



**FACULTAD 1**

---

**Subsistema de visualización y clasificación automática de  
noticias Sistema de Análisis y Clasificación Automática de  
Noticias (SACAN)**

Trabajo de diploma para optar por el título de  
**Ingeniero en Ciencias Informáticas**

**Autor:**

Ramón García Castell

**Tutores:**

MSc. Sahilyn Delgado Pimentel

Ing. Nolberto Rojas Pérez

---

La Habana, junio de 2018

## **Declaración de autoría**

Declaro ser autor de la presente tesis y ofrezco a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma. Para que así conste firmo la presente a los \_\_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_.

---

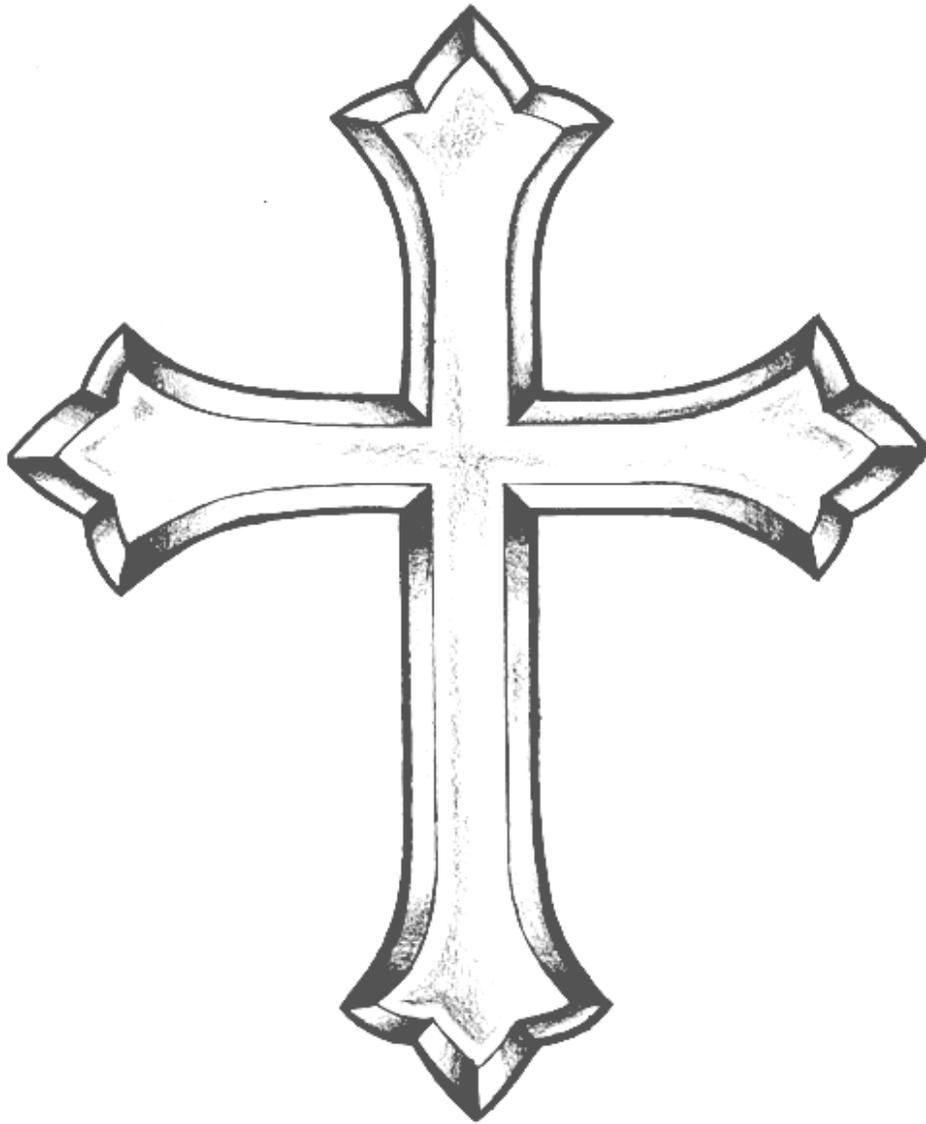
Firma del autor  
Ramón García Castell

---

Firma del tutor  
MSc. Sahilyn Delgado Pimentel

---

Firma del tutor  
Ing. Nolberto Rojas Pérez



*“Mas buscad primeramente el reino de Dios y su justicia,  
y todas estas cosas os serán añadidas.”*

*Jesucristo*

## *Agradecimientos:*

*A Dios primeramente por tomarme de la mano y guiarme durante este largo y duro proceso. A mi amada madre Ana Moraima Castell Pin que se ha sacrificado en gran manera, a mi papá Ramón García Rabasa, a mi padrastro Rafelito, a mi hermana Marínés y su novio Ariel. A mi tía Ilda, Mayte, a mis primos Daniel y Javier. A mi querida abuela Nelís, a Mina y Magaly. A mi hermano Reynol por su apoyo incondicional. A mi hermana Iviana y su esposo Staysys.*

*A mis tutores Sahilyn Delgado Pimentel y Nolberto Rojas Pérez que me ayudaron mucho y fueron pieza importante para lograr este objetivo.*

*A mi oponente Yuneldís Reyes Velázquez por su trabajo.*

*A mis hermanos en Cristo: Yoandri, Daniel, Víctor, Alejandro, Andy, Dargel, Yúnior y al resto de toda la familia que tengo aquí en la UCI. También a todos los hermanos en mi iglesia, Rocío, Danay, Andy, Ricardito, Belena, Rebeca, Amanda, Lester, María y todos los hermanos de la iglesia de Almendares.*

*A Yoan Antonio López Rodríguez, Alexander García O'reilly y Miguel Angel Chavez Alonso. Gracias por su valiosa ayuda.*

*A mis compañeros de edificio: Osvaldo, Vladimir, Raúl, Mauricio, Jorge A., Juan Carlos, Manuel, Michel, Carlitos, Alejandro, Diwel, Alfred y David.*

*A mis compañeros de aula y de todo el año en especial a: Lázaro, Sandra, Elena, Lisbet, Ana, David, Cesar, Evelyn, Heikel, Esteban, Beatriz, Harry y Yúnior.*

*A todos los profesores que me impartieron clases durante los 5 cursos, en especial a Walber Manega, Alejandro M. Castellini, Yanet Salazar, María C. Hidalgo, Gendry Alfonso, Graciela González, Vismar Fernández, Israel Portuondo, Aneyty Martin, Alexander Rodríguez M., Dailiany Soler, Julio J. García Coste, José C. Pérez Zamora y Briseis A. Godínez.*

*A Yailen, Anietsy y Mariana por ayudarme en distintos momentos de mi carrera.*

*A todos aquellos que contribuyeron a mi formación profesional.*

*A mi amada Alíaly.*

*A la Revolución, a Fidel y a la UCI por darme la oportunidad de realizar este sueño.*

*Y los que me faltaron de todo corazón: ¡Muchas Gracias!*

***Dedicatoria:***

*Dedico el presente Trabajo de Diploma:*

*A Dios por sostenerme en los momentos más difíciles y mi madre que ha dado su vida  
para que cumpla este sueño.*

## Resumen

El presente trabajo consiste en el desarrollo de un subsistema de visualización y clasificación automática de noticias. Se estudiaron varios sistemas existentes que visualizan y clasifican noticias. Se analizaron varias tecnologías de las reportadas en la bibliografía y se seleccionó las ontologías como vía para la clasificación automática. Para la implementación se utilizó Drupal en su versión 7.59, y como guía del proceso la metodología de desarrollo de software AUP-UCI. Se emplearon como lenguajes de programación PHP 5, HTML 5 y CSS 3. La solución cuenta MySQL versión 5 como gestor de bases de datos. El modelado se realizó con la herramienta Visual Paradigm 8, el lenguaje de modelado fue UML y en las pruebas de seguridad se empleó la herramienta Acunetix. El subsistema permite gestionar las noticias que han sido previamente almacenadas en la base de datos y las clasifica en dependencia del área geográfica de incidencia. Las pruebas de software arrojaron que el módulo es funcional y estable.

**Palabras clave:** clasificación automática, Drupal, subsistema, visualización.

**Capítulo 1: Visualización y clasificación de información. .... 5**

1.1 La recuperación de información..... 5

1.2 Tecnologías de procesamiento de información..... 5

1.3 Tecnologías para la visualización de información ..... 7

1.4 Estudio de los sistemas existentes ..... 8

Conclusiones del estudio de homólogos..... 8

1.5 Metodología de desarrollo de software ..... 9

1.6 Herramientas y tecnologías ..... 9

Conclusiones parciales..... 12

**Capítulo 2: Análisis y diseño del subsistema de visualización y clasificación automática de noticias.  
..... 13**

2.1 Modelo conceptual ..... 13

2.2 Descripción de la propuesta de solución ..... 13

2.3 Requisitos funcionales del sistema ..... 15

2.4 Casos de uso ..... 18

2.5 Arquitectura de Drupal..... 21

    2.5.1 Patrones de diseño en Drupal..... 22

2.6 Diagrama de datos ..... 24

2.8 Modelo de despliegue ..... 25

Conclusiones parciales..... 26

**Capítulo 3: Implementación y prueba del subsistema de visualización y clasificación de noticias  
..... 27**

3.1 Diagrama de componentes..... 27

3.2 Estándares de codificación..... 28

3.3 Validación de la propuesta de solución.....	32
3.3.1 Integración.....	33
3.3.2 Pruebas de seguridad.....	34
Resultados de las pruebas de seguridad.....	34
3.4 Interfaces de la solución.....	35
Conclusiones parciales.....	36
<b>Conclusiones.....</b>	<b>37</b>
<b>Recomendaciones.....</b>	<b>38</b>
<b>Bibliografía.....</b>	<b>39</b>
<b>Anexos.....</b>	<b>42</b>
Anexo 1: Especificación de los requisitos funcionales.....	<b>¡Error! Marcador no definido.</b>
Anexo 2: Descripción de los casos de uso.....	<b>¡Error! Marcador no definido.</b>
Anexo 3: Interfaz de usuario.....	53
<b>Figura 1. Modelo de dominio. Elaboración propia.....</b>	<b>13</b>
<b>Figura 2. Representación gráfica de la ontología utilizada. Elaboración propia. ....</b>	<b>14</b>
<b>Figura 3: Diagrama de casos de uso. Elaboración propia. ....</b>	<b>19</b>
<b>Figura 4. Arquitectura de Drupal 7. Estructura en 5 capas (Vandyk, 2011). ....</b>	<b>22</b>
<b>Figura 5. Diagrama de clases de diseño del CMS Drupal. ....</b>	<b>24</b>
<b>Figura 6. Modelo de datos. Elaboración propia. ....</b>	<b>25</b>
<b>Figura 7. Modelo de despliegue. Elaboración propia.....</b>	<b>25</b>
<b>Figura 8. Diagrama de componentes. Elaboración propia.....</b>	<b>27</b>
<b>Figura 9. Resultados de las pruebas funcionales.....</b>	<b>33</b>
<b>Figura 10. Noticias sin revisar.....</b>	<b>35</b>

<b>Figura 11. Generación de reportes.</b> .....	<b>36</b>
<b>Figura 12. Iniciar sesión.</b> .....	<b>53</b>
<b>Figura 13. Noticias revisadas.</b> .....	<b>54</b>
<b>Tabla 1. Especificación de los requisitos funcionales</b> .....	<b>15</b>
<b>Tabla 2. Resumen de la etapa de captura de los requisitos funcionales</b> .....	<b>16</b>
<b>Tabla 3. Casos de uso del sistema</b> .....	<b>18</b>
<b>Tabla 4. Descripción del CU8. Generar boletín.</b> .....	<b>19</b>
<b>Tabla 5. Descripción del CU10. Clasificar noticia.</b> .....	<b>20</b>
<b>Tabla 5. Descripción de caso de uso 1. Autenticar usuario.</b> .....	<b>42</b>
<b>Tabla 6. Descripción del CU2. Gestionar usuario.</b> .....	<b>43</b>
<b>Tabla 7. Descripción del CU3: Gestionar permisos a usuarios.</b> .....	<b>46</b>
<b>Tabla 8. Descripción del CU4. Gestionar roles de usuario</b> .....	<b>47</b>
<b>Tabla 9. Descripción del CU5. Gestionar cuenta.</b> .....	<b>50</b>
<b>Tabla 10. Descripción del CU8. Generar boletín.</b> .....	<b>51</b>

## Introducción

Internet, el proyecto que comenzó como Arpanet en 1969 es hoy un recurso indispensable para el hombre. Cada año el número de sitios web aumenta, así también las redes sociales disponibles y los usuarios. En el año 2017, dos investigadores de Cumulus Media midieron la cantidad de información que se comparte en la red en tan sólo un minuto. Según el informe se envían: 452.000 tuits, se suben 46.000 fotos a Instagram, se crean 1,8 millones de snaps en Snapchat, se envían más de 15.000 archivos GIF a través de Facebook Messenger, se reproducen 40.000 horas de audio en Spotify, se mandan 156 millones de correos electrónicos y 16 millones de SMS, se pasan 990.000 vistazos de perfiles en Tinder, se reproducen 4,1 millones de vídeos en YouTube y se realizan unos 3,5 millones de búsquedas en Google. Con todo este cúmulo de datos en Internet, encontrar la información que el usuario desea se vuelve una labor difícil. Por lo cual existen en la actualidad tecnologías útiles para la recuperación de información. Entre las que están, los sistemas de recuperación de información (SRI).

Estos sistemas se encargan de la búsqueda de información en documentos electrónicos y cualquier tipo de colección documental digital (BAEZA-YATES, 1999). Los SRI se clasifican dependiendo del modelo. Existen cinco modelos (DOMINICH, 2000): modelos clásicos (el cual incluye al modelo booleano), el de espacio vectorial y el probabilístico, modelos alternativos (estos están basados en la Lógica Difusa), modelos lógicos (los cuales definen la recuperación de información como un proceso), modelos basados en la interactividad (incluyen posibilidades de expansión del alcance de la búsqueda y hacen uso de retroalimentación por la relevancia de los documentos recuperados), modelos basados en la Inteligencia Artificial (los cuales pueden utilizar bases de conocimiento, redes neuronales, algoritmos genéticos y procesamiento del lenguaje natural).

Según el sitio oficial de la Biblioteca perteneciente a la universidad de Alcalá, algunos ejemplos de SRI son: catálogos de biblioteca, bases de datos bibliográficas y buscadores de Internet. Los SRI no se limitan a estos tres grupos. Existen sistemas informáticos que para su correcto funcionamiento dependen de información recopilada de internet, como son los sistemas de clasificación automáticas de noticias.

En la Universidad de las Ciencias Informáticas (UCI), creada en 2002, el principal objetivo es contribuir a la informatización de la sociedad cubana. En ella radica en el Centro de Ideoinformática (CIDI), el cual desarrolló el Sistema de Almacenamiento, Clasificación y Análisis de Noticias (SACAN), el mismo permite entre otras funcionalidades clasificar las noticias y gestionarlas; este sistema es actualmente utilizado por la Oficina de Información del Comité Central del Partido Comunista de Cuba que se encarga de

monitorear, recopilar y analizarlas noticias que son publicadas en la red.

En el sistema actual se han detectado una serie de deficiencias que dificultan la correcta toma de decisiones de los usuarios que lo utilizan. La recopilación de la información actualmente se realiza de forma manual, siendo complejo el trabajo teniendo en cuenta la cantidad de sitios web de noticias que existen en la actualidad y que son explorados por usuarios; de manera similar ocurre con la clasificación de las noticias recolectadas, las cuales deben ser clasificadas e introducidas al sistema de forma manual, aumentando el tiempo de análisis de la noticia e incidiendo en la generación de los reportes. Por otra parte, el modo en que se presentan los mismos puede ser mejorado, en aspectos referentes a la visualización como es el diseño y la arquitectura de información; actualmente esto afecta el uso del sistema, ya que los usuarios necesitan más tiempo al encontrar la información que desean y el diseño inquieta el estado de ánimo de los usuarios al presentar una interfaz. El sistema actual tampoco permite la personalización de los reportes y la exportación de los mismos, a otros formatos diferentes de (.pdf), impidiendo que se ajusten a las necesidades del momento.

Por lo anteriormente planteado se define como **problema de investigación**: ¿Cómo contribuir a la visualización y clasificación automática de noticias del sistema SACAN para apoyar la toma de decisiones?

Se enmarca el **objeto de estudio** en el proceso de visualización y clasificación automática de información. Se selecciona como **campo de acción** en el proceso de visualización y clasificación automática de noticias.

Se plantea como **objetivo general**: Desarrollar un subsistema para la visualización y clasificación automática de noticias del sistema SACAN para contribuir a la correcta toma de decisiones.

Por consiguiente, los **objetivos específicos** son los siguientes:

1. Describir los referentes teóricos fundamentales que sustentan la investigación.
2. Diagnosticar el estado actual de la visualización y clasificación automática de noticias en sistemas de clasificación de noticias.
3. Diseñar las funcionalidades del subsistema para la visualización y clasificación automática de noticias.
4. Implementar las funcionalidades del subsistema para la visualización y clasificación automática de noticias.
5. Validar la propuesta solución.

Se plantea como **Idea a defender** que el desarrollo de un subsistema para la clasificación automática y visualización de noticias permitirá un apoyo a la toma de decisiones a través del sistema SACAN.

Para dar solución a los objetivos específicos y dar cumplimiento a la idea a defender se definen las siguientes **tareas de la investigación**:

1. Realización de un estudio sobre la visualización y clasificación automática de noticias.
2. Selección de las tecnologías, metodología, herramientas y estándares que se necesitan para implementar la propuesta de solución.
3. Definición de los requisitos funcionales y no funcionales de la propuesta de solución.
4. Implementación de la propuesta de solución.
5. Realización de las pruebas de seguridad, funcionales e integración con la documentación respectiva.

Para el desarrollo de la investigación fueron empleados varios **métodos científicos**, entre los que se encuentran los teóricos, empíricos, los cuales guiaron el avance de la misma. Estos son:

**Métodos teóricos:**

Análítico-Sintético: para el estudio de las fuentes bibliográficas y conceptos existentes sobre las herramientas de visualización y clasificación automática de noticias, con el objetivo de poder demostrar la necesidad de la investigación.

Análisis Histórico-Lógico: utilizado para realizar un estudio de cómo se comportan las herramientas de visualización y clasificación automática de noticias, así posibilitar la creación del marco teórico.

Inductivo-Deductivo: para el estudio de las principales iniciativas de visualización y clasificación automática de noticias y los algoritmos utilizados para lograrlo con el objetivo de determinar cuáles son las alternativas viables a incorporar en la presente investigación.

**Métodos empíricos:**

Observación: utilizado para obtener información sobre los sistemas de visualización y clasificación automática de noticias existentes. Observar los metadatos que utilizan, los algoritmos empleados y el tipo de dato con que trabajan.

Análisis documental: se empleó para la consulta de la literatura especializada en las temáticas afines a la investigación.

Modelación: para la representación de la solución propuesta a través de la modelación de diagramas que permiten modelar el proceso a seguir para desarrollar la solución creando una abstracción con el objetivo

de explicar la realidad.

El presente documento se estructura en tres capítulos, además de las secciones de introducción, conclusiones, recomendaciones, bibliografía y anexos.

### **Capítulo 1: Visualización y clasificación de información.**

En este capítulo se abordará los fundamentos teóricos relacionados con la clasificación y presentación de la información en la web. Análisis del procesamiento de información seleccionando las técnicas y herramientas, constituye un referente en el proceso de clasificación y presentación. Las tecnologías, herramientas y metodología de desarrollo de software empleadas.

### **Capítulo 2: Análisis y diseño del subsistema de visualización y clasificación automática.**

Se describe la propuesta de solución. Se especifican los requisitos funcionales y no funcionales. Se realiza el análisis y diseño a través de la transformación de los requisitos en casos de uso del sistema. Se muestran los diferentes diagramas de clases del diseño, diagramas de secuencia y de casos de uso.

### **Capítulo 3: Implementación y prueba del módulo análisis y visualización de datos.**

Se realiza la implementación del subsistema y se especifican las pruebas realizadas al mismo para comprobar que cumple con los requisitos planteados y asegurar la calidad del software, para ello se utilizan: diagramas de diseño, de despliegue, de componentes, así como los diferentes casos de prueba y resultados de los mismos.

## **Capítulo 1: Visualización y clasificación de información.**

En este capítulo se abordarán los fundamentos teóricos relacionados con la visualización y clasificación automática de noticias, realizando una lectura al estado del arte y plasmando las principales características atendidas a los sistemas homólogos. Quedará reflejada la base tecnológica utilizada, así como la metodología seguida en el proceso de desarrollo de software.

### **1.1 La recuperación de información**

Según Lluís Codina, 2000 “La recuperación de información es un procedimiento que consiste en extraer información de un fondo documental en base a comparar las necesidades de información de los usuarios con los documentos existentes en ese fondo”. Este proceso consta de cuatro etapas: rastreo, indexación, procesamiento y presentación. La primera etapa se encarga de explorar de forma automática la web, recopilando páginas y documentos en general que habrán de ser indizados (FIGUEROLA, 2010). El objetivo de la indexación es seleccionar y remarcar todas las palabras importantes que conforman el documento. Dentro de la red, estas palabras son conocidas como metadatos (GRANADOS, 2016).

El procesamiento de datos es una técnica que consiste en la recolección de datos primarios de entrada, que son evaluados y ordenados, para obtener información útil, que luego serán analizados por el usuario final, para que pueda tomar las decisiones o realizar las acciones que estime conveniente. La presentación es la forma de mostrar los resultados de las tres etapas anteriores, de un modo agradable e intuitivo. La presente investigación está enfocada en la etapa de procesamiento y en la presentación o visualización de información.

En el procesamiento de la información intervienen muchos elementos a tener en cuenta como son, tecnologías, métodos, algoritmos y técnicas. Todas estas son utilizadas para satisfacer las necesidades de los usuarios.

### **1.2 Tecnologías de procesamiento de información**

La primera de las tecnologías estudiadas es la **base de datos léxica**. Ejemplo, Wordnet. Esta es una base de datos léxica en línea diseñada para usarse bajo el control del programa. Los sustantivos, verbos, adjetivos y adverbios en inglés están organizados en conjuntos de sinónimos, cada uno representa un concepto lexicalizado. Las relaciones semánticas vinculan los conjuntos de sinónimos (MILLER, 1990). Así como Wordnet es usada en el idioma inglés, también existen base de datos léxicas para otros idiomas. Estos se utilizan principalmente para la desambiguación de palabras, esto significa que en muchas

ocasiones una misma palabra puede tomar varios significados.

También existen otras tecnologías, como son las **ontologías**, que intervienen en el procesamiento de información, muy usadas por los motores de búsqueda. Una ontología es una descripción explícita y formal de conceptos en un dominio de discurso (NOY,2005). Las ontologías son usadas como una base de datos donde se consulta información.

La **minería de datos** es otra técnica utilizada consiste en la extracción implícita de información potencialmente útil, de los datos que están previamente desconocidos (WITTEN, 2016); la minería de datos es una esfera de las ciencias de la computación muy amplia, esta incluye la minería web; la minería Web es la extracción de interesantes patrones potencialmente útiles e información implícita de artefactos o actividades relacionadas con la *World Wide Web* (PAMNANI, 2010). La minería web utiliza herramientas de la minería de datos para extraer información tanto del contenido de las páginas, de su estructura de relaciones (enlaces) y de los registros de navegación de los usuarios.

Otro elemento que se puede implementar en la etapa de procesamiento es la clasificación de la información recolectada. Esta se puede llevar a cabo con tecnologías de la inteligencia artificial.

En las últimas décadas las **Redes Neuronales Artificiales** (ANN) han recibido un interés particular como una tecnología para minería de datos, puesto que ofrece los medios para modelar de manera efectiva y eficiente problemas grandes y complejos. Los modelos de ANN son dirigidos a partir de los datos, es decir, son capaces de encontrar relaciones (patrones) de forma inductiva por medio de los algoritmos de aprendizaje basado en los datos existentes más que requerir la ayuda de un modelador para especificar la forma funcional y sus interacciones (SALAS,2004).

Las ANN son un método de resolver problemas, de forma individual o combinadas con otros métodos, para aquellas tareas de clasificación, identificación, diagnóstico, optimización o predicción en las que el balance datos/conocimiento se inclina hacia los datos y donde, adicionalmente, puede haber la necesidad de aprendizaje en tiempo de ejecución y de cierta tolerancia a fallos. En estos casos las RNAs se adaptan dinámicamente reajustando constantemente los “pesos” de sus interconexiones (SALAS,2004).

Otra de las técnicas para la clasificación es el razonamiento **basado en casos**. Este es el proceso de solucionar nuevos problemas basándose en las soluciones de problemas anteriores (AAMODT,1994).

El método de los **k vecinos más cercanos** es un método de clasificación supervisada (Aprendizaje, estimación basada en un conjunto de entrenamiento y prototipos) que sirve para estimar la función de densidad de las predictores por cada clase (FIX,1989).

En la etapa de procesamiento existen algoritmos que utilizan buscadores para ordenar los documentos antes de presentarlos al usuario. Uno de los buscadores más utilizados en la actualidad es Google, el mismo utilizaba un algoritmo llamado PageRank, este es una familia de algoritmos utilizados para asignar de forma numérica la relevancia de los documentos (páginas web) indexados por un motor de búsqueda. A cada página web se le asigna un número en función del número de enlaces de otras páginas que la apuntan.

Otro de los algoritmos existentes es el Mozrank, este representa un puntaje de popularidad del enlace, refleja la importancia de cualquier página web dada en Internet. Las páginas ganan MozRank por el número y la calidad de otras páginas que enlazan a ellos, cuanto mayor sea la calidad de los enlaces entrantes, mayor será MozRank. MozRank calcula esta puntuación en una escala logarítmica entre 1 y 10 (GEETHA,2012).

### **1.3 Tecnologías para la visualización de información**

Todo lo planteado anteriormente incide en la etapa de procesamiento de información. La siguiente sección abordará lo referente a la visualización de la información. Algunas de las tecnologías usadas son:

**HTML5:** provee básicamente tres características: estructura, estilo y funcionalidad. Nunca fue declarado oficialmente, pero, incluso cuando algunas APIs (Interface de Programación de Aplicaciones) y la especificación de CSS3 por completo no son parte del mismo, HTML5 es considerado el producto de la combinación de HTML, CSS y Javascript. Estas tecnologías son altamente dependientes y actúan como una sola unidad organizada bajo la especificación de HTML5. HTML está a cargo de la estructura, CSS presenta esa estructura y su contenido en la pantalla y Javascript le brinda funcionalidades(GAUCHAT,2012).

**Bootstrap:** es un marco de trabajo web o conjunto de herramientas de código abierto para diseño de sitios y aplicaciones web. Contiene plantillas de diseño con tipografía, formularios, botones, cuadros, menús de navegación y otros elementos de diseño basado en HTML y CSS, así como, extensiones de JavaScript opcionales adicionales (COCHRAN, 2012).

**CMS** son las siglas de Content Management System, que se traduce directamente al español como Sistema Gestor de Contenidos. Como su propio nombre indica, es un sistema que permite gestionar contenidos. En líneas generales, un CMS permite administrar contenidos en un medio digital y gestionar los contenidos de una web. El desarrollo desde un CMS es más rápido. Así se optimiza el factor tiempo, algo importante a tener en cuenta para el desarrollo. Permiten desarrollos escalables. Esto

significa se puede orientar el desarrollo desde un CMS para añadir nuevas funcionalidades no previstas en el futuro sin que la web sufra cambios traumáticos mediante la incorporación de módulos (GUTIÉRREZ, 2006).

#### **1.4 Estudio de los sistemas existentes**

Por lo tanto, en la siguiente unidad hablaremos de sistemas que realizan el procesamiento o la visualización de la información:

##### **Sacan**

Este sistema fue desarrollado por la Universidad de Ciencias Informáticas. Es un Sistema de Análisis y Clasificación de Noticias; fue desarrollado utilizando el marco de desarrollo Symfony, brinda facilidades para la recopilación de noticias, ofrece estadísticas sobre las noticias compiladas, realiza la mayoría de los procesos de manera manual, no posee mecanismo de rastreo, ni indexación, ni procesamiento de información y la forma de presentar la información es obsoleta.

##### **Sistema de Clasificación de noticias en línea del portal de noticias Indian**

El sistema clasifica el texto de entrada o el documento en clases predefinidas como noticias de entretenimiento, políticas, de negocios, relacionadas con la salud, entre otras. Estos documentos no estructurados, después del proceso de clasificación, son convertidos en información estructurada como una tabla. El sistema utiliza las redes neuronales para darle solución a la clasificación automática de documentos, con una tasa de error de promedio de 5.298 y la precisión promedio de 94.702. Utiliza MATLAB como entorno de desarrollo.

##### **Sistema de clasificación de títulos de noticias**

El marco probabilístico es el utilizado para clasificar los titulares de las noticias. El proceso de clasificación de los titulares de noticias se divide en tres módulos, módulo de pre procesamiento, módulo de aprendizaje de probabilidad y titulares de noticias módulo de clasificación. Según la disponibilidad de variedad de titulares se clasifica cada título de noticias a su predefinida categoría calculando su probabilidad máxima en esa categoría. El trabajo ha sido realizado utilizando el enfoque de bolsa de palabras donde cada titular se divide en palabras y cada palabra tiene una cierta probabilidad. El sistema tiene una precisión promedio de 86.66.

##### **Conclusiones del estudio de homólogos**

Después de realizado el estudio de homólogos se puede concluir que:

1. Ninguno de los sistemas estudiados da solución al problema de investigación.

2. La clasificación automática tiene un mayor grado de efectividad con el uso de tecnologías referentes a la Inteligencia Artificial.

### **1.5 Metodología de desarrollo de software**

Una metodología es un conjunto de procedimientos, técnicas, herramientas y un soporte documental que ayuda a los desarrolladores a realizar un nuevo software. Puede seguir uno o varios modelos de ciclo de vida, es decir, el ciclo de vida indica qué es lo que hay que obtener a lo largo del desarrollo del proyecto, pero no cómo hacerlo (MÉNDEZ, 2010).

Para el desarrollo de la solución se usará AUP en variación de la metodología para la Universidad (AUP-UCI), por ser la metodología empleada en el centro CIDI para el desarrollo de proyectos.

#### **AUP-UCI**

Es una variación de la metodología de desarrollo AUP diseñada por la UCI. Fue creada con el objetivo de adaptarse al ciclo de vida definido para la actividad productiva de la Universidad. Describe de una manera simple y fácil de entender, la forma de desarrollar aplicaciones de software de negocio usando técnicas ágiles, esta define tres fases para componer el ciclo de vida de un proyecto:

- Inicio: se mantiene de AUP, pero se modifica el objetivo; se llevan a cabo las actividades relacionadas con la planeación del proyecto.
- Ejecución: se unifican las tres restantes fases que propone AUP (Elaboración, Construcción y Transición). Se ejecutan las actividades requeridas para obtener el software.
- Cierre: esta fase se incorpora a la metodología en relación a AUP. Se analizan los resultados del proyecto, su ejecución y se llevan a cabo actividades formales de cierre del proyecto.

### **1.6 Herramientas y tecnologías**

En la actualidad existen un conjunto de tecnologías potentes para el desarrollo de aplicaciones web como los sistemas de gestión de contenido, mencionados anteriormente. Estos permiten crear y mantener el sistema con mayor facilidad pues con ellos se pueden realizar tareas en un menor espacio de tiempo.

#### **Drupal CMS**

Drupal es un sistema de administración de contenidos que posibilita crear, clasificar y publicar información en un portal web. Distribuido bajo la licencia de GNU *Public License* (GPL), dispone de varias versiones de su núcleo siendo las ramas 7 y 8 de sus últimas liberaciones, presenta una estructura modular que le permite mediante un amplio repositorio de módulos añadirle nuevas funcionalidades.

Presenta un gran catálogo de módulos y temas gráficos para un sin número de prestaciones gratuitos

de pago. Entre sus beneficios podemos encontrar (Drupal.org, 2018):

- Código abierto: presenta una comunidad de más de un millón de usuarios y desarrolladores.
- Posee herramientas para organizar, estructurar, encontrar y reutilizar el contenido.
- Permite la categorización de contenidos.
- Creación automática de URL amigables.
- Opciones de creación de cuentas de usuarios y roles.
- Maneja gran número de tipos de contenido, como video, texto, *blogs*, *podcasts*, manejo de menú, estadísticas en tiempo real y control de revisiones opcional.
- Posee más de dieciséis mil módulos disponibles.
- Multiplataforma.
- Internacionalización.
- Soporta varios gestores de bases de datos (SQLite, PostgreSQL, MySQL, MariaDB o equivalente).
- Gran variedad de temas.

### **Lenguaje Hypertext Pre-processor 5 (PHP)**

Lenguaje de código abierto adecuado para el desarrollo web que puede ser incrustado en el HTML y permite la creación de páginas web dinámicas. Entre las características por las cuales se selecciona este lenguaje se encuentran:

- Drupal está implementado en PHP.
- Multiplataforma.
- Presenta soporte para bases de datos como: MySQL, PostgreSQL, Oracle, ODBC, DB2, Microsoft SQL Server, Firebird y SQLite.

### **Lenguaje HyperText Markup Language 5 (HTML)**

Profundizando en las características de HTML anteriormente mencionadas, algunas de estas son:

- Estructura del cuerpo: permite agrupar todas estas partes de una web en nuevas etiquetas que representarán cada una de las partes típicas de una página.
- Etiquetas para contenido específico: utiliza etiquetas específicas para cada tipo de contenido en particular, como audio, video, etcétera.
- Bases de datos locales: permite el uso de una base de datos local, con la que se puede trabajar en una página web por medio del cliente y a través de un API.

- Aplicaciones web *offline*: permite la creación de aplicaciones web que funcionen sin necesidad de estar conectados a internet.
- Geolocalización: permite localizar geográficamente las páginas web por medio de un API.

### **Lenguaje Cascading Style Sheets 3 (CSS)**

Es un lenguaje que describe la presentación de los documentos estructurados en hojas de estilo utilizado para especificar el aspecto de una página web. Se basa en reglas que rigen el comportamiento del estilo de los elementos (MUÑOZ VALLEJO, 2015).

### **Sistemas gestores de bases de datos (SGBD)**

Los SGBD se definen como el conjunto de programas que administran y gestionan la información contenida en una base de datos (ALVAREZ, 2007).

#### **MySQL 5**

Sistema de gestión de bases de datos ampliamente utilizado en aplicaciones web, como Drupal. Funciona sobre plataformas como: AIX, BSD, FreeBSD, HP-UX, GNU/Linux, Mac OS X, Windows, etcétera. Permite (Mysql.org, 2018):

- Amplio subconjunto del lenguaje SQL.
- Disponibilidad en gran cantidad de plataformas y sistemas.
- Posibilidad de selección de mecanismos de almacenamiento que ofrecen diferentes velocidades de operación, soporte físico, capacidad y distribución geográfica.
- Transacciones y claves foráneas.
- Conectividad segura.
- Replicación.
- Búsqueda de indexación de campos de texto.

### **NetBeans 8.2**

Es un entorno de desarrollo gratuito y de código abierto. Permite el uso de un amplio rango de tecnologías de desarrollo tanto para escritorio, como para aplicaciones web, o para dispositivos móviles. Da soporte a las siguientes tecnologías, entre otras: Java, PHP, Groovy, C/C++ y HTML5. Además, posee una muy importante característica, es multiplataforma (GOOVY, 2014).

### **Lenguaje de modelado UML**

Lenguaje unificado de modelado del inglés *Unified Modeling Language* (UML) es el más conocido y utilizado actualmente. Se utiliza para visualizar, especificar, construir y documentar los artefactos de un software.

UML utiliza diagramas y una semántica bien definida para elaborar los artefactos del sistema en las distintas etapas de su ciclo de vida, fundamentalmente en el análisis y diseño (STEVENS, 2007). Brinda el lenguaje de modelado para:

- Modelo de proceso de negocios con casos de uso.
- Modelado de clases con objetos.
- Modelado de componentes.
- Modelado de distribución y despliegue.

El lenguaje ayuda al usuario a comprender bien el funcionamiento del software y reflexionar antes de invertir grandes cantidades de dinero en un proyecto que no le sea viable. El modelado ayuda además a mejorar la capacidad del equipo para gestionar la complejidad del software.

### **Visual Paradigm 8.0**

Herramienta que soporta el ciclo de vida completo de desarrollo de software. Soporta el lenguaje unificado de modelado (UML). Sencillo de utilizar, fácil de instalar y actualizar. Permite la generación de código a partir de diagramas para varios lenguajes como .Net, Java, PHP. Posibilita la representación gráfica de diagramas como: componentes, despliegue, secuencia, casos de uso; clase, actividad, estado, entre otros. Se integra con diversos sistemas gestores de bases de datos. Además, permite ver las relaciones entre los componentes del diseño y mejora la comunicación entre los miembros del equipo usando el lenguaje gráfico (LARMAN, 1999).

### **Conclusiones parciales**

Como resultado de la investigación realizada en este capítulo se puede concluir que:

- El estudio de los referentes teóricos permitió entender el problema, el ámbito de la investigación, los conceptos asociados y profundizar en las tecnologías disponibles en la actualidad para solucionar el problema planteado.
- El estudio de los sistemas homólogos permitió identificar la tecnología más adecuada para la solución del problema planteado.
- La selección de la metodología, tecnologías y herramientas libres seleccionadas, posibilitó definir una base tecnológica que permitirá un adecuado desarrollo de la solución.

## Capítulo 2: Análisis y diseño del subsistema de visualización y clasificación automática de noticias.

En el presente capítulo se caracteriza la propuesta de solución, a partir de su funcionamiento, teniendo en cuenta los estilos arquitectónicos y los patrones de diseños aplicables al sistema de gestión de contenidos seleccionado: Drupal. Se muestra a través del modelo de dominio, los principales objetos que se manejan en el contexto del sistema y se identifican los requisitos funcionales y no funcionales que deben estar presentes en la propuesta de solución.

### 2.1 Modelo conceptual

Un **modelo del dominio o modelo conceptual** es una representación visual de las clases conceptuales u objetos del mundo real en un dominio de interés. (FOWLER,1997)

A continuación, se muestra el modelo conceptual del negocio.

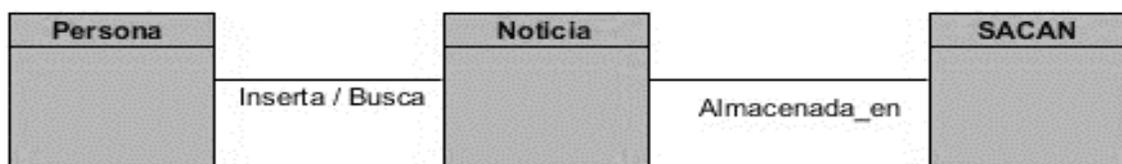


Figura 1. Modelo de dominio. Elaboración propia.

Especificación de los elementos que intervienen en el modelo conceptual:

**Persona:** Objeto que interactúa con el sistema y con la noticia.

**Noticia:** Informe que se da acerca de un hecho o un suceso reciente.

**SACAN:** Sistema informático creado para gestionar las noticias.

### 2.2 Descripción de la propuesta de solución

El subsistema de visualización y clasificación automáticas de noticias depende del subsistema de rastreo, indexación y procesamiento. Este subsistema se encarga de limpiar del texto las palabras que carecen de concepto como por ejemplo las preposiciones, este proceso se llama procesamiento de lenguaje natural, cada noticia a clasificar transita por este proceso y el resultado es utilizado por el subsistema de visualización y clasificación automáticas de noticias.

El subsistema desarrollado está enfocado en dos grandes aristas. La primera es la clasificación automática de noticias, para lograr una clasificación efectiva existen tecnologías relacionadas con la inteligencia

artificial, estas se mencionaron en el capítulo anterior; la utilizada para la implementación de la propuesta de solución fueron las ontologías, las mismas son empleadas para representar el conocimiento. La clasificación que se lleva a cabo es en dependencia del área geográfica, para esto se creó una ontología que representa el conocimiento geográfico que los usuarios finales utilizan en la clasificación de noticias.

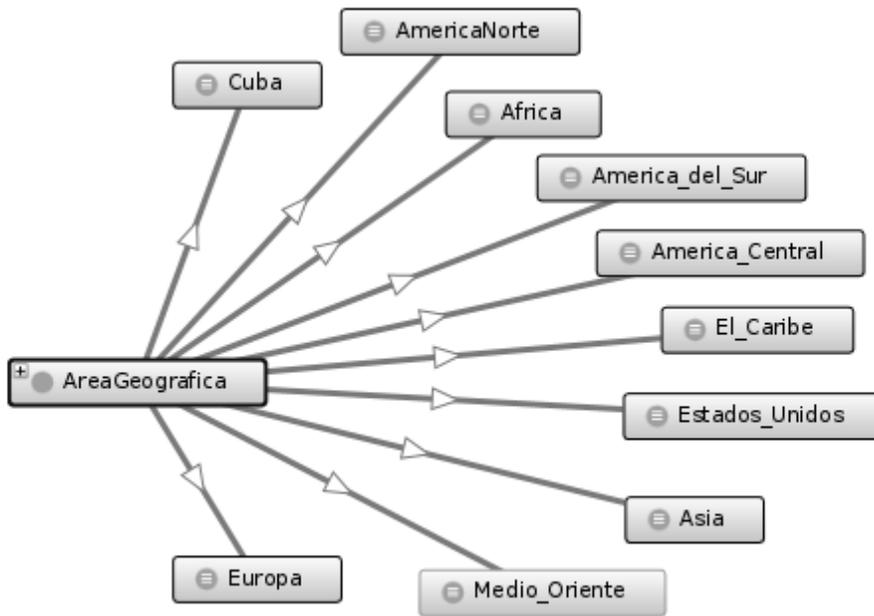


Figura 2. Representación gráfica de la ontología utilizada. Elaboración propia.

Con esta forma de representar el conocimiento el sistema es capaz de diferir a que área geográfica pertenece cada noticia.

La segunda esfera es la visualización de las noticias, esta brinda facilidades al usuario para revisar las clasificaciones realizadas por el subsistema, también brinda una interfaz intuitiva y mejora aspectos de la arquitectura de información.

### 2.3 Especificación de requisitos de software

Los requerimientos para un sistema son la descripción de los servicios proporcionados por el sistema y sus restricciones operativas. Estos requerimientos reflejan las necesidades de los clientes de un sistema que ayude a resolver algún problema. (PRESSMAN, 2011)

Se recomienda, establecer diferentes niveles de especificación del sistema, los que deben estar en correspondencia con los lectores a los que va dirigido para una mayor comprensión. En correspondencia a lo antes planteado, en la presente investigación se tratarán en la descripción de requisitos solo dos niveles, requisitos funcionales y no funcionales, detallados a continuación.

### 2.3 Requisitos funcionales del sistema

Los requisitos funcionales son declaraciones de los servicios que debe proporcionar el sistema, de la manera en que éste debe reaccionar a entradas particulares y de cómo se debe comportar en situaciones particulares. En algunos casos, los requerimientos funcionales de los sistemas también pueden declarar explícitamente lo que el sistema no debe hacer. Estos requerimientos dependen del tipo de software que se desarrolle de los posibles usuarios del sistema y del enfoque general tomado por la organización al redactar requerimientos (PRESSMAN, 2011).

En el proceso de levantamiento de requisitos para el subsistema de visualización y clasificación automáticas de noticias del sistema SACAN, fueron detectados utilizando las técnicas de entrevista y observación un total de 27 requisitos funcionales.

Tabla 1. Especificación de los requisitos funcionales

Nº	Nombre	Descripción	Prioridad	Complejidad
RF1	Autenticar usuario	El sistema debe permitir autenticar el usuario antes de acceder al sistema.	Alta	Baja
RF2	Añadir usuario	El sistema debe permitir añadir un usuario.	Alta	Baja
RF3	Modificar usuario	El sistema debe permitir editar los datos del usuario.	Alta	Baja
RF4	Mostrar usuario	El sistema debe permitir mostrar los datos de un usuario.	Alta	Baja
RF5	Listar usuario	El sistema debe permitir listar los usuarios.	Alta	Baja
RF6	Eliminar usuario	El sistema debe permitir eliminar un usuario.	Baja	Baja
RF7	Asignar permisos a usuarios	El sistema debe permitir Asignar permisos a usuarios.	Alta	Baja
RF8	Listar permisos a usuarios	El sistema debe permitir mostrar una lista de los posibles permisos.	Media	Baja

RF9	Crear roles para los usuarios	El sistema debe permitir crear rol para los usuarios.	Media	Baja
RF10	Eliminar roles para los usuarios	El sistema debe permitir eliminar un rol de usuario.	Media	Baja
RF11	Listar los roles de usuarios	El sistema debe permitir listar los roles de los usuarios.	Media	Baja
RF12	Modificar rol de usuario	El sistema debe permitir editar los datos de un rol de usuario.	Media	Baja
RF13	Mostrar cuenta	El sistema debe permitir mostrar los datos del usuario autenticado.	Baja	Baja
RF14	Modificar cuenta	El sistema debe permitir a un usuario modificar los datos de su cuenta.	Baja	Baja
RF15	Enviar mensaje a usuario	El sistema debe permitir enviar mensajes a un usuario.	Baja	Media
RF16	Mostrar mensaje recibido	El sistema debe permitir mostrar el mensaje recibido	Baja	Media
RF17	Eliminar mensaje recibido	El sistema debe permitir eliminar mensaje recibido.	Baja	Media
RF18	Buscar noticia	El sistema debe permitir buscar noticias.	Alta	Media
RF19	Modificar noticia	El sistema debe permitir modificar los datos de la noticia.	Media	Media
RF20	Listar noticia	El sistema debe permitir listar las noticias.	Alta	Media
RF21	Eliminar noticia	El sistema debe permitir eliminar las noticias.	Alta	Media
RF22	Mostrar noticia	El sistema debe permitir mostrar los datos de la noticia.	Alta	Media
RF23	Clasificar noticia	El sistema debe brindar una clasificación a la noticia.	Media	Alta
RF24	Seleccionar noticias para boletín	El sistema de debe permitir seleccionar las noticias para boletín.	Media	Media
RF25	Generar boletín	El sistema debe generar un documento .pdf con las noticias seleccionadas.	Alta	Media
RF26	Mostrar filtros para reporte	El sistema debe mostrar filtros para realizar un reporte	Media	Media
RF27	Generar reporte	El sistema de debe generar tablas en dependencia del filtro seleccionado.	Media	Media

A continuación, se realiza un resumen de los requisitos funcionales:

Tabla 2. Resumen de la etapa de captura de los requisitos funcionales

Prioridad para el cliente	Cantidad de requisitos funcionales
<b>Alta</b>	11

<b>Media</b>	10
<b>Baja</b>	6
<b>Total</b>	27

### 2.3.2 Requisitos no funcionales

Los requisitos no funcionales son restricciones de los servicios o funciones ofrecidos por el sistema. Incluyen restricciones de tiempo, sobre el proceso de desarrollo y estándares. Los requerimientos no funcionales a menudo se aplican al sistema en su totalidad y surgen de las necesidades del usuario, debido a las restricciones en el presupuesto, a las políticas de la organización, a la necesidad de interoperabilidad con otros sistemas software o hardware, o a factores externos como regulaciones de seguridad o legislaciones sobre privacidad (PRESSMAN, 2011).

#### Apariencia o Interfaz externa

**RNF 1:** La interfaz gráfica de la aplicación debe concebirse con un ambiente amigable, sencillo y de navegación fácil para el usuario.

##### ➤ Seguridad

**RNF 3:** En caso de que el sistema presente alguna falla, los errores deben mostrar la menor cantidad de detalles posible, de forma tal, que se evite dar información que comprometa la seguridad e integridad del sistema. Sólo se mostrarán detalles ampliados del error a usuarios con privilegios de administración.

**RNF 4:** Se garantizará la integridad de la información mediante mecanismos de control de acceso, utilizando, usuario, contraseña y niveles de accesos para cada usuario, de manera que cada uno pueda tener disponible solamente las opciones que se encuentran en correspondencia con su actividad.

##### ➤ Requisitos de licencia

**RNF 5:** Se requiere el uso de licencia GNU/GPL versión 2 para el CMS Drupal.

##### ➤ Hardware

**RNF 6.** El ordenador donde se ejecute el módulo debe tener una memoria *RAM* de 2GB o superior y un procesador a una velocidad 2.10 GHz o superior.

##### ➤ Software

**RNF 7.** El ordenador donde se ejecute el módulo debe tener instalada la distribución de *GNU/Linux Ubuntu* 14.04 o superior.

**RNF 8.** Se requiere un servidor de base de datos *MySQL*.

## 2.4 Casos de uso

Un modelo de caso de uso es una descripción de los pasos o las actividades que deberán realizarse para llevar a cabo algún proceso (JACOBSON, 1992).

A continuación, se definen los casos de uso que se corresponden con los requisitos identificados, los cuales serán una de las bases durante las fases de análisis, diseño e implementación del subsistema.

Tabla 3. Casos de uso del sistema

Casos de Uso(CU)	Referencias
<b>CU1: Autenticar usuario</b>	RF1
<b>CU2: Gestionar usuario</b>	RF2, RF3, RF4, RF5, RF6
<b>CU3: Gestionar permisos a usuarios</b>	RF7, RF8
<b>CU4: Gestionar roles de usuarios</b>	RF9, RF10, RF11, RF12
<b>CU5: Gestionar cuenta</b>	RF13, RF14
<b>CU6: Gestionar mensajes</b>	RF15, RF16, RF17
<b>CU7: Gestionar noticia</b>	RF18, RF19, RF20, RF21, RF22
<b>CU8: Generar boletín</b>	RF24, RF25
<b>CU9: Generar reporte</b>	RF26, RF27
<b>CU10: Clasificar noticia</b>	RF23

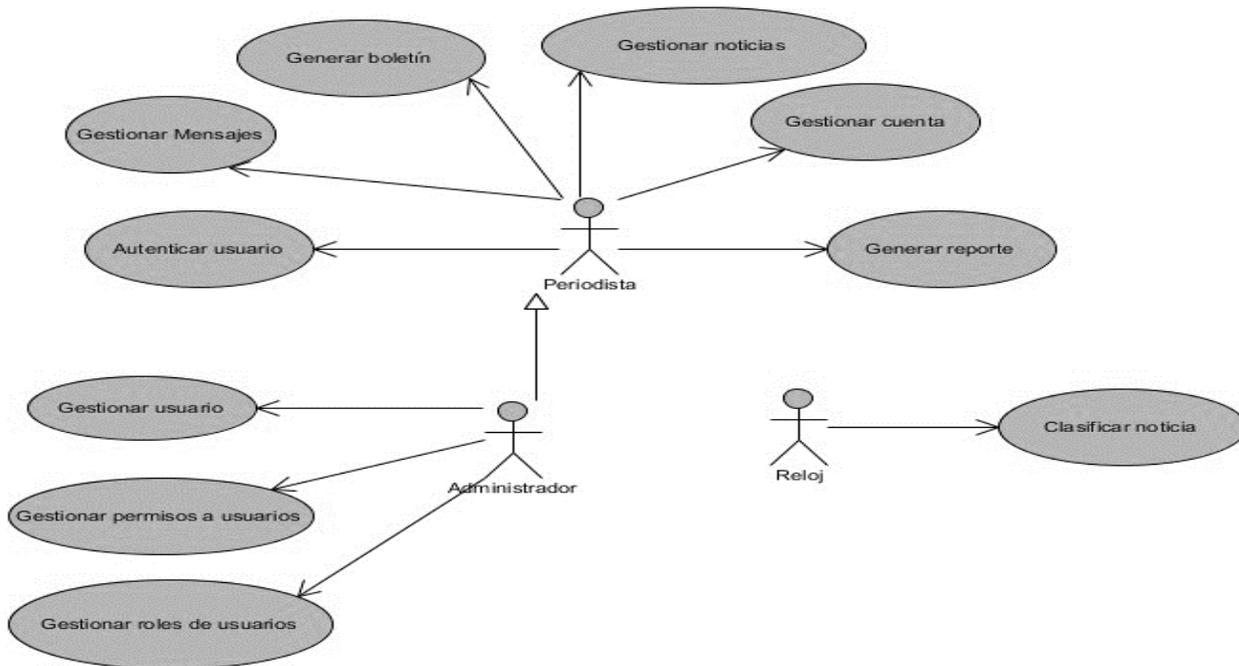


Figura 3: Diagrama de casos de uso. Elaboración propia.

El objetivo fundamental de la especificación de los casos de uso es un mejor entendimiento de las funcionalidades que abarcan. A continuación, se describen dos de ellos:

Tabla 4. Descripción del CU8. Generar boletín.

<b>Objetivo</b>	Generar un documento pdf, con las noticias seleccionadas
<b>Actores</b>	Usuario básico
<b>Resumen</b>	Este CU describe cómo se genera un pdf
<b>Complejidad</b>	Media
<b>Prioridad</b>	Alta
<b>Precondiciones</b>	Que el usuario esté autenticado en el sistema.
<b>Postcondiciones</b>	
<b>Flujo de eventos</b>	

<b>Flujo básico &lt;Generar Boletín&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Se autentica en el sistema.	
2.		Muestra una interfaz con una lista de noticias
3.	El usuario selecciona las noticias que desea	
4.		Genera un documento pdf con las noticias seleccionadas
<b>Flujos alternos</b>		
1. No se puede autenticar usuario		
	<b>Actor</b>	<b>Sistema</b>
1.		Muestra un mensaje de error donde se le comunica al usuario las causas del mismo.

Tabla 5. Descripción del CU10. Clasificar noticia.

<b>Objetivo</b>	Clasificar automáticamente la noticia en dependencia del área geográfica.
<b>Actores</b>	Relej
<b>Resumen</b>	Este CU describe como se toma el cuerpo de cada noticia y se clasifica en dependencia del área geográfica.
<b>Complejidad</b>	Alta
<b>Prioridad</b>	Alta

<b>Precondiciones</b>		
<b>Postcondiciones</b>		
<b>Flujo de eventos</b>		
<b>Flujo básico &lt;Clasificar noticia&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Ejecuta el hook_cron	
2.		Clasifica la noticia.

## 2.5 Arquitectura de Drupal

Drupal posee una arquitectura estructurada en diferentes capas, por lo que para el desarrollo del subsistema se propone el trabajo con la arquitectura n-capas, específicamente 5 capas que son descritas a continuación (Vandyk, 2011):

**Plantillas (*templates*):** Esta capa establece la apariencia gráfica que se le muestra al usuario. Esta separación entre la información y los estilos permite cambiar la apariencia del portal web sin necesidad de modificar los contenidos.

**Vistas (*views*):** Es la capa encargada de mostrar en los temas los cambios realizados a través de los módulos.

**Entidades (*entities*):** Representa las entidades, que engloban los nodos, los usuarios, las taxonomías y los comentarios. Esta nueva estructura permite que sea posible añadirle campos a todo aquello que sea una entidad.

**Módulos (*modules*):** Engloba los elementos que operan sobre los nodos otorgando funcionalidades a Drupal. Permiten incrementar sus capacidades o adaptarlas a las necesidades de cada portal web.

**Base de Datos (*database*):** Esta capa es la encargada de gestionar el acceso a la información almacenada referente al funcionamiento del sistema y a los contenidos que serán mostrados a través del tema activo.

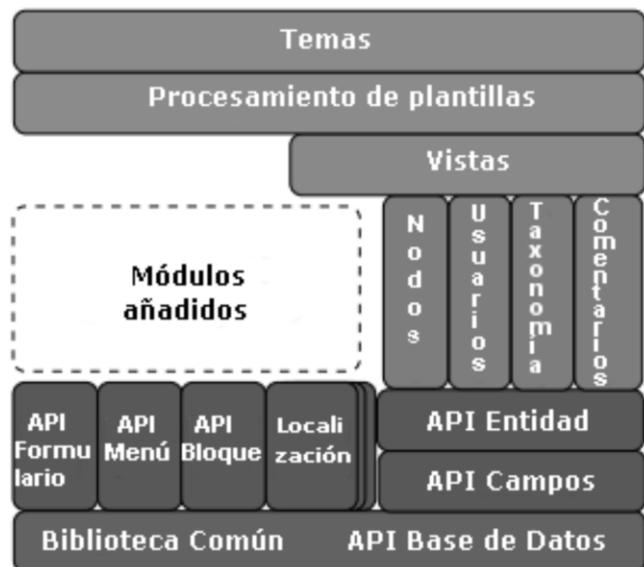


Figura 4. Arquitectura de Drupal 7. Estructura en 5 capas (VANDYK, 2011).

### 2.5.1 Patrones de diseño en Drupal

Drupal hace uso de patrones de diseño como: *singleton* (sencillo o instancia única), *decorator* (decorador), *observer* (observador), *bridge* (puente), *chain of responsibility* (cadena de responsabilidad) y *command* (comando). Estos permiten diseñar sistemas seguros y que a su vez cumplan con los estándares de diseño establecidos por normas internacionales para el desarrollo de aplicaciones web. Los patrones usados para el desarrollo del subsistema se evidencian de la forma siguiente:

#### **Observer (Observador)**

Los módulos que implementan un *hook* de inserción o actualización de una entidad, son declarados como observadores de las entidades con las que interactúan. En el caso del subsistema desarrollado, el patrón se evidencia en la actualización del atributo: “clasificación” de la entidad noticia.

#### **Chain of responsibility (Cadena de responsabilidades)**

El sistema de menú de Drupal es la evidencia del patrón cadena de responsabilidades, en cada petición de una página, el sistema de menú de Drupal determina si hay algún módulo para responder la petición, si el

usuario tiene acceso al recurso solicitado y que función se debe llamar para procesar la petición. En este proceso se trasmite el mensaje de la petición por cada uno de los componentes que se encuentran inmersos.

### **Diagrama de clases de diseño**

Al ser Drupal el sistema de gestión de contenidos utilizado en el desarrollo de la propuesta de solución, es necesario analizar su estructura y funcionamiento, para tener el conocimiento necesario de cómo se confeccionan las clases del diseño web. Drupal depende de un conjunto de clases o módulos que le posibilita al usuario final interactuar con el software, estos se encuentran estructurados de acuerdo a sus funcionalidades. Este CMS se encuentra estructurado mediante 4 paquetes principales: *Themes*, *Includes*, *Modules* y *Scripts*, que a su vez están conformados por diversos archivos. Con el fin de lograr una mejor comprensión de la aplicación, se caracteriza cada uno de estos componentes.

**Themes:** Contiene los temas que vienen con la distribución del CMS. Al igual que los módulos no es recomendable modificar directamente estos temas o subir temas adicionales a esta carpeta. Los módulos adicionales se deben guardar en la carpeta `/sites/all/themes` o `/sites/default/themes`, definida en el archivo de instalación del CMS.

**Includes:** contiene un conjunto de librerías en forma de archivos PHP con extensión `.inc`, que incluyen funciones comunes del sistema, indispensables para su correcto funcionamiento, como son las conexiones a la base de datos.

**Modules:** contiene todos los módulos del núcleo que permiten la integración de disímiles funcionalidades. No es recomendable modificar directamente estos módulos o subir módulos adicionales a esta carpeta. Los módulos adicionales se deben guardar en la carpeta `/sites/all/modules` o `/sites/default/modules`, definida en el archivo de instalación del CMS.

**Scripts:** Contiene utilidades adicionales que no utiliza Drupal directamente, pero que se puede utilizar desde la línea de comandos del Shell.

A continuación, se presenta el diagrama de clase de diseño que utiliza Drupal por defecto:

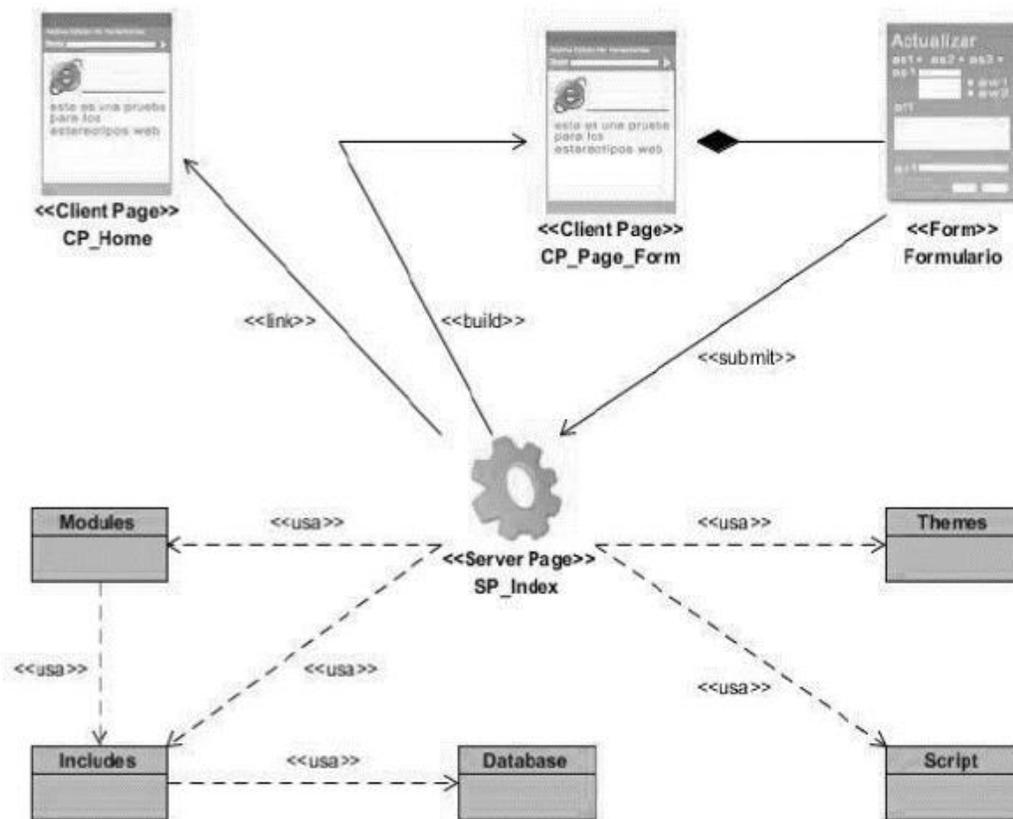


Figura 5. Diagrama de clases de diseño del CMS Drupal.

## 2.6 Diagrama de datos

Un modelo de datos está orientado a representar los elementos que va a procesar el sistema, la composición y atributos de los mismos. Además de dónde se encuentran almacenados actualmente dichos objetos, la relación entre ellos y los procesos que los transforman (PRESSMAN, 2011).

El modelo de datos del subsistema contiene 110 tablas. A continuación, se muestra dos de las tablas más utilizadas en el modelo de datos que se propone para el desarrollo del sistema.

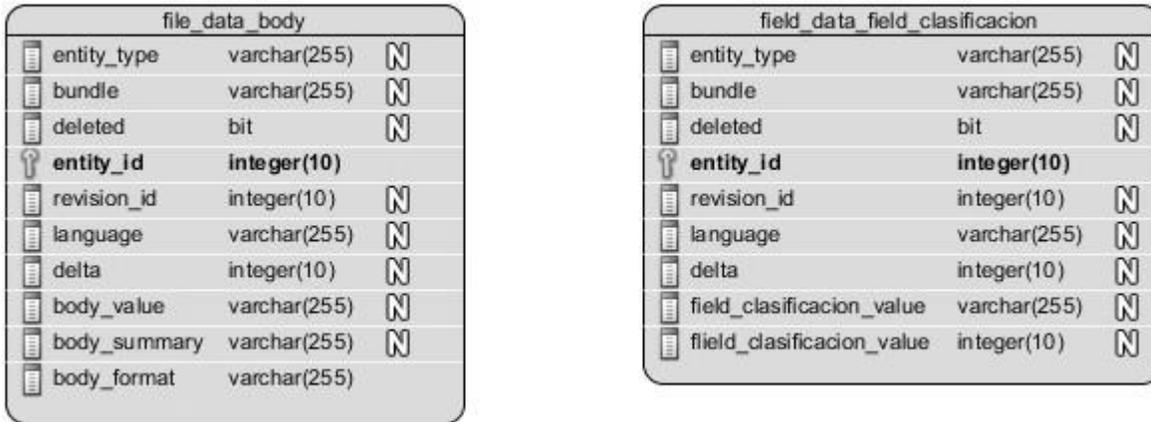


Figura 6. Modelo de datos. Elaboración propia.

## 2.8 Modelo de despliegue

El diagrama de despliegue es uno de los tipos de diagramas que aparecen cuando se modelan los aspectos físicos de los sistemas orientados a objetos. Muestra la configuración de los nodos que participan en la ejecución y de los componentes que residen en ellos. Son utilizados para modelar la vista de despliegue estática de un sistema. La siguiente figura muestra el diagrama de despliegue correspondiente a la propuesta de solución (SPARXSYSTEMS, 2014).

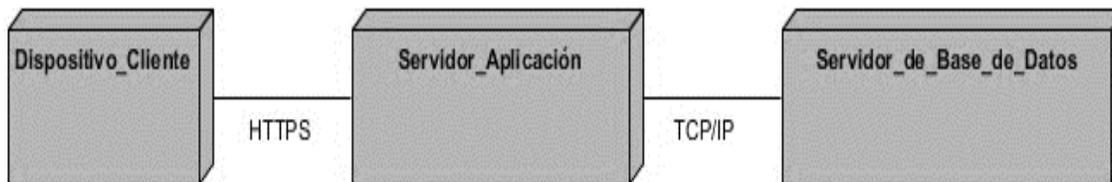


Figura 7. Modelo de despliegue. Elaboración propia.

### Descripción de los elementos de interface y comunicación

**HTTPS:** Protocolo para establecer la conexión segura entre la PC cliente y el servidor de aplicaciones a través del puerto.

**TCP/IP:** Protocolo para establecer la conexión entre el servidor de aplicaciones y el servidor de base de datos a través del puerto definido para el gestor de base de datos. La conexión entre estos servidores permitirá ejecutar un conjunto de órdenes y obtener rápidamente respuesta a las mismas.

### **Conclusiones parciales**

En este capítulo se han abordado elementos del análisis y diseño del subsistema desarrollado arribando a las siguientes conclusiones:

- La elaboración del modelo de dominio y su descripción permitió una mayor comprensión del negocio.
- La identificación de los requisitos funcionales y no funcionales, permitió agrupar las funcionalidades del sistema en casos de usos y conocer las características que debe cumplir la solución.
- La elaboración del diagrama de despliegue permitió identificar la distribución física de los componentes del sistema propuesto.

### Capítulo 3: Implementación y prueba del subsistema de visualización y clasificación de noticias

En el presente capítulo se presentan los componentes y los estándares de codificación que sustentan la implementación del subsistema de visualización y clasificación de noticias. Se describe y fundamenta el proceso de validación de la solución implementada, mediante la utilización de los casos de pruebas.

#### 3.1 Diagrama de componentes

Los diagramas de componentes muestran los elementos de diseño de un sistema de software. Permiten visualizar con mayor facilidad la estructura general del sistema y el comportamiento del servicio que estos componentes proporcionan y utilizan a través de interfaces (Microsoft Developer Network, 2014).

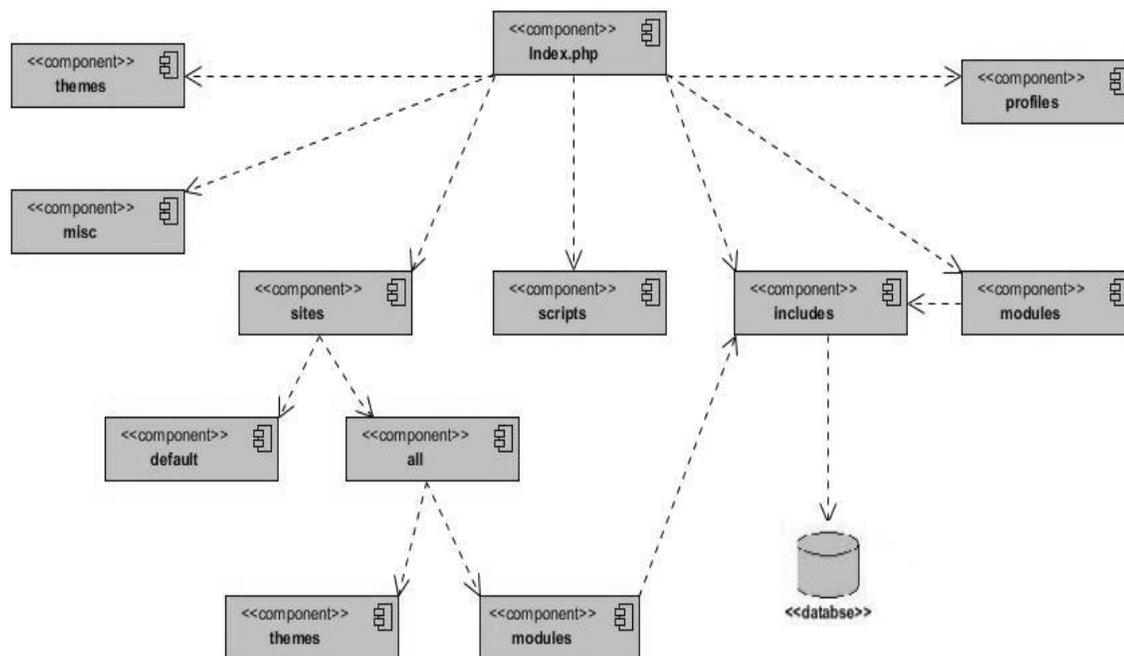


Figura 8. Diagrama de componentes. Elaboración propia.

A continuación, se describen los componentes presentes en el diagrama:

- **Index.php:** Este componente es el punto de inicio de la aplicación. A partir de esta entrada se solicitan los diferentes módulos del sistema de gestión de contenidos.
- **Themes:** En este componente se incluyen los temas que vienen con la distribución de Drupal.

- **Misc:** Incluye los archivos JavaScript e imágenes requeridas por el sistema.
- **Includes:** Contiene los ficheros indispensables para el funcionamiento de Drupal.
- **Script:** Proporciona utilidades adicionales que no emplea Drupal directamente, pero que pueden ser utilizadas a través de las líneas de comandos de Shell.
- **Modules:** En este componente se encuentran los módulos contenidos en el core de Drupal.
- **Profiles:** Contiene los perfiles de instalación de Drupal.
- **Sites:** Se encuentran elementos extras y modificaciones que se añaden al sistema original.

### 3.2 Estándares de codificación

Los estándares de codificación constituyen un factor primordial en el desarrollo de proyectos. Garantizan que el código obtenido sea fácil de leer, entendido y modificado independientemente de quien haya sido el desarrollador del producto. Son una guía para el equipo de desarrollo, permiten asegurar que el código presente calidad y no contenga errores. Drupal proporciona a sus desarrolladores un conjunto de normas para fomentar el código de una forma uniforme para todos (GIL, 2012).

A continuación, se describen los estándares de codificación utilizados en la implementación del subsistema.

#### Identación

La Identación consiste en insertar espacios en blanco o tabuladores en determinadas líneas de código para facilitar su comprensión. En programación se utiliza Identación para anidar elementos. En Drupal se debe indentar con 2 espacios, nunca con tabuladores. Además, no se debe dejar espacios en blanco al final de cada línea. A continuación, se muestra un ejemplo.

```
function exportar_pdf(&$form, &$form_state) {
  $array_id = array();
  foreach($form['views_row_selector'] as $Array => $are
  // echo "<pre>"; print_r($are); echo "</pre>";

  if(isset($are['#checked'])) {
    if ($are['#checked'] == 1) {
      $array_id[] = $are['#return_value'];
      //var_dump($array_id);
    }

  }
}
```

## Etiquetas de apertura y cierre

Cuando se escribe en PHP, siempre se deben utilizar las etiquetas `<?php y ?>`, y en ningún caso la versión corta `<? y ?>`. En general se omite la etiqueta de cierre de PHP (`?>`) al final de los archivos `.module` y `.inc`. Esta convención evita que se puedan quedar olvidados espacios no deseados al final del archivo (después de la etiqueta de cierre `?>`), que serían identificados como salida HTML y podrían provocar un error muy típico, *“Cannot modify header information-headers already sent by ...”*. Por tanto, la etiqueta de cierre final del archivo (`?>`) es opcional en Drupal. A continuación, se muestra un ejemplo de la utilización de las etiquetas apertura y cierre, contenido en el archivo `page.tpl.php`.

```
<div align="justify">
  <?php
    echo "<br>";
    print ($node->body['und']['0']['safe_value']);
  ?>
</div>
```

## Operadores

Los operadores binarios, que se utilizan entre dos valores, deben separarse de estos valores, a ambos lados del operador, por un espacio. Por ejemplo, `$numero = 3`, en el lugar de `$numero=3`. Esto se aplica a operadores como `+`, `-`, `*`, `/`, `=`, `==`, `!=`, `>`, `<`, `.` (Concatenación de cadenas), `.=`, `+=`, `-=`, etc. Los operadores unarios como `++`, `--` no deben tener separación. Por ejemplo: `$suma++`.

## Uso de comillas

Se pueden utilizar tanto las comillas simples como la (‘cadena’) como las comillas dobles (“cadena”) para delimitar las cadenas de caracteres. Las comillas dobles son necesarias si se desean incluir variables dentro de las cadenas de texto. Por ejemplo, `<h2> $title </h2>`. También se recomienda el uso de las comillas dobles cuando el texto puede incluir alguna comilla simple.

## Uso de punto y coma (;) en código PHP

Aunque PHP permite escribir líneas de código individuales sin el terminador de línea (`;`), como por ejemplo `<?php print $title ?>`. En Drupal es siempre obligatorio: `<?php print $title; ?>`.

- ✓ Correcto: `<?php print $title; ?>`.
- Incorrecto: `<?php print $title ?>`.

## Estructuras de control

Con respecto a las estructuras de control, hay que tener en cuenta las siguientes normas:

- ✓ Debe haber un espacio entre el comando que define la estructura (*if*, *while*, *for*, *etc*) y el paréntesis de apertura. Esto es así para no confundir las estructuras de control con la nomenclatura de las funciones.
- ✓ La llave de apertura (`{`) se situará en la misma línea que la definición de la estructura, separada por un espacio.
- ✓ Se recomienda usar siempre las llaves `{}` aun en los casos en que no sea obligatorio su uso (una sola "línea" de código dentro de la estructura de control).
- ✓ Las estructuras *else* y *elseif* se escribirán en la línea siguiente al cierre de la sentencia anterior.

## Funciones

Los nombres de las funciones deben estar escritos en minúsculas y las palabras separadas por guion bajo. Además, se debe incluir siempre como prefijo el nombre del módulo o tema, para evitar así duplicidad de funciones. En su declaración, después del nombre de la función, el paréntesis de inicio de los argumentos debe ir sin espacio. Cada argumento debe ir separado por un espacio, después de la coma del argumento anterior. Ejemplo:

```
function forum_help($path, $arg){
```

En la llamada a la función se aplican las mismas reglas anteriores con respecto a los parámetros, como se muestra a continuación:

```
$field =field_info_instance('node', 'taxonomy_forums', $node->type);
```

Como excepción, es posible usar más de un espacio antes de una asignación (`=`) para mejorar la presentación, cuando se estén realizando varias asignaciones en bloque:

```
$numero1 = foo($a, $type);
```

```
$primer_valor = foo2($b);
```

```
$i = foo3();
```

## Arrays

Los valores dentro de un *array* (o matriz) se deben separar por un espacio (después de la coma que los separa). El operador => debe separarse por un espacio a ambos lados. Cuando la línea de declaración del *array* supera los 80 caracteres, cada elemento se debe escribir en una única línea, indentándolo una vez (2 espacios). En este último caso, la coma de separación del último elemento también se escribirá, aunque no existan más elementos. De esta forma se evitan errores al añadir nuevos elementos al vector.

```
function privatemsg_menu() {
    $items['messages'] = array(
        'title'           => 'Messages',
        'title callback'  => 'privatemsg_title_callback',
        'page callback'   => 'privatemsg_list_page',
        'page arguments'  => array('list'),
        'file'            => 'privatemsg.pages.inc',
        'access callback' => 'privatemsg_user_access',
        'type'            => MENU_NORMAL_ITEM,
        'menu_name'       => 'user-menu',
    );
}
```

### Nombres de archivos

Los nombres de archivos deben escribirse siempre en minúscula. La única excepción son los archivos de documentación, que tendrán extensión .txt y el nombre en mayúscula. Por ejemplo README.txt.

### Comentar el código

Para la realización de comentarios suelen emplear */\** para comentarios en varias líneas y *//* para comentarios de una única línea. Se deben escribir frases completas, comenzándolas con mayúscula y terminándolas con un punto. En caso de que en el comentario se haga referencia a una constante, esta deberá escribirse en mayúscula (por ejemplo TRUE o FALSE).

```

<?php
/**
 * Implements hook_form_alter
 */
function generar_boletin_form_alter(&$form,&$form_state,$form_id) {

    if($form_id == 'views_form_noticias_revisadas_page'){
        $form['actions']['submit']['#value'] = 'Generar Boletin';
        $form['#submit'][] = 'exportar_pdf';

//echo "<pre>";print_r($form);echo "</pre>";
    }
}

```

### 3.3 Validación de la propuesta de solución

Los procesos de pruebas de software se destacan como el proceso de establecer datos de entrada al producto de software e inspeccionar los resultados obtenidos, permitiendo determinar si el software funciona realmente como se espera. Las pruebas de software responden fundamentalmente a dos interrogantes, ¿se ha obtenido un buen producto?, ¿se ha desarrollado de forma correcta? Este concepto da lugar al proceso de verificación y validación del software (ZAMORA, 2011).

Existen varios enfoques de pruebas que suelen ser utilizadas, dentro de las que se pueden mencionar:

#### Pruebas de Caja Blanca

Dirigidas a las funciones internas del sistema. Consisten en una verificación técnica del software que los desarrolladores pueden usar para examinar si su código trabaja como es esperado. Se realizan probando la lógica de la aplicación y comprobando el estado del software en varios puntos, verificando que el resultado de dicho estado coincida con lo esperado.

#### Pruebas de Caja Negra

Se desarrollan sobre la interfaz visual del software y se centran en los requisitos funcionales de la aplicación, sin adentrarse en el funcionamiento interno de la aplicación. A través de su realización se pueden encontrar errores de interfaz, funciones incorrectas, errores de salida y problemas con el acceso a datos.

#### Pruebas de Sistema

Las pruebas de sistema tienen como objetivo verificar el sistema de software para comprobar si el mismo cumple con sus requisitos. Cuenta con distintos tipos de pruebas, las cuales algunas son funcionales, de usabilidad, de rendimiento, de seguridad, entre otras.

### Resultados de las pruebas funcionales

Se probaron todos los casos de prueba que responden a las funcionalidades del subsistema de visualización y clasificación de noticias. En total fueron detectadas 30 no conformidades de las cuales todas fueron resueltas, los principales errores detectados se correspondían a errores de redacción, ortográficos y de idioma.

A continuación, se muestra una gráfica donde se desglosan las no conformidades detectadas en las tres iteraciones realizadas. En la primera iteración se recoge un total de 30 no conformidades de las cuales todas fueron resueltas, en la segunda iteración no arrojaron no conformidades.

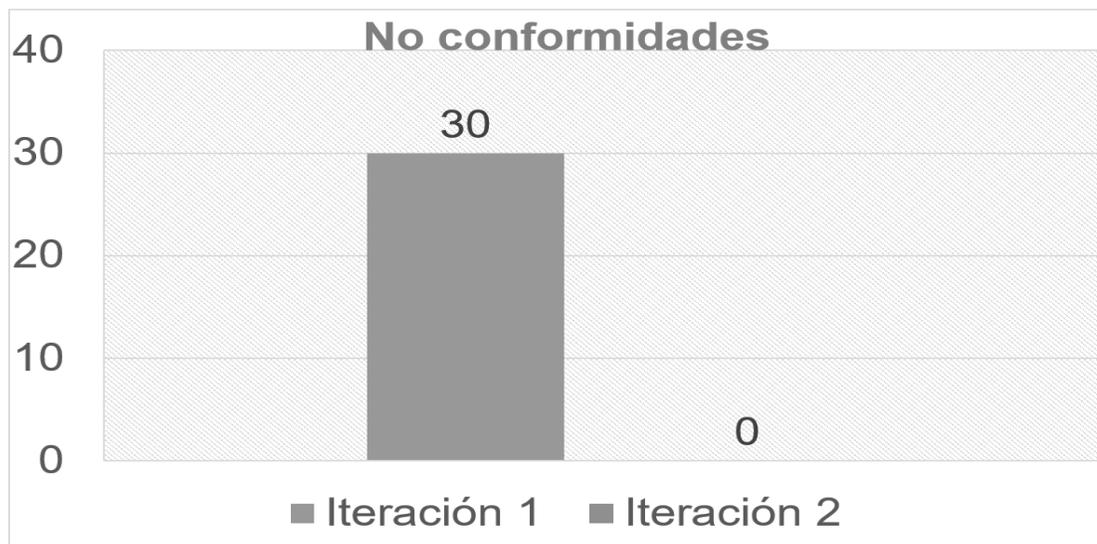


Figura 9. Resultados de las pruebas funcionales

#### 3.3.1 Integración

Las pruebas de integración están divididas en dos aristas. Primeramente, los módulos integrados durante el desarrollo del subsistema fueron: Link, Botón-cancelar, Date, Privatmsg y Requiere-loguin; también se probó el subsistema integrado con el subsistema de rastreo e indexación en ambientes similares al de despliegue. El desarrollo de las pruebas garantizó que el subsistema presentara un correcto funcionamiento.

### **3.3.2 Pruebas de seguridad**

Las pruebas de seguridad están diseñadas para probar las vulnerabilidades en el ambiente del lado del cliente, las comunicaciones de red que ocurren mientras los datos pasan del cliente al servidor de vuelta, y el ambiente del lado del servidor. Cada uno de estos dominios pueden recibir ataques, y es labor de quien prueba la seguridad descubrir las debilidades que pueden explotar quienes tengan la intención de hacerlo (PRESSMAN, 2011).

#### **Resultados de las pruebas de seguridad**

Las pruebas de seguridad se aplican en dos iteraciones. Se desarrollaron pruebas mediante la herramienta *Acunetix Web Vulnerability Scanner 8.0*, la cual ha sido considerada como la herramienta pionera en la tecnología de análisis de seguridad.

En la primera iteración se detectan 26 no conformidades, de ellas 3 alertas de tipo alta, 11 alertas tipo media y 12 de tipo baja. En la segunda iteración se solucionan todos los problemas encontrados que pudieran comprometer la seguridad e integridad de la aplicación con la instalación de módulos de seguridad.

### 3.4 Interfaces de la solución

The screenshot displays the XILEMA sacan news interface. At the top left is the logo for XILEMA sacan. To the right are navigation buttons for 'Noticias', 'Boletín', and 'Reportes'. On the left side, there are two menu sections: 'Mensajes privados' with a link to 'Escribir nuevo mensaje' and 'Mensajes', and 'Menu de usuario' with links for 'Mi cuenta' and 'Cerrar sesión'. The main content area features a search bar with 'Buscar' and 'Palabras Claves' labels, and an 'Aplicar' button. Below the search bar, three news items are listed, each with a title, source, and a brief description. The first item is 'México en alerta tras terremoto en el estado de Guerrero', the second is 'Ministra de Cooperación de Emiratos Árabes Unidos visita Cuba', and the third is 'ETECSA y compañía de Luxemburgo firman acuerdo para servicio'. Each item includes a source link and 'eliminar' and 'editar' options.

**Mensajes privados**

- [Escribir nuevo mensaje](#)
- [Mensajes](#)

**Menu de usuario**

- [Mi cuenta](#)
- [Cerrar sesión](#)

**Buscar** **Palabras Claves**

**México en alerta tras terremoto en el estado de Guerrero**  
Fuente: <http://www.cubadebate.cu>  
La activación de la alarma sísmica provocó el desalojo de viviendas y centros de asistencia mexicana. [eliminar](#) [editar](#)  
Hasta el momento no se reportan víctimas. De acuerdo con el Servicio Sismológico Nacional, el movimiento telúrico se originó a las hora...

---

**Ministra de Cooperación de Emiratos Árabes Unidos visita Cuba**  
Fuente: <http://www.cubadebate.cu>  
Ricardo Cabrisas Ruiz, Vicepresidente del Consejo de Ministros y Ministro de Economía y Planificación de Cuba, recibió a Reem Bint Ebrahim Al Hashimy, Ministra de Estado de Cooperación Internacional de los Emiratos Árabes Unidos.... [eliminar](#) [editar](#)

---

**ETECSA y compañía de Luxemburgo firman acuerdo para servicio**  
Fuente: <http://www.cubadebate.cu>  
Esto será posible mediante un servicio satelital de órbita media de alta capacidad muy similar en algunos parámetros de calidad a la fibra, garantizando capacidades adicionales que soportan el [eliminar](#) [editar](#)

Figura 10. Noticias sin revisar

Mensajes privados

- [Escribir nuevo mensaje](#)
- [Mensajes](#)

Menu de usuario

- [Mi cuenta](#)
- [Cerrar sesión](#)

Inicio » Reportes

## Reportes

Clasificación Fuente

Todas



Figura 11. Generación de reportes.

### Conclusiones parciales

- ✓ La confección de los diagramas de componentes y despliegue permitió establecer una mayor comprensión de la arquitectura y funcionamiento de los elementos del sistema.
- ✓ Las pruebas realizadas al software permitieron determinar y corregir errores y fallos en el funcionamiento de la aplicación, lo que permite asegurar que la propuesta de solución es estable y segura.

## **Conclusiones**

Con la investigación realizada se logró el desarrollo del subsistema de visualización y clasificación automática de noticias. El mismo perseguía brindar un sistema para la gestión noticias y capaz de clasificar las en dependencia del área geográfica de incidencia. Los objetivos propuestos fueron cumplidos con satisfacción, generando cada uno de ellos los siguientes resultados:

- La elaboración del marco teórico permitió identificar la necesidad de crear el subsistema de visualización y clasificación automática de noticias, determinando la selección de las herramientas, metodología y tecnologías factibles a utilizar en el desarrollo del subsistema.
- Los artefactos generados durante el flujo de trabajo de análisis y diseño sirvieron para conformar la primera visión de la implementación del sistema, defendiéndose durante la misma una estrategia para lograr una construcción flexible y robusta a través de la utilización de una arquitectura afín, patrones de diseño bien establecidos y estándares de codificación.
- Mediante la realización de pruebas de software para evaluar las funcionalidades se probó la conformidad de los requisitos especificados y se comprobó el cumplimiento del objetivo general trazado.

## **Recomendaciones**

A modo de recomendación se propone:

- Diseñar una ontología que abarque un área más profunda en el conocimiento que representa.

## Bibliografía

- AAMODT, Agnar; PLAZA, Enric. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI communications*, 1994, vol. 7, no 1, p. 39-59.
- ALVAREZ, S. Sistemas gestores de bases de datos. 2014. Available from <http://www.desarrolloweb.com/articulos/sistemas-gestores-bases-datos.html>.
- BAEZA-YATES, Ricardo, et al. *Modern information retrieval*. New York: ACM press, 1999.
- CAMPOS, O. Breve introducción a CSS3. 2011. Available from Internet:<<http://www.genbetadev.com/desarrollo-web/breve-introduccion-a-css3>>.
- COCHRAN, David. *Twitter bootstrap web development how-to*. Packt Publishing Ltd, 2012.
- CODINA, Lluís. Evaluación de recursos digitales en línea: conceptos, indicadores y métodos. *Revista española de documentación científica*, 2000, vol. 23, no 1, p. 9-44.
- DOMINICH, Sándor. A unified mathematical definition of classical information retrieval. *Journal of the Association for Information Science and Technology*, 2000, vol. 51, no 7, p. 614-624.
- DRUPAL.ORG. Drupal CMS Benefits. 2018. Available from Internet:<<https://www.drupal.org/features>>.
- FIGUEROLA, Carlos G., et al. Proyecto 7: un motor de recuperación de información web colaborativo. *Scire: representación y organización del conocimiento*, 2010, vol. 16, no 2, p. 53-59.
- FIX, Evelyn; HODGES, Joseph Lawson. Discriminatory analysis. Nonparametric discrimination: consistency properties. *International Statistical Review/Revue Internationale de Statistique*, 1989, vol. 57, no 3, p. 238-247.
- FOWLER, Martin. *Analysis patterns: reusable object models*. Addison-Wesley Professional, 1997.
- GAUCHAT, Juan Diego. *El gran libro de HTML5, CSS3 y Javascript*. Marcombo, 2012.
- GEETHA, S.; SATHIYAKUMARI, K. Backlink Analysis Using Mozrank Algorithm of Blogs. *International Journal of Computer Science*, 2012, vol. 1, no 9.

GIL, F. 2012. *Experto en Drupal. Curso de creación y gestión de portales web con Drupal 7*. [en línea]. 1.1. España: s.n. Disponible en: <https://www.forcontu.com/libros/experto-en-drupal-7-nivel-inicial>.

GOOVY, M., 2014. *Netbeans IDE de desarrollo*. [en línea]. 2014. S.l.: s.n. Disponible

GRANADOS CERVANTES, Antonio de Jesús. Elementos básicos para la indexación de artículos en la red: el caso de SciELO. *Innovación en TIC*, 2016.

GUTIÉRREZ, Mario Pérez-Montoro. Gestión del conocimiento, gestión documental y gestión de contenidos. En *Tendencias en documentación digital*. Trea, 2006. p. 110-134.

JACOBSON, I., et al. *Ingeniería de Software Orientada a Objetos-Un acercamiento a través de los casos de uso*. 1992.

LARMAN, C. *UML y patrones: introducción al análisis y diseño orientado a objetos*. In., 1999.

MÉNDEZ, A.V., 2010. *Metodologías de desarrollo de Software*. diciembre 2010. S.l.: s.n.

MICROSOFT DEVELOPER NETWORK 2014. *Diagramas de componentes de UML: Referencia*. [en línea]. Disponible en: <https://msdn.microsoft.com/es-es/library/dd409390.aspx>.

MILLER, George A., et al. Introduction to WordNet: An on-line lexical database. *International journal of lexicography*, 1990, vol. 3, no 4, p. 235-244.

MUÑOZ VALLEJO, Javier. Desarrollo de un portal de información y venta de vehículos siguiendo la metodología J. December. 2015.

NOY, Natalya F.; MCGUINNESS, Deborah L. Desarrollo de Ontologías-101: guía para crear tu primera ontología. traducido del inglés por: E. Antezana,, [http://protege.stanford.edu/publications/ontology\\_development/ontology101-es.pdf](http://protege.stanford.edu/publications/ontology_development/ontology101-es.pdf), 2005.

PAMNANI, Rajni; CHAWAN, Pramila. Web Usage Mining: A research area in Web mining. *Proceedings of ISCTE*, 2010, p. 73-77.

PETRAKIS, Euripides GM, et al. X-similarity: Computing semantic similarity between concepts from different ontologies. *Journal of Digital Information Management*, 2006, vol. 4, no 4.

PRESSMAN, R. S. *Ingeniería del Software Un enfoque práctico*, Séptima edición ed. 2010.

QUIALA, I.B.S. 2014. *Aplicación de soporte a la Metodología de desarrollo de distribuciones GNU/Linux, Nova-OpenUp*. Trabajo de diploma para optar por el título de ingeniero en Ciencias Informáticas. La Habana: Universidad de las Ciencias Informáticas.

SALAS, Rodrigo. Redes neuronales artificiales. *Universidad de Valparaíso. Departamento de Computación*, 2004, vol. 1.

SPARXSYSTEMS. *Diagrama de Despliegue UML 2*. [En línea]. Sparx Systems - Tutorial UML 2 - Diagrama de Despliegue, 2014. [Citado el: 20 de febrero de 2017.] Disponible en:

STEVENS, P. AND POOLEY, R. Utilización de UML. 2007.

VANDYK, J. An Introduction to Drupal Architecture. In DrupalCamp Des Moines. Iowa, 2011.

WITTEN, Ian H., et al. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

WORLD WIDE WEB CONSORTIUM, et al. Guía Breve de Web Semántica. *línea*. Available: <http://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica>. [Último acceso: 20 Abril 2014], 2014.

## Anexos

**Tabla 5.** Descripción del CU1. Autenticar usuario.

<b>Objetivo</b>	Autenticar al usuario	
<b>Actores</b>	Usuario básico	
<b>Resumen</b>	El CU se inicia cuando intentan acceder al sistema. Termina cuando el usuario accede al sistema.	
<b>Complejidad</b>	Baja	
<b>Prioridad</b>	Alta	
<b>Precondiciones</b>	Debe ser un usuario básico que esté operando en el sistema.	
<b>Postcondiciones</b>		
<b>Flujo de eventos</b>		
<b>Flujo básico &lt;Autenticar usuario&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Introduce los datos del usuario.	
2.		Muestra la interfaz principal.
3.		Termina el CU.
<b>Flujos alternos</b>		
1. Los datos no son válidos.		
	<b>Actor</b>	<b>Sistema</b>
3		Muestra un mensaje de error donde se le comunica al usuario las causas del mismo.
<b>Relaciones</b>	<b>CU Incluidos</b>	
	<b>CU Extendidos</b>	
<b>Requisitos funcionales</b>	<b>no</b>	
<b>Asuntos pendientes</b>		

**Tabla 6.** Descripción del CU2. Gestionar usuario.

<b>Objetivo</b>	Crear, eliminar, modificar, mostrar usuario y listar usuarios	
<b>Actores</b>	Administrador	
<b>Resumen</b>	Este CU describe cómo crear, eliminar, modificar, mostrar usuario y listar usuarios	
<b>Complejidad</b>	Baja	
<b>Prioridad</b>	Media	
<b>Precondiciones</b>	Que el usuario esté autenticado en el sistema.	
<b>Postcondiciones</b>		
<b>Flujo de eventos</b>		
<b>Flujo básico &lt;Gestionar usuario&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Se autentica en el sistema.	
2.		Muestra una interfaz para que escoja la opción de: <ul style="list-style-type: none"> <li>• Crear usuario</li> <li>• Eliminar usuario</li> <li>• Modificar usuario</li> <li>• Mostrar usuario</li> <li>• Listar usuarios</li> </ul>
3.	Si se escoge la opción: <ul style="list-style-type: none"> <li>• Crear usuario ir a la interfaz crear usuario</li> <li>• Eliminar usuario ir a la interfaz eliminar usuario.</li> <li>• Modificar ir a la interfaz modificar usuario.</li> <li>• Mostrar ir a la interfaz mostrar usuario.</li> <li>• Listar ir a la interfaz listar usuarios.</li> </ul>	
<b>Flujos alternos</b>		
1. No se puede autenticar usuario		
	<b>Actor</b>	<b>Sistema</b>
4		Muestra un mensaje de error donde se le comunica al usuario las causas del mismo.

<ul style="list-style-type: none"> <li>• <b>Sección 2: “Crear usuario”</b></li> </ul>		
<b>Flujo básico &lt;Gestionar usuario: Crear usuario&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Selecciona la opción “Añadir usuario”.	
2.		Muestra la interfaz crear usuario.
3.	Introduce los datos correspondientes.	
4.		Comprueba que los datos cumplan con los requisitos necesarios para que sea válido.
5.		Comprueba que no existan los datos en la base de datos.
<b>Flujos alternos</b>		
4. Los datos no son válidos.		
	<b>Actor</b>	<b>Sistema</b>
1.		Muestra un mensaje de error donde se le comunica al usuario las causas del mismo.
<b>Flujos alternos</b>		
5. Los datos existen.		
	<b>Actor</b>	<b>Sistema</b>
1.		Muestra un mensaje de error donde se le comunica al usuario las causas del mismo.
<ul style="list-style-type: none"> <li>• <b>Sección 3: “Eliminar usuario”</b></li> </ul>		
<b>Flujo básico &lt; Gestionar usuario: Eliminar usuario &gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Selecciona la opción “cancelar cuenta”.	
2.		Muestra la interfaz eliminar usuario.
3.	Selecciona usuario a eliminar y hace clic en el botón “Eliminar cuenta”.	
4.		Busca usuario seleccionado en la base de datos y lo elimina.
<b>Flujos alternos</b>		

	Actor	Sistema
1.		.
<ul style="list-style-type: none"> <li>• <b>Sección 4: “Modificar usuario”</b></li> </ul>		
<b>Flujo básico &lt; Gestionar usuario: Modificar usuario &gt;</b>		
	Actor	Sistema
1.	Selecciona la opción “editar”.	
2.		Muestra la interfaz modificar usuario.
3.	Introduce los datos correspondientes.	
4.		Busca usuario seleccionado en la base de datos y lo modifica.
<b>Flujos alternos</b>		
	Actor	Sistema
1.		.
<ul style="list-style-type: none"> <li>• <b>Sección 5: “Mostrar usuario”</b></li> </ul>		
<b>Flujo básico &lt; Gestionar usuario: Mostrar usuario &gt;</b>		
	Actor	Sistema
1.	Selecciona al usuario	
2.		Muestra la interfaz con los datos del usuario.
<b>Flujos alternos</b>		
	Actor	Sistema
1.		
<ul style="list-style-type: none"> <li>• <b>Sección 6: “Listar usuarios”</b></li> </ul>		
<b>Flujo básico &lt; Gestionar usuario: Listar usuarios &gt;</b>		
	Actor	Sistema
1.	Selecciona la opción “usuarios”.	
2.		Muestra la interfaz listar usuarios.
<b>Flujos alternos</b>		
	Actor	Sistema
1.		

**Tabla 7.** Descripción del CU3: Gestionar permisos a usuarios.

<b>Objetivo</b>	Asignar o quitar permisos a los usuarios	
<b>Actores</b>	Administrador	
<b>Resumen</b>	Este CU describe cómo se asigna y quitan permisos a los usuarios	
<b>Complejidad</b>	Baja	
<b>Prioridad</b>	Media	
<b>Precondiciones</b>	Que el usuario esté autenticado en el sistema.	
<b>Postcondiciones</b>		
<b>Flujo de eventos</b>		
<b>Flujo básico &lt;Gestionar permisos a usuario&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Se autentica en el sistema.	
2.		Muestra una interfaz para que escoja la opción de: <ul style="list-style-type: none"> <li>• Permisos</li> </ul>
3.	Si se escoge la opción: <ul style="list-style-type: none"> <li>• Permisos</li> </ul>	
<b>Flujos alternos</b>		
1. No se puede autenticar usuario		
	<b>Actor</b>	<b>Sistema</b>
5		Muestra un mensaje de error donde se le comunica al usuario las causas del mismo.
<ul style="list-style-type: none"> <li>• <b>Sección 2: “Asignar permisos”</b></li> </ul>		
<b>Flujo básico &lt;Gestionar permisos a usuarios: Asignar permisos&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Selecciona la opción “permisos”.	
2.		Muestra la interfaz muestra la interfaz de los permisos.

3.	Selecciona los permisos al usuario correspondiente	
4.		Guarda la configuración y habilita una parte del sistema deshabilitada para el usuario correspondiente
5.		
<b>Flujos alternos</b>		
	<b>Actor</b>	<b>Sistema</b>
1.		
<ul style="list-style-type: none"> <li>• <b>Sección 3: “Quitar permisos”</b></li> </ul>		
<b>Flujo básico &lt;Gestionar permisos a usuarios: Asignar permisos&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Selecciona la opción “permisos”.	
2.		Muestra la interfaz muestra la interfaz de los permisos.
3.	Desmarca los permisos al usuario correspondiente	
4.		Guarda la configuración y deshabilita una porción del sistema que estaba habilitada.
5.		
<b>Flujos alternos</b>		
	<b>Actor</b>	<b>Sistema</b>
1.		

**Tabla 8.** Descripción del CU4. Gestionar roles de usuario

<b>Objetivo</b>	Crear, eliminar, modificar y listar roles de usuarios
<b>Actores</b>	Administrador
<b>Resumen</b>	Este CU describe cómo crear, eliminar, modificar y listar roles de usuarios
<b>Complejidad</b>	Baja
<b>Prioridad</b>	Media

<b>Precondiciones</b>	Que el usuario esté autenticado en el sistema.	
<b>Postcondiciones</b>		
<b>Flujo de eventos</b>		
<b>Flujo básico &lt;Gestionar usuario&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Se autentica en el sistema.	
2.		Muestra una interfaz para que escoja la opción de: <ul style="list-style-type: none"> <li>• Roles</li> </ul>
3.	Si se escoge la opción: <ul style="list-style-type: none"> <li>• Roles</li> </ul>	
<b>Flujos alternos</b>		
1. No se puede autenticar usuario		
	<b>Actor</b>	<b>Sistema</b>
6		Muestra un mensaje de error donde se le comunica al usuario las causas del mismo.
<ul style="list-style-type: none"> <li>• <b>Sección 2: “Crear rol de usuario”</b></li> </ul>		
<b>Flujo básico &lt;Gestionar roles de usuario: Crear rol de usuario&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Escribe el rol nuevo para añadir	
2.		Crea el rol nuevo
3.		
4.		
<b>Flujos alternos</b>		
5. El rol existe.		
	<b>Actor</b>	<b>Sistema</b>

1.		Muestra un mensaje de error donde se le comunica al usuario las causas del mismo.
<ul style="list-style-type: none"> <li>• <b>Sección 3: “Eliminar rol usuario”</b></li> </ul>		
<b>Flujo básico &lt; Gestionar usuario: Eliminar usuario &gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Selecciona la opción “editar”.	
2.		Muestra la interfaz del rol del usuario.
3.	Hace clic en el botón eliminar rol	
4.		Busca el rol en la base de datos y lo elimina
<b>Flujos alternos</b>		
	<b>Actor</b>	<b>Sistema</b>
1.		.
<ul style="list-style-type: none"> <li>• <b>Sección 4: “Modificar rol de usuario”</b></li> </ul>		
<b>Flujo básico &lt; Gestionar rol de usuarios: Modificar rol de usuarios&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Selecciona la opción “editar”.	
2.		Muestra la interfaz modificar rol de usuario.
3.	Introduce los datos correspondientes.	
4.		Busca usuario seleccionado en la base de datos y lo modifica.
<b>Flujos alternos</b>		
	<b>Actor</b>	<b>Sistema</b>
1.		.
<ul style="list-style-type: none"> <li>• <b>Sección 6: “Listar rol de usuarios”</b></li> </ul>		
<b>Flujo básico &lt; Gestionar usuario: Listar usuarios &gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Selecciona la opción “Roles”.	
2.		Muestra la interfaz roles de los usuarios.

Flujos alternos		
	Actor	Sistema
1.		

**Tabla 9.** Descripción del CU5. Gestionar cuenta.

<b>Objetivo</b>	Mostrar y modificar los datos de la cuenta del usuario	
<b>Actores</b>	Usuario básico	
<b>Resumen</b>	Este CU describe cómo se muestran y modifican los datos de una cuenta.	
<b>Complejidad</b>	Baja	
<b>Prioridad</b>	Media	
<b>Precondiciones</b>	Que el usuario esté autenticado en el sistema.	
<b>Postcondiciones</b>		
<b>Flujo de eventos</b>		
<b>Flujo básico &lt;Gestionar cuenta&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Se autentica en el sistema.	
2.		Muestra una interfaz para que escoja la opción de: <ul style="list-style-type: none"> <li>• Mi cuenta</li> </ul>
3.	Si se escoge la opción: <ul style="list-style-type: none"> <li>• Mi cuenta</li> </ul>	
<b>Flujos alternos</b>		
1. No se puede autenticar usuario		
	<b>Actor</b>	<b>Sistema</b>
7		Muestra un mensaje de error donde se le comunica al usuario las causas del mismo.
<ul style="list-style-type: none"> <li>• <b>Sección 2: “Mostrar cuenta”</b></li> </ul>		
<b>Flujo básico &lt;Gestionar cuenta: Mostrar cuenta&gt;</b>		

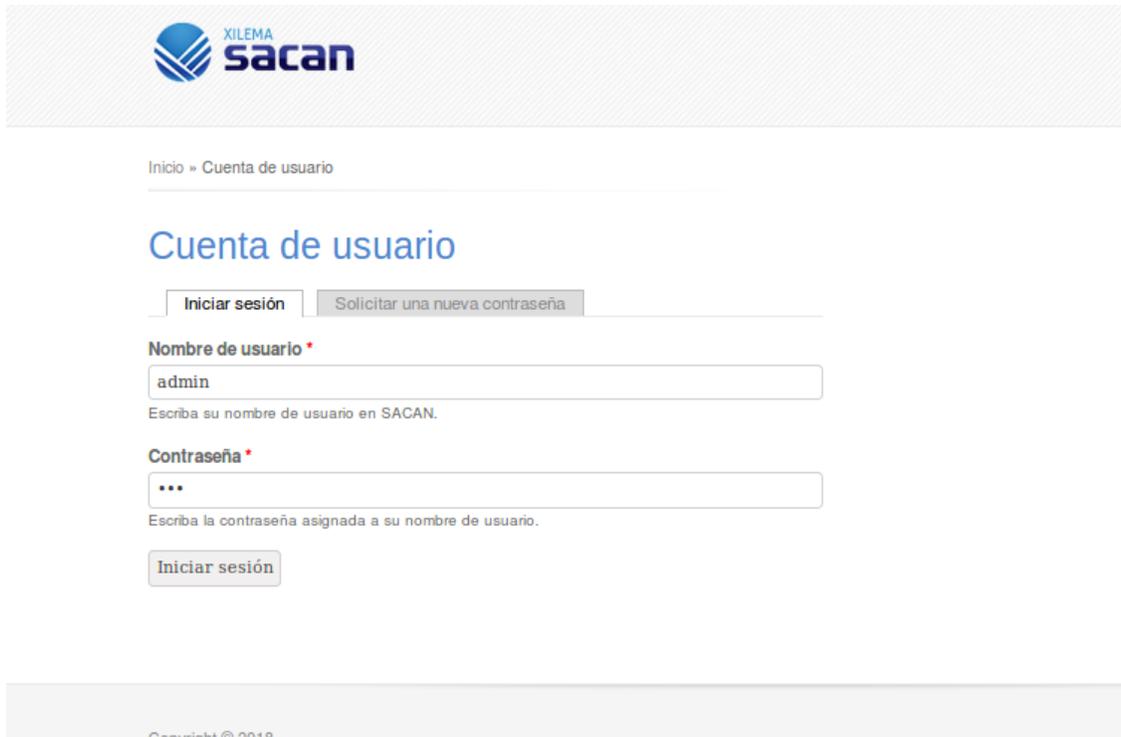
	Actor	Sistema
1.	Selecciona la opción "Mi cuenta".	
2.		Muestra la interfaz con los dato del usuario
5.		
<b>Flujos alternos</b>		
	Actor	Sistema
1.		
<ul style="list-style-type: none"> <li>• <b>Sección 3: "Modificar cuenta"</b></li> </ul>		
<b>Flujo básico &lt;Gestionar cuenta: Modificar cuenta&gt;</b>		
	Actor	Sistema
1.	Selecciona la opción "Mi cuenta".	
2.		Muestra la interfaz con los datos del usuario
3.	Da clic en el botón editar	
4.		Muestra la interfaz para editar la cuenta.
5.	Entra los datos pertinentes	
6.		Modifica los datos del usuario en la base de datos

**Tabla 10.** Descripción del CU8. Generar boletín.

<b>Objetivo</b>	Generar un documento pdf, con las noticias seleccionadas
<b>Actores</b>	Usuario básico
<b>Resumen</b>	Este CU describe cómo se genera un pdf
<b>Complejidad</b>	Media
<b>Prioridad</b>	Alta
<b>Precondiciones</b>	Que el usuario esté autenticado en el sistema.
<b>Postcondiciones</b>	
<b>Flujo de eventos</b>	

<b>Flujo básico &lt;Generar Boletín&gt;</b>		
	<b>Actor</b>	<b>Sistema</b>
1.	Se autentica en el sistema.	
2.		Muestra una interfaz con una lista de noticias
3.	El usuario selecciona las noticias que desea	
4.		Genera un documento pdf con las noticias seleccionadas
<b>Flujos alternos</b>		
1. No se puede autenticar usuario		
	<b>Actor</b>	<b>Sistema</b>
8		Muestra un mensaje de error donde se le comunica al usuario las causas del mismo.

## Anexo 2: Interfaz de usuario



The screenshot shows the user login page for XILEMA SACAN. At the top left is the logo with the text "XILEMA sacan". Below the logo is a breadcrumb trail: "Inicio » Cuenta de usuario". The main heading is "Cuenta de usuario". There are two buttons: "Iniciar sesión" (highlighted) and "Solicitar una nueva contraseña". Below these are two input fields: "Nombre de usuario \*" with the value "admin" and "Contraseña \*" with masked characters "•••". Below the password field is a button "Iniciar sesión". At the bottom of the page, there is a copyright notice: "Copyright © 2019".

Figura 12. Iniciar sesión.

Mensajes privados

- [Escribir nuevo mensaie](#)
- [Mensajes](#)

Menu de usuario

- [Mi cuenta](#)
- [Cerrar sesión](#)

Inicio » Noticias revisadas

## Noticias revisadas

Buscar

Aplicar

Seleccionar todos los elementos

**El oficialismo se prepara para una batalla campal en la...**

Las imágenes de las tropas de choque del oficialismo cubano en la Cumbre de las Américas en Panamá, en 2015, se quedarán cortas este abril. El...

Fuente: <http://www.14ymedio.com/>

**Cuba se mantiene como uno de los países que aplica...**

La reciente sugerencia del residente norteamericano Donald Trump de restablecer la pena de muerte en todos los estados en caso de tráfico de...

Fuente: <https://www.cibercuba.com/>

**Cuba reitera en ONU su compromiso con la protección de los...**

El embajador de Cuba ante la ONU en Ginebra, Pedro Luis Pedroso, reiteró el compromiso de su país con la promoción y la protección de los derechos...

Figura 13. Noticias revisadas.