conocimiento.

Clasificar y recuperar fuentes de conocimiento es la parte donde más se hace hincapié. Para clasificar la fuente de conocimiento se trabajan cuatro tipos de extensiones: texto, imagen, vídeo o audio. Con respecto a imagen, vídeo o audio, el usuario experto es quien determina a que proceso principal pertenece cada archivo, mientras que la clasificación de fuentes textuales se realiza de manera automática, mediante el análisis de su contenido.

El sistema desarrollado es una herramienta para apoyar a las organizaciones, ya que de manera sencilla puede ayudar a los miembros de una organización a clasificar la información que ésta maneja y poder recuperarla de manera rápida cuando sea necesario.

## **ABSTRACT**

This paper proposes an information Web system for classification and retrieval of knowledge information sources: text, image, audio and video. A review of art state was conducted and the following reviewed themes were: Classification techniques, supervised classification algorithms, unsupervised classification algorithms, semi-supervised classification algorithms and classification of the different sources of knowledge.

It is proposed a methodology that consists in four stages: Identification of key processes, Description of main information processes, Classification of the knowledge source and Knowledge source recovery.

Classify and retrieve knowledge sources is the most important part in the present work. To classify the source of knowledge four types of different extensions are use: Text, image, video or audio. With regard to image, video or audio, the expert user is who determines the main process of each file, while the classification of textual sources is performed automatically by analyzing its content.

The developed Web system is a useful tool for organizations, because it can help members of an organization to classify information and recover it quickly when need it.

## **DEDICATORIA**

A Dios por darme la salud y ánimos para llevar a cabo mis estudios. A mis padres Miguel Ángel y Alma Rosa, que con su apoyo y dedicación me han conducido por este camino hasta llegar a donde me encuentro.

A mi hermana Myriam Michelle, por estar conmigo siempre que la necesito. Mi novia y gran amiga Krisbel, quien me ha apoyado en esos momentos difíciles en el transcurso de este trabajo. Mi tía María Guadalupe Romero, por ser como una segunda madre para mí.

A toda mi familia y amigos que siempre están ahí cuando los necesito.

## **AGRADECIMIENTOS**

Primero que nada agradezco a Dios por mantenerme activo dentro de mis estudios de maestría así como a mis padres Miguel Ángel y Alma Rosa quienes han sido mis pilares, familiares, especialmente a mi tía Guadalupe, amigos y maestros, en especial a mi director de tesis el Dr. Mario Barceló Valenzuela que con su apoyo y comentarios fue “retroalimentando” mi trabajo de tesis. Al Dr. Gerardo Sánchez Schmitz por estar al pendiente, brindarme su amistad y ayuda durante mi trabajo, así como al Dr. Ricardo Rodríguez Carvajal. Todos ellos fueron quienes me motivaron a pertenecer a un estudio de posgrado por lo cual me encuentro muy agradecido. Por último y no menos importante doy las gracias a CONACYT y al Programa integral de fortalecimiento (PIFI) ya que gracias a su ayuda pude ayudarme a solventar mis gastos durante la estancia de posgrado

## ÍNDICE GENERAL

RESUMEN .....	ii
ABSTRACT .....	iii
DEDICATORIA.....	iv
AGRADECIMIENTOS .....	v
ÍNDICE DE FIGURAS .....	viii
ÍNDICE DE TABLAS.....	ix
1. INTRODUCCIÓN.....	1
1.1. Planteamiento del problema.....	4
1.2. Objetivo general .....	5
1.3. Objetivos específicos.....	5
1.4. Hipótesis.....	5
1.5. Alcances y delimitaciones .....	5
1.6. Justificación .....	5
2. MARCO DE REFERENCIA .....	6
2.1. Conocimiento.....	6
2.1. Gestión de conocimiento .....	7
2.2. Sistemas de gestión de conocimiento (KMS) .....	8
2.2.1. Procesos clave .....	9
2.2.2. Repositorio de conocimiento .....	10
2.3. Fuentes de conocimiento .....	11
2.4. Técnicas de clasificación manual .....	13
2.5. Minería de datos (DM).....	13
2.6. Clasificación .....	15
2.6.1. Algoritmos de clasificación supervisados .....	15
2.6.2. Algoritmos de clasificación no supervisados .....	20
2.6.3. Algoritmos de clasificación semi supervisados.....	23
2.7. Clasificación de la Fuente de Conocimiento.....	26
2.7.1. Clasificación de texto.....	26

2.7.2. Clasificación de Medios de comunicación .....	27
2.7.3. Clasificación multimedia .....	28
2.8. Software .....	30
2.8.1. Ejemplos de software utilizado para la clasificación .....	30
2.9. Estudios previos .....	31
3. METODOLOGÍA.....	36
3.1. Identificación de procesos principales .....	37
3.2. Describir información de procesos principales .....	38
3.3. Clasificar la fuente del conocimiento .....	39
3.4. Recuperar el conocimiento .....	40
4. IMPLEMENTACIÓN .....	43
4.1. Identificación de procesos principales .....	43
4.1.1. Entrevistas cara a cara.....	43
4.1.2. Obtener procesos principales y el medio donde estos se almacenarán.....	44
4.2. Describir información de procesos principales .....	45
4.2.1. Acreditación de Programas de Estudio .....	46
4.2.2. Programación Académica .....	47
4.3. Clasificar la fuente del conocimiento .....	51
4.4. Recuperar el conocimiento .....	60
5. RESULTADOS .....	65
6. CONCLUSIONES.....	67
6.1. Conclusiones.....	67
6.2. Recomendaciones.....	68
6.3. Trabajos futuros.....	68
7. REFERENCIAS .....	70
8. ANEXOS .....	78



## ÍNDICE DE FIGURAS

Figura 2.1 Matriz de conocimiento de Wiig (1993).....	14
Figura 3.1 Metodología propuesta.....	36
Figura 3.2 Identificación de procesos clave.....	37
Figura 3.3 Identificación de tareas en base a procesos clave.....	39
Figura 3.4 Estructura de agrupación de la fuente de conocimiento: imagen/audio/vídeo.....	39
Figura 3.5 Estructura de agrupación de la fuente de conocimiento: texto.....	40
Figura 3.6 Proceso de recuperación de conocimiento propuesto.....	41
Figura 4.1 Valoración final de los procesos clave detectados.....	44
Figura 4.2Caso de uso: Captura y clasificación de información.....	52
Figura 4.3Sistema para clasificar la fuente de conocimiento según su extensión.....	52
Figura 4.4Diagrama de secuencia: Clasificación de fuentes no textuales.....	55
Figura 4.5Diagrama de secuencia: Clasificación de fuentes textuales.....	59
Figura 4.6Caso de uso: Recuperación de la fuente de conocimiento .....	60
Figura 4.7Diagrama de secuencia: Recuperación de la fuente de conocimiento.....	61
Figura 4.8Estructura de tabla clasificación en la base de datos.....	62
Figura 4.9Resultados y opc. de filtrado para la recuperación de conocimiento.....	63
Figura 8.1 Orden de programación académica de la Universidad de Sonora. (CCT, 2013-2014).....	79

## ÍNDICE DE TABLAS

Tabla 1.1 Clasificación de divisiones y departamentos .....	4
Tabla 2.1 Algunos medios disponibles para almacenar/clasificar conocimiento. ....	12
Tabla 2.2 Análisis comparativo de algoritmos para clasificación tipo supervisados. ...	20
Tabla 2.3 Algunos algoritmos de clasificación semi supervisado. ....	26
Tabla 4.1 Procesos principales detectados en el dpto. de estudio. ....	45
Tabla 4.2 Lista de tareas y agentes para el proceso principal: acreditación de programas de estudio. ....	46
Tabla 4.3 Lista de tareas, agentes y recursos para el proceso principal: programación académica. ....	47
Tabla 4.4 Palabras clave por proceso principal y repeticiones. ....	56
Tabla 4.5 Peso $W_i$ en base a repeticiones $P_i$ . ....	57

# 1. INTRODUCCIÓN

Dentro de una organización se cuenta con una serie de factores que están directamente relacionados con su producto o servicio. Dentro de éstos, el conocimiento es vital para facilitar las tareas diarias a los empleados en el área de trabajo, propiciando con ello la disminución de tiempos y mejoramiento de la productividad al realizar sus funciones. Una base de conocimiento puede ayudar a plasmar nuestro pensamiento desde la etapa de captura en base a ciertos criterios de clasificación y posteriormente poder recuperar todo lo almacenado cuando sea solicitado. Por lo anterior, es importante identificar y aprovechar los activos de conocimiento con el fin de obtener ventajas competitivas. Administrar conocimiento es una tarea difícil, por lo que se ha llegado a proponer el uso de sistemas de información que ayuden a gestionar el conocimiento organizacional. En este capítulo se da una breve presentación de la Universidad de Sonora así como las distintas Divisiones por las que está conformada, hasta llegar y profundizar en el Departamento de estudio: Ingeniería Industrial y de Sistemas. Posterior a ello se plantea el problema, el objetivo general así como los específicos, hipótesis, alcances y limitaciones y por último la justificación.

Actualmente, nuestra sociedad enfrenta grandes retos. La globalización, el crecimiento de las empresas, el avance científico y tecnológico han propiciado que las universidades realicen un esfuerzo por estar a la vanguardia de las nuevas tendencias tecnológicas. Asimismo, dentro del contexto económico y político mundial, donde el cambio es el común denominador, se demandan nuevos retos al sistema de educación superior, exigiendo la formación de profesionales capaces de canalizar estos avances, basándose en el dominio y manejo de la información, por lo que se requiere estudiantes con una actitud de mayor inventiva y creatividad para enfrentar estos retos (Universidad de Sonora, 2004).

La Universidad de Sonora es una Institución de Educación Superior autónoma y de servicio público, con personalidad jurídica y capacidad para autogobernarse, elaborar

sus propios estatutos, reglamentos y demás aspectos normativos, así como para adquirir y administrar sus bienes y recursos. Es una institución de educación superior y ejerce la libertad de enseñanza, investigación y difusión de la cultura. La Universidad de Sonora fue fundada el 12 de octubre de 1942 y se le conoce como el más valioso patrimonio cultural y científico del estado de Sonora, por la magnitud y calidad de los recursos humanos y materiales, el número de estudiantes, la presencia de sus egresados, y por ser partícipe de la historia regional (Instituciones Públicas de Educación Superior, 2004).

Su ubicación geográfica y estratégica le permite desempeñar un papel esencial e impulsar nuevas y diversificadas opciones educativas, generar y aplicar nuevos descubrimientos ante los retos que presenta el desarrollo global y el presente milenio. La Universidad de Sonora está dividida en unidades académicas regionales, la unidad regional centro con sede en la ciudad de Hermosillo, la unidad regional sur con sede en Navojoa y la unidad regional norte con sede en la Ciudad de Caborca, Sonora, integrada con los planteles ubicados en las ciudades de Caborca, Santa Ana y Nogales (Congreso del Estado de Sonora).

Con el fin de brindar una educación de calidad, la universidad comprende una gran gama de programas de estudios clasificados en departamentos, y a su vez estos en divisiones. La unidad regional centro cuenta con 20 departamentos en seis divisiones (Tabla 1.1).

La División de Ingeniería de la Unidad Regional Centro se formó en Noviembre de 1991, sus antecedentes datan desde 1958, cuando inician las carreras de Ingeniería Civil, Ingeniería Química, Ingeniería Industrial en 1962 e Ingeniería de Minas en 1970. Actualmente la División de Ingeniería la conforman los departamentos de Ingeniería Civil y Minas, Ingeniería Industrial, Ingeniería Química y Metalurgia, e Investigación en Polímeros y Materiales. Este trabajo se enfoca en el Departamento de Ingeniería Industrial quedata de 1962. Su creación fue parte medular del Plan de Desarrollo

promovido por el Gobierno del Estado en ese año, como una estrategia para absorber la oferta de mano de obra resultante al terminar el programa de internación temporal de trabajadores mexicanos a EEUU (braceros). En ese momento, es cuando nace el proyecto de maquiladoras en la zona fronteriza, la cual necesitaba de profesionales en el área de supervisión y control de esa modalidad productiva. El plan curricular de Ingeniero Industrial fue la mejor opción para responder a esa necesidad.

A partir del semestre 2003-2, el Departamento de Ingeniería Industrial y de Sistemas atiende dos programas de licenciatura y uno de postgrado: Licenciatura en Ingeniería Industrial y de Sistemas, Licenciatura en Ingeniería en Sistemas de Información y Especialidad en Desarrollo Sustentable. Para el año 2007 se anexó al Departamento el postgrado de Ingeniería Industrial compuesto por dos niveles: maestría y especialidad. El programa a nivel maestría se le dio el nombre de Maestría en Ciencias de la Ingeniería: Ingeniería Industrial; mientras que el programa a nivel de Especialidad con el nombre de Especialidad en Ingeniería Industrial. El nombre de la maestría tuvo modificaciones, quedando en el año de 2012 con el nombre de Maestría en Ingeniería: Ingeniería en Sistemas y Tecnología (Valenzuela, 2009).

Dentro del departamento se realizan diversas actividades para mantener un control de las funciones que desempeñan, sin embargo cuando se quiere tener acceso al conocimiento de manera inmediata, es imposible obtenerlo, ya sea por diferentes factores como: no se sabe que es realmente lo que se está buscando, no se encuentra la persona “experta” en dicha temática, confusiones sobre lo que realmente se desea obtener, entre otros.

El Jefe de Departamento y Secretario Académico mencionan que desde los inicios del departamento de Ingeniería Industrial los tiempos de búsqueda de información actuales son muy largos y por otro lado siempre se depende de alguna persona para poder realizar algún proceso dentro del departamento, es decir, si algún empleado encargado de cierta tarea se ausenta, cualquier otro empleado que consulte el conocimiento almacenado del proceso llevado a cabo por su compañero de trabajo,

podría llevar a cabo ésta tarea. Otros problemas podrían ser que existiera duplicidad de funciones, pérdida de tiempo, etc.

<b>UNIVERSIDAD DE SONORA: UNIDAD REGIONAL CENTRO</b>	
<b>DIVISIÓN</b>	<b>DEPARTAMENTO</b>
<b>1. División de Ciencias Exactas y Naturales</b>	<ul style="list-style-type: none"> <li>a) Dpto. de Física.</li> <li>b) Dpto. de Matemáticas.</li> <li>c) Dpto. de Geología.</li> </ul>
<b>2. División de Ingeniería.</b>	<ul style="list-style-type: none"> <li>a) Dpto. de Ingeniería Civil y Minas.</li> <li>b) Dpto. de Ingeniería Química y Metalurgia.</li> <li>c) Dpto. de Ingeniería Industrial.</li> </ul>
<b>3. División de Ciencias Biológicas y de la Salud.</b>	<ul style="list-style-type: none"> <li>a) Dpto. de Agricultura y Ganadería.</li> <li>b) Dpto. de Ciencias Químico-Biológicas.</li> <li>c) Dpto. de Enfermería.</li> </ul>
<b>4. División de Ciencias Sociales.</b>	<ul style="list-style-type: none"> <li>a) Dpto. de Derecho.</li> <li>b) Dpto. de Sociología y Administración Pública.</li> <li>c) Dpto. de Psicología y Ciencias de la Comunicación.</li> <li>d) Dpto. de Trabajo Social.</li> <li>e) Dpto. de Historia y Antropología.</li> </ul>
<b>5. División de Ciencias Económicas y Administrativas.</b>	<ul style="list-style-type: none"> <li>a) Dpto. de Economía.</li> <li>b) Dpto. de Contabilidad.</li> <li>c) Dpto. de Administración.</li> </ul>
<b>6. División de Humanidades y Bellas Artes.</b>	<ul style="list-style-type: none"> <li>a) Dpto. de Letras y Lingüística.</li> <li>b) Dpto. de Bellas Artes.</li> <li>c) Dpto. de Lenguas Extranjeras.</li> </ul>

**Tabla 1.1** Clasificación de divisiones y departamentos. Universidad de Sonora (Ley número 4, 1991)

## 1.1. Planteamiento del problema

El realizar un gran número de actividades relacionadas con cuestiones académicas, de investigación y de vinculación son las tareas gestionadas día a día por el departamento de Ingeniería Industrial. El desempeñar estas tareas implica conocer acerca de los procesos y el conocimiento involucrado en los mismos, lo cual actualmente no se cumple en su totalidad.

## **1.2. Objetivo general**

Diseñar e implementar una estrategia para categorizar el conocimiento organizacional considerando los procesos principales del departamento, así como poder obtenerlo de manera rápida y confiable.

## **1.3. Objetivos específicos**

- Elaborar una propuesta para sistematizar los formatos tipo: texto, audio, video e imagen, al almacenar el conocimiento organizacional.
- Elaborar una propuesta para clasificar de manera automática archivos en texto plano.
- Diseñar una estructura tecnológica para categorizar el conocimiento organizacional.
- Diseñar una estructura tecnológica para recuperar el conocimiento puntual que se requiera.

## **1.4. Hipótesis**

El categorizar y recuperar el conocimiento, agilizará los tiempos en que los miembros de la organización organizan y recuperan información clave.

## **1.5. Alcances y delimitaciones**

Dentro de la División de Ingeniería de la Universidad de Sonora se cuentan con seis licenciaturas: Química, Industrial, Sistemas de Información, Minas, Civil y Mecatrónica. Dentro del presente proyecto se trabajará con la categorización y almacenamiento de procesos principales con relación a la División de Ingeniería pero exclusivamente con el Departamento de Ingeniería Industrial y con al menos una de sus carreras (Ing. Industrial, Ing. en Sistemas de Información e Ing. Mecatrónica).

## **1.6. Justificación**

El presente proyecto de investigación lleva como finalidad plantear, desarrollar e implementar una estrategia para aprovechar el conocimiento organizacional. Para ello, se profundizará en el estudio del conocimiento que se genera en los procesos principales de la organización. Actualmente no se cuenta con ningún mecanismo que gestione toda esa información (conocimiento) en el área de estudio.

## 2. MARCO DE REFERENCIA

Dentro de este apartado se aborda la literatura relacionada al tema de estudio de esta investigación. Se mencionan conceptos básicos para la investigación como son: Conocimiento, gestión de conocimiento (GC) y los sistemas de gestión de conocimiento (KMS). Antes de dar a conocer los diferentes tipos de algoritmos que pueden ayudar a llevar a cabo la clasificación de información se tratará brevemente acerca de la clasificación manual y la minería de datos. Es importante comprender los diferentes algoritmos que existen para clasificar información: no supervisados, semi-supervisados y supervisados, por lo que se mencionan ejemplos y funcionalidades de cada uno de ellos. Finalmente se presentan dos estudios de aplicación relacionados al campo de esta investigación.

### 2.1. Conocimiento

Hoy en día, el conocimiento que se encuentra dentro de las organizaciones, aparte de ser un recurso, se está convirtiendo en un importante activo estratégico para ellas (Sîrbu et al., 2009). La capacidad con la que cuentan estas organizaciones para utilizar eficientemente este conocimiento y distribuirlo a todos sus departamentos, cada vez se está volviendo más importante como determinante para la ventaja competitiva y cada vez este será más crítico para el éxito y la supervivencia de éstas (Ichijo y Kohlbacher, 2006). Por lo general, el conocimiento siempre implica a un usuario o a un agente, quien lo utiliza para realizar las acciones necesarias para alcanzar una meta. El conocimiento puede y debe ser evaluado por las decisiones o las acciones a las cuales conduce (Davenport, 1998).

Según Beckman (1997) "el conocimiento es razonar sobre datos e información para garantizar la operación, resolución de problemas, procesos de toma de decisiones, y el aprendizaje en modo activo". Por otro lado Davenport y Prusak (1998) mencionan que "El conocimiento es una mezcla fluida de la experiencia enmarcada, de valores, de la información del contexto y de la percepción del experto, que proporciona un



marco para evaluar e incorporar nuevas experiencias e información. Se origina y se aplica en la mente de los conocedores”.

La capacidad que poseen las organizaciones para crear ventajas competitivas sostenibles en el tiempo, determina el éxito de estas ante sus competidores. El aprendizaje organizacional es un elemento importante dentro de ellas, ya que es un mecanismo de adquisición de conocimiento, fuente de dichas ventajas e inspiración para concebir mejores soluciones (Pérez y Cortés, 1999).

El conocimiento no es un concepto que difiere radicalmente de lo que es información, sino que la información se convierte en conocimiento una vez que se procesa en la mente del individuo: conocimiento tácito (Polanyi, 1962). De nuevo este conocimiento será convertido en información: conocimiento explícito una vez que se comunica a los demás en forma de texto, salida en pantalla, audios, u otros medios (Nonaka, 1994).

## **2.1. Gestión de conocimiento**

Una de las estrategias para aprovechar el conocimiento que poseen las organizaciones, es utilizando la gestión de conocimiento (GC), la cual para Wiig et al. (1997) es la construcción y aplicación sistemática, explícita y deliberada de conocimiento para maximizar la efectividad organizacional con respecto al conocimiento al usar sus activos de conocimiento. Según Quintas et al. (1997) la GC es el proceso de manejar críticamente el conocimiento para alcanzar las necesidades existentes, para identificar y dar un mejor uso a los activos del conocimiento actual como del adquirido, y para desarrollar con ello nuevas oportunidades. Uriarte (2008) define a la GC como el proceso mediante el cual las organizaciones agregan valor a sus activos utilizando el conocimiento, haciendo evidente que la GC está directamente relacionada con el proceso de identificación, adquisición y mantenimiento de los conocimientos que son esenciales para la organización.

Mientras que las metodologías y herramientas de la GC se encuentran en continua evolución, los conceptos fundamentales de acceso al conocimiento y la recuperación de información se mantienen sin cambios. No importa si lo que se solicita es de una

tabla de registros, de un modelo de elementos finitos, de una red neuronal, de un documento de texto, de imágenes etc., la operación básica requerida para obtener la información y/o conocimiento de cualquiera de las fuentes puede llegar a ser modelada como una función que convierte un conjunto de argumentos de entrada (solicitud de información) en un conjunto de argumentos de salida (información solicitada). Cualquiera de las funciones analizadas puede ser definida como fuente de conocimiento (Bless et al., 2012).

## **2.2. Sistemas de gestión de conocimiento (KMS)**

Alavi et al. (1999) hace mención que lo que es nuevo y emocionante dentro del área de la GC es la posibilidad de poder utilizar tecnologías de la información, por ejemplo: internet, intranets, navegadores, almacén de datos, filtrado de datos y agentes de software. El conocimiento ha sido considerado un factor clave dentro de una organización, por lo que su efectiva administración es fundamental para desarrollar nuevas ventajas competitivas. Para poder llevar a cabo lo anterior, un gran número de organizaciones ha comenzado a participar en iniciativas de GC así como en invertir en sistemas de gestión del conocimiento (KMS) (O'Brien et al. 2006; Hahn et al. 2009). La gran demanda de KMS incluye la capacidad de las organizaciones para poder ser flexible y responder rápidamente a las cambiantes condiciones del mercado, así como su potencial para una mejor toma de decisiones y productividad (Harris, 1996).

La implementación de un sistema de GC generalmente requiere de herramientas que den soporte a este proceso, facilitando el flujo de información entre los diferentes elementos que conforman un grupo de trabajo, por ello, la selección e implementación de estas herramientas tecnológicas se convierte en un factor importante para apoyar la GC (Ramírez y Martín, 2003). Con un KMS funcional, las administraciones pueden maximizar sus recursos relacionados a conocimiento organizacional debido a que serían capaces de utilizar, acumular, compartir y crear conocimiento de gran importancia (Kuo et al. 2011). Un KMS debe de proporcionar la información /conocimientos adecuados para cumplir con claridad las necesidades de tareas de los distintos tipos de usuario. (Kuo et al. 2009; Liu et al. 2008).

### **2.2.1. Procesos clave**

En las metodologías no está claro y explícito el origen de donde se debe de partir para hacer una iniciativa de gestión de conocimiento. No queda explicado si el proceso de la GC se debe aplicar a toda la organización, a una parte de ella o a sus procesos o funciones más relevantes. Al llevar a cabo la GC a partir de los factores claves de éxito, se determinarán los procesos claves de conocimiento y con ello, se detectarán las áreas que son más prioritarias para llevar a cabo el proceso de GC, por ello es importante comprender los aspectos relacionados a un proceso. Los procesos del conocimiento y los procesos de la organización se ligan juntos y por lo tanto se deben mirar juntos. Los procesos de la organización se pueden modelar y analizar extensivamente con métodos bien conocidos y establecidos. La simple cartografía del conocimiento, no satisface los requerimientos de una aproximación comprensiva de los procesos orientados a la GC (Barceló, 2008).

Para la definición de los procesos claves de conocimiento, otros objetivos que deben ser considerados para mejorar el trabajo de conocimiento son (Davenport et al. 1996):

- Hacer explícitos y consistentes los procesos implícitos de conocimiento.
- Agregar conocimiento a los procesos para agregar valor a los procesos del cliente.
- Involucrar al cliente en los procesos para mejorar su satisfacción con el resultado.
- Compartir de manera más efectiva el conocimiento a través de procesos.
- Mejorar la ejecución de las iniciativas y los programas.

En resumen, los procesos clave de conocimiento (Barceló, 2008) es una metodología mediante la cual es posible el gestionar el conocimiento organizacional. Al gestionar el conocimiento en base a factores claves de éxito se detectarán todas aquellas áreas que cuenten con procesos críticos para un correcto desarrollo laboral. Lo anterior genera que aparte de tratar de obtener y almacenar el conocimiento debemos de

comprender a detalle los procesos desarrollados en el departamento que queramos gestionar el conocimiento.

Es común que las organizaciones, estén conformadas por divisiones o departamentos, en los cuales se tienen y se llevan a cabo diferentes procesos, los cuales, aunque no tienen el alcance de un proceso clave, si constituyen el proceso o procesos principales que atañen a esa parte de la organización, por lo tanto estos procesos no cumplen con los objetivos que menciona Davenport (1996).

En general los procesos principales conllevan a las empresas a salir adelante, cuando se trabaja en departamentos o divisiones de una empresa, podemos verlos como un sub-grupo de los procesos clave, ya que éstos no tienen el alcance global de la organización.

### **2.2.2. Repositorio de conocimiento**

Los repositorios de conocimiento o memoria organizacional, pueden llegar a considerarse un producto estratégico de la GC, los cuales son apoyados en tecnologías de información para potencializar su uso. Nevo et al. (2005) hacen mención que un repositorio puede ser descrito como la forma en que las organizaciones almacenan conocimiento del pasado para dar soporte a las actividades del presente. Walsh (1991) se refiere al repositorio como toda aquella información que se tenga a través de la historia de la organización, la cual pueda ser recuperada para soportar las decisiones del presente. Para Stein (1995), el repositorio es la forma mediante la cual el conocimiento del pasado es trasladado a las actividades del presente, obteniendo resultados de niveles de efectividad organizacional más altos. Por último, Ackerman et al. (2000) consideran como un repositorio único a aquel que contiene información de cierto tipo para toda la organización. Al momento de considerar los procesos de captura y recuperación de conocimiento, el repositorio de conocimiento, debe poseer una estructura que facilite estos procesos.

## 2.3. Fuentes de conocimiento

Para gestionar y almacenar el conocimiento dentro de las organizaciones se puede recurrir a distintas fuentes de conocimiento (Tabla 2.1).

Fuente de conocimiento	Descripción	Autor
Documentos (texto, hojas electrónicas, diapositivas, etc.)	La clasificación de texto, implica la asignación de un conjunto de palabras dentro de un documento a una serie de clases pre-definidas de manera automática, mediante la utilización de alguna técnica de aprendizaje máquina. Se necesita etiquetar los datos para poder clasificar.	Dharmadhikari (2012)  Dalal et al. (2011)
Caso de uso	Ayudan a realizar el análisis, diseño, implementación y prueba de procesos. Aparecen con el desarrollo UML para estructurar conceptos y requerimientos de usuario para aplicaciones.	Dolques (2012)  Jacobson (2004)
E-mail	Ocupa un lugar primordial fundamentalmente por su bajo costo, amplia difusión (clientes, proveedores, público, empresas, etc.), facilidad de uso, diversidad de usos. Se utiliza ampliamente en negocios y comercio electrónico. Puede ser clasificado con ayuda de diversos algoritmos.	Park (2010)  Zhu (2008)
	Permiten y favorecen el poder publicar y compartir información, el autoaprendizaje, el trabajo en equipo, la comunicación, la retroalimentación, el acceso a otras fuentes de información que apoyen sobre cierto tema	

Redessociales (Facebook, Twitter, Google+, chat, Linked in, etc)	e incluso facilitan el aprendizaje constructivista y colaborativo así como mantenerse en contacto con los expertos.	Imbernón et al. (2011)
Imágenes	Representación visual, mental o verbal de un objeto real o imaginario. Atraviesan por tres procesos: producción, circulación y recepción. Su significado llega a ser el resultado de múltiples relaciones sociales, morales, religiosas, entre otras.	Soto (2012) Paklone (2011)
Audio	Es cualquier grabación, propia o ajena, realizada de forma magnética o digital presentada en distintos soportes (CD, MP3, entre otros). Puede ser “puro” o contener una serie de sonidos mezclados.	Yu (2009) Lin (2005)
Vídeo	Pueden ser definidos como la luz eléctrica que, al generarse, contiene en sí misma información visual por lo que sería posible afirmar que el simple hecho de que un texto se vea en una pantalla por medio de señales eléctricas que contienen información visual constituye una forma del video. Pueden aplicarse distintos métodos para llevar a cabo su clasificación.	Dimitrova (2011) Pérez (2010)

**Tabla 2.1** Algunos medios disponibles para almacenar/clasificar conocimiento.

En la tabla 2.1 se muestran distintas fuentes para gestionar y almacenar el conocimiento, una breve descripción del mismo y los autores consultados.

## 2.4. Técnicas de clasificación manual

La clasificación es un procedimiento utilizado para identificar categorías y asignar entidades a dichas categorías, considerando sus atributos. Por ejemplo un evento determinado, según el énfasis de acentuación, en GC lo podemos clasificar como: dato (número, letra, signo o símbolo sin significado alguno), información (conjunto de datos significativos) o conocimiento (razonamiento sobre datos e información). Una estrategia de clasificación muy utilizada en GC es la taxonomía. Estas son sistemas de clasificación básicas que permiten describir conceptos y sus dependencias en forma jerárquica. Dentro de las organizaciones se tienen distintos conceptos que pueden ser vistos como “bloques” de construcción de conocimiento y experiencia. Una vez que los conceptos principales han sido identificados y capturados, se pueden organizar dentro de una jerarquía que se conoce como una taxonomía de conocimiento estructurado, éstas permiten que el conocimiento pueda ser representado gráficamente de tal manera que refleje a la organización la información con que se cuenta. Cuanto más alto se coloque el concepto, se considera más general, cuanto más bajo el concepto se coloca, será más específico con respecto a una instancia de categorías de nivel superior. Como se observa en la figura 2.1, la construcción de una taxonomía implica identificar, definir, comparar, y agrupar los elementos de manera manual, por lo que todos los actores organizacionales tienen que ponerse de acuerdo sobre el sistema de clasificación que se utilizará para obtener la clasificación deseada (Dalkir, 2011).

## 2.5. Minería de datos

Existe un gran número de aplicaciones para el aprendizaje máquina, de los cuales, la más significativa es la minería de datos (Vaithyanathan et al. 2013).

La minería de datos involucra el uso de varias herramientas de análisis de datos sofisticadas para poder llegar a descubrir “lo desconocido”, patrones válidos y las relaciones entre grandes conjuntos de datos. Estas herramientas son los métodos de aprendizaje máquina, modelos estadísticos y algoritmos matemáticos. La minería de datos no solo está compuesta por una colección y manejo de datos, sino que también

incluye el análisis y predicción de los mismos. Fayad (1996) menciona que es un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos. Las técnicas de clasificación dentro de la misma son capaces de procesar una amplia variedad de datos.



**Figura 2.1** Taxonomía de conocimiento. Wiig (1993).

Según Chauhan (2010) existen diferentes técnicas para llevar a cabo la DM: pre procesamiento, asociación, clasificación, reconocimiento de patrones y agrupación. De acuerdo con Rajeswari et al. (2011, 2012) las técnicas más populares son la clasificación y la asociación las cuales sirven para predecir el interés de los usuarios y la relación entre todos aquellos elementos de datos que han sido utilizados por los mismos.

En la actualidad se manejan grandes cantidades de información, esto hace que sea inevitable el uso de herramientas robustas para garantizar la fiabilidad del sistema, a pesar de esto, dentro de esta inmensa cantidad de datos todavía existe una enorme masa de información oculta, de gran importancia, a la que no se puede acceder por técnicas clásicas de recuperación de información. El descubrimiento de esta información oculta es posible gracias a la DM, que permite la elaboración de resultados apropiados, pues durante la misma se aplica el algoritmo automático encargado de extraer el conocimiento inherente a los datos (Solarte et al. 2009).



La aplicación de algoritmos de minería ha permitido detectar patrones en los datos y, por ende, crear modelos que sustenten la toma de decisiones, lo que contribuye al mejoramiento de los índices de competitividad o del problema en particular (Roddick y Lees, 2001).

## **2.6. Clasificación**

La clasificación es una técnica de la DM con amplias aplicaciones, la cual, clasifica a los datos de diversos tipos. Se puede utilizar en todos los ámbitos de nuestra vida, la cual ayuda a categorizar cada elemento de datos dentro de un conjunto predefinido de clases o grupos (Vaithyanathan et al. 2013).

Con la llegada de “la era de la información”, el crecimiento de la obtención de recursos infinitos de información presentó un gran desafío a la hora de querer procesar los datos. Por un lado, los recursos de información aumentan de manera muy acelerada y por el otro, las personas que demandan información valiosa también aumenta (Wei et al. 2012).

Hernández (2004) define la clasificación como un proceso de categorización de información. Para establecer las categorías, se requiere elaborar una base de entrenamiento previa, la cual contiene la información de las diferentes clases. Si se cuenta con un objeto y se desea saber si pertenece a una de las clases, se identifican los atributos más significativos y se evalúa qué tan semejantes son a los atributos de la base de entrenamiento.

Dentro de la DM existe una serie de técnicas para llevar a cabo la clasificación de diferentes fuentes de conocimiento, entre ellas se tienen:

### **2.6.1. Algoritmos de clasificación supervisados**

Para Dharmadhikari et al. (2011) este tipo de algoritmos son aquellos que utilizan los datos de entrenamiento, los cuales son aquellos donde cada documento se encuentra etiquetado en cero o más categorías, con el fin de aprender a clasificar en base a un categorizador para los nuevos textos. Aquí, un documento se considera como un ejemplo positivo para todas las categorías en las cuales se encuentre etiquetado, y

como un ejemplo negativo para todas aquellas donde no lo esté. Para Ozgur (2004) la tarea de un algoritmo de entrenamiento para clasificación es encontrar “un vector de peso” el cual categorice de la mejor manera nuevos contenidos textuales.

Algunos de los algoritmos de clasificación supervisada más utilizados son:

**a) Clasificador “el vecino más cercano-k”**

Es conocido por ser un buen algoritmo de reconocimiento de patrones. Teniendo un documento de prueba, este algoritmo encuentra a “k- vecinos” más cercanos entre los documentos de entrenamiento y utiliza las categorías de los k-vecinos con el fin de asignar a una clase (Narayana, 2003). Para Dharmadhikari et al. (2011) este algoritmo también se basa en el supuesto de que las características de los miembros de la misma clase deben ser similares.

Ventajas: Es un método eficaz, simple, no paramétrico y fácil de implementar.

Desventajas: Su principal inconveniente es que se vuelve lento cuando el tamaño del conjunto de entrenamiento crece, por lo que la presencia de características irrelevantes degrada severamente su exactitud.

**b) Clasificador “red neural artificial (ANN)”**

El método consiste en multi-capas de alimentación en la cual existe al menos una capa de entrada, una oculta y una de salida las cuales son basadas en propagación de retroceso estándar para su aprendizaje. El aprendizaje se produce mediante el ajuste de todos aquellos “pesos” en los nodos para minimizar todas aquellas diferencias entre la activación de nodos de salida con su salida normal (Atkinson et al., 1997).

Ventajas: Alto nivel de reconocimiento de patrones.

Desventajas: Alto nivel de aprendizaje y computacional.

**c) Clasificador “paralelepípedo”**

Perumal et al. (2010) mencionan que este clasificador consiste en dos bandas de imagen que son utilizadas para determinar el “área de entrenamiento” de los píxeles basándose en valores máximos y mínimos. Muchos de los píxeles suelen dejarse sin clasificar debido a que puede existir superposición entre cada uno de los píxeles. Otro detalle acerca de este tipo de clasificación es que los valores de los datos de cada píxel son comparados con los límites superior e inferior establecidos.

Ventajas: Dentro de los clasificadores supervisados suele ser uno de los más sencillos.

Desventajas: Aunque se considere de las técnicas de clasificación más exactas, no es el más utilizado, puede dejar fuera de la clasificación gran cantidad de píxeles.

**d) Clasificador “distancia mínima”**

Se basa en la regla de decisión de distancia mínima la cual calcula la distancia espectral entre el vector de medición para el píxel candidato a aquella clase que tenga la distancia mínima con el patrón de comparación. Los descriptores de textura son útiles para distinguir entre diferentes tipos de superficie, aquí los métodos de análisis más comunes se basan en analizar cualquier tono “gris espacial” o en análisis de ocurrencia (Parrinello et al., 2006).

Ventajas: Fácil implementación.

Desventajas: Las clases podrían no llegar a ser bien generadas debido a que éstas surgen en base a los píxeles los cuales en ocasiones no son todos analizados.

**e) Clasificador “NaiveBayes”**

Este método es aquel que hace supuestos de independencia de datos (Narayana, 2003). Por otro lado Khan et al. (2010) mencionan que este clasificador realiza las predicciones mediante la lectura de una serie de atributos tomados como ejemplo para la representación para después aplicar el teorema de Bayes para estimar las

probabilidades. Los supuestos de las características de independencia hacen que toda aquella presencia de características irrelevantes no afecte a la hora de realizar la clasificación.

Ventajas: Requiere una pequeña cantidad de datos de entrenamiento para estimar los parámetros necesarios para la clasificación. Los clasificadores basados en este algoritmo exhiben una alta precisión y velocidad a la hora de aplicarse a bases de datos de gran tamaño.

Desventajas: Funciona bien solo si las características asumidas son independientes por lo que, cuando hay dependencia existirá bajo rendimiento.

#### **f) Clasificador “Árbol de decisión”**

En este método, la clasificación de los documentos de entrenamiento se hace mediante la construcción de definiciones falso/verdadero bien definidos dentro de la estructura del árbol. Las hojas representan la categoría correspondiente a querer clasificar y las ramas las conjunciones de las características que conducen a las distintas categorías (Khan et al. 2010).

Ventajas: Se adapta a cualquier tipo de datos. Suele ser rápido, incluso en grandes cantidades de atributos.

Desventajas: Su mayor riesgo es que solo se adapta a los datos de entrenamiento mediante la ocurrencia de un árbol alternativo.

#### **g) Clasificador “Reglas de decisión”**

En este método se utiliza la inferencia basada en reglas para poder llegar a clasificar los documentos dentro de sus categorías (Khan et al. 2010). Estos clasificadores son útiles para el análisis de datos no estandarizados. Otra característica es que se construye un conjunto de reglas que describen el perfil de cada una de las categorías. Las reglas presentan la estructura: “Si la condición luego la conclusión”, donde las condiciones serán aquellas que correspondan a las características de la categoría y la conclusión mediante el nombre de las categorías o alguna otra regla a ser probada (Dharmadhikari et al. 2011).

Ventajas: Capacidad de un buen análisis de desempeño sistemático.

Desventajas: Su principal inconveniente es la necesidad de la participación de personas expertas para la construcción del conjunto de reglas.

#### **h) Clasificación “Soporte máquina vector (SVM)”**

En este algoritmo se aborda el problema de aprender a discriminar entre los factores positivos y negativos de una clase dada de n-vectores. Se basa en el principio de la teoría de aprendizaje computacional para la minimización de riesgos estructurales.

Este tipo de clasificación necesita de ambos tipos de entrenamiento: positivos y negativos, por lo que comúnmente no se utiliza este tipo de clasificadores. El rendimiento de la clasificación se mantiene sin cambios incluso si documentos que ya no pertenecen al grupo de vectores de soporte son retirados del conjunto de datos de entrenamiento, lo que suele ser una de sus principales ventajas (Dharmadhikari et al. 2011).

Ventajas: Suele ser uno de los métodos de clasificación de texto más eficaces, ya que es capaz de manejar grandes cantidades de características así como una alta habilidad de generalización (Wang et al. 2006).

Desventajas: Suele ser un algoritmo muy complejo, lo que hace que se requiera un alto tiempo y consumo de memoria durante la etapa de formación de las clases.

En la tabla 2.2, se muestra un listado de otros algoritmos que utilizan técnicas clasificaciones supervisadas:

<b>Métodos de clasificación Supervisados</b>				
<b>Algoritmo</b>	<b>Ventajas</b>	<b>Desventajas</b>	<b>Base</b>	<b>Fuente</b>
BR	Se adapta al cálculo de valores de matrices en diagonal.	No realiza correlaciones de etiquetas de forma explícita.	Matriz Gaussiana	Geist (1999)
	Excelente para clasificar. Mejor	Al generalizar resulta en		

Adaboost	precisión.	disminución de rendimiento.	Algebraica	La et al. (2012)
Propagación de retroceso	Aprendizaje en forma iterativa. Mayor capacidad de generalización.	Computacionalmente complejo por el algoritmo que presenta.	Algebraica	Pradeep et al. (2011)
C4.5	Basado en árboles de decisión, mejora la precisión y predicción. Fácil de entender, popular y poderoso.	No toma en cuenta correlaciones entre clases.	Algebraica	Jantan et al. (2010)

**Tabla 2.2** Análisis comparativo de algoritmos para clasificación tipo supervisados.

Como se puede observar en la tabla 2.2 cada algoritmo cuenta con una base distinta para su desarrollo, la cual depende del tipo de información que se pretenda clasificar.

### 2.6.2. Algoritmos de clasificación no supervisados

Dharmadhikari et al. (2011) mencionan que en este tipo de algoritmos no se cuenta con una colección de documentos etiquetados. El objetivo es agrupar los documentos sin contar con conocimiento o intervención adicional de tal manera que los documentos que estén dentro de un grupo cuenten con las mismas similitudes. Este tipo de metodología es clasificada en dos grandes grupos: particionados y jerárquicos. Los algoritmos anidados son aquellos que producen particiones anidadas mediante la división de datos (enfoque divisivo) o por fusión (enfoque de aglomerados) de grupos basados en similitudes. Por otro lado, los algoritmos particionados agrupan los datos dentro de particiones no superpuestas lo que hace que los criterios de agrupamiento sean optimizados.

Algunos de los algoritmos de clasificación no supervisados más utilizados son:

**a) Clasificador “Clúster jerárquico”**

Este tipo de algoritmo produce una jerarquía de clúster en forma de estructura de árbol llamado “dendrograma”. La raíz del árbol consiste en un solo grupo que contiene todas las observaciones, y las hojas corresponden a observaciones individuales (Berkhin, 2002).

Ventajas: Simplicidad y capacidad para capturar la información correctamente.

Desventajas: No se cuenta con grupos discretos y se cuentan con diferentes agrupaciones para diferentes grupos de estudio.

**b) Clasificador “Clúster divisible jerárquico”**

Este tipo de algoritmos comienzan con un grupo a la vez. Durante cada iteración se divide por grupos hasta encontrar el más correspondiente a ciertos criterios como un número de requisición o algún valor “k” alcanzado. En cada paso de esta técnica el grupo con el tamaño más grande es dividido. Como se utiliza similitud de documentos en lugar de distanciamiento de patrones, el grupo es dividido en aquel que contenga menos similitudes con los documentos base. Después se eliminan del grupo todos aquellos documentos con similitudes menos significantes para formar un nuevo grupo más enfocado a lo solicitado (Dharmadhikari et al. 2011).

Ventajas: Se tiene una mayor similitud promedio dentro de los documentos del nuevo clúster.

Desventajas: Mientras comience a incrementar el tamaño de grupos suele tornarse complejo.

**c) Clasificador “Clúster aglomerativo jerárquico”**

En este tipo de algoritmos se comienza con cada documento de forma individual en un grupo aparte, en cada una de las iteraciones realizadas se fusionan todos aquellos grupos que presenten similitudes hasta que se cumple el criterio de búsqueda. Se clasifican principalmente como: enlace único, enlace completo y enlace promedio dependiendo del método definido en los grupos de similitud (Mitchell, 2006).

- Enlace único: Define la similitud de dos grupos  $C_i$  y  $C_j$  como los dos grupos más parecidos.

- Enlace completo: Define la similitud de dos grupos  $C_i$  y  $C_j$  como los dos grupos menos similares.
- Enlace promedio: Define la similitud de dos grupos  $C_i$  y  $C_j$  como el promedio de similitudes entre ambos grupos.

#### **d) Clasificador “Clúster particionado”**

En este tipo de agrupamiento las clases son mutuamente excluyentes. Cada objeto es miembro de donde cuente con mayor similitud, por lo que cada objeto puede ser clasificado dentro de un grupo representativo. Una vez dado el número de clusters  $k$ , una partición inicial será construida; posteriormente la solución de grupo es refinada al mover los documentos de un grupo a otro (Ozgur, 2004).

Ventajas: Solo se requiere de un repaso a través del conjunto de datos por lo que es relativamente rápido.

Desventajas: Las agrupaciones resultantes no son independientes con respecto al orden en que los documentos son procesados.

#### **e) Clasificador “De red organizacional Kohonen”**

Dharmadhikari et al. (2011) mencionan que en este tipo de algoritmo se utiliza un tipo especial de red neuronal la cual se denomina “red de auto organización Kohonen”. La novedad de este método es que detecta de forma automática el número de clases presente en un conjunto de documentos dados para posteriormente colocarlo en su clase correspondiente. Inicialmente utiliza la red de auto organización para explorar la ubicación de posibles grupos dentro del espacio de características. Posterior a ello, se comprueba si alguno de los grupos puede ser fusionado sobre alguna base de agrupación que resulte adecuada, los cuales representan las diferentes clases presentes en el conjunto de documentos dados.

Ventajas: Suelen ser más fácil para los seres humanos ver las relaciones que hay entre grandes cantidades de datos.

Desventajas: Computacionalmente complejo y extenso.



### 2.6.3. Algoritmos de clasificación semi supervisados

Aquí se utilizan datos no etiquetados en conjunto con algunos que si lo están para poder llegar a clasificar nuevos documentos que aún no han sido etiquetados. Dentro de la clasificación de texto la mayor parte de las ocasiones existe limitación de datos que presenten etiquetas, por lo que querer llevarlo a cabo resulta altamente costoso para las organizaciones. En base a lo anterior, los algoritmos semi supervisados son una buena solución a esta situación debido a que su marco de aplicación contempla la agrupación y agrupamiento (Pise, 2008).

Algunos de los algoritmos de clasificación semi supervisados más utilizados son:

#### **a) Clasificador “Co-entrenamiento”**

Este tipo de algoritmo suele ser aplicado cuando un conjunto de datos tiene una división natural de sus características y se requieren dos puntos de vista de los datos (Subramanya, 2008). Se supone que cada conjunto proporciona un conjunto de características las cuales son condicionalmente independientes (Dharmadhikari et al. 2011). Primeramente se trabaja con un clasificador para cada vista usando algunos ejemplos ya etiquetados. Las predicciones de confianza de cada clasificador son utilizadas entonces para llevar a cabo el etiquetado de los datos que no lo están y así, construir datos de entrenamiento adicionales (Nigam et al. 2006).

Ventajas: Simplicidad y aplicabilidad a casi todos los clasificadores existentes.

Desventajas: En algunas aplicaciones de división de características de manera natural podría no estar disponible ya que su rendimiento en este tipo de casos suele ser bajo.

#### **b) Clasificador “Basado en expectativas de maximización (EM)”**

También conocido como algoritmo de maximización expectativa, suele ser utilizado para entrenar clasificadores mediante la estimación de parámetros de un modelo generativo de expectativas de maximización expectativa (Nigam et al. 2006).

Ventajas: Es una representación muy simple con respecto a la complejidad de un texto, otra ventaja que muestra es que con las clases establecidas y las grandes cantidades de datos sin etiquetar, se puede llegar a encontrar un modelo más probable que si solo usamos datos etiquetados.

Desventajas: No garantiza máximos globales dentro del modelo de espacio de probabilidad que abarca.

**c) Clasificador “Basado en gráficos”**

Dharmadhikari et al. (2011) mencionan que en este tipo de algoritmos se trabaja sobre un conjunto de datos de prueba el cual es mostrado al momento de “entrenamiento”, donde se supone que los datos son insertados dentro de un colector de baja dimensión que se expresa por un gráfico. Cada muestra de datos es representada por un vértice dentro de un grafo que es ponderado por los pesos lo que proporciona una medida de similitud entre los vértices.

Ventajas: No solo se proporciona una solución binaria sino que también para multiclases y se puede medir la incertidumbre.

Desventajas: Requieren de cálculos excesivos.

Dado que hay una gran variedad de métodos de clasificación, en la tabla 2.3, se ilustran otros métodos semi-supervidados que con frecuencia se utilizan.

<b>Métodos de clasificación Semi-Supervisados</b>				
<b>Algoritmo</b>	<b>Ventajas</b>	<b>Desventajas</b>	<b>Base</b>	<b>Fuente</b>
Clasificador de multi-etiquetado mediante factorización de matrices no-negativas.	Capaz de adaptarse a entornos semi-supervisados en conjunto con la representación de documentos de bajo rango utilizando la matriz de factorización no-negativa. Se adapta a un entorno con un número pequeño de datos de entrenamiento y un gran número de etiquetas de clase.	Existe una gran influencia generada por dos parámetros sobre el rendimiento: variables latentes y parámetros de sintonización. El valor incorrecto elegido reducirá notablemente el rendimiento.	Análisis de matriz de características	Liu, R. and Yang, L. (2006)
Basado en gráficos SSL con multi-etiquetado.	Efectiva utilización en grandes cantidades de datos no etiquetados así como capacidad de explotar las relaciones entre etiquetas.	La mayor parte de las veces es utilizado para archivos de video. No se adapta bien a textos	Algebraica	Zha,Z., Mie, T. Wang, Z. Hua X. (2008)
Aprendizaje multi-etiquetado mediante el aprendizaje de dependencia de etiquetas.	Mejora de precisión mediante la configuración de SSL	Incremento de tiempo para grandes conjuntos de datos	Análisis de relaciones, método de validación cruzada	Wei, Yang, Zhu y Wang (2011)

Aprendizaje multi-etiqueta mediante aprendizaje silvester (SIAM).	Utilización de grandes cantidades de datos no etiquetados, así como también la capacidad de explotar la relación entre etiquetas. Mejora significativa en la precisión	Podría tornarse lento al utilizar grandes conjuntos de datos	Gráficos para introducción de datos, análisis de nodos	Chen , Song , Wang y Zhang (2008)
Matriz de factorización no negativesemi supervisada.	Uso de NMF en conjunto con SSL permite la extracción de las características más discriminantes que si solo se usara NMF	Alta complejidad computacional	Matriz de datos, factorización conjunta de datos, matriz de características	Lee , Yoo y Choi (2009)

**Tabla 2.3** Algunos algoritmos de clasificación semi supervisado.

Para la elaboración de la tabla 2.3, se hizo una revisión de la literatura para finalmente elegir los distintos tipos de algoritmos que ayuden a gestionar una clasificación semi-supervisada desde las diferentes perspectivas de los autores mencionados.

## 2.7. Clasificación de la Fuente de Conocimiento

A continuación se presentan las fuentes más utilizadas para almacenar conocimiento en un repositorio electrónico organizacional y su relación con las técnicas de clasificación automática.

### 2.7.1. Clasificación de texto (Documentos, hojas electrónicas, diapositivas, casos de uso)

La gran disponibilidad de documentos de texto en forma digital ha proporcionado una amplia variedad de información para las personas que, para su correcta utilización, debe de estar organizada de manera sistemática. Llevar a cabo lo anterior, facilita un mejor almacenamiento de la información, búsqueda y recuperación de contenidos

relevantes de texto según las necesidades de quien lo busque (Dharmadhikari et al. 2011).

La clasificación automática de texto (documentos y casos de uso) es un problema importante desde diferentes ámbitos que tiene muchas aplicaciones como lo son: la indexación automática de artículos científicos, el filtrado de correo no deseado, la identificación de género de algún documento en particular, atribuciones de autoría, ensayos de clasificación automatizada, codificación de una encuesta, clasificación de noticias, entre otras (Fabrizio et al. 2002).

Las técnicas de aprendizaje automático tienen la gran ventaja de que se comprenden de una mejor manera desde el punto de vista teórico, permitiendo así que los usuarios confíen en el rendimiento y configuración de ciertos parámetros para llevarla a cabo. Dichas técnicas se dividen en tres grandes grupos los cuales son: aprendizaje supervisado, no supervisado y semi supervisado (Wang et al. 2006).

Wei (2012) menciona que la clasificación de texto es el enfoque de todas aquellas áreas dedicadas al procesamiento de información, donde un gran conjunto de archivos preestablecidos se asignan a cierta clase de atributo predefinido, en la cual, su tarea será el dividir los archivos de hipertexto en varias categorías en función de los contenidos predefinidos por lo que en este caso, lo que se estudia es el contenido de un archivo de texto. En el sistema de clasificación de texto, el clasificador es un factor clave, ya que la calidad del rendimiento del clasificador está directamente relacionada con el efecto del clasificador y la eficacia del texto

Algunos de los algoritmos más utilizados para este tipo de clasificación son los siguientes: vecino k más cercano, NaiveBayes, Árbol de decisión, Reglas de decisión, SVN, ANN, Clúster particional, Basado en gráficos::Dharmadhikari et al. (2011), Khan (2010), Wang et al. (2006), Narayana (2003).

### **2.7.2. Clasificación de Medios de comunicación (E-mail, redes sociales, chat)**

Para Taiwo et al. (2010) el uso de correo electrónico se ha convertido en uno de los más rápidas y eficaces formas de comunicación lo que impacta de manera

considerable en la transferencia de conocimiento dentro de un grupo determinado de personas. Sin embargo, el aumento de usuarios que lo utilizan y los grandes volúmenes de mensajes que se tienen se puede llegar a obtener correo sin estructurar, congestión, sobrecarga, pérdida de prioridad, entre otros. Lo anterior, resulta en un impacto directo por querer contar con herramientas que ayuden a gestionar y clasificar el contenido de este medio de comunicación.

El envío masivo de correo electrónico no deseado, también conocido como Spam, ha generado la necesidad de contar con filtros antispam fiables. Por ello, Awad et al. (2011) mencionan que el uso de un clasificador basado en técnicas de aprendizaje automático podría filtrar de manera automática el correo spam o incluso cualquier tipo de información que se desee.

Las redes sociales (Facebook, Twitter, Google+, Linked in, etc.) en conjunto, hacen que dentro de una organización el aprendizaje sea más interactivo y significativo ya que facilitan una mayor interacción con el personal que labora y su comunicación en sí. En base a ello es importante poder analizar a detalle y de ser posible llegar a clasificar la información que se maneja para así facilitar y mejorar el desempeño laboral. Hacer un uso apropiado de lo anterior generará un ambiente organizacional más dinámico (Imbernón et al., 2011).

Algunos de los algoritmos más utilizados para este tipo de clasificación son los siguientes: Basado en gráficos, NaiveBayes, Vecino k más cercano., ANN, SVM, matriz de factorización no negativa, Clúster particional, Propagación de retroceso: Awad et al. (2011), Pradeep et al. (2011), Park (2010), Taiwo et al. (2010).

### **2.7.3. Clasificación multimedia (Imágenes, audio, vídeo)**

Para Palaniswani et al. (2006) la clasificación de imágenes es un proceso para “ensamblar” grupos de píxeles idénticos que pueden llegar a encontrarse en datos obtenidos mediante la detección de características dentro de una imagen en base a clases preestablecidas, las cuales corresponden a categorías solicitadas por cierto tipo de usuario en base a comparación de píxeles entre sí y los de identidad ya conocida

(clases). La identificación de imágenes ha llegado a impactar de manera considerable al brindar significados e información importante a las organizaciones ya que es un medio de captura muy común que se utiliza hoy en día. Este tipo de clasificación se lleva a cabo mediante la identificación de píxeles y la comparación con otros ya asignados a distintos tipos de clases con el fin de clasificar todos aquellos píxeles que tengan una identidad desconocida. Un ejemplo de píxel de identidad conocida es todo aquel que se encuentre localizado dentro de las “áreas de entrenamiento” de la clase. Las características estadísticas de las clases estimadas a partir de píxeles de entrenamiento, dependen del algoritmo que vaya a ser utilizado para su clasificación (Parrinello et al., 2006). Solo un pequeño número de algoritmos de clasificación de imágenes han demostrado contar con una buena precisión con respecto a la clasificación y detección de imágenes. Hoy en día es necesario un clasificador que sea eficaz a la hora de querer clasificar los contenidos existentes de cualquier tipo de imagen (Perumal et al., 2010).

Para recuperar la información (conocimiento) de usuarios en grandes contenidos de datos multimedia, la clasificación automática de audio-video juega un papel importante para llevarlo a cabo. Ambos pueden ser clasificados dentro de una base de datos bien organizada para que posteriormente se tenga una recuperación rápida y precisa de los mismos. Para clasificar los datos de audio es necesario utilizar coeficientes de onda de sonidos representados mediante imágenes y para videos las características de histogramas de color son extraídas de las imágenes de los video clips las cuales son utilizadas para obtener las características visuales (Subashini et al. 2012).

Para poder tener una recuperación automática de video antes que nada se requiere que este haya sido clasificado. Brezeale (2008) ha estudiado y realizado encuestas con respecto a las características de clasificación automáticas como el texto, variedad visual, color de histograma, movimientos y texturas, entre otros.

Para la clasificación de audio un gran esfuerzo ha sido dedicado a investigar las características relevantes de los diversos tipos de sonido con que se cuenta, así como

los algoritmos que pueden ser aplicados para su clasificación, lo cual se realiza mediante el estudio de las ondas que generan en el dominio del tiempo (Yu, 2009).

Algunos de los algoritmos más utilizados para este tipo de clasificación son los siguientes: ANN, Parallepiped, Minimumdistance: Atkinson et al. (1997), Perumal et al. (2010), Parrinello et al. (2006), SVM, Audio segmentation: Subashini et al. (2012), Yu (2009), Dhanalakshmi et al. (2008).

## **2.8. Software**

Software es el conjunto de instrucciones y demás datos necesarios para hacer funcionar una computadora. Todo lo que podemos hacer con computadoras requiere de software específico para ello. Hay software para jugar, dibujar, sacar fotos, escuchar y grabar música, comunicarnos con otros, escribir, hacer cálculos, software para que funcione otro software, y software para construir nuevo software (Salvador,2006). El conocimiento interno puede residir dentro de las cabezas de la gente; embebidos en comportamientos, procedimientos, software y equipo; y puede ser registrado en varios documentos; o almacenado en bases de datos y depósitos en línea. Por ende, la adquisición, almacenamiento y recuperación de conocimiento puede “descansar” en agentes de software en lugar de recaer sobre las personas. Bajo esta perspectiva, Davies et al. (2003) proponen una nueva arquitectura para la GC que aborda todos los aspectos del ciclo de vida de la GC, desde la identificación y almacenamiento, hasta la transferencia y aplicación, todo mediante la ayuda de un software.

### **2.8.1. Ejemplos de software utilizado para la clasificación**

- **Weka:** es una herramienta que permite la experimentación de análisis de datos mediante la aplicación, análisis y evaluación de las técnicas más relevantes de análisis de datos, principalmente las provenientes del aprendizaje automático, sobre cualquier conjunto de datos del usuario. Para agrupar información se basa en una serie de algoritmos con el fin de poder llegar a obtener diferentes clasificaciones según corresponda la información (García et al. 2010)



• **FaCT (FastClassification of Terminologies) (FACT, WWW)**: desarrollado por Ian Horrocks (Horrocks, 1998), permite reglas transitivas, inversas, restricciones cualificadas, jerarquías entre otros. Es lo suficientemente expresivo para soportar y razonar sobre cualquier base de conocimiento. Escrito en CommonLisp, fácilmente ejecutable por cualquier programa Lisp de forma local, tiene escrita una versión servidor FaCT para ser usada vía interfaz CORBA sobre cualquier sistema con acceso a la red. Es el razonador empleado por defecto por el editor de ontologías OilEd para clasificar los conceptos en una jerarquía según las descripciones que tengan.

## 2.9. Estudios previos

Debido a que durante la revisión de literatura de la presente no fue posible encontrar estudios previos relacionados directamente con esta investigación, se hará mención a estudios con un enfoque parecido a esta investigación.

El estudio “**El impacto de las tecnologías de información y los sistemas de memoria en la compartición y aplicación de conocimiento organizacional**” realizado por Choi et al. (2010) hace mención acerca del impacto que tienen las tecnologías de información dentro de la gestión del conocimiento, así como su aplicación al trabajar en equipo. En él se presenta una metodología en la cual se analiza el impacto generado dentro de las organizaciones cuando se trabaja en equipo y se codifica, almacena y posteriormente se recupera el conocimiento que haya sido detectado. La investigación anterior fue aplicada dentro de dos empresas situadas en Korea del sur, la cuales, según Choi tienen establecidas buenas prácticas de trabajo así como de administración del conocimiento.

La investigación estuvo compuesta por un total de ocho hipótesis:

**Hipótesis 1.** El uso de tecnologías de información (TI) para brindar soporte a las prácticas de GC dará lugar a un mejor desempeño en los equipos de trabajo.

**Hipótesis 2.** El desarrollo de una memoria organizacional permitirá una efectiva compartición de conocimiento entre equipos de trabajo organizacional.

**Hipótesis 3.** El desarrollo de una memoria organizacional permitirá una efectiva aplicación de conocimiento entre equipos de trabajo organizacional.

**Hipótesis 4.** El uso de TI para gestionar prácticas de conocimiento permitirá una efectiva compartición de conocimiento entre los miembros del equipo de trabajo.

**Hipótesis 5.** El uso de TI para gestionar prácticas de conocimiento permitirá una efectiva aplicación del conocimiento que existe entre los miembros de la organización.

**Hipótesis 6.** El compartir el conocimiento, conllevará a la organización a un alto rendimiento en equipo.

**Hipótesis 7.** Aplicar el conocimiento conllevará a un alto rendimiento en equipo.

**Hipótesis 8.** Compartir el conocimiento permitirá tener un alto nivel con respecto al conocimiento aplicado en forma grupal dentro de una organización.

Para realizar el estudio se llevó a cabo una encuesta entre los miembros de cada organización la cual se aplicó de forma separada a los líderes de proyectos y sus empleados. Como resultados se obtuvo que existe una verdadera importancia al querer gestionar conocimiento organizacional mediante la ayuda de tecnologías de la información. No solamente basta con tener una memoria organizacional sino que el ayudar a gestionar su información mediante la ayuda de sistemas tecnológicos será de vital ayuda para ayudar a aplicar, compartir y diseminar el conocimiento de una manera más fácil. Por último cabe mencionar que en este estudio no se aborda nada relacionado con respecto a cómo clasificar mediante la ayuda de sistemas de TI la información que esté contenida dentro de la memoria organizacional lo cual es uno de los enfoques principales de este trabajo.

El siguiente estudio relacionado al **“Diseño de un sistema de memoria organizacional para la mejora de los procesos de servicio de un departamento de tecnología”** es un proyecto en el que se diseña e implementa un sistema de memoria organizacional a través de una metodología que involucra tanto el aspecto organizacional a través de una metodología como el aspecto tecnológico de una organización enfocada a los servicios de tecnología, en la ciudad de Hermosillo,

Sonora(López, 2012). López (2012) hace mención que desde el punto de vista tecnológico, es importante contar con una plataforma tecnológica que facilite la retención, recuperación, compartición y distribución del conocimiento generado dentro de las áreas laborales. Para esto, la utilización de tecnologías de información facilitará a la organización, el poder almacenar y recuperar el conocimiento generado por los técnicos que laboran dentro del departamento de estudio al atender las órdenes de servicio, así como la interacción con los usuarios y dispositivos. Considerando lo anterior, se concibe a este modelo de memoria organizacional como un enfoque integral de trabajo, orientado hacia el aprovechamiento del conocimiento existente dentro del departamento de tecnología, en particular en las áreas de soporte técnico, redes y telecomunicaciones. Por tanto, constituye un apoyo fundamental para mejorar la calidad del servicio prestado por el departamento, gestionando adecuadamente el conocimiento presente y futuro en las mencionadas áreas. Los elementos que se tomaron en cuenta para poder llevar a cabo el estudio fueron los siguientes:

**Problema:** Se entiende como problema cuando algún usuario presenta un error en el funcionamiento del dispositivo informático.

**Usuario:** Personal de la institución que hace uso de la infraestructura informática gestionada por el departamento de tecnología.

**Atención a usuarios:** Es el área del departamento de tecnología encargada de atender todos los problemas presentados a los usuarios de la infraestructura informática.

**Caso:** Son generados después de que un usuario presenta un problema y este se comunica con el área de atención a usuarios del departamento de tecnología.

**Casos anteriores relacionados:** Son casos atendidos con anterioridad que presentaron una problemática similar a algún caso que se encuentra activo, estos casos pueden servir de gran ayuda para que los técnicos resuelvan la nueva problemática.

**Notificación:** Son mensajes en forma de correo electrónico que se envían al usuario para mantenerlo al tanto del progreso de atención del caso generado.

**Técnico:** Recurso humano del departamento de tecnología que se encarga de dar seguimiento a los casos de servicio relacionados con servicio técnico preventivo y correctivo a equipos de cómputo y servicios de red.

**Dispositivo:** Equipos Informáticos como computadoras de escritorio, laptops, tablets pc, dispositivos móviles, entre otros.

**Procedimiento:** Se dividen en tres tipos, procedimientos en observación, procedimientos aprobados y procedimientos nuevos. Los procedimientos en observación son una secuencia de pasos propuestos por algún técnico que pueden dar solución a algún problema, pero que aún no han sido aprobados oficialmente por el departamento de tecnología aunque de igual manera son documentados. Los procedimientos aprobados son los que ya han sido revisados por el departamento y han sido aprobados como procedimientos oficiales que cumplen con las normas de seguridad establecidas. Finalmente, los procedimientos nuevos son procedimientos que no se encuentran documentados, generados a través de la atención de un caso por un técnico del departamento, basados en información obtenida de las fuentes externas y la interacción con las herramientas de comunicación en tiempo real, agregados a la memoria organizacional y pasando a ser procedimientos en observación.

**Bitácora de usuario y dispositivos:** Información generada a partir de los seguimientos de casos por parte de los técnicos donde se almacena información como problema visto, procedimientos utilizados y sugerencias.

**Conocimiento de fuentes externas:** Procedimientos y técnicas compartidas y sugeridas por los mismos proveedores de hardware y software, personas que laboran en instituciones externas con casos similares, capacitaciones impartidas y consultores externos.

**Herramientas de comunicación en tiempo real:** Son herramientas basadas en TI de apoyo para la comunicación en tiempo real de un técnico con una o varias personas expertas en algún tema, con la finalidad de que estas puedan asesorar eficientemente al técnico para que este pueda solucionar alguna problemática presentada.

**Experto:** Individuo perteneciente al departamento de tecnología que tiene suficientes conocimientos e información para poder asesorar a algún técnico en la resolución de alguna problemática presentada.

**Solución al problema:** Conjunto de información generada a través de la resolución de alguna problemática presentada como usuario y dispositivos atendidos, tipo de problemática presentada y procedimientos utilizados, mismos que serán agregados al caso atendido y a la bitácora del usuario y dispositivo atendidos.

Al igual que en el estudio previo, no se hace mención acerca de técnicas que ayuden a categorizar el conocimiento que se tenga dentro de la memoria organizacional. Por lo anterior el enfoque de este trabajo se basará en la investigación para poder llegar a clasificar los distintos tipos de fuentes de conocimiento existentes dentro de una base de conocimiento organizacional.

### 3. METODOLOGÍA

En el capítulo anterior se describieron diferentes tipos de algoritmos para llevar a cabo la clasificación de información/conocimiento. Un punto importante para poder clasificar la información organizacional es identificar los procesos principales que se tienen, con el fin de analizar a detalle las tareas o procedimientos necesarios que conllevan, de tal manera que se pueda comprender lo relacionado a cada uno de ellos. Además de comprender los procesos principales y sus respectivas tareas, es importante saber si éstos se encuentran contenidos en algún medio digital, de tal forma que al querer hacer uso de ellos sea fácil encontrarlos o mejor aún, no generar dependencia hacia las personas que cuentan con el conocimiento organizacional. Dentro de cualquier empresa, puede llegar a existir un gran número de procesos principales y no necesariamente las tareas que conlleva cada uno de ellos es “propia” del mismo, por lo anterior, mismos tipos de tareas/procedimientos pueden pertenecer a distintos procesos. Recuperar la información de forma organizada no es algo fácil para las organizaciones debido a diversos factores: mala organización, no se sabe quién la posee, entre otros, lo que provoca grandes vacíos de conocimiento para cuando surge la necesidad de obtenerlo de manera inmediata. En base a la problemática identificada, se propone la metodología de enfoque cualitativo mostrada en la figura 3.1, la cual está conformada por cuatro etapas.

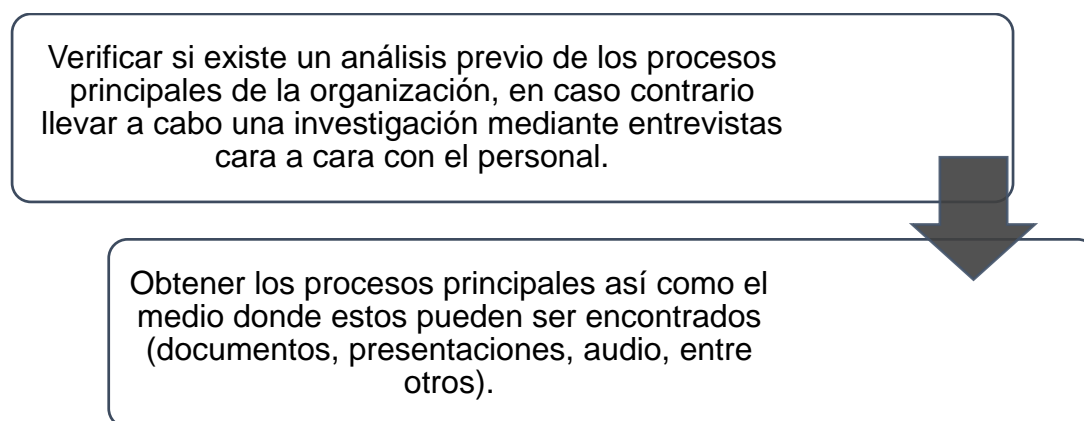


**Figura 3.1** Metodología propuesta

A continuación se describe cada uno de los pasos de la metodología propuesta:

### 3.1. Identificación de procesos principales

Es necesario identificar cuáles son los procesos que se llevan a cabo y para ello, es pertinente realizar una investigación acerca de los procesos o funciones que se llevan a cabo dentro del departamento de una organización, detectando las actividades o tareas asociadas a cada una de ellas. La secuencia para la identificación de procesos principales es mostrada en la figura 3.2.



*Figura 3.2 Identificación de procesos principales*

•**Análisis previo de procesos principales y/o entrevistas cara a cara:** Para identificar cuáles son los procesos realizados por un departamento de cualquier organización, será necesario establecer conversaciones cara a cara con los empleados de la organización y así, determinar si existe un análisis previo de los mismos o no. Cuando en el área de estudio se tenga conocimiento sobre sus procesos clave, se hará uso de ellos para continuar con las siguientes etapas de la metodología propuesta. Para los casos donde no se tenga conocimiento alguno sobre los procesos clave, se deberá de realizar una serie de entrevistas con el personal que participa en los procesos organizacionales. El punto central de las entrevistas cara a cara será el poder llegar a reunir información directamente de las personas encargadas de gestionar información vital de la empresa. Con éste instrumento, se espera poder comprender cuales son los procesos principales que se

tienen en su espacio de trabajo. La entrevista estará basada en un formato en el cual se pueda detectar quién fue la persona que haya respondido a ella, el puesto de trabajo y las actividades más importantes que realiza en relación a su puesto. Para evitar perder detalles de la misma, es recomendable grabar la entrevista mediante un dispositivo digital, con el fin de analizarse a detalle posteriormente.

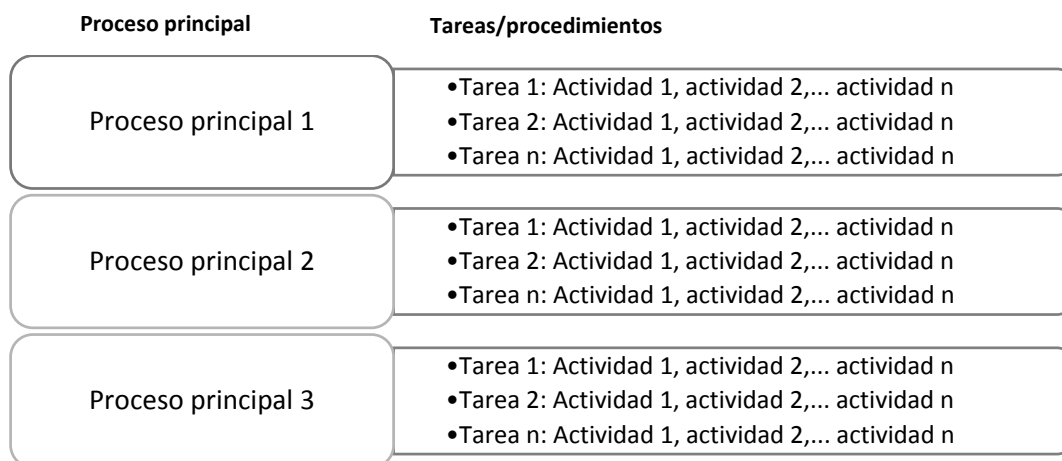
•**Obtener procesos principales y el medio donde estos se almacenan:** En base a la información obtenida del punto anterior, ya sea que, haya existido información previa o toda aquella que resulte ser nueva, será posible poder identificar las actividades/procesos que son de suma importancia para el puesto del departamento. Adicionalmente a ello, se pretende saber si existe la manera de poder encontrar información relacionada a los procesos en algún medio digital, de manera que en la entrevista realizada en esta etapa será necesario el cuestionarlo a la persona encargada.

### **3.2. Describir información de procesos principales**

Cada proceso principal puede conllevar a una serie de tareas/procedimientos distintos lo cuales será necesario describir. Un proceso es una secuencia de actividades para desarrollar algo, en general están constituidos por un conjunto de tareas y actividades, por lo que el punto central de esta etapa es el comprender cómo se realizan las tareas/procedimientos que se llevan a cabo en cada proceso, esto se puede fortalecer al analizar a detalle la entrevista realizada al agente de conocimiento. Un punto a considerar dentro de la descripción de información que conlleva cada proceso es el conocer si existe algún medio digital donde se encuentre contenido, con el fin de poder acceder a ellos cuando sea requerido. En base a lo anterior, será necesario crear agrupaciones con respecto al medio digital donde se encuentra almacenado cada proceso y las tareas que éstos conllevan. Los medios digitales más utilizados son: presentaciones, documentos de texto, vídeos, audios, entre otros. En la figura 3.3 se puede observar que puede existir un cierto número de procesos, los cuales están constituidos por un número de tareas distintas, y que para llevar a cabo estas tareas se necesita cumplir con una serie de actividades diferentes. El analizar los audios tomados de las entrevistas a los agentes de



conocimiento, así como el análisis a estudios previos de procesos principales (en caso de existir), ayudará a comprender a fondo la descripción de los procesos identificados en la etapa previa.

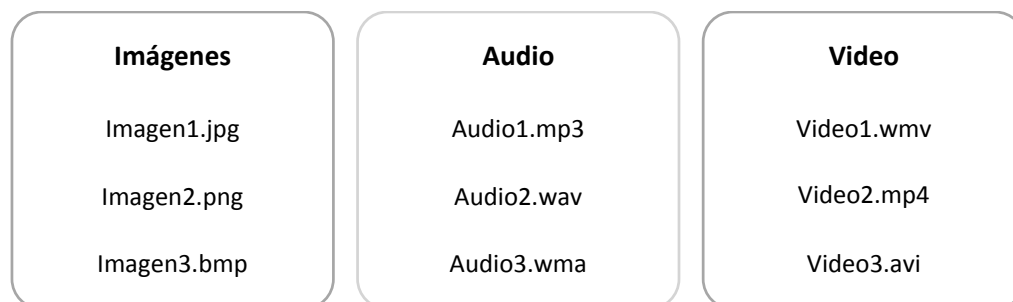


**Figura 3.3** Identificación de tareas/procedimientos en base a procesos principales

### 3.3. Clasificar la fuente del conocimiento

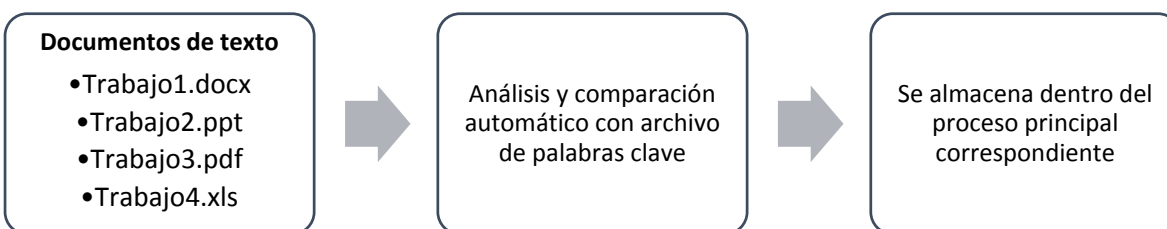
Para clasificar la fuente de conocimiento, se deberá de implementar un algoritmo el cual ayude a poder identificar los diferentes tipos de archivos en donde se encuentre contenida la información de los procesos principales. La información que conlleva cada uno de los procesos, puede estar presente en diferentes medios digitales: texto, imágenes, audio o vídeo. El algoritmo se basará en la identificación del tipo de extensión que contengan los diferentes medios digitales.

**Clasificación de imagen, audio o vídeo:**Adicional a la clasificación por tipo de extensión, para archivos tipo imagen, audio y vídeo, el usuario experto, el cual es una persona con conocimientos sobre el tema, capaz de poder seleccionar el proceso principal perteneciente, podrá seleccionar el proceso principal al cual corresponda. En la figura 3.4 se muestra una posible estructura de cómo se pretende clasificar la información en base al tipo de extensión del archivo.



**Figura 3.4** Estructura de agrupación de la fuente de conocimiento: imagen/audio/vídeo

**Clasificación defuentes textuales:**Para el caso de fuentes textuales, la asignación de proceso principal será llevada a cabo de forma automática. Para asignar el proceso al que pertenece cada archivo, el contenido de éste deberá ser comparado con una serie de palabras clave establecidas. El proceso de clasificación es mostrado en la figura 3.5.



**Figura 3.5** Estructura de agrupación de la fuente de conocimiento: texto

Como se observa, pueden llegar a existir archivos con extensiones distintas (.ppt, .docx) pero aun así, llegar a estar dentro de la misma categoría (texto).

### 3.4. Recuperar el conocimiento

Será necesario el desarrollo de un sistema, en el cual estarán contenidos los diferentes tipos de archivos que ayudan a la ejecución de los procesos principales (etapa 2). El sistema deberá de tener los archivos agrupados según su tipo de formato y, para ello, se basará en el algoritmo de agrupamiento propuesto en la

etapa anterior. Por otro lado, el sistema contará con una interfaz gráfica, la cual permitirá al usuario común (persona que sólo consulta información y no necesita conocer acerca de los procesos principales) visualizar lo que éste desee encontrar. La forma en como el sistema de información llevará a cabo la recuperación de información basada en las peticiones del usuario es mostrada en la figura 3.6.



**Figura 3.6** Proceso de recuperación de conocimiento propuesto

**Paso 1: Petición por parte del usuario.** En este paso se debe hacer una petición de la información que desee encontrar. Dependiendo de la necesidad o requerimientos de información que necesite el usuario común, será la palabra o palabras clave que se introduzcan en el buscador del sistema de información.

**Paso 2: Base de datos.** Una de las características principales de la base de datos con la que trabajará el sistema, es que ésta debe ser diseñada con el fin de poder almacenar los diferentes tipos de archivos contenidos en los medios digitales mostrados en las etapas previas. El sistema de información a utilizar, comparará las palabras introducidas por el usuario común con las diferentes tareas/procedimientos contenidas en la base de datos.

**Paso 3: Encontrar coincidencias con respecto a la petición.** Como se revisó en el marco de referencia, existen distintos tipos de algoritmos que permiten llevar a cabo el análisis de información contenida en medios digitales. El sistema de información descifrará el contenido de las fuentes “textuales” mediante análisis automático de su contenido, con el fin de poder encontrar patrones que serán comparados con la

búsqueda del usuario común, y así, regresar solamente resultados relacionados con lo solicitado. Para el análisis de audio, vídeo e imágenes el patrón de comparación se llevará a cabo analizando el nombre que contenga el archivo en que están contenidos.

**Paso 4: Mostrar resultados en pantalla.** En base a las coincidencias encontradas en el paso previo, el sistema de información mostrará en pantalla los resultados de búsqueda al usuario común, de tal manera que si se encuentra más de uno y en distintos medios digitales, éste podrá seleccionar entre un conjunto de recursos según lo desee.

**Paso 5: Reutilización y aprovechamiento.** En base a los resultados obtenidos en el paso 4, cuando el usuario común recupere el conocimiento solicitado, éste podrá retroalimentar el sistema, la retroalimentación podrá llevarse a cabo al actualizar tareas/procedimientos que se necesitan llevar a cabo en los diferentes procesos principales. Por otro lado, el conocimiento que haya sido de utilidad podrá ser reutilizado cuando le sea necesario. Al retroalimentar el sistema, será necesario actualizar la base de datos, de tal forma que cuando algún otro usuario realice una petición de búsqueda, los resultados obtenidos sean los más indicados.

## **4. IMPLEMENTACIÓN**

El estudio se llevó a cabo en el Departamento de Ingeniería Industrial de la Universidad de Sonora Unidad Regional Centro. Para su elaboración fue necesario llevar a cabo pláticas formales con personal del departamento y, posteriormente analizar y utilizar herramientas tecnológicas para el desarrollo de la metodología propuesta. A continuación se presenta el detalle de las etapas propuestas.

### **4.1. Identificación de procesos principales**

Estos procesos son aquellos que son indispensables para el departamento de estudio, en este caso, no hubo enfoque en procesos que resultaran ser irrelevantes. Para llegar a la determinación de procesos principales fue necesario llevar a cabo una investigación previa con el personal involucrado en los mismos.

#### **4.1.1. Entrevistas cara a cara**

Para la elaboración de este apartado fue necesario realizar una serie de entrevistas dirigidas con el personal: Jefe de Departamento, Secretario Académico y Coordinadores de Programa. Las preguntas de la entrevista son mostradas en el anexo 1.

Al realizar la entrevista al Dr. Gerardo Sánchez (secretario académico del Departamento de Ing. Industrial), comentó que ya existe un análisis previo de los procesos principales del área de estudio. Es por ello que para esta sección se ha retomado la investigación realizada por Sánchez (2012), donde se determinan los procesos principales dentro del departamento de Ingeniería Industrial (figura 4.1).

<b>Valoración Final de los Procesos del Proyecto</b>				
<b>Estrategias de GC</b>				
<b>Procesos</b>				
Descripción	Identificar	Recuperar	Compartir	Utilizar
PAD: Programación académica	12.1951	15.8536	13.8293	30.8293
PAD: Administrativos de oficina	5.26828	6.58536	12.8537	13.7317
PES: Acreditación del programa de estudios	13.2683	13.561	7.41463	14.0732
PFO: Titulación	5.60974	7.51218	13.6829	18.2439
PAD: Seguimiento de competencias	5.65852	7.60974	6.68292	10.7805
PEX: Participación en el Bufete de Desarrollo	4.14633	7.95119	4.48781	8.7561
PFO: Orientación general de servicios de la institución	2.43902	5.41463	6.68294	10.1707
PAD: Relaciones con empleadores	3.41462	3.95121	3.21952	7.12195
PEX: Seguimiento de egresados	2.19512	2.68292	3.34146	4.26829

**Figura 4.1** Valoración final de los procesos principales detectados. Sánchez (2012)

En la figura 4.1, se puede observar que la programación académica es el proceso más crítico así como sus respectivas puntuaciones para la identificación, recuperación, compartición y utilización del conocimiento, siendo esta última la que tiene una puntuación más relevante con respecto a las demás.

Para fortalecer la determinación de los procesos principales, se llevó a cabo una entrevista con el Jefe de Departamento y Coordinadores de Programa. Posterior a las entrevistas fue necesario analizar los archivos de audio donde se almacenaron cada una de las entrevistas, y así, obtener cada detalle acerca de los procesos principales y la manera en cómo se lleva a cabo cada uno de ellos.

#### **4.1.2. Obtener procesos principales y el medio donde estos se almacenarán**

A continuación en la tabla 4.1 se presentan a detalle las funciones/tareas que conlleva realizar cada uno de los procesos más importantes en el departamento de Ingeniería Industrial.

Nombre de proceso	Medio donde se encuentra información relacionada
Organización de cursos de titulación	Documento oficial
Seguimiento de competencias	Documento oficial
Administrativos de oficinas	Documento oficial
Acreditación del programa de estudios	Documento oficial, PDF, Documentos en texto digital
Relaciones con empleadores	Documento oficial, PDF, Documentos en texto digital
Seguimiento de egresados	PDF, Documentos en texto digital, Hojas de Cálculo
Programación académica	Documento oficial, PDF, Documentos en texto digital
Orientación general de servicios de la institución	Documento oficial, PDF, Documentos en texto digital
Participación en el bufete del desarrollo	Documento oficial, PDF, Documentos en texto digital

*Tabla 4.1 Procesos principales detectados en el departamento de estudio.*

## 4.2. Describir información de procesos principales

Con la información obtenida de la tabla 4.1, así como las entrevistas realizadas en la sección previa, se determinó que los dos procesos más importantes son: Acreditación de programas de estudio y Programación académica. **En el sistema Web utilizado en este trabajo, se encuentran listados todos los procesos principales del Departamento de estudio, pero las pruebas (entrenamiento) del sistema sólo fueron llevadas a cabo haciendo uso de los dos procesos principales más importantes, los cuales son descritos a continuación.**

#### 4.2.1. Acreditación de Programas de Estudio

En la tabla 4.2 se observan las tareas/procedimientos relativos al proceso de “Acreditación de programas de estudio”, así como los agentes encargados de realizarlas.

<b>Tareas/actividades</b>	<b>Agente</b>
Secretaría General Académica ordena se inicie el proceso de acreditación del programa académico	Secretario general académico
En reunión especial: Coordinador de programa, academia, Jefe de departamento y Director de División realizan el análisis previo para valorar la solicitud	Director de división
Se nombra una comisión especial	Director de división
Primero se debe evaluar por el Comité Interinstitucional para la Evaluación de la Educación Superior (CIEES), por lo que se trabaja bajo los criterios de este organismo	Agente genérico (es nombrado para procesos)
Se reúne la documentación, las evidencias, se realiza el pago y se programa la visita de los evaluadores	Agente genérico (es nombrado para procesos)
Se organiza la visita y se coordinan actividades	Jefe de departamento. Secretario académico.
Se lleva a cabo la visita de evaluación	
Se esperan resultados	
Si se obtiene nivel 1, se inicia el proceso para ser acreditado	
Se define el Organismo Acreditador y se realizan los procedimientos para agendar el proceso de acreditación	
Se nombra una comisión especial	Director de división.
Se reúnen evidencias y se envía la información al Organismo Acreditador	Agente genérico (es nombrado para procesos)
Se prepara visita de acreditación	Jefe de departamento. Agente genérico (es nombrado para procesos)



Se atiende a los acreditadores en visita	Agente genérico (es nombrado para procesos)
Se esperan resultados	

**Tabla 4.2** Lista de tareas y agentes para el proceso principal: acreditación de programas de estudio.

Las tareas mostradas en la tabla 4.2 son presentadas en orden de ejecución de las mismas.

#### 4.2.2. Programación Académica

Dado que éste proceso es el más crítico de todos los que se presentaron, en seguida se presenta una descripción con mayor detalle del mismo (tabla 4.3). Se muestran las tareas por las que está compuesto, así como los agentes de conocimiento y recursos utilizados para elaborarlo.

Tareas/actividades	Agente	Recursos utilizados
Secretaría General Administrativa inicia el proceso enviando una circular	Secretaría general administrativa.	* Uso de correo electrónico. *Calendario escolar. *Contrato colectivo de trabajo. *Criterios para la programación académica.
Los Jefes de Departamento y los Coordinadores de programa acuerdan fechas locales	Jefe de departamento.	*Coordinación de reuniones.
Los Coordinadores de programa realizan los pronósticos de grupos requeridos para el siguiente semestre	Coordinadores de programa.	*Bases de datos de información estadística. *Criterios para la programación académica. *Procedimientos de sistema de gestión de calidad. *Habilidad de análisis. Procesamiento de la información.
Se informa al departamento responsable de la infraestructura sobre las necesidades de aulas y	Coordinadores de programa.	*Coordinación de reuniones. *Uso de correo electrónico.

laboratorios		
El departamento libera espacios físicos a los programas	Secretario académico.	*SIIA: Sistema de Información Administrativa.
Los Coordinadores de programa dan de alta los grupos en el SIIA y los envían a autorización	Coordinadores de programa.	*SIIA: Sistema de Información Administrativa.
Las Divisiones autorizan la creación de grupos	Director de división.	*SIIA: Sistema de Información Administrativa.
Una vez autorizados los grupos, los Departamentos inician el procedimiento de asignación de carga académica en Primera Etapa	Jefe de departamento.	*SIIA: Sistema de Información Administrativa.
Se verifica y autoriza la programación de profesores con la Comisión Verificadora	Jefe de departamento. Agente genérico (Sindicato).	
Se realiza la Segunda Etapa de la programación, y se definen los grupos que se someterán a concurso curricular	Jefe de departamento.	*SIIA: Sistema de Información Administrativa. *Procesamiento de información.
Se verifica y autoriza la programación de profesores con la Comisión Verificadora	Agente genérico (Sindicato). Agente genérico (Comisión especial). Jefe de departamento.	
Se solicita a la Comisión		

Dictaminadora de la División se nombre Jurado para las materias a concursar	Director de división.	*Documento formal (oficio).
Se preparan las convocatorias y los jurados: los que son aprobados por el Consejo Divisional	Jefe de departamento.	*Relación de áreas académicas. *Documentos bases para concursos.
Se reciben documentos de los candidatos	Secretaria.	*Habilidad social de atención a usuarios. *Documento formal (oficio).
Se entregan los expedientes al Presidente del Jurado	Secretaria.	Equipo de cómputo.
Los Jurados evalúan utilizando como referencia el Estatuto de Personal Académico	Agente genérico (comisión especial).	*Estatuto de personal académico. *Criterios de evaluación de concursos determinados.
Cada Secretario del Jurado entrega al departamento un documento con el resultado del concurso	Agente genérico (es nombrado para procesos).	Documento formal (oficio).
Se envían los resultados a Recursos Humanos, y al consejo, como notificación oficial del concurso	Secretaria.	Documento formal (oficio).
Se cita a los ganadores para que elijan su carga entre las materias que de cada área salieron a concurso, respetando el máximo de 25 horas	Jefe de departamento.	Coordinación de reuniones.
Si quedan grupos después de esta etapa, se cita a los		

segundos y demás lugares ganadores, siempre y cuando hayan sido declarados aptos, para que elijan entre los grupos sobrantes	Jefe de departamento.	Documento formal (oficio).
Se notifica al Sindicato y al Consejo sobre los resultados de la asignación a ganadores	Secretario académico.	Coordinación de reuniones.
Si aún quedan grupos por asignar, se ofrecen a los MHS por antigüedad, respetando el perfil necesario, hasta terminar de asignarlos, y se notifica nuevamente al Sindicato y Consejo	Jefe de departamento.	Documento formal (oficio).
Se envía a Consejo la Programación definitiva aprobada por el Sindicato para su aprobación final.	Jefe de departamento.	Documento formal (oficio).

**Tabla 4.3** Lista de tareas, agentes y recursos para el proceso principal: programación académica.

Este último proceso suele ser el más importante ya que de él depende la carga semestral que existe cada semestre en todos los departamentos de la Universidad de Sonora. Para el análisis detallado del mismo, se procedió a estudiar el Contrato Colectivo de Trabajo (CCT) del Sindicato de Trabajadores y Académicos de la Universidad de Sonora (STAUS), la investigación previa de los procesos principales del área de estudio llevada a cabo por Sánchez (2012), así como las grabaciones

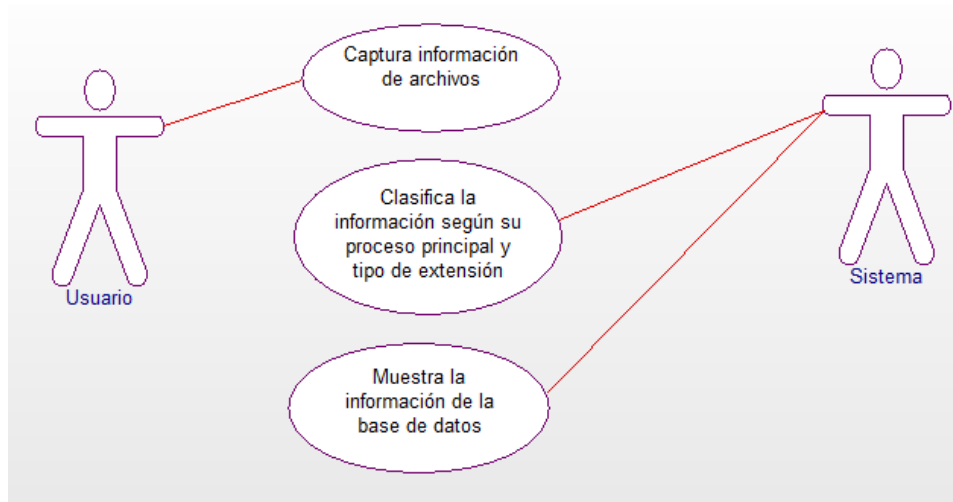
obtenidas en las entrevistas con el personal. La normativa relacionada a este proceso se presenta en el anexo 2.

Debido a la gran cantidad de actividades que se llevan a cabo en la Programación Académica, suele ser difícil gestionar toda la información que ayuda a su elaboración, ya que es un gran problema querer localizar exactamente dónde se encuentra la documentación que conlleva este proceso. En base a esto surge la necesidad por clasificar las distintas fuentes de información con el fin de poder encontrarlas de una manera más fácil.

### **4.3. Clasificar la fuente del conocimiento**

Para llevar a cabo esta etapa fue necesario elaborar un sistema vía Web utilizando el lenguaje de programación php, en el cual los agentes que vayan a utilizar el sistema pueden seleccionar un tipo de archivo para ser cargado en el sistema: texto, imagen, audio o vídeo. Para modelar el sistema se utilizó UML.

El lenguaje unificado de modelado (UML), representa de manera gráfica los diseños de un sistema de información. Son utilizados para visualizar, construir y documentar sistemas de software (Larman, 2008). Su representación suele ser mediante un caso de uso, el cual es una interacción típica entre un usuario y el sistema de cómputo, en general representan lo que el usuario hace con el sistema. Por otro lado, la manera para mostrar grupos de objetos que pertenecen al sistema y la interacción que hay entre ellos, es mediante un diagrama de secuencia. En la figura 4.2 se muestra el caso de uso que representa la captura de información en el sistema utilizado.



**Figura 4.2** Caso de uso: *Captura y clasificación de información*

El usuario experto es quien captura la información relacionada a los tipos de archivo permitidos (textos, imágenes, vídeos y audios), debe ser una persona que conozca el área de trabajo, ya que con ayuda de sus conocimientos se capturará en el sistema la información de manera adecuada. Cuando el usuario experto guarde el archivo, el sistema determinará de manera automática el tipo de extensión con que cuenta.

La manera en cómo se lleva a cabo lo que se mencionó anteriormente será explicado más adelante, la figura 4.3 muestra la interfaz utilizada para la clasificación de la fuente de conocimiento.

**Figura 4.3** Sistema para clasificar la fuente de conocimiento según su extensión

La interfaz mostrada en la figura 4.3 cuenta con diferentes elementos:

**Descripción:** Aquí el usuario experto debe de colocar una descripción sobre el tipo de archivo que se vaya a cargar en el sistema. En este campo se describe a detalle información relacionada directamente con el archivo que se está cargando. La persona encargada de subir archivos, debe de conocer acerca de los procesos que se llevan a cabo en el área, ya que al colocar la descripción del mismo, ésta debe ser lo más descriptiva posible, generando así una mayor confiabilidad al querer recuperar información mediante la utilización de este campo. El proceso de recuperación será mencionado en la siguiente etapa.

**Archivo:** Al dar clic sobre “seleccionar un archivo” el usuario experto puede navegar por los archivos de su computadora y seleccionar el que desee cargar en el servidor, éstos deben de ser de tipo texto, imagen, vídeo o audio. Si el tipo seleccionado no es permitido, se muestra un aviso.

**Proceso principal:** Este campo ayudará a realizar la clasificación de los archivos cargados en el sistema. El sistema muestra todos los procesos principales con que cuenta el Departamento de estudio, pero las pruebas y entrenamiento del mismo, se realizaron sobre los dos procesos críticos mencionados en la etapa anterior, en especial en la programación académica. Para los casos de imagen/audio/vídeo, el usuario experto deberá seleccionar de manera manual el proceso principal al cual pertenezcan. En el caso de fuentes textuales, de manera automática, el sistema hará un análisis del contenido y lo comparará con una serie de palabras clave establecidas en el servidor, con el fin de establecerlo dentro del proceso principal más adecuado. Las palabras clave son aquellas que describen de manera puntual cada uno de los procesos principales, éstas pueden variar para cada uno de los procesos.

**Guardar:** Por último el usuario experto puede guardar los cambios sobre el archivo que haya seleccionado y cargado en el servidor, el sistema identifica el tipo de extensión con que cuenta (doc, pdf, xls, jpg, mp3, etc.) y posteriormente lo coloca en la carpeta correspondiente dentro del servidor.

La manera en cómo el sistema lleva a cabo la clasificación de las fuentes de conocimiento se muestra a continuación.

### **Clasificación de imagen, audio o vídeo**

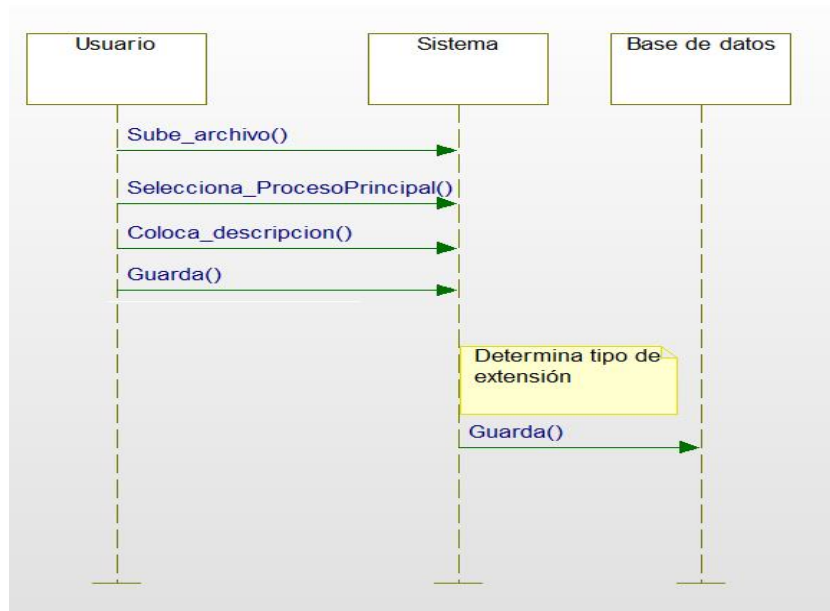
En esta parte del sistema, la clasificación es de manera semi-supervisada, ya que el sistema Webmuestra la lista de procesos principales disponibles y el usuario experto con conocimiento será quien lo seleccione de manera manual. El tipo de extensión del archivo es determinado de manera automática por el sistema.

A continuación se muestra un fragmento de código utilizado para colocar dentro de la carpeta de "imagen" los diferentes tipos de extensiones más comunes que pueden tener los archivos de tipo imagen:

```
if (($extension=="jpg" || $extension=="jpeg" || $extension=="png" || $extension=="
bmp") && $type=="image/jpeg" || $type=="image/bmp" || $extension=="gif" ||
$extension=="tiff" || $extension=="exif" &&$size<$max_size){ $location =
"Documentacion/Imagenes/"; $descripcion="imagen"; $tipo="Imagen"; }
```

En la figura 4.4 se muestra el diagrama de secuencia que representa como se lleva a cabo el proceso del código anterior. Para los tipos de archivo texto, audio y vídeo la secuencia lógica es la misma. El usuario experto es quien carga en el sistema un tipo de archivo: texto, imagen, audio o vídeo, él mismo será quien seleccione el proceso principal al cual se desee que pertenezca, coloca una descripción al mismo y por último, guarda la información capturada anteriormente. De manera automática el sistema determina el tipo de extensión y guarda dentro de la base de datos en la categoría correspondiente.





**Figura 4.4** Diagrama de secuencia: Clasificación de fuentes no textuales

Para establecer a qué proceso principal pertenecen las fuentes textuales, el sistema analiza de manera automática el contenido y lo compara con una lista de palabras clave. La lista de palabras clave se obtiene en base a un análisis completo del contenido textual de archivos que pertenezcan al proceso principal analizado, enfocándose en las cinco palabras de mayor aparición, así como con apoyo del usuario experto para ayudar a identificarlas y listarlas de la mejor forma posible.

### **Clasificación de fuentes textuales**

En esta parte del sistema, la clasificación está desarrollada en base a algoritmos de clasificación supervisados, los cuales son aquellos que utilizan ciertos datos para su entrenamiento. El sistema Web se basa en la lista de palabras clave contenidas en la base de datos del servidor. Para llevar a cabo esta parte, el sistema se parte de la lógica del “algoritmo clasificador vecino k más cercano”, el cual es usado para comparar las palabras clave contenidas en el servidor con el archivo de texto a clasificar. Suele ser un algoritmo eficaz y fácil de implementar.

Por otro lado, también se hace uso de la lógica del “algoritmo clasificador Reglas de decisión”, el cual ayuda a establecer el conjunto de reglas que describe a cada uno de los procesos principales. En ésta investigación se presentan las pruebas sobre los procesos principales “programación académica” y “acreditación de programas de estudio”.

Para obtener las palabras clave relacionadas a cada proceso, se hizo uso del analizador de texto implementado en el sistema Web, con el fin de contabilizar las palabras que aparecen con más frecuencia. Debido a que en la literatura revisada no se encontró ningún tipo de fuente que indicara el número de palabras que tienen que seleccionarse para ayudar a llevar a cabo la clasificación, se tomaron cinco palabras clave por proceso principal, lo anterior en base a que en artículos revisados para este trabajo, generalmente se utilizan cinco palabras clave. Por otra parte se consultó a personas del área de estadística para tomar la decisión acerca del número de repeticiones  $P_i$  necesarias para cada Palabra Clave  $i$ . Con ello, y con ayuda de un usuario experto, se obtuvieron las palabras que describen a cada proceso principal:

<b>Palabras clave por Proceso principal</b>			
	<b>Acreditación de programas de estudio</b>	<b>Programación académica</b>	<b>Repeticiones (Pi)</b>
Palabras clave (i)			
1	Acreditación	Programación	$i \geq 20$
2	CIEES	Concurso	$15 \geq i \leq 19$
3	Evaludador	Profesores	$11 \geq i \leq 14$
4	Visita	Antigüedad	$6 \geq i \leq 10$
5	Organismo	Evaluación	$1 \geq i \leq 5$

**Tabla 4.4** Palabras clave por proceso principal y repeticiones

La tabla 4.4 muestra los rangos superior e inferior asignados para las repeticiones de cada palabra. La columna repeticiones  $P_i$  muestra la frecuencia propuesta en que una palabra clave pudiera aparecer en el texto analizado. Debido a que el objetivo central de este trabajo no es el determinar el conjunto de palabras clave así como su rango de repeticiones  $P_i$ , se propone como trabajo futuro elaborar una propuesta para ello.

Pudiera existir el caso en que al cargar un documento hubiera confusión sobre en qué proceso principal clasificarlo, por ello, a cada conjunto de repeticiones ( $P_i$ ) se asignó un peso " $W_i$ ", con el fin de determinar las puntuaciones al clasificar un documento dentro de su respectivo proceso principal. Lo anterior se llevó a cabo tomando como base la lógica del algoritmo "vecino-k más cercano", donde los vecinos "k-cercanos" son la serie de palabras clave establecidas dentro del servidor. La distancia a "k-vecino" es representada por el número de repeticiones  $P_i$  de la palabra con respecto a los límites superiores e inferiores establecidos. En la tabla 4.5 se puede observar el peso " $W_i$ " que se dio a cada rango de repeticiones " $P_i$ ".

Repeticiones ( $P_i$ )	Peso ( $W_i$ )
$i \geq 20$	0.38
$15 \geq i \leq 19$	0.28
$11 \geq i \leq 14$	0.21
$6 \geq i \leq 10$	0.11
$1 \geq i \leq 5$	0.02
$\Sigma$ límite inferior= 53	$\Sigma W_i = 1$

**Tabla 4.5** Peso  $W_i$  en base a repeticiones  $P_i$

La columna Peso ( $W_i$ ) de la tabla 4.5 se compone de dos valores:

**Numerador:** Valor del límite inferior permitido para cada conjunto  $P_i$ .

**Denominador:** Suma total de los límites inferiores (igual a 53).

Al subir un archivo al servidor, el sistema realiza un análisis de su contenido y éste es comparado con los diferentes conjuntos de palabras clave almacenados en el sistema. Para cada conjunto de palabras clave, el sistema lleva a cabo una sumatoria de puntuaciones individuales de cada proceso principal. La ecuación 4.1 representa la sumatoria para cada proceso principal:

$$P_{c_k} = \sum_{i=1}^n P_i W_i$$

(4.1)

Dónde:

$P_{c_k}$ : Peso del proceso principal  $i$ .

$P_i$ : Número de veces que se repite la palabra clave  $i$ .

$W_i$ : Peso correspondiente según en el rango de repetición que se encuentre la palabra clave  $i$ .

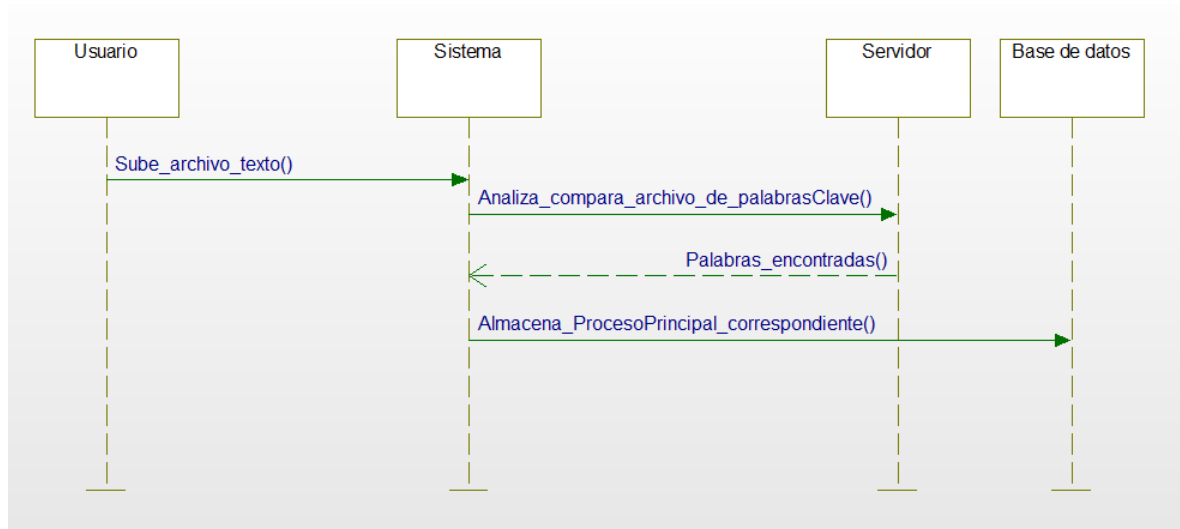
Una vez realizados los cálculos necesarios, de manera automática, el sistema clasifica al archivo dentro del proceso principal que tiene el peso más alto.

A continuación se muestra un fragmento de código utilizado para el análisis automático de fuentes textuales:

```
$texto = file_get_contents('conocimiento.txt');
$keywords = array_map('trim', file('palabras_clave.txt'));
$Arreglo_palabras = preg_split("/[\s,\\"]+/", $texto);
$words = array_count_values($Arreglo_palabras);
```

```
foreach ($words as $key_words => $value){
    if ($value>=num_palabra_clave)
```

En la figura 4.5 se muestra el diagrama de secuencia que representa cómo se lleva a cabo el proceso del código anterior:



**Figura 4.5** Diagrama de secuencia: Clasificación de fuentes textuales

Como se puede observar en la figura 4.5, el usuario experto es quien carga en el sistema el archivo de texto, así como colocar una descripción del mismo antes de guardar. Posterior a ello, el sistema analiza y compara el archivo cargado por el usuario experto con otro que se encuentra dentro del servidor, el cual contiene las palabras clave para establecer el proceso principal. El servidor da respuesta al sistema en el caso de haber encontrado o no coincidencias con el archivo de palabras clave. Por último, el sistema almacena dentro del proceso principal correspondiente.

La implementación de análisis automático de texto dentro del presente trabajo es otra aportación para quienes gestionan la información en el área de estudio, debido a que ayuda a agilizar tiempos a la hora de querer organizar la documentación con que se cuenta. La utilidad que conlleva este tipo de clasificación es realmente significativa: el

usuario experto en el tema solo es requerido para elaborar la serie de palabras clave con las cuales el proceso principal será comparado, después de ello, el sistema es quien hace la clasificación de manera automática sin necesitar de esta persona (a pesar de ello, se recomienda que el usuario experto participe en el proceso de captura de archivos). Por otro lado, ayuda a los empleados a encontrar la información cuando ésta sea requerida y sobre todo, localizar lo que realmente se está buscando. Cabe mencionar que un tipo de archivo puede llegar a pertenecer a distintas categorías (procesos principales), y con ayuda del sistema Web la clasificación es llevada a cabo de una manera más rápida y sobre todo sencilla.

#### 4.4. Recuperar el conocimiento

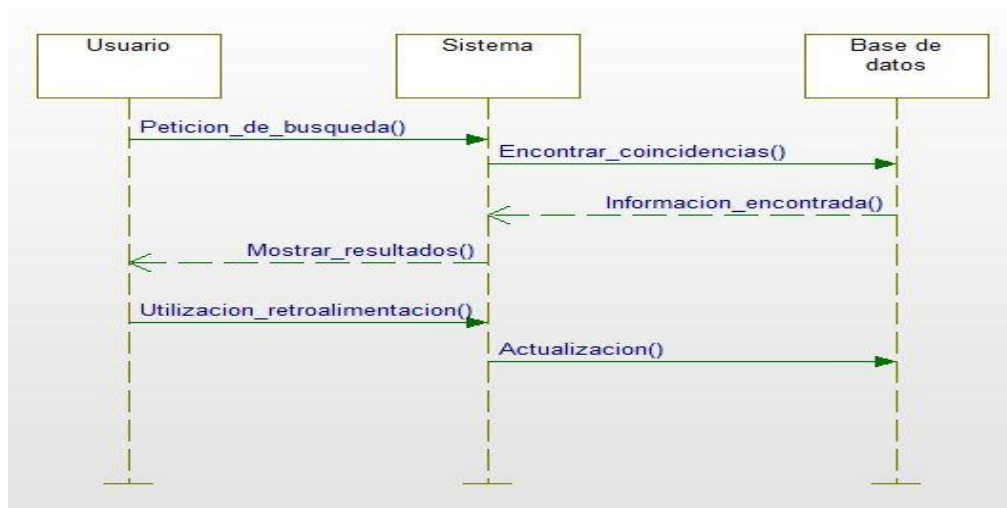
Para la elaboración de esta etapa fue necesaria la implementación de un buscador en la sección de “consulta de archivos existentes”, dentro del sistema Web utilizado.

Como se mencionó en el capítulo anterior, la recuperación de conocimiento consta de una serie de pasos en los que el usuario común tiene contacto con el sistema.



**Figura 4.6** Caso de uso: Recuperación de la fuente de conocimiento

El caso de uso que representa la recuperación de conocimiento es mostrado en la figura 4.6, el usuario común hace una petición de búsqueda, posteriormente el sistema identifica si existen coincidencias en lo que se tiene almacenado y por último muestra los resultados en pantalla.



**Figura 4.7** Diagrama de secuencia: Recuperación de la fuente de conocimiento.

La secuencia de actividades que conlleva el proceso de recuperación de la fuente de conocimiento se muestra en la figura 4.7, la recuperación consiste en la petición de búsqueda por parte del usuario común, el sistema se encarga de encontrar coincidencias en la base de datos comparando los patrones de búsqueda establecidos. La información encontrada es mostrada en pantalla al usuario común y finalmente éste puede utilizar y retroalimentar el sistema, en dado caso de ocurrir esto último, el sistema actualiza de manera automática la base de datos. Para el caso de retroalimentación de los archivos del sistema, el usuario debe de ingresar a las opciones de “archivo” → “actualizar” dentro del sistema Web (figura 4.9).

### **Paso 1: Petición por parte del usuario.**

El buscador ofrece al usuario común tres maneras distintas de hacer una petición cuando se deseen encontrar los archivos solicitados:

**Por proceso principal:** Se puede seleccionar de entre los procesos principales disponibles, el sistema compara la información contenida en la base de datos del servidor y posteriormente muestra las coincidencias (si es que las hay) al usuario.

**Por tipo:** Otra opción disponible para hacer los filtrados de búsqueda es seleccionar el tipo de archivo que se desee encontrar: texto, imagen, audio o vídeo. Esta columna es determinada de manera automática por el sistema al detectar la extensión con que cuentan los archivos cargados en el servidor.

**Por descripción:** El usuario tiene la opción de escribir lo que desee encontrar, el sistema compara las palabras introducidas con la información contenida en la base de datos del sistema, cabe mencionar que la comparación es con respecto al campo de “descripción” que contiene cada archivo.

Para hacer filtrados con un mayor nivel de detalle, el usuario puede usar de manera combinada las tres opciones de búsqueda mencionadas anteriormente.

## Paso 2: Base de datos.

La base de datos utilizada en el sistema Web fue implementada con el gestor de base de datos “My SQL”, la cual cuenta con una tabla llamada “clasificación”. En la figura 4.6 se muestra la estructura que presenta la tabla.

#	Nombre	Tipo	Cotejamiento
<input type="checkbox"/> 1	<b>ID</b>	int(11)	
<input type="checkbox"/> 2	<b>DESCRIPCION</b>	varchar(100)	latin1_swedish_ci
<input type="checkbox"/> 3	<b>Tipo</b>	varchar(20)	latin1_swedish_ci
<input type="checkbox"/> 4	<b>CATEGORIA</b>	varchar(100)	latin1_swedish_ci
<input type="checkbox"/> 5	<b>ARCHIVO</b>	varchar(100)	latin1_swedish_ci

**Figura 4.8** Estructura de tabla clasificación en la base de datos

El campo “ID” es utilizado para llevar un control interno de los archivos que se localizan en el servidor, los campos “descripción, tipo, categoría y archivo” ya fueron descritos al inicio de esta sección.



### Paso 3: Encontrar coincidencias con respecto a la petición.

El sistema compara la petición de búsqueda por parte del usuario común con respecto a la información contenida en la base de datos. En caso de encontrar coincidencias, éstas son mostradas en pantalla. A continuación se muestra un fragmento de la codificación usada en el sistema para llevar a cabo las búsquedas solicitadas.

```
$query_Recordset1 = "SELECT * FROM clasificacion WHERE CATEGORIA LIKE '%".$busqueda_categoria."%' AND Tipo LIKE '%".$busqueda_tipo."%' AND DESCRIPCION LIKE '%". $busqueda_Descripcion."%'";
```

El código anterior incluye las tres diferentes opciones de búsqueda explicadas en el paso 1: proceso principal, tipo de extensión y descripción.

### Paso 4: Mostrar resultados en pantalla.

Para cada uno de los archivos el usuario común puede ver su descripción, el tipo de extensión (.doc,.jpg,.pdf, etc.), y el proceso principal al cual fue asignado. La figura 4.9 muestra la interfaz utilizada.

The screenshot shows a search interface titled "Buscador de archivos: por proceso principal, tipo o descripción". It includes a search filter section with dropdowns for "Proceso principal" (set to "Todas") and "Tipo" (set to "Todos"), and a text input for "Descripción" with a "Buscar" button. Below this is a table with the following structure:

DESCRIPCION	Tipo de archivo según su extensión	Categoría (proceso principal) al cual fue asignado	Opciones para cada archivo		
	Tipo	Proceso principal	ARCHIVO		
Ejemplo de imagen	Imagen	Alta de grupos en SIA	Consultar	Actualizar	Eliminar
En el presente periodo se estuvo trabajando con la cuarta etapa de la metodología, la cual tiene que ver con la recuperación del conocimiento.	Texto	Recepción documentos de candidatos	Consultar	Actualizar	Eliminar
ejemplo de video	Video	Inicio de proceso	Consultar	Actualizar	Eliminar
ejemplo de docx	Texto	Informes de infraestructura requerida	Consultar	Actualizar	Eliminar
Ejemplo de texto pdf	Texto	Resultados de concursos	Consultar	Actualizar	Eliminar
ejemplo de texto ppt	Texto	Informes de infraestructura requerida	Consultar	Actualizar	Eliminar
Ejemplo de un audio con actualización y cambio de categoría	Audio	Inicio de proceso	Consultar	Actualizar	Eliminar

At the bottom left, it says "Total de archivos: 7" and "Ver todos". An arrow points from the text "Descripción dada por el usuario experto" to the search input field.

**Figura 4.9** Resultados y opciones de filtrado para la recuperación de conocimiento

## **Paso 5: Reutilización y aprovechamiento.**

Como se observa en la figura 4.9, para cada archivo el usuario común tiene la opción de: consultar, actualizar o eliminar.

**Consultar:** Esta opción permite visualizar el contenido del archivo, en caso de ser una extensión soportada por el navegador de internet, su contenido podrá ser visto en línea. Para los tipos de extensión no soportados por el navegador, éstos serán descargados al dispositivo del usuario, y se deberá de seleccionar un programa para visualizarlo.

**Actualizar:** Esta opción permite realizar una modificación con respecto a la información con que cuenta el archivo.

**Eliminar:** Cuando el usuario considere que un archivo ya no es necesario, se tendrá la opción de eliminar la información contenida en el servidor. Un mensaje de confirmación antes de la eliminación permanente será mostrado.

## 5. RESULTADOS

En esta sección se presentan los principales resultados obtenidos en este proyecto de investigación, donde se trabajó con la problemática detectada en el Departamento de estudio: No contar con información de manera organizada para posteriormente recuperarla.

Las carencias en el área de estudio eran muy notables, ya que los diferentes tipos de archivo donde manejan su información no se encontraban contenidos en un mismo lugar, es decir, el jefe de departamento tenía cierto documento, la secretaria podía tener otro, etc. En el momento de querer recuperar información, no se sabía quién era la persona que la tenía, por lo que en gran parte de las ocasiones se optaba por volver a hacer un archivo nuevo.

Se tomó como base y referencia al trabajo desarrollado por Sánchez (2012), Quién propuso un modelo para identificar los diferentes procesos principales del Departamento de estudio, pero no se contaba con una herramienta que ayudara a gestionar todos los archivos referentes a cada uno de los procesos. Es por ello que se hizo la propuesta de clasificación y recuperación desarrollada en esta investigación y que es un complemento a la propuesta por Sánchez (2012), ya que fue posible identificar los procesos que son de mayor relevancia para el Departamento, y así, clasificar y recuperar los archivos cargados en el sistema por el usuario experto.

Lo anterior fue posible gracias a la programación de un sistema Web, utilizando algoritmos que ayudan a la clasificación de los medios digitales donde almacenan la información: texto, imagen, audio y vídeo. Una de las características interesantes de este proyecto fue la clasificación automática de fuentes de información textual, ya que mediante la identificación de contenido de texto plano, es posible comparar un documento de texto con una serie de palabras principales dadas por el usuario experto, y así, llevar a cabo la clasificación automática de la misma. Este tipo de clasificación automática es un verdadero impacto para quienes laboran en la organización, ya que ayuda a agilizar tiempos a la hora de querer organizar la

documentación con que se cuenta. Por otro lado, si la información se encuentra clasificada, ayudará en la recuperación de la misma, ya que mediante la implementación de filtrados de búsqueda, se puede restringir sobre lo que se desee obtener.

Con la implementación de un sistema donde los empleados pudiesen almacenar de manera conjunta todos los diferentes tipos de archivos con que se cuentan, el Departamento adquiere beneficios, por ejemplo: No crear dependencia hacia ninguno de los empleados, contar con un repositorio de información de forma organizada en el que se sepa a dónde pertenece cada archivo, así como poder recuperarla de manera rápida y confiable cuando ésta sea solicitada.

Los usuarios del Departamento no tenían de manera establecida los diferentes conjuntos de palabras clave que describen a cada proceso principal, por lo que con ayuda del sistema Web se analizaron diferentes documentos relacionados a cada proceso con el fin de determinar cuáles eran las más importantes (el usuario experto también tiene un papel importante, ya que él ayudó a confirmar o reacomodar las palabras con respecto a su importancia).

La metodología propuesta para clasificar y recuperar información, puede ser utilizada para cualquier tipo de organización que quisiera gestionar su información. Cabe mencionar que para cada una de las organizaciones donde desee ser implementada, debe de estudiarse a detalle cuales son los procesos principales con que se cuentan, así como determinar la lista de palabras clave asociadas a cada uno de ellos, con ayuda de un usuario experto en el área y el sistema vía Web implementado. Lo anterior ayudará al sistema a clasificar de la mejor manera posible los archivos, de modo que cuando se quiera recuperar alguno de ellos, sea lo que en realidad se está buscando. Un elemento clave para el éxito de este proyecto, fue la participación de los miembros del área donde se llevó cabo el mismo, ya que desde un principio estuvieron de acuerdo en llevar a cabo la implementación de un sistema que ayudara a la gestión de toda la documentación con que se cuenta.

## 6. CONCLUSIONES

En este capítulo se presentan las conclusiones, recomendaciones y posibles trabajos futuros en relación al procesamiento automático de fuentes de conocimiento de cualquier naturaleza.

### 6.1. Conclusiones

El contar con fuentes de información (conocimiento) de manera desorganizada, resulta ser una tarea tediosa para quienes laboran en una organización, ya que cuando se desea recuperar de manera inmediata puede llegar a transcurrir mucho tiempo antes de localizarlo. En base a lo anterior surge el interés por desarrollar este trabajo, con el fin de ayudar a clasificar y recuperar información de manera eficaz.

En este trabajo de investigación, se desarrolló una metodología que puede adaptarse a cualquier tipo de organización sin problema alguno, solo que dependiendo de su naturaleza, la identificación de los procesos principales puede variar.

Para desarrollar una herramienta que ayudara a mantener de forma organizada el conocimiento con que cuenta una organización, se tuvo que comprender a detalle cuál es la información/procesos importantes que ésta desarrolla. Para ello se estudiaron diferentes temas (presentados en el marco de referencia), así como mantener pláticas con los agentes de conocimiento de la organización de estudio.

Se elaboró una propuesta para sistematizar los formatos de tipo texto, audio, video e imagen, al almacenar el conocimiento organizacional. Lo anterior fue elaborado en base a algoritmos de clasificación con el fin de tener la información de manera organizada. Una propuesta para la clasificación de fuentes textuales fue elaborada con el fin de asignar dentro de un proceso principal correspondiente cada uno de los archivos cargados por un usuario experto. Por último se diseñó una estructura tecnológica para categorizar y recuperar el conocimiento organizacional, basada en el lenguaje de programación php.

Una gran ventaja de esta propuesta, es que cuando los miembros de la organización quieran recuperar información de manera rápida, bastará con ingresar al sistema Web y realizar la búsqueda en el módulo de “archivos existentes”. El usuario común puede realizar sus búsquedas mediante las distintas opciones de filtrado con que cuenta el sistema, evitando así perder tiempo al querer encontrar un archivo en específico.

Aunque en el sistema propuesto aún se requiere trabajar con la clasificación automática de imagen, audio y vídeo, éste tiene la capacidad de ayudar a clasificar y recuperar de manera eficaz lo que los usuarios gestionan en su área de trabajo.

## **6.2. Recomendaciones**

Este trabajo fue realizado con el fin de ayudar a una organización a tener una forma organizada de la información con que se cuenta. Al momento de querer aplicar la metodología planteada, se debe poner mucho énfasis en la identificación de los procesos principales de las empresas, ya que en base a ello, se “entrena” el sistema Web. Si el estudio donde se lleva a cabo la investigación de los posibles procesos principales no es correcto, la recuperación de las distintas fuentes de conocimiento con que cuente la organización, no será la correcta cuando éstas quieran ser recuperadas.

## **6.3. Trabajos futuros**

Como trabajos futuros a este trabajo, se podría llevar a cabo la implementación de clasificación automática para imágenes, audio o vídeo, mediante el análisis de contenido con ayuda de algoritmos de clasificación especializados en ello. El implementar lo anterior al sistema Web desarrollado, ayudaría a evitar que el usuario experto tuviera que capturar de manera manual el proceso principal al cual los archivos pertenecen. Cabe mencionar que para la clasificación automática de imagen, audio o vídeo, tendría que llevarse a cabo una investigación a fondo para obtener los patrones con los cuales será comparado lo que se encuentre en el análisis de los mismos. Otro posible estudio es elaborar un análisis literario de posibles sistemas de información que ayuden a llevar a cabo la gestión de documentos en una

organización, con el fin de poder contrastar las ventajas y desventajas con respecto al sistema Web propuesto en este trabajo de investigación. Por último se tiene el realizar un estudio a fondo para ayudar a establecer las palabras clave que describen a diferentes procesos principales, así como qué rangos (número de repeticiones) deben cumplirse para considerarlas dentro del proceso principal correspondiente.

## 7. REFERENCIAS

Ackerman, M. y Halverson, C. 2000. Re-examining organizational memory. Communications of the ACM.

Alavi, M. y Leidner D. 1999. Knowledge management systems: Issues, challenges, and benefits. Smith School of Business. Volume 1, Article 7.

Atkinson, P. y Tatnall, A. 1997. Introduction to neural networks in remote sensing, International Journal of Remote sensing, volume 11, pp. 699-709.

Awad, W. y ELseuofi, S. 2011. Machine Learning methods for E-mail Classification. Math.&Comp.Sci.Dept., Science faculty, Port Said University Information. SystemDept., Ras El Bar High inst

Barceló, M. (2008). Una propuesta metodológica para la gestión de conocimiento centrada en los procesos claves de la organización. Tesis Doctoral, Universidad de Murcia, España. (2008).

Barceló, M., Sánchez, G., Perez, A., Martin, F., Palma, J. (2008). Defining the Problem: Key Element for the Success of Knowledge Management, Journal of Knowledge Management Research & Practice (KMRP).

Beckman, T.: A Metodology of Knowledge Management. In Proceedings of the International Association of Science and Technology for Development (IASTED) AI and Soft Computing Conference. Banff, Canada (1997)

Berkhin, P. 2002. Survey of Clustering Data Mining Techniques. Research paper, Accrue Software, <http://www.accrue.com/products/researchpapers.html>, 2002.

Bless, P., Klabjan, D. y Chang, S. 2012. Automated knowledge source selection and service composition. Computational optimization and applications.

Brezeale, D. y Diane, J. 2008. Automatic video classification: A Survey of the literatura. IEEE Transactions on systems, man, and cybernetics-part c: application and reviews, vol. 38, no. 3, pp. 416-430.

Chauhan, R., Kaur, H y Alam, M. 2010. Data Clustering Method for Discovering Clusters in Spatial Cancer Databases. International Journal of Computer Applications Volume 10– No.6

Chen, G., Song, Y. y Zhang, C. 2008. Semi-supervised Multi-label Learning by Solving a Sylvester Equation. In SDM.



Choi, Sue Young; Lee, Heeseok; and Yoo, Youngjin. 2010. "The Impact of Information Technology and Transactive Memory Systems on Knowledge Sharing, Application, and Team Performance: A Field Study," *MIS Quarterly*, (34: 4) pp.855-870.

Congreso del Estado de Sonora, 1991 Ley Orgánica de la Universidad de Sonora. [online] <<http://docs.mexico.justicia.com.s3.amazonaws.com/estatales/sonora/ley-orgánica-de-la-universidad-de-sonora.pdf>>[Accesed: 10 may 2013].

Dalal, M. y Zaveri, M. 2011. Automatic text classification: A technical review. *International Journal of Computer Applications*, volumen 28- No. 2.

Dalkir, K. 2011. *Knowledge management in Theory and Practice*, Second edition. The MIT Press Cambridge, Massachussets. London, England.

Davenport, T.H., Prusak L.: *Working Knowledge: How Organizations Manage What They Know*, Boston: Harvard Business School Press (1998)

Davenport T., Jarvenpaa S. and Beers M. (1996). *Improving Knowledge Work Processes*, Sloan Management Review, Summer.

Davies J., Fensel D., Harmelen F. (2003). *Towards the Semantic Web, Ontology-driven Knowledge Management*. John Wiley & Sons Ltd.

Dimitrova, N. y Agnihotri, L. 2011. Video classification using object tracking.

Dhanalakshmi, P., Palanivel. S. y Ramaligam. V. 2008. Classification of audio signals using SVM and RBFNN. In Elsevier, *Expert systems with application*, Vol. 36, pp. 6069–6075.

Dharmadhikari, S., Ingle, M. y Kulkarni, P. 2011. Empirical Studies on Machine Learning Based Text Classification Algorithms. *Advanced Computing: An International Journal (ACIJ)*, Vol.2, No.6.

Dharmadhikari, S., Ingle, M., y Kulkarni,P. 2012. Analysis of semi supervised learning methods towards multi label text classification. *International Journal of Computer Applications*. Volumen 42- No. 16.

Dolques, X. 2012. Fixing generalization defects in UML use case diagrams. Campus universitaire de Beaulieu, 35042 Rennes, France.

Fabrizio, S. 2002. Machine Learning in Automated Text Categorization, *ACM Computing Surveys*, Vol. 34, No. 1.

- Fayyad, U. 1996. Data mining and Knowledge discovery in databases: Applications in Astronomy and Planetary Science. Microsoft Research.
- Geist, G., Howell, G. y Watkins, D. 1999. The Br eigenvalue algorithm. Society for industrial and applied mathematics. Vol. 20, No. 4, pp. 1083-1098.
- Gronau, N., Weber, E. (2004). Management of Knowledge Intensive BusinessProcesses. In (Desel, J., Pernici, B., Weske, M. Hrsg.): Business Process Management, Springer (Heidelberg), pp 1-17.
- Hahn, J. y Wang, T. 2009. Knowledge management systems and organizational knowledge processing challenges: a field experiment. Decision Support Systems, Vol. 47 No. 4, pp. 332-42.
- Harris, D. 1996. Creating a Knowledge Centric Information Technology Environment, <http://www.htca.com/ckc.htm>
- Hernández, J., Ramírez, M., y Ramírez F. 2004. Introducción a la Minería de Datos.
- Horrocks, I. (1998). The FaCT system, Automated Reasoning with Analytic Tableaux and Related Methods 307-312 , Springer Berlin/Heidelberg.
- Ichijo, K. y Kohlbacher, F. 2006. Global Knowledge Creation – The Toyota Way. Int. J. Automotive Technology and Management, 7, pp. 116-134
- Imbernón, F., Silva, P. y Guzmán, C. 2011. Competencias en los procesos de enseñanza-aprendizaje virtual y semipresencial. Co -municar, 36; 107-114.
- Instituciones Públicas de Educación Superior, 2004. Contraloría USON. [online] <<http://contraloria.uson.mx/wp-content/uploads/2012/02/resultadosauditoria2004.pdf>>[Accesed: 15 may 2013].
- Jacobson, I. 2004. Use cases –Yesterday, today, and tomorrow. SoftwSystModel : 210–220 / Digital Object Identifier (DOI) 10.1007/s10270-004-0060-3
- Jantan, H., Razak, A. 2010. Human Talent Prediction in HRM using C4.5 Classification Algorithm. HamidahJantan et al. / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 08, 2010, 2526-2534.
- Khan, A., Baharudin, B., y Lee, L. 2010. A Review of Machine Learning Algorithms for Text- Documents Classification. Journal Of Advances in Information Technology, Vol. 1 , No. 1.

- Kuo, R., Lai, M. y Lee, G. 2011. The impact of empowering leadership for KMS adoption. *Management Decision*.
- Kuo, R. y Lee, G. 2009. KMS adoption: the effects of information quality. *Management Decision*, Vol. 47 No. 10, pp. 1633-51.
- La, L., Guo, Q., Yang, D. y Cao, Q. 2012. Multiclass boosting with adaptive group-based kNN and its application in text categorization. Hindawi Publishing Corporation *Mathematical Problems in Engineering* Volume 2012, Article ID 793490, 24 pages
- Larman, G. 2008. *UML y patrones: una introducción al análisis y diseño orientado a objetos y al proceso unificado*. Pearson Prentice Hall.
- Lee, Yoo y Choi. 2009. Semi-supervised Nonnegative Matrix factorization. *IEEE*.
- Lin, R. y Chen, L. 2005. A new approach for classification of generic audio data. Department of computer and Information Science. National Chiao Tung University.
- Liu, Y., Jin, R. y Yang, L. 2006. Semi-supervised Multi-label Learning by Constrained Non-Negative Matrix Factorization .In: *AAAI*.
- Liu, D. y Wu, I. 2008. Collaborative relevance assessment for task-based knowledge support. *Decision Support Systems*, Vol. 44 No. 2, pp. 524-43.
- López, M 2012. Diseño de un sistema de memoria organizacional para la mejora de los procesos de servicio de un departamento de tecnología. Tesis de Maestría, Universidad de Sonora, Hermosillo, México.
- Mitchell, T. 2006. *The Discipline of Machine Learning* , CMU-ML-06-108, July 2006.
- Narayana, K. 2003. Advances in Automatic text categorization. *DRTC Workshop on Semantic Web*, Bangalore, India, 8-10 December, 2003.
- Nevo, D. y Wand, Y. 2005. Organizational memory information systems: a transactive memory approach. *Decision Support Systems* 39, 549-562.
- Nigam, K., McCallum, A., y Mitchell. T. 2006. *Semi-supervised Text Classification Using EM*. MIT Press.
- Nonaka, I. 1994. "A Dynamic Theory of Organizational Knowledge Creation" *Organization Science*, (5) 1, pp. 14-37.
- O'Brien, J. y Marakas, G. 2006. *Management Information Systems*, 7th ed., McGraw-Hill International, New York, NY.

Ozgun, A. 2004. Supervised and unsupervised machine learning techniques for text document categorization. Thesis submitted in Department of Computer Science, Bogaziki University.

Paklone, I. 2011. Conceptualization of visual representation in urban planning. ISSN 2029-7475.

Palaniswami, C., Upadhyay, A. K., &Maheswarappa, H. P. 2006. Spectral mixture analysis for sub pixel classification performance of multispectral images. Journal of Computing, 2(2), 124-129.

Park, S. y Dong, U. 2010. Automatic E-mail classification using Dynamic category hierarchy and semantic features. School of information technology, Chonbuk national university, South Korea.

Parrinelo, T., Vaughan,R. 2006. On comparing multifractal and classical features in minimum distance classification of AVHRR imagery. International Journal of Remote Sensing Vol. 27, No. 18, 3943-3959.

Perez, J. 2010. Internet's Information as a Video Signal and its Editor Vol. 29 Issue 57, p178-191, 14p.

Pérez, Z. J. y Cortéz, R. J. A.: Aprendizaje organizacional como ventaja competitiva. Magazine for managers CLEDEA, pp. 32-22 (1999)

Perumal, K. y Bhaskaran, R. 2010. Supervised Classification Performance of Multispectral Images. JOURNAL OF COMPUTING, VOLUME 2, ISSUE 2, FEBRUARY 2010, ISSN 2151-9617

Pise, N y Kulkarni, P. 2008. A survey of Semi-Supervised Learning Methods". IEEE International conference on Computational Intelligence and Security. 2008.30-34.

Polanyi, M. 1962. Personal Knowledge: Toward a Post-Critical Philosophy, New York, NY: Harper Torchbooks.

Pradeep, T., Srinivasu, P., Avadhani, P.y Murthy, Y. 2011. Comparison of variable learning rate and Levenberg-Marquardt back-propagation training algorithms for detecting attacks in Intrusion Detection Systems. Tummala Pradeep et al. / International Journal on Computer Science and Engineering (IJCSE).

Quintas, P., Lefrere, P. y Jones, G. 1997. Knowledge management: a strategic agenda. Journal of Long Range Planning, Vol. 30, No. 3, pp. 385-91.

Rajeswari, K y Vaithiyannathan, V. 2012. Mining Association Rules Using Hash Table. International Journal of Computer Applications.

Rajeswari, K. y Vaithiyannathan, V. 2011. Heart Disease Diagnosis: An Efficient Decision Support System Based on Fuzzy Logic and Genetic Algorithm. International Journal of Decision Sciences, Risk and Management by Inderscience Publications.

Ramírez, P. N. y Martín, M. A., 2003. Herramientas para la gestión del conocimiento. Categoría Administración, Gestión del Conocimiento, pp. 1-34.

Roddick, J. y Lees, B. 2001. Paradigms for spatial and spatio-temporal data mining. In MILLER, H. y HAN, J. Geographic data mining and knowledge discovery. London: Taylor & Francis.

Salvador, R. (2006). Software. Escuela superior de comercio "Libertador Gral. San Martín, 7 año de la EGB.

Sánchez, G. (2012). Estructuración tecnológica de sistemas de gestión del conocimiento para procesos clave intensivos en conocimiento. Tesis Doctoral, Universidad de Murcia, España.

Sirbu, M., Doinea, O. y Goirgiana, M.: Knowledge based economy - the basis for insuring a sustainable development. Annals of the University of Petroșani, Economics, 9, pp. 227-232 (2009)

Solarte, G y Ocampo, C., 2009. Técnicas de Clasificación y análisis de representación del conocimiento para problemas de diagnóstico. Scientia et Technica Año XV, No 42 Agosto de 2009. Universidad Tecnológica de Pereira.

Soto, J. 2012. Images, society and its decoding. ISSN: 1578-8946.

Stein E. 1995. Organizational Memory: Review of Concepts and Recommendations for Management: International Journal of Information Management, Vol 15, No 2, pp. 17-32.

Subashini, K., Palanivel, S. y Ramalingam, V. 2012. Audio-video based classification using SVM and AANN. International Journal of Computer Applications (0975 – 8887).

Subramanya, A. y Bilmes, J. 2008. Soft-Supervised Learning for text classification. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Pages: 1090-1099.2008.

Taiwo, A., Shikun, Z. y Rinat, K. 2010. Email Classification Using Back Propagation Technique. Department of Electronics and Computer Engineering University of Portsmouth, United Kingdom.

Universidad de Sonora, 2004. Departamento de Ingeniería Industria. [online] Available at: <<http://www.industrial.uson.mx/departamento.php>> [Accesed: 7 April 2013].

Universidad de Sonora, 1991. Ley orgánica número 4. [online] Available at: <[http://www.uson.mx/institucional/marconormativo/leyesyestatutos/ley\\_num4\\_organica.htm](http://www.uson.mx/institucional/marconormativo/leyesyestatutos/ley_num4_organica.htm)> [Accesed: 20Mayo 2014].

Uriarte, F. A., 2008, Introduction to Knowledge Management. Jakarta, Indonesia: ASEAN Foundation.

Vaithyanathan, V., Rajeswari, K., Tajane, K. y Pitale, R. 2013. Comparison of different classification techniques using different datasets. International Journal of Advances in Engineering & Technology.

Valenzuela, J., 2009. División de Ingeniería. [online] Available at: <[http://www.ingenierias.uson.mx/wp-content/uploads/2010/08/propuesta\\_09.pdf](http://www.ingenierias.uson.mx/wp-content/uploads/2010/08/propuesta_09.pdf)> [Accesed: 18 may 2013].

Walsh, P. 1991. Organizational Memory, Academic of Management Review. 1,pp. 57-91.

Wang, Z., Sun, X., Zhang, D. 2006. An optimal Text categorization algorithm based on SVM. Communications, Circuits and Systems Proceedings,

Wei, O. Yang, J. y Wang 2011. Semi-supervised Multi-label Learning Algorithm using dependency among labels. In IPCSIT vol. 3.

Wei, L., Wei, B., y Wang B. 2012. Text Classification Using Support Vector Machine with Mixture of Kernel. A Journal of Software Engineering and Applications, 2012, 5, 55-58

Wiig, K., Towe B. y Pizziconi V. 1997. Knowledge Management: Where Did It Come From and Where Will It Go?. Expert Systems with Applications, Vol. 13, pp 1.14.

Wiig, K. 1993. Knowledge management foundations: Thinking about thinking. How people and organizations create, represent and use knowledge. Arlington, TX: Schema Press.

Yu, G. y Slotine, J. 2009. Audio classification from time-frequency texture.

Zha, Z. Mie, T., Wang, Z. y Hua, X. 2008. Graph-Based Semi-Supervised Learning with Multi-label. In ICME. page 1321-1324.

Zhu, Z. 2008. An email classification model based on rough set and support vector machine. In the Fifth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 5, pp. 236-40.

## 8. ANEXOS

### Anexo 1

Encuesta para obtener información de los procesos principales

- Nombre del entrevistado:
- Puesto que desempeña en la institución/división/departamento.
- Describa en general como define a su puesto.
- Hay muchos procesos que involucran a su puesto: ¿Cuáles considera Usted que son los procesos principales que desarrollan y que atañen a su puesto?
- De los procesos principales mencionados, ¿Cuáles considera que son los tres más importantes y por qué?
- De favor ¿Clasifique por orden de importancia a los procesos, donde se observe una jerarquía en función del impacto que tiene a las actividades de su puesto?
- De los tres procesos principales mencionados, ¿Describa las actividades, funciones o tareas, que se deben llevar a cabo en cada uno de ellos?
- ¿Las tareas son realizadas por usted mismo? Si no es así, ¿Quién las realiza?
- ¿Existe información documentada en papel relacionada a los procesos principales? Si es así, ¿está a su disposición o necesita solicitar acceso a la misma?
- ¿Existe información almacenada en medios digitales relacionada a los procesos principales? ¿En qué tipo de formato se encuentra almacenada? Si es así, ¿está a su disposición o necesita solicitar acceso a la misma?



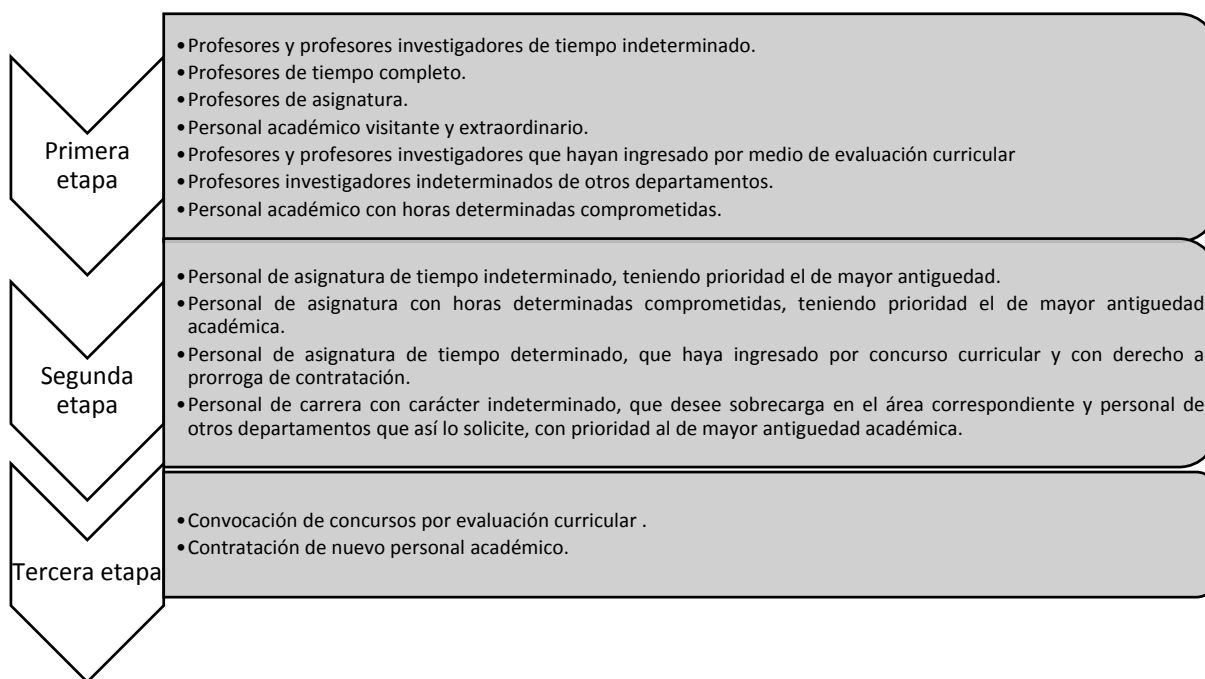
## Anexo 2

### Programación académica semestral y los horarios (Cláusula 82 del CCT)

La programación semestral de la carga académica y horarios de trabajo corresponde ser realizada por los Jefes de Departamento de acuerdo al procedimiento establecido en la presente cláusula, en los términos de la Ley orgánica, del EPA, el CCT y las disposiciones reglamentarias emitidas por el Colegio Académico.

El orden de programación de profesores en la Universidad de Sonora es basada en la categoría que tenga cada profesor y por otro lado la antigüedad del mismo. Por ello, la asignación de grupos es asignada en base a tres posibles etapas.

El orden de programación en cada área será la siguiente:



**Figura 8.1** Orden de programación académica de la Universidad de Sonora. (CCT, 2013-2014)