



Propuesta para la recuperación de información  
proveniente de Internet para el Sistema de Información  
de Mercadotecnia de Albet S.A.

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

**Autor:**

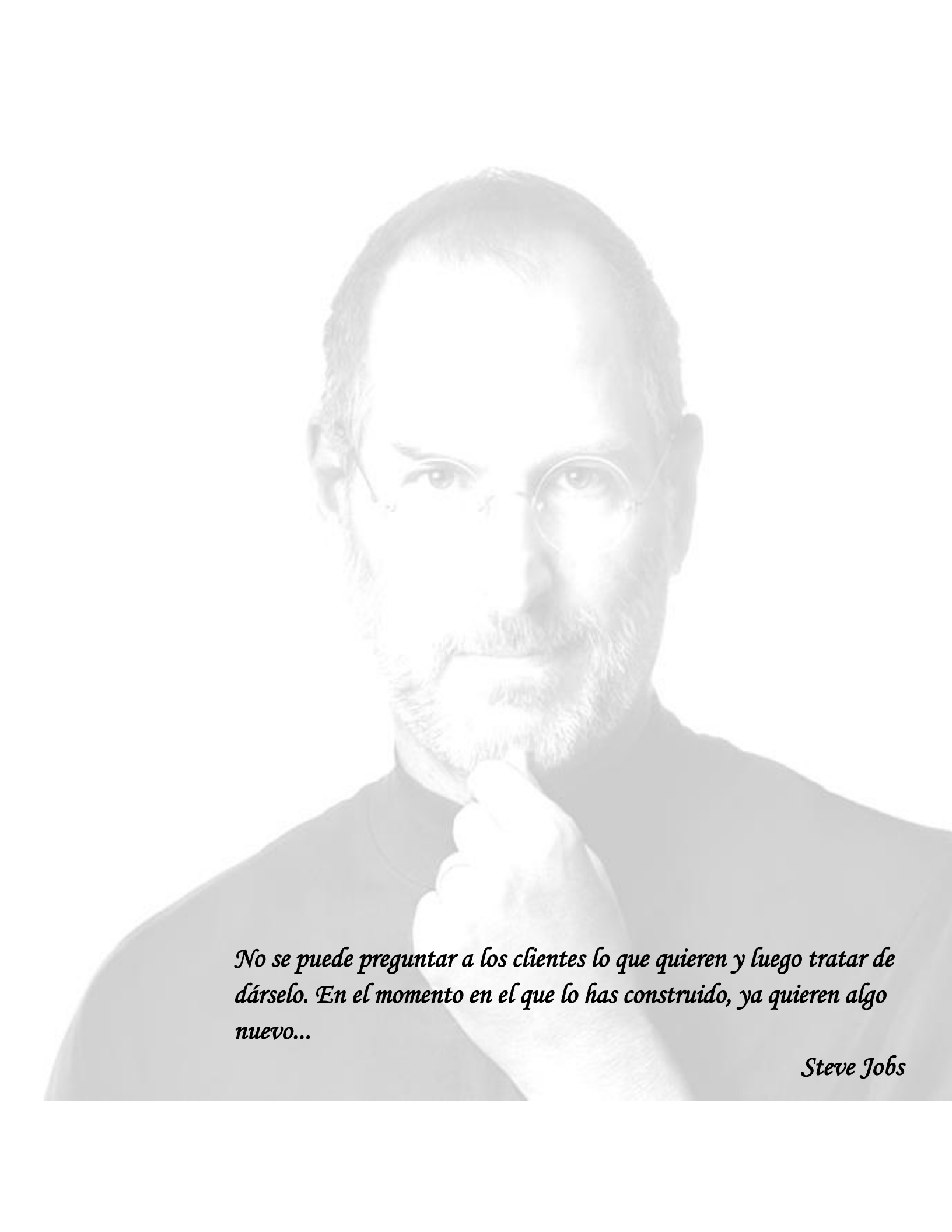
**Sandy Hernández Carballo**

**Tutores:**

**Ing. Yenny Leal González**

**Ing. Sergio René Vazquez Rodriguez**

**Ing. Oscar Lázaro Garcés Pérez**



*No se puede preguntar a los clientes lo que quieren y luego tratar de dárselo. En el momento en el que lo has construido, ya quieren algo nuevo...*

*Steve Jobs*

Declaro que soy el único autor de este trabajo y autorizo a la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

Para que así conste firmo la presente a los \_\_\_\_ días del mes de \_\_\_\_\_ del año \_\_\_\_\_.

Sandy Hernández Carballo

\_\_\_\_\_

Firma del autor

Ing. Yenny Leal González

\_\_\_\_\_

Firma del tutor

Ing. Sergio R. Vazquez Rodriguez

\_\_\_\_\_

Firma del tutor

Ing. Oscar L. Garcés Pérez

\_\_\_\_\_

Firma del tutor

**Ing. Yenny Leal González**

Graduada de Ingeniera en Ciencias Informáticas en el curso 2007- 2008. Posee la categoría docente de Instructora. Actualmente se desempeña como especialista de mercado en la empresa Albet S.A. Ingeniería y Sistemas como miembro del Grupo de Mercadotecnia.

Correo: [ylgonzalez@uci.cu](mailto:ylgonzalez@uci.cu)

**Ing. Sergio René Vazquez Rodriguez:**

Graduado de Ingeniero en Ciencias Informáticas en el curso 2009-2010. Se ha desempeñado como desarrollador y jefe en proyectos relacionados con la gestión de procesos en la Universidad de las Ciencias Informáticas. Actualmente es Jefe de la Línea Soluciones para Ciudad Digital, del Centro de Informatización Universitaria (CENIA).

Correo: [svazquez@uci.cu](mailto:svazquez@uci.cu)

**Ing. Oscar Lázaro Garcés Pérez**

Graduado de Ingeniero en Ciencias Informáticas en el curso 2010-2011. Su trabajo de diploma fue la propuesta de arquitectura de un Sistema de Información de Mercadotecnia para Albet S.A. Actualmente es analista de información en el departamento de Ingeniería Social Universitaria del Centro de Ideoinformática (CIDI).

Correo: [olgarcés@uci.cu](mailto:olgarcés@uci.cu)

*En la vida siempre nos vemos influenciados de una manera u otra por todos los que nos rodean, positiva o negativamente.*

*Pero sin duda alguna, mi mayor influencia ha sido mi madre, quien me ha apoyado siempre en todas mis decisiones, incluso con las que no ha estado de acuerdo, ahí ha estado cuando regreso y le tengo que decir: “Mami, tenías razón”. Siempre me ha dado aliento para seguir adelante y superarme, a tal punto que llego aquí hoy, finalizado mi trabajo de diploma, gracias a ella.*

*La familia siempre nos apoya, aunque siempre también unos más que otros, mi “abuelita Julita” y mi “abuelito Turino”, los adoro, mi abuela Deysi, la imagen de mi madre, igualitas las dos aunque lo nieguen.*

*Mi padre, por supuesto que ha puesto su granito de arena, de una forma u otra también ha contribuido. Mis hermanas, poniéndome en la obligación de servir de ejemplo como hermano mayor, y mi hermano, que aunque está lejos, nos mantenemos cerca.*

*Agradecimiento especial para **mí**, por haber llegado hasta aquí aunque no conforme, pude haberme esforzado más.*

*En fin, ni para cuando acabar, mi piquete de viciosos (mayormente influencia negativa) – Sandy salió el capítulo de tal o mas cual serie, oye, un partido de dota, montaron un servidor de WOW-. Pero sin dudas aprendí mucho con ellos, se podría decir que lo que de verdad aprendí de informática, lo aprendí ahí, en el ocio, en los retos que nos imponíamos unos a otros para lograr algún programita para hacer algo, conectarse a alguna PC, generalmente también éramos los mismos TEEN para las copas de programación o cualquier evento que se presentaba. Ahí estábamos siempre.*

*No puedo dejar de poner en el lugar que van a mis tutores, de veras me han apoyado bastante, gracias a ellos he sacado esta tesis en menos de 4 meses, la jefa Yenny, subjefe Sergio y compañero Oscar.*

*En fin, se podría hacer todo una tesis de agradecimientos, pero hay que parar pues son solo 80 hojas. Gracias a todos los que me rodean por influir de una manera u otra, ya sea positiva o negativamente...*

---

## RESUMEN

A la hora de implementar un Sistema de Información de Mercadotecnia en una empresa, existe la problemática de que el manejo de información que hay en Internet es muy engorroso por su diversidad y gran volumen. Una herramienta que se encargue de la recuperación de información proveniente de Internet sirve a los especialistas en sus investigaciones haciendo, más eficaz el proceso de toma de decisiones. Este documento presenta una propuesta que apoye la recuperación de información proveniente de Internet para un Sistema de Información de Mercadotecnia en la empresa Albet S.A. Se incluyen los elementos teóricos que apoyan la investigación, las herramientas y tecnologías a utilizar a lo largo del proceso de desarrollo y se documentan los artefactos que se generaron siguiendo un proceso de desarrollo con la metodología ágil SXP.

**PALABRAS CLAVE:** Internet, marketing, recuperación de información, requerimientos de software

## ABSTRACT

*When implementing a marketing information system in a company, there is the problem that managing information available on the Internet is very difficult due to its diversity and high volume. A tool that handles retrieving information from the Internet would be ideal, to support the marketing specialists' to execute more effective decision-making process. This paper presents a proposal that supports Internet's information retrieval for a Marketing Information System in Albet S.A. It includes the theoretical elements that support the research, the tools and technologies to be used throughout the development process and documents the artifacts generated by using the agile methodology: SXP.*

**Keywords:** Internet, information retrieval, marketing, software requirement

<b>INTRODUCCIÓN</b> .....	1
<b>CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA</b> .....	5
1.1. Introducción.....	5
1.2. Internet y mercadotecnia.....	6
1.3. Información .....	7
1.4. Observación y vigilancia de Internet.....	7
1.5. Cibermetría.....	9
1.6. Recuperación de Información.....	13
1.7. Modelos de los SRI.....	16
1.8. Soluciones existentes .....	18
1.9. Estudio de otros <i>Spiders</i> .....	21
1.10. Conclusiones del capítulo.....	25
<b>CAPÍTULO 2: ENTORNO DE DESARROLLO Y OBTENCIÓN DE REQUISITOS</b> .....	26
2.1. Introducción.....	26
2.2. Entorno desarrollo para la descripción de la propuesta .....	26
2.3. Propuesta de entorno de desarrollo.....	31
2.4. Arquitectura .....	38
2.5. Flujo actual de los procesos del negocio .....	41
2.6. Descripción general de la propuesta solución.....	42
2.7. Estrategias para el levantamiento de requerimientos .....	44
2.8. Principales funcionalidades.....	48
2.9. Requisitos no funcionales .....	50

---

2.10.	Lista reserva del producto .....	51
2.11.	Conclusiones del capítulo.....	52
<b>CAPÍTULO 3: ESPECIFICACIÓN Y VALIDACIÓN DE REQUISITOS .....</b>		<b>54</b>
3.1.	Introducción.....	54
3.2.	Planificación .....	54
3.3.	Descripción de las Historias de Usuario .....	55
3.4.	Validación de requerimientos .....	58
3.5.	Conclusiones del capítulo.....	63
<b>CONCLUSIONES .....</b>		<b>64</b>
<b>RECOMENDACIONES .....</b>		<b>65</b>
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>		<b>66</b>
<b>GLOSARIO .....</b>		<b>70</b>
<b>ANEXOS.....</b>		<b>73</b>



---

## **INTRODUCCIÓN**

Mercadotecnia, según Martin L Bell (1987) "son las actividades integradas de un negocio dirigidas hacia la satisfacción de las necesidades de los clientes". (1) O sea, son estudios que realiza una empresa sobre las necesidades y preferencias de los clientes con el objetivo de ser elegidos por estos como satisfactor de sus necesidades.

Eso era antes, el mundo de hoy es cada vez más competitivo, los mercados más globales y los consumidores más exigentes e informados. La mercadotecnia no se puede limitar solamente al estudio del cliente, sino que tiene que ampliar su campo de acción hacia el mercado en general, estudiar competencias, clientes, posibles clientes, tendencias tecnológicas, etc.

Hoy el internet se ha convertido en el medio de comunicación, entretenimiento, e información más importante del mundo, convirtiéndola en un mercado global. Las empresas han aprovechado este potencial como medio de difusión y muchas cuentan con páginas web, dominios propios, correo electrónico, entre otras.

En esta etapa se le otorga la mayor importancia a la distribución, a la innovación de productos orientada hacia el consumidor. Pero, con las ventas como objetivo principal, las empresas no logran adaptarse a la gran cantidad de información con que cuenta el cliente. Por tanto, se necesita convertir este volumen de información útil en datos manejables por el personal que lo necesite, para poder alcanzar niveles máximos de calidad y satisfacción al cliente, lanzar nuevos productos, y consolidar los ya existentes.

Para enfrentar estos retos hay que cubrir necesidades de información mayores cada día, aunque la estadística, la computación y la ingeniería están muy desarrolladas, casi nunca se cuenta con toda la información necesaria para tomar correctas decisiones de mercadotecnia.

En la actualidad crece el número de empresas cubanas que desarrollan procesos relacionados con la mercadotecnia, aunque el número sigue siendo insuficiente y los estudios siguen siendo llevados a cabo de forma manual, por lo que terminar una investigación requiere de mucho tiempo y recursos. Albet S.A. cuenta con un grupo de trabajo que se encarga de esta actividad, pero lo hace de forma semi-automatizada. La situación problemática en Albet S.A. es la siguiente:

- La información útil en Internet es diversa y varía para cada producto, servicio o empresa que se estudia, por lo que los especialistas deben manejar un gran volumen de información dispersa en Internet, la cual tienen que buscar manualmente y mucha puede pasar desapercibida a la vista.
- Los especialistas almacenan la información de forma personal y en herramientas ofimáticas, por lo que puede duplicarse una investigación.
- Las direcciones de sitios útiles en las investigaciones, no están disponibles todo el tiempo, ni se tiene constancia de la información anterior si esta varía en Internet.
- No se tiene constancia inmediata de cambios o actualizaciones en sitios importantes para una investigación de mercados, hasta que el especialista visita las páginas.
- El proceso de buscar, clasificar y asociar información proveniente de Internet se realiza de forma manual, por lo que los especialistas invierten mucho esfuerzo y tiempo en este proceso.

La implantación de un Sistema de Información de Mercadotecnia (SIM) sería oportuna en este caso, cuya finalidad es recabar, clasificar, analizar, evaluar y distribuir información pertinente, oportuna y precisa que servirá a quienes toman decisiones de mercadotecnia para mejorar la planeación, ejecución y control. En dicho sistema la primera función consistiría en evaluar las necesidades de información de las personas encargadas de la toma de decisiones y buscarla, este proceso a su vez necesita ser automatizado pues es inmenso el volumen de información existente en la red y muy dinámico como para poder realizarlo manualmente. Ya se encuentra desarrollado el SIM para Albet S.A. pero falta proveer dicho sistema de la información necesaria.

Para darle solución a la situación antes expuesta se plantea el siguiente **problema a resolver**: ¿Cómo gestionar información diversa y actual proveniente de Internet para apoyar las investigaciones de mercado que realizan los especialistas de Albet S.A.?

El **objeto de estudio** de este trabajo son los procesos de recuperación de información proveniente de Internet.

El **objetivo general** de este trabajo es modelar una herramienta que apoye las investigaciones de mercado en Albet S.A., mediante la incorporación a su SIM de funcionalidades que recuperen información diversa y actual proveniente de Internet.

### Tareas de la investigación:

1. Caracterización de la recuperación de información proveniente de Internet.
2. Análisis de sistemas que incluyan funcionalidades relacionadas con la recuperación de información de internet.
3. Análisis de herramientas para la obtención de nuevas páginas a partir de los vínculos que esta posee (*spiders*).
4. Descripción de las metodologías de desarrollo, tecnologías y lenguajes de programación para el desarrollo de las funcionalidades.
5. Definición de los requerimientos funcionales y no funcionales a cumplir.
6. Validación de los requerimientos definidos.
7. Elaboración de los artefactos establecidos por la metodología seleccionada.
8. Diseño de los prototipos de interfaz de usuario para las funcionalidades de la herramienta.

**Posibles resultados:** la modelación de las funcionalidades para la recuperación de información que apoyará a los especialistas de mercadotecnia de Albet S.A. en la creación de investigaciones de mercado sobre productos, servicios, empresas u otros elementos. La integración de estas funcionalidades en el SIM reducirá el tiempo y costos (en uso de Internet) en la creación de estas investigaciones, evitando duplicación y reutilizando la información en varios estudios al mismo tiempo. Igualmente ayudará a contar con una mayor cantidad de información actualizada proveniente de Internet, por lo que deben ser realizados estudios de mayor calidad, logrando la creación de una estrategia de mercadotecnia superior para los elementos comercializables en Albet S.A.

### Métodos teóricos:

- **Análisis histórico – lógico:** Permitió analizar las características de la automatización de los procesos de recuperación y obtención de la información proveniente de Internet, asistiendo en la comprensión del objeto de estudio y el análisis del estado del arte.
- **Modelación:** Se utilizó a la hora de confeccionar el modelo de análisis con sus clases y atributos, el modelo de datos, el diagrama de despliegue y el de componentes.
- **Analítico-sintético:** Se utilizó en el proceso de análisis de la bibliografía, realizando una síntesis de la misma.

---

### Métodos empíricos:

- **Observación:** Se utilizó en la investigación para determinar el problema a resolver, y comprender las necesidades del cliente y los procesos que se llevan a cabo actualmente para el proceso de obtención y recuperación de información proveniente de Internet.

### El presente documento de tesis está estructurado de la siguiente forma:

**Capítulo 1: Fundamentación teórica:** Se presentan los elementos teóricos que sirven de base a la investigación del problema planteado. Se realiza un estudio del estado del arte de las herramientas que sustentan la recuperación de información de internet.

**Capítulo 2: Entorno de desarrollo y obtención de requisitos:** Se define la tecnología sobre la cual se va a desarrollar el sistema, y la metodología que se utilizará para llevar a cabo todo el proceso de desarrollo. Se realiza la descripción de los procesos de negocio y de aquellos que serán automatizados. Una propuesta general del sistema así como requisitos funcionales y no funcionales.

**Capítulo 3: Especificación y validación de requisitos:** Se presenta la lista de reserva del producto así como la descripción de las funcionalidades a través de las historias de usuarios. Se propone el plan de iteraciones con el cálculo del esfuerzo asociado a las historias de usuarios, además de los métodos utilizados para la validación de requisitos.

---

## **CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA**

### **1.1. Introducción**

En el mundo de hoy, para lograr ventajas competitivas, se necesita una mentalidad orientada hacia el mercado. Las empresas no pueden sobrevivir por el simple hecho de producir pues existe gran variedad de opciones de compra y servicios en comparación con la cantidad de clientes que hay, quienes buscan mayor calidad al elegir el principal satisfactor de sus necesidades.

El marketing o mercadotecnia según Kotler (1989), "es la técnica de administración empresarial que permite anticipar la estructura de la demanda del mercado elegido, para concebir, promocionar y distribuir los productos y/o servicios que la satisfagan y/o estimulen, maximizando al mismo tiempo las utilidades de la empresa". (2) Y de acuerdo con la Asociación Americana de Marketing (2007), es "la actividad, conjunto de instituciones y procesos para crear, comunicar, entregar e intercambiar ofertas que tienen valor para los clientes, usuarios, socios y la sociedad en general". (3)

En ambos casos se detecta la necesidad para la empresa de realizar una investigación de mercado, la cual proporcione información precisa sobre la condición real del mercado, que brinde al especialista en mercadotecnia datos valiosos en cuanto a conocimiento sobre los clientes, competidores y otras fuerzas del mercado, de forma que cumpla con su tarea de proporcionar información relevante, confiable, actualizada y exacta para ayudar a la toma de decisiones.

De acuerdo con Kotler (1991) "Investigación de Mercados es la función que vincula al consumidor, al cliente y al público con el comerciante a través de la información, esta se utiliza para identificar y definir las oportunidades y problemas de mercadotecnia, generar, refinar y evaluar las acciones de mercadotecnia, vigilar su desarrollo y mejorar su comprensión como un proceso". (4) Según la Asociación Americana de Mercadotecnia (2007) "La Investigación de mercados especifica qué información se requiere para tratar estos asuntos, diseña el método para recopilar la información, administra e instrumenta el proceso de recolección de datos, analiza los resultados y comunica sus hallazgos e implicaciones". (3)

Investigar el mercado y conocer los clientes siempre ha sido importante desde que surgió el comercio, pero hoy más que nunca la mejor posición en la preferencia del consumidor, la consigue el que mayor información tenga de él.

Gestionar esta información es complejo, este proceso se puede realizar utilizando un Sistema de Información de Marketing (SIM) que ayude a desarrollar una respuesta rápida y eficaz ante una toma de decisión. Pero con la llegada de Internet y el gran volumen de información que en esta se maneja, el proceso de la investigación de mercado se hace muy difícil si se tiene que buscar manualmente. Además, si no se almacena, se pueden perder datos valiosos y si no se mantiene actualizándose sistemáticamente, se pueden tomar decisiones basándose en datos obsoletos.

Por ello la necesidad de que un SIM cuente con una herramienta que ayude en este proceso, que observe la red, que convierta el ruido, la inundación de datos y noticias en información útil que permita anticipar las nuevas oportunidades de negocio, sus riesgos e incertidumbres, mediante la difusión de información sobre nuevas tecnologías y tendencias del sector.

En el presente capítulo se tratan los principales conceptos relacionados con el tema de los sistemas de obtención y recuperación de información proveniente de Internet, así como sistemas o herramientas que apoyan este proceso y selección de algunas que puedan abreviar y optimizar las funcionalidades del sistema que se propone.

## **1.2. Internet y mercadotecnia**

Para comenzar, es necesario definir y caracterizar el ambiente sobre el que se trabaja, o sea Internet; de dónde se extrae la información necesaria y para qué se utiliza.

### **1.2.1. Internet**

Según La Real Academia Española, Internet es: “red informática mundial, descentralizada, formada por la conexión directa entre computadoras u ordenadores mediante un protocolo especial de comunicación”. (5) Terceiro (1996) define Internet como “un medio universal de comunicación y búsqueda de información a muy bajo coste”. (6)

Por tanto, es un conjunto descentralizado de redes de comunicación interconectadas de forma tal que las redes físicas que la componen funcionan como una red lógica única de alcance mundial. Esto la convierte en la mayor fuente de información de hoy en día como una gran biblioteca común pero en formato digital.

### **1.2.2. Mercadotecnia en Internet**

Internet ha beneficiado la mercadotecnia dándole posibilidades de distribución a muy bajo costo de información a una audiencia global aprovechando la cualidad del medio de poder dar respuestas instantáneas o de motivar al visitante a hacer determinada acción.

Mercadotecnia en Internet no significa hacer un sitio y esperar que este atraiga al cliente por sí solo, sino que tienen que tener en cuenta las características de este nuevo mercado donde la competencia es muy exigente, por la gran cantidad de empresas presentes en este medio. Esto ha dado lugar a que, sobre este medio, tengan que realizarse las investigaciones que apoyen a los especialistas en la toma de decisiones.

En Internet hay más de 555 millones de páginas web, (7) por lo tanto la necesidad de informarse de las debilidades y fortalezas de la competencia, es primordial para el éxito y la conquista del mercado.

### **1.3. Información**

Está claro que lo más importante para una buena investigación de mercado es la calidad de la información que se logre obtener y convertir en conocimiento, pero, ¿qué es información?

La Real Academia Española considera que información es "Comunicación o adquisición de conocimientos que permiten ampliar o precisar los que se poseen sobre una materia determinada". (5)

Definiendo información en el contexto de una empresa, Emery (1977) la considera como: "los conocimientos necesarios para los distintos procesos de decisión que determinan el comportamiento de la organización". (8) Dándole el valor que tiene para una empresa en sus decisiones y planificación.

### **1.4. Observación y vigilancia de Internet**

El futuro de cualquier empresa depende de su capacidad de reacción ante el cambio, o sea, a las innovaciones tecnológicas y de mercado; lo que obliga a las empresas a investigar acerca de las limitaciones y oportunidades que se tienen a la hora de adaptarse a la evolución de la ciencia y la tecnología.

Cuando el objeto a investigar es Internet, el volumen de información es enorme, manualmente es muy engorroso hacerlo. Tampoco se disponen de herramientas informáticas tan inteligentes que solo de decirle lo que se necesita, lo hagan. Por tanto, se hace necesario la adopción de herramientas especializadas en la captura y estructuración documental de información relevante para la empresa.

#### **1.4.1. Observatorio tecnológico**

Según Legite (2003) un Observatorio Tecnológico es “un centro encargado de gestionar el conocimiento de las organizaciones a través de la Vigilancia Tecnológica, función principal del mismo, permitiendo generar nuevos conocimientos y ofrecerlos a la sociedad en general. Además, permite establecer vínculos con otras organizaciones con el fin de compartir y recibir información necesaria para la toma de decisiones”. (9)

Permite anticipar las nuevas oportunidades de negocio, sus riesgos e incertidumbres, mediante la difusión de información sobre nuevas tecnologías y tendencias del sector. Su principal tarea es convertir el “ruido”, “la inundación” de datos y noticias de Internet, en información con sentido.

Un Observatorio Tecnológico es una vía para la potenciación de la capacidad de detección de cambios y avances tecnológicos, de su grado de maduración y de oportunidades de mercado. Se basa en el empleo de técnicas y procedimientos de Vigilancia Tecnológica. (9)

#### **1.4.2. Vigilancia tecnológica**

La Vigilancia Tecnológica es un proceso organizado, selectivo y permanente, de captar información del exterior y de la propia organización sobre ciencia y tecnología, seleccionarla, analizarla, difundirla y comunicarla, para convertirla en conocimiento para tomar decisiones con menor riesgo y poder anticiparse a los cambios. (10)

Es el arte de descubrir, recolectar, tratar, almacenar informaciones y señales pertinentes, débiles y fuertes, que permitirán orientar el futuro y proteger el presente y el futuro de los ataques de la competencia. Transfiere conocimientos del exterior al interior de la empresa. (11)



## Capítulo 1: Fundamentación teórica

O sea, es la forma sistemática de captación y análisis de información científico-tecnológica para intentar comprender y explicar la evolución de la tecnología. Las empresas, podrán anticiparse a los efectos negativos que sobre su actividad puede tener dicha evolución, aprovechar las oportunidades que la misma ofrece y apoyar a la toma de decisiones, ayudando a la identificación de los escenarios más probables y al estudio del impacto previsible sobre la actividad de la empresa de las tecnologías emergentes y los desarrollos tecnológicos que en el inmediato futuro se deriven de su evolución.

### **A través de los estudios de vigilancia tecnológica:**

- Se detectan fuentes de información esenciales para hacer frente a las decisiones tecnológicas.
- Se extrae información relevante sobre tendencias tecnológicas, novedades, invenciones, potenciales socios o competidores, aplicaciones tecnológicas emergentes.
- Se contemplan aspectos regulatorios y de mercado que pueden condicionar el éxito de una innovación tecnológica.

Toda esta información codificada y analizada brinda a un decisor, ya sea una empresa o institución científica, la posibilidad de trazar planes y formular estrategias tecnológicas, minimizando la incertidumbre.

### **Entre los beneficios de una buena vigilancia tecnológica están: (12)**

- Conocer cambios de las tecnologías y cambios en los mercados próximos al entorno.
- Reducción de riesgos de toma de decisiones, al conocer mejor qué posición ocupar con las estrategias seleccionadas.
- Conocer hacia dónde avanzar, porque se podrán conocer las nuevas necesidades de los clientes.
- Llevar los esfuerzos organizacionales hacia nuevos terrenos y tendencias clave del avance en todos los aspectos organizativos, innovar hacia procesos productivos, productos, capital humano.
- Conocer la competencia, búsqueda de alianzas con nuevos socios o asesoramiento de expertos.

Todo este proceso de captura y análisis de información se convierte en conocimiento para la empresa y su aprovechamiento proporciona las investigaciones de mayor calidad y por consiguiente, reduce el riesgo de incertidumbre en la toma de decisiones.

### **1.5. Cibermetría**

Estas investigaciones están centradas en el estudio de información de Internet, pero ¿cómo está dicha información en Internet?

Las páginas web son en su mayoría un documento electrónico que contiene información específica de un tema en particular. Las web se almacenan en algún sistema de cómputo que se encuentre conectado a la red mundial de información denominada Internet, de tal forma que este documento pueda ser consultado por cualquier persona que se conecte a esta red y cuente con los permisos apropiados para hacerlo. Una página web es la unidad básica del WWW<sup>1</sup>.

El continuo crecimiento de Internet y el excesivo volumen de información que posee, condicionó la necesidad de una ciencia que estudiara detalladamente el comportamiento cuantitativo y cualitativo de la web y del ciberespacio en general, la Cibermetría.

**La Cibermetría** estudia la aplicación de las técnicas informétricas<sup>2</sup> a cualquier tipo de información disponible en Internet. A su vez, formando parte de esta, la Webmetría se basa en la aplicación de la Informetría, y otras técnicas nuevas de medida, específicamente a la información disponible a través de la WWW, estudiando con profundidad a la web. Es el estudio de los aspectos cuantitativos de la construcción y uso de los recursos de información, estructuras y tecnologías en Internet, desde perspectivas bibliométricas<sup>3</sup> e informétricas. (13)

Entre los principales objetivos de la Cibermetría se encuentran la construcción de indicadores. (14) Estos permiten el estudio de aspectos tales como la evolución del tamaño de la web y la forma en que la misma pudiera hacer algún tipo de cambio; aunque no necesariamente en el tamaño sino en ámbitos como la estructura y el contenido.

Uno de los frentes abiertos en el campo de la Cibermetría es el estudio de aspectos importantes de la web, a través de indicadores cibernéricos por ser una medida de relevancia en la toma de decisiones,

---

1 WWW: Acrónimo de Red informática mundial (*World Wide Web* en inglés )

2 Informetría: Es el estudio de los aspectos cuantitativos de la información en cualquier forma

3 Bibliometría: Es el estudio de los aspectos cuantitativos de la producción, diseminación y utilización de la información registrada.

que cuantifica aspectos de creación, difusión y aplicación de la ciencia y la tecnología en la medida en que están representadas en Internet. (13)

### 1.5.1. Indicadores bibliométricos

Los indicadores, en términos generales, representan una medición agregada y compleja que permite describir o evaluar un fenómeno, su naturaleza, estado y evolución. (15)

Con los indicadores bibliométricos, se podrán determinar, entre otros, los siguientes aspectos: (14)

- El crecimiento de cualquier campo del saber.
- El envejecimiento de los campos científicos.
- La evolución cronológica de la producción científica.
- La productividad de los autores o instituciones.
- La colaboración entre científicos e instituciones.
- El impacto o visibilidad de las publicaciones en la comunidad científica.
- La dispersión de las publicaciones científicas entre las diversas fuentes.

### 1.5.2. Aplicaciones de la Cibermetría

La gran masa de información que brinda Internet, ha proporcionado argumentos para poder realizarle estudios profundos a la web partiendo que es su componente más usado: (16)

- Clasificación de las páginas web por tipos de documentos.
- Estadísticas de uso y usuarios de las páginas web en un periodo de tiempo determinado.
- Ordenar los sitios web más visitados y páginas personales según el tipo de documento.
- Los tipos de colecciones electrónicas disponibles en cada sede web.
- Factor de impacto de la web y productividad de los autores.
- Análisis del contenido de las páginas web.
- Clasificar los sitios por protocolos utilizados por los enlaces y dominios a los que pertenecen.
- Clasificar los sitios por recursos que utilicen.

Los indicadores cibernéticos ya han sido incorporados a los estudios de descripción y evaluación de la actividad científica, como por ejemplo en las caracterizaciones de diversos entornos, como países, o sencillamente instituciones o empresas.

De forma general, los indicadores cibernéticos pueden agruparse en tres grandes grupos o tipos de medida: (16)

- **Medidas descriptivas:** Como su nombre indica, miden fundamentalmente el tamaño o número de objetos encontrados en cada sede, la riqueza de las páginas, ficheros medianos o ricos en contenido, densidad de enlaces totales y únicos. Se utilizan para medir la penetración de Internet desde el punto de vista de los contenidos que por países, regiones, organizaciones o grupos de individuos pueden presentar.
- **Medidas de visibilidad e impacto:** Se basan en el carácter hipertextual de la web y exploran los patrones de enlace entre páginas y sedes de distintas procedencias. El número y diversidad de enlaces externos recibidos, su volumen respecto a los contenidos y objetos de enlace (llamado Factor de Impacto Web) o índices que se construyen de acuerdo con el peso relativo de las sedes de origen de los enlaces, como el *PageRank* de Google, aunque en otras instituciones se puede visualizar como *Ranking*.
- **Medidas de popularidad:** El consumo de información, medido en términos numéricos y características de las visitas desde la web, resulta un atractivo, aunque extremadamente complejo de implementar, es un método de evaluación, que ayuda a saber la popularidad que puede presentar la información o la forma de representarla.

### 1.5.3. Herramientas cibernéticas

Las herramientas en las que se basa la Cibermetría son los robots de búsqueda para la extracción de datos cuantitativos de las sedes web previamente identificadas, permitiendo obtener datos complementarios para la realización de estudios de este tipo. Los robots de búsqueda forman parte de los motores de búsqueda. Para la tarea de realizar recolección de datos para un estudio webmétrico es recomendable utilizar los robots de búsqueda.

Un robot es un programa que atraviesa una estructura de hipertexto recuperando ese enlace y el resto de los que están referenciados allí. Los robots son usualmente llamados "*Web Wanderers*", "*Web Crawlers*" o "*Spiders*" (arañas de búsqueda) y se suele imaginar que se mueven entre los sitios como si fuesen virus, lo que no es el caso, un robot simplemente visita los sitios y extrae los enlaces que están incluidos dentro de estos. (13) Son programas que inspeccionan las páginas del WWW de forma metódica y automatizada.

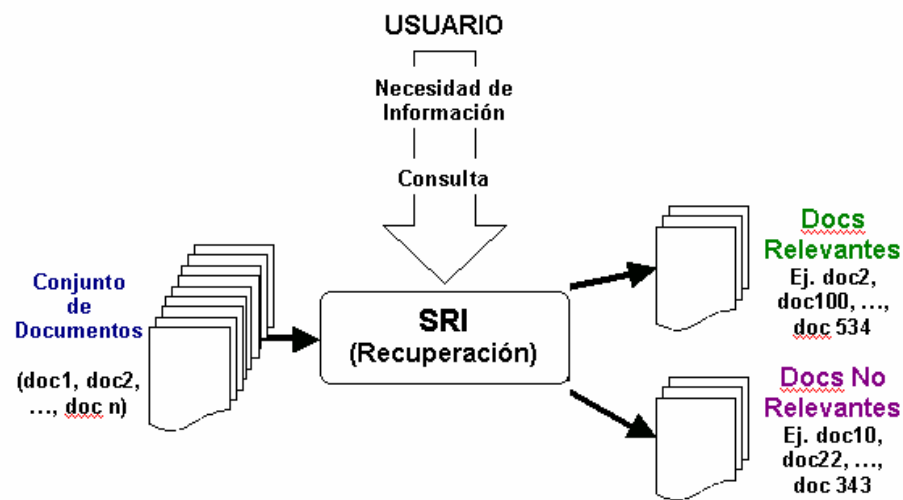
## **1.6. Recuperación de Información**

Resulta necesario tratar con toda información electrónica que se puede obtener para que sirva a los especialistas. En las ciencias de computación existe un área, la Recuperación de Información (RI) que estudia y propone soluciones a este escenario planteando modelos, algoritmos y heurísticas. Esta no es un área nueva, sino que viene desarrollándose desde finales de la década de 1950. Sin embargo, en la actualidad adquiere un papel mucho más importante, se puede plantear que disponer o no de la información justa en tiempo y forma puede resultar en el éxito o el fracaso de cualquier operación. (17) De aquí la importancia de los Sistemas de Recuperación de Información (SRI) que pueden manejar esta situación de manera eficaz.

Baeza-Yates (2011) define "la Recuperación de Información trata con la representación, el almacenamiento, la organización y el acceso a ítems de información". (18) Cabe aclarar que en la definición anterior, el elemento información es no estructurado, tal como documentos de texto libre (por ejemplo, un archivo de texto que contenga La Biblia) o semi-estructurado, como lo son las páginas web.

### **1.6.1. La problemática de la Recuperación de Información**

De forma general según Baeza-Yates (2011) "el problema de la RI puede ser estudiado desde dos puntos de vista: el computacional y el humano. El primer caso tiene que ver con la construcción de estructuras de datos y algoritmos eficientes que mejoren la calidad de las respuestas. El segundo caso corresponde al estudio del comportamiento y de las necesidades de los usuarios". (18)



**Figura 1: La problemática de la RI. Tomado de Baeza-Yates 2011.**

Si se analiza la problemática de la RI desde un alto nivel de abstracción (Figura anterior) se puede establecer que:

- Existe una colección de documentos que contienen información de interés (sobre uno o varios temas).
- Existen usuarios con necesidades de información, quienes las plantean al SRI.
- Como respuesta, el sistema retorna referencias a documentos “relevantes”, es decir, aquellos que satisfacen la necesidad expresada, generalmente en forma de una lista ordenada.

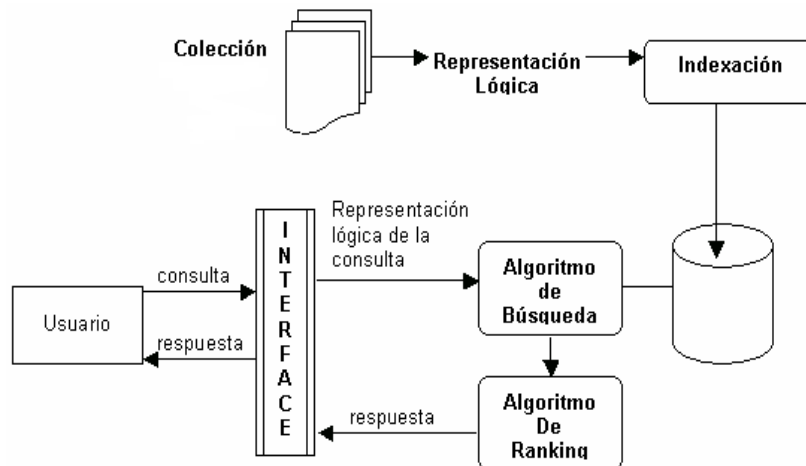
La respuesta “ideal” de un SRI está formada solamente por documentos relevantes a la consulta, pero, en la práctica, esto no se logra del todo, debido a que existe el problema de compatibilizar la expresión de la necesidad de información y el lenguaje y de los documentos. Entonces, el SRI recupera la mayor cantidad posible de documentos relevantes, minimizando la cantidad de documentos no relevantes en la respuesta. En términos de eficiencia, se plantea la idea de precisión de la respuesta, es decir, cuando más documentos relevantes contengan el conjunto solución más precisa será.

Para cumplir con sus objetivos, un SRI debe realizar algunas tareas básicas, las cuales se encuentran planteadas en cuestiones computacionales: (19)

- Representación lógica de los documentos y, opcionalmente, almacenamiento del original.
- Representación de la necesidad de información del usuario en forma de consulta.

- Evaluación de los documentos respecto a la consulta para establecer la relevancia de cada uno.
- Prioridad de los documentos considerados relevantes para formar el conjunto solución o respuesta.
- Presentación de la respuesta al usuario.
- Retroalimentación o refinamiento de las consultas (para aumentar la calidad de las respuestas).

En la figura siguiente se puede apreciar con mayor detalle la arquitectura básica de un SRI, el tratamiento de los documentos y la interacción con el usuario.



**Figura 2: Arquitectura básica de un SRI. Tomado de Bordignon 2007.**

Se inicia desde un conjunto de documentos de texto, los cuales están compuestos por sucesiones de palabras que forman estructuras gramaticales (por ejemplo, oraciones y párrafos). Tales documentos están escritos en lenguaje natural y expresan ideas de su autor sobre un determinado tema. Para poder realizar operaciones sobre estos documentos, es necesario obtener primero una representación lógica de todos, la cual puede consistir en un conjunto de términos, frases u otras unidades (sintácticas o semánticas).

A partir de la representación lógica, existe un proceso de indexación que llevará a cabo la construcción de estructuras de datos (normalmente denominadas índices) que la almacene y soporte búsquedas eficientes.

El algoritmo de búsqueda acepta como entrada una expresión de consulta de un usuario y verificará en el índice cuáles documentos pueden satisfacerlo. Luego, un algoritmo de *ranking* determinará la relevancia

de cada documento y retornará una lista con la respuesta. Se establece que el primer ítem de dicha lista corresponde al documento más relevante respecto a la consulta y así sucesivamente en orden decreciente.

La interfaz de usuario permite que este especifique la consulta mediante una expresión escrita en un lenguaje preestablecido (palabras claves) y además sirve para mostrar las respuestas retornadas por el sistema. (19)

## **1.7. Modelos de los SRI**

Existen varios modelos matemáticos o técnicas empleadas en la recuperación de información, cada uno de ellos tiene sus ventajas e inconvenientes. Los principales modelos clásicos de recuperación de información son: modelo booleano, modelo espacio vectorial y modelo probabilístico.

### **1.7.1. Modelo booleano**

Es un modelo de recuperación simple, basado en la teoría de conjuntos y el álgebra booleana. Dada su inherente simplicidad y su pulcro formalismo ha recibido gran atención, siendo adoptado por muchos de los primeros sistemas bibliográficos comerciales. Su estrategia de recuperación está basada en un criterio de decisión binario (pertinente o no pertinente) sin ninguna noción de escala de medida, sin noción de un emparejamiento parcial en las condiciones de la pregunta. (20)

En un sistema de este tipo, los documentos se encuentran representados por conjuntos de palabras claves (términos). La indización se realiza asociando un peso binario a cada término del índice: 0 si el término no aparece en el documento y 1 si aparece al menos una sola vez. Las búsquedas consisten en un conjunto de palabras claves conectadas por los operadores lógicos *AND*, *OR*, *NOT*. El grado de similitud entre un documento y una consulta también será binario y un documento será relevante cuando su grado de similitud sea igual a 1, de lo contrario, el documento no tendrá ninguna relevancia en cuanto a la consulta. (21)

### **1.7.2. Modelo probabilístico**



El modelo probabilístico de recuperación de información también conocido como *Binary Independence Retrieval* intenta resolver el problema de la recuperación de información desde una óptica probabilística. Se basa fundamentalmente en estimar la probabilidad de que un documento sea relevante para una consulta dada, o sea, dada una consulta, existe un conjunto de documentos que contienen exactamente los documentos relevantes y no otros. (20)

Matemáticamente, la similitud entre un documento (d) y una consulta (q) es:

$$sim(d,q) = \frac{P\left(\frac{R}{d}\right)}{P\left(\frac{R'}{d}\right)}$$

Donde:

$P(R|d)$  es la probabilidad de que el documento d sea relevante en el conjunto de documentos relevantes R y  $P(R'|d)$  es la probabilidad de que el documento d no sea relevante en el conjunto de documentos relevantes R, por lo que:

Un documento será relevante si  $P(R|d) > P(R'|d)$ .

La probabilidad de que un documento sea relevante para una consulta dada, depende de la forma en que se representan los documentos en el sistema. Una vez calculada la probabilidad de los documentos relevantes para la consulta dada, son ordenados descendientemente y mostrados al usuario.

### 1.7.3. Modelo espacio vectorial

El modelo de espacio vectorial, es uno de los modelos que más se emplean en los sistemas de recuperación de información en la actualidad. En este modelo, cada documento se representa como un vector t-dimensional, donde t corresponde a la cantidad de términos asociados al documento. Cada elemento del vector, tomará un valor de relevancia (peso) dependiendo de cuán representativo o no sea el término en el documento. Un término que no aparezca en el documento, tendrá un peso igual a 0. (21)

Existen varias formas para calcular el peso asociado a un término dentro de un documento, siendo la más extendida la ponderación **TF.IDF** que consiste en multiplicar dos factores que reflejan la importancia de los términos en el documento:

El primer factor **TF** (abreviatura de *Term Frequency*), pretende reflejar la importancia de los términos en los documentos, concediendo mayor importancia a los términos que aparecen con mayor frecuencia en los documentos. El segundo factor **IDF** (abreviatura de *Inverse Document Frequency*), o Inverso de la Frecuencia de Documentos, dará mayor importancia a un término cuanto menor sea el número de documentos de la colección en los que aparece dicho término. Es decir, el IDF de un término es inversamente proporcional al número de documentos en que aparece dicho término. (21)

Si un término aparece mucho en un documento su peso aumentará, pero si aparece en muchos documentos, disminuirá, pues este término no es muy útil para distinguir un documento de otro (32). Así mismo, se crea un vector para cada una de las consultas introducidas por los usuarios con las mismas características de los vectores por cada documento de la colección. Esto permite calcular, fácilmente un valor de similitud entre el vector de una consulta y los vectores de cada uno de los documentos. La similitud entre dos documentos será dada por la distancia coseno.

$$\text{sim}(d_i, d_j) = \vec{d}_i \cdot \vec{d}_j = \sum_{r=1}^k w_{r,i} \times w_{r,j}$$

Así mismo, se puede calcular la relevancia de los documentos para una consulta de acuerdo a la siguiente fórmula:

$$R(\vec{q}, \vec{d}_i) = \cos(\vec{q}, \vec{d}_i) = \frac{\langle \vec{q}, \vec{d}_i \rangle}{|\vec{q}| \cdot |\vec{d}_i|} = \frac{\sum_{j=1}^n w_{j,q} \cdot w_{j,i}}{\sqrt{\sum_{j=1}^n w_{j,q}^2} \cdot \sqrt{\sum_{j=1}^n w_{j,i}^2}}$$

Lo cual permite establecer un orden en que se muestran los documentos recuperados por el sistema de recuperación de información.

## 1.8. Soluciones existentes

Un estudio de estos procesos, los de recuperación de información, estudios cibernéticos, vigilancia tecnológica, llevaron a algunos sistemas o servicios como:

### 1.8.1. C@BI

C@bi (Cámaras Búsqueda de Información) es una empresa que ofrece un servicio de información sobre Comercio Exterior en base a la existente en Internet. Para ello, C@bi cuenta con una serie de expertos camerales que, empleando una serie de páginas web clasificadas y adecuadas para el objetivo señalado y utilizando una metodología específica que facilita la búsqueda, ofrecen de forma rápida y precisa la información que la empresa cliente le solicita. (22)

### **¿Cómo funciona?**

La empresa cliente solicita información al servicio C@bi. El experto C@bi, si lo estima oportuno, se pondrá en contacto con la empresa con el fin de analizar su consulta y ofrecer una mejor respuesta, verificando si, desde su experiencia en Comercio Exterior, lo que solicita la empresa es lo que realmente necesita. “Una correcta respuesta exige una buena pregunta.” (22)

El experto C@bi busca por la empresa cliente y le ofrece en un breve plazo de tiempo la información solicitada. La comunicación entre la empresa y la Cámara podrá ser por medio de e-mail, fax, teléfono o mediante entrevista personal con el técnico especialista C@bi.

### **1.8.2. Vigitec**

Es un Grupo de Vigilancia Tecnológica de la Dirección de Información de la Universidad de Ciencias Informáticas (UCI), que pone a disposición numerosos productos y servicios de Vigilancia Tecnológica e Inteligencia Empresarial, cuya finalidad fundamental es garantizar una toma de decisiones documentada. (23)

Dichos productos y servicios están diseñados para apoyar todos los procesos desarrollados en la UCI: Docente, Productivo e Investigativo. Varios de los productos que se brindan en la Dirección, son Estudios de Tendencias, Investigaciones de Mercado, Perfiles Estratégicos (país, compañía y sector) y Alertas Tecnológicas, entre otros. (23)

Además de la elaboración de dichos productos, los especialistas de la Dirección, están capacitados para adiestrar a otros especialistas de la UCI, en los temas de Vigilancia Tecnológica, Inteligencia Empresarial, Alfabetización Informacional y Gestión Documental, a través de la realización de cursos y talleres sobre

estos temas, así como, brindar asesoría para la implementación de dichos procesos o sistemas en las facultades, centros productivos, departamentos y direcciones de la Universidad. (23)

Anteriormente en Vigitec cuando se trabajaba sobre la plataforma Windows, contaba con algunas herramientas (endnote x, ucinet, bibexcel, pajek, entre otras) que los apoyaban los procesos de vigilancia tecnológica, pero estas no son libres. Desde que se migró a la plataforma Linux, los procesos de vigilancia tecnológica los realizan basados en servicios RSS y/o boletines que brindan algunos sitios identificados.

### 1.8.3. UCITRAVELFIVE

Es un estudio que se le realizó con una lista de partida de 240 sitios web de la UCI para analizar la evolución de dichos sitios. Se realizó un análisis de datos exhaustivos a estos sitios web ofreciendo datos comparativos entre cada estudio, lo que constituye la base fundamental para medir la evolución de la web. (16)

Fue realizado en la UCI en el polo Soluciones Informáticas para Internet (SINI), por el proyecto Generador de Estudios Webmétricos (GEWEB) del Grupo de Proyectos de Cibermetría Aplicada (CIBA). Para el estudio se utilizó un *spider*, el WIRE, desarrollado en el Centro de Investigación de la Web (CIW) de Chile. (16)

Utilizando este *spider* se recogen datos estadísticos de la web que permiten establecer pronósticos y tendencias que ayuden a tener un control de las tecnologías que actualmente son utilizadas y establecer líneas de trabajo en función de mejorar el uso de las mismas y acelerar este proceso.

### 1.8.4. Conclusiones del estudio

A través de este estudio se pudo identificar algunos procesos que realizan o servicios que brindan algunas empresas, como C@bi, pero en esta los especialistas realizan las investigaciones manualmente, en sitios que ya tienen definidos.

Por otra parte, en Vigitec toda la gestión de la información y los reportes los gestiona el Equipo de Vigilancia Tecnológica, o sea los especialistas, que cada cierto tiempo hacen levantamiento de información y actualizan sus bases de datos.

Con el estudio desarrollado por GEWEB se pudo identificar la necesidad de incluir en el sistema que se propone desarrollar, una herramienta que incorpore funcionalidades de análisis estadísticos brindando los beneficios que aporta a una investigación de mercado y a realizar vigilancia tecnológica. Es por esta razón que se propone el uso del *spider* WIRE del que se describen algunas características a continuación:

**WIRE Crawler:** Este proyecto chileno WIRE es un esfuerzo iniciado por el Centro de Investigación Web (*Center for Web Research*) dirigido por el Dr. Ricardo Baeza-Yates, para crear una aplicación que permita la recuperación de información; diseñada para ser utilizada en la web. (24)

**WIRE incluye:**

1. Un formato simple para almacenar una colección de documentos web.
2. Un rastreador web.
3. Herramientas para la extracción de las estadísticas de la colección.
4. Herramientas para la generación de informes acerca de la colección.

Las principales características del software WIRE son las siguientes: (24)

- **Escalabilidad:** diseñado para trabajar con grandes volúmenes de documentos, ha sido probado con varios millones de documentos.
- **Prestaciones:** programado en C/C++ para un alto rendimiento.
- **Configurable:** todos los parámetros para el rastreo y la indexación se pueden configurar a través de un archivo XML.
- **Análisis:** incluye varias herramientas para analizar, extraer estadísticas y la generación de informes sobre sub-conjuntos de la web, por ejemplo: la web de un país o de una gran intranet.
- **De código abierto:** el código está libremente disponible.

Además el sistema está diseñado para centrarse en la evaluación de la calidad de la página, utilizando diferentes estrategias de rastreo y la generación de datos web para la caracterización de los estudios.

### 1.9. Estudio de otros *Spiders*

Hasta el momento se pudo identificar la necesidad de realizar vigilancia tecnológica y la forma de automatizarla con el uso del *spider* WIRE; pero uno de los problemas plantea la necesidad de automatizar

los procesos de recuperación de información. Por esto, se realiza un estudio de otros *spider* especializados en estas funciones.

### 1.9.1. Apache Lucene

Lucene es una novedosa herramienta que permite tanto la indización como la búsqueda de documentos. Creada bajo una metodología orientada a objetos e implementada completamente en Java, no se trata de una aplicación que pueda ser descargada, instalada y ejecutada sino de una API<sup>4</sup> flexible, muy potente y fácil de utilizar, a través de la cual se pueden añadir, con pocos esfuerzos de programación, capacidades de indización y búsqueda a cualquier sistema que se esté desarrollando. En septiembre de 2001 pasó a formar parte de la familia de Apache. A partir del año 2005 este SRI tomó gran auge en el mundo y comenzó a ser usado en algunos sitios y aplicaciones. (25)

### 1.9.2. Minion

Este sistema está diseñado para ser altamente configurable y se propone que sea usado en investigaciones así como en entornos empresariales. Presenta indización y recuperación concurrente. En mayo de 2008 fue publicado su código fuente y la versión desarrollada hasta el momento es parte del proyecto de Tecnologías para la Búsqueda Avanzada en los laboratorios de la compañía Sun Microsystems/Oracle. Actualmente es utilizado en el sistema Aura, desarrollado en los laboratorios de Sun así como en aplicaciones web de la misma compañía. (26)

### 1.9.3. MnoGoSearch

MnoGoSearch es un motor de búsqueda completo de código abierto y basado en SQL. Consiste en dos partes. La primera parte es un mecanismo de indización el cual se mueve a través de vínculos de hipertexto HTML y almacena información acerca de los documentos en la base de datos. La segunda parte es una interfaz web la cual muestra en el navegador un formulario HTML y los resultados de búsquedas. (27)

**Entre sus principales características destacan:**

---

4 API: Acrónimo de Interfaz de Programación de Aplicaciones (*Application Programming Interface* en inglés)

## Capítulo 1: Fundamentación teórica

- Soporte para diversos protocolos: HTTP, HTTPS, FTP, NNTP.
- Analizadores incorporados para diversos formatos de archivo: text/html, text/xml, text/plain y audio/mpeg.
- Soporte para autenticación de proxy.
- Indización multihilo.
- Interfaces web CGI, Perl y PHP.
- Lenguaje de consulta booleano.
- Soporte para la mayoría de los conjuntos de caracteres modernos.
- Soporte para múltiples bases de datos: MySQL, PostgreSQL, SQLite, Mimer, Virtuoso, Interbase, Oracle, MS SQL, DB2, Sysbase.
- Posee una API externa para PHP.
- Configuración a través de ficheros.
- Manejo de clústeres de base de datos.

De estos últimos tres se realiza un análisis más detallado buscando aquel que preste mayor beneficio al sistema que se propone; evaluando sus características más importantes en una tabla comparativa, mostrando sus ventajas y desventajas.

**Tabla 1: Comparación entre motores de búsqueda.**

Características	Apache Lucene	Minion	MnoGoSearch
Documentos que indiza	HTML, TXT, PDF, XML, PPT, DOC, RTF	Texto plano	HTML, TXT, PDF, XML, PPT, DOC, RTF, JPG, GIF, PNG, ODT
Índice incremental	Si	Si	Si

<b>Modelo RI</b>	Combinación del espacio vectorial y el booleano	Combinación del modelo booleano, consultas relacionales y de proximidad	Combinación del espacio vectorial y el booleano
<b>Operadores de búsqueda</b>	Booleanos, frases, términos <i>wildcard</i> , campos	Booleanos, proximidad, campos	Booleanos, operadores +/-, frases, proximidad, campos
<b>Creación de resumen</b>	Si	Si	Si
<b>Lenguajes de programación</b>	Java, C++, C#, Python, Perl, Ruby, PHP	Java	Java, C++, C#, Python, Perl, Ruby, PHP
<b>Configuración</b>	Interfaz	Interfaz(en desarrollo)	Ficheros de configuración
<b>Plataforma</b>	Windows, Mac Os, GNU/Linux, Unix	Windows, Mac Os, GNU/Linux, Unix	Windows, Unix GNU/Linux.
<b>Licencia</b>	Apache Software Foundation (ASF)	GNU Public licence GPL	GNU Public licence GPL

### 1.9.1. Conclusiones del estudio de los *spider*

Se propone utilizar el MnoGoSearch ya que además de las características y ventajas anteriormente expuestas, su indizador crea un sistema de índice invertido que consiste en una lista ordenada de todas y cada una de las palabras que aparecen en los distintos documentos asociadas a los documentos concretos en los que aparecen. También es un sistema ya probado en la UCI formando parte del buscador interno de la universidad, ORION.



---

### **1.10. Conclusiones del capítulo**

Con la investigación llevada a cabo en este capítulo se puede arribar a las siguientes conclusiones:

- Se realizó un análisis de las condiciones y los problemas que caracterizan el objeto de estudio ayudando a un mejor entendimiento de lo que se necesita desarrollar.
- Se investigó acerca de los diferentes conceptos asociados a Cibermetría, Webmetría, observación tecnológica, mercadotecnia y sistemas de recuperación de información necesarios para una mejor comprensión del sistema que se propone desarrollar; la cual dará solución a la situación problemática existente actualmente en la empresa Albet S.A.
- Un estudio sobre sistemas o soluciones homólogas, no arrojó resultados satisfactorios sobre alguna herramienta que satisfaga las necesidades del cliente. Aunque si se identificaron vías para darle solución al problema planteado.

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

### CAPÍTULO 2: ENTORNO DE DESARROLLO Y OBTENCIÓN DE REQUISITOS

#### 2.1. Introducción

Con el objetivo de lograr un producto que no sólo solucione los problemas existentes sino que además logre unificar calidad y eficiencia, se adoptaron las herramientas, lenguajes y metodologías del trabajo de diploma “Propuesta de arquitectura para el sistema de gestión de información de mercadotecnia para Albet S.A.” En este capítulo se presenta el estudio realizado del entorno de desarrollo así como las características de la propuesta de solución; los elementos relacionados al modelado del negocio y se definen los flujos de proceso. Se definen los métodos para la obtención de requisitos así como una descripción de las principales funcionalidades, además se define la lista de reserva del producto y los requisitos no funcionales.

#### 2.2. Entorno desarrollo para la descripción de la propuesta

##### 2.2.1. Metodologías de desarrollo de software

Una metodología es el conjunto de fases, procedimientos, reglas, técnicas, herramientas, documentación y aspectos de formación para los desarrolladores de un sistema. (28)

El desarrollo de software es un proceso difícil y lleno de riesgos donde la metodología define quién debe hacer qué, cuándo y cómo debe hacerlo. El fracaso en el desarrollo de un sistema está estrechamente relacionado con una mala selección de la misma. No existe un proceso de desarrollo universal, aplicable a todo proyecto. Las características del equipo de desarrollo, el dominio de aplicación, el tipo de contrato, la complejidad y envergadura del proyecto, son factores que hacen necesario la adopción de una u otra metodología de desarrollo. (29)

Actualmente existen metodologías tradicionales y ágiles: (30)

- Las metodologías tradicionales están basadas en la planificación inicial de todo el trabajo a realizar. Establecen estrictamente las actividades involucradas, los roles definidos, los artefactos que se deben producir, las herramientas y notaciones que serán utilizadas así como el modelado y la documentación detallada.
- Las metodologías ágiles dan un mayor valor al personal responsable del desarrollo del software, a la colaboración con el cliente y al desarrollo incremental con iteraciones muy cortas donde se entregan

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

versiones funcionales del sistema a los interesados. Son capaces de responder a los cambios que puedan surgir a lo largo del proyecto

Teniendo en cuenta que el cliente (Dirección de Mercadotecnia de la empresa Albet S.A.) se encuentra en el campus universitario y las condiciones actuales de desarrollo requieren entregas de productos funcionales en cortos períodos de tiempo, se propone el uso de las metodologías ágiles para el desarrollo del sistema.

Las metodologías Scrum y Programación extrema (*eXtreme programming*) son unas de las más usadas en el proceso de planificación y construcción de proyectos pequeños, en dichos proyectos el cliente forma parte del equipo de desarrollo.

### 2.2.1. SXP (Scrum y Programación extrema)

SXP, es un híbrido cubano de metodologías ágiles, que ofrece una estrategia tecnológica, a partir de la introducción de procedimientos ágiles que permitan actualizar los procesos de software para el mejoramiento de la actividad productiva fomentando el desarrollo de la creatividad, aumentando el nivel de preocupación y responsabilidad de los miembros del equipo, ayudando al líder del proyecto a tener un mejor control del mismo. Consiste en una programación rápida o extrema, cuya particularidad es tener como parte del equipo, al usuario final, pues es uno de los requisitos para llegar al éxito del proyecto. (31)

Consta de 4 fases principales:

- **Planificación-Definición:** donde se establece la visión, se fijan las expectativas y se realiza el aseguramiento del financiamiento del proyecto.
- **Desarrollo:** es donde se realiza la implementación del sistema hasta que esté listo para ser entregado.
- **Entrega:** puesta en marcha.
- **Mantenimiento:** donde se realiza el soporte para el cliente.

De cada una de ellas se despliegan 7 flujos de trabajo: concepción inicial, captura de requisitos, diseño con metáforas, implementación, prueba, entrega de la documentación, soporte e investigación, el cual se

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

utiliza por el equipo de desarrollo cuando sea necesario, es decir, es un flujo que se puede mover y utilizarlo en cualquier parte del ciclo de vida del proyecto. (31)

De estos flujos se realizan numerosas actividades tales como el levantamiento de requisitos, la priorización de la Lista de Reserva del Producto, definición de las Historias de Usuario, diseño, implementación, planificación de las iteraciones y las actividades que se van a realizar para lograr el producto, pruebas, además de las tareas necesarias para realizar las investigaciones para documentar todo el proceso. (31)

La integración entre estas dos metodologías define un proceso de software que se ilustra en la siguiente figura:

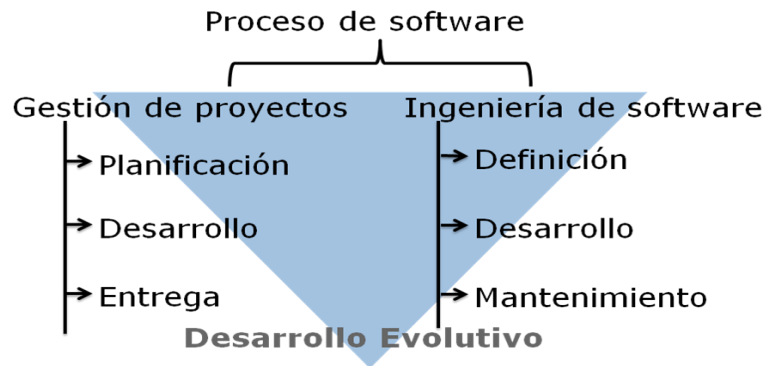


Figura 3: Estructura de la metodología SXP. Tomado de Peñalver Romero 2007.

### 2.2.2. Lenguaje Unificado de Modelado (UML) 2.0

El Lenguaje Unificado de Modelado (UML) es una de las innovaciones en el mundo tecnológico de desarrollo de software que más expectativas y entusiasmo han generado en muchos años, expectativas que se han cumplido y a su vez generado muchas más. (32)

UML es un lenguaje de modelado visual que se usa para especificar, visualizar, construir y documentar artefactos de un sistema de software. Captura decisiones y conocimiento sobre los sistemas que se deben construir. Se usa para entender, diseñar, hojear, configurar, mantener, y controlar la información sobre tales sistemas. (32)

#### Propiedades de UML como lenguaje de modelado estándar:

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

- Concurrencia, es un lenguaje distribuido y adecuado a las necesidades de conectividad actuales y futuras.
- Ampliamente utilizado por la industria desde su adopción por OMG (*Object Management Group*).
- Reemplaza a decenas de notaciones empleadas con otros lenguajes.
- Modela estructuras complejas.
- Las estructuras más importantes que soportan tienen su fundamento en las tecnologías orientadas a objetos, tales como objetos, clase, componentes y nodos.
- Emplea operaciones abstractas como guía para variaciones futuras, añadiendo variables si es necesario.
- Comportamiento del sistema: casos de uso, diagramas de secuencia y de colaboraciones, que sirven para evaluar el estado de las máquinas.

### Como funciones principales tiene:

- Visualizar: UML permite expresar de una forma gráfica un sistema de forma que cualquiera lo puede entender.
- Especificar: permite especificar cuáles son las características de un sistema antes de su construcción.
- Construir: a partir de los modelos especificados se pueden construir los sistemas diseñados.
- Documentar: los propios elementos gráficos sirven como documentación del sistema desarrollado que pueden servir para su futura revisión.

UML ayuda a los usuarios a entender la realidad desde un punto de vista de la tecnología y la posibilidad de que reflexione antes de invertir y gastar grandes cantidades de dinero en proyectos que no estén seguros en su desarrollo, reduciendo el costo y el tiempo empleado en la construcción de los módulos que construirán el software. (32)

### 2.2.3. Herramienta CASE

Las herramientas CASE (*Computer Aided Software Engineering*) se puede definir como un conjunto de métodos, utilidades y técnicas que facilitan la automatización del ciclo de vida del desarrollo del software, completamente o en alguna de sus fases. (33)

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

### Visual Paradigm 6.4

El *Visual Paradigm Enterprise Edition* es una herramienta fácil de utilizar que apoya las últimas notaciones de UML, ingeniería inversa, y generación de diagramas UML a partir del código. Tiene versiones para licencias propietarias y para distribución libre.

Se ejecuta sobre las plataformas: Windows (98, 2000, XP, Vista, 7), Linux, Mac OS X y Solaris. Además se integra con IDEs (Eclipse, JBuilder, NetBeans, IntelliJ IDEA, JDeveloper and webLogic Workshop) para soportar la fase de implementación de desarrollo de software. (34) Brinda la posibilidad de:

- Realizar análisis textual identificando clases candidatas para la creación de los diferentes diagramas.
- Permite a los desarrolladores de un mismo equipo trabajar sobre un mismo proyecto.
- Facilita el modelado visual permitiendo personalizar las notaciones y formas importándolas de la biblioteca o de archivos.

Se propone utilizar Visual Paradigm ya que es una herramienta que corre sobre cualquier plataforma y se integra con varios IDE. Es fácil de usar y permite modelar bases de datos y transformar diagramas de Entidad-Relación en tablas posibilitando la creación y diseño del modelo de datos.

### 2.2.4. Herramienta para la creación de prototipos

Las herramientas para la creación de prototipos permiten la representación esquemática de un sistema, sin elementos gráficos que muestren contenido. Son realmente útiles como herramientas de comunicación entre programadores, diseñadores y clientes.

### Evolus Pencil 1.3.4

Es una herramienta libre y de código abierto para crear diagramas y prototipos de interfaz gráfica de usuario que todos puedan usar. (35)

Características principales:

- Construcción de prototipos.
- Conexión entre páginas.
- Exportación a formatos HTML, PNG, documento Word y PDF.

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

- Multiplataforma: Puede ser instalado tanto en Windows como Linux, además puede agregarse como complemento para el navegador Mozilla Firefox.
- Tipo de funcionamiento: arrastrar y soltar.

### 2.3. Propuesta de entorno de desarrollo

#### 2.3.1. Framework

Un framework es una estructura conceptual y tecnológica de soporte definida, normalmente, con artefactos o módulos de software concretos, con base en la cual otro proyecto de software puede ser organizado y desarrollado. Típicamente, puede incluir soporte de programas, bibliotecas y un lenguaje interpretado entre otros programas para ayudar a desarrollar y unir los diferentes componentes de un proyecto. (36)

#### **Symfony 1.4.15**

Symfony es un marco de trabajo que tiene como objetivo fundamental automatizar los patrones más utilizados en la elaboración de sistemas web; además de obligar a clarificar a los programadores el código, establece un estándar de código legible, encapsula operaciones complejas en simples líneas de código, ahorrando mucho más tiempo a la hora de mostrar datos directamente de la base de datos. (36)

Este marco de trabajo está realizado o implementado bajo el lenguaje PHP 5, ha sido utilizado en varias aplicaciones web obteniéndose de esta manera resultados satisfactorios. Es compatible con la mayoría de gestores de base de datos existentes, se puede comentar sobre MySQL, PostgreSQL, Oracle y SQL Server de Microsoft. También otorga facilidades en diferentes plataformas. (36)

Symfony provee una abstracción de la base de datos, que permite una mayor facilidad de obtención de los datos haciendo a la vista y a las acciones independientes del gestor de base de datos. La vista tiene *helpers* que facilitan el trabajo para los diseñadores en el código HTML de la aplicación, aunque es necesario crear una nueva vista, mantiene al modelo y al controlador original; se encarga de mantener alejados a la vista y al modelo de los detalles del protocolo utilizado. (36)

Se propone utilizar el framework Symfony ya que además de las características ventajosas que tiene la capacidad de poner en funcionamiento una gran cantidad de procesos en una sola línea de comando, está

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

desarrollado con PHP. Permite generar de forma automática los formularios de acceso a dato utilizando la base de datos definida para el sistema a desarrollar, formularios que incluyen validación automatizada y relleno automático de datos (“*repopulation*”), lo que asegura la obtención de datos correctos y mejora la experiencia de usuario. Además, esta versión en específico es con la que se desarrolló el SIM de Albet S.A., entonces para asegurar la compatibilidad se restringe a utilizar la versión 1.4.15.

### 2.3.2. Lenguajes del lado del cliente

#### ***Hypertext Markup Language (HTML) 5.0***

HTML es un lenguaje de composición de documentos y especificación de ligas de hipertexto que define la sintaxis y coloca instrucciones especiales que no muestra el navegador, aunque sí le indica cómo desplegar el contenido del documento, incluyendo texto, imágenes y otros medios soportados. Unas de sus principales ventajas es su facilidad de uso porque puede ser editado desde cualquier editor de texto. Tiene dos características esenciales: hipertexto y universalidad. Hipertexto significa que puede crear un vínculo en una página web hacia otra mientras que la universalidad significa que, puesto que los documentos HTML se guardan como archivos de solo texto, cualquier ordenador puede leer una página web. (37)

#### **JavaScript**

Es un lenguaje interpretado por lo que no es necesario compilar los programas para ejecutarlos, se pueden probar en cualquier navegador sin necesidad de procesos intermedios., es orientado a objetos.

JavaScript es sumamente utilizado en el mundo del desarrollo web por ser muy versátil y potente, tanto para la realización de pequeñas tareas como para la gestión de complejas aplicaciones. Con su utilización se puede evitar un poco la sobrecarga del lado del servidor, debido a que con funciones como la validación de formularios (que antes se hacían del lado del servidor), se pueden realizar con este lenguaje y de una manera rápida. (38)

De sus ventajas se puede decir que: (38)

- No requiere tiempo de compilación.
- Los scripts pueden desarrollarse en un período de tiempo relativamente corto.
- Posee características de interfaz, que son gestionadas por el navegador y por el código HTML.



## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

- Los programas JavaScript pueden ser pequeños y compactos.
- No requieren mucha memoria ni tiempo adicional de transmisión.

### AJAX

AJAX<sup>5</sup> se trata de la unión de varias tecnologías que se desarrollan de forma autónoma y que se integran de formas nuevas y sorprendentes. (39)

Las tecnologías que forman AJAX son: (39)

- XHTML y CSS, para crear una presentación basada en estándares.
- DOM, para la interacción y manipulación dinámica de la presentación.
- XML, XSLT y JSON, para el intercambio y manipulación de información.
- XMLHttpRequest, para el intercambio asíncrono de información.
- JS, para unir todas las demás tecnologías.

Para el desarrollo de aplicaciones AJAX se requiere de un conocimiento avanzado de todas y cada una de las tecnologías anteriores. Pero ¿cómo funciona una aplicación AJAX?

En una aplicación web tradicional las acciones del usuario en la página desencadenan llamadas al servidor. Una vez procesada la petición del usuario, el servidor devuelve una página HTML al navegador del usuario. Esta técnica tradicional para desarrollar aplicaciones web funciona correctamente, pero no muestra una buena sensación al usuario. Cuando se realizan peticiones continuas al servidor el usuario tiene que esperar que se recargue la página con los cambios solicitados. (39)

AJAX se ha desarrollado para mejorar la interacción entre el usuario y la aplicación, evitando las constantes recargas de la página, debido a que el intercambio de información con el servidor se produce en un segundo plano. Este segundo plano es la creación de un elemento intermedio, una nueva capa que mejora la respuesta de la aplicación, y que no permite que el usuario encuentre una ventana del navegador en blanco, esperando la respuesta del servidor. (39)

### JQuery 1.8

---

<sup>5</sup> AJAX: Acrónimo de *Asynchronous JavaScript + XML* en inglés, que se puede traducir como “JavaScript Asíncrono + XML”.

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

jQuery es una biblioteca de JavaScript que permite simplificar la manera de interactuar con los documentos HTML, manipular el árbol DOM, manejar eventos, desarrollar animaciones y agregar interacción con la técnica AJAX a páginas web. Incluye soporte para CSS 1. Es un software libre y de código abierto. Compatible con los navegadores Mozilla Firefox 2.0+, Internet Explorer 6+, Safari 3+, Opera 10.6+ y Google Chrome 8+. (40)

jQuery, al igual que otras bibliotecas, ofrece una serie de funcionalidades basadas en JavaScript que de otra manera requerirían de mucho más código, es decir, con las funciones propias de esta biblioteca se logran grandes resultados en menos tiempo y espacio. (40)

### Hojas de estilo en cascada (CSS) 3.0

Las hojas de estilos en cascada (*Cascading Style Sheets*, CSS) es un componente de HTML dinámico, que especifica las características de formato de la página, así como el color de la fuente, el espaciado entre letras, los márgenes o la imagen de fondo, que pueden aplicarse a una o a un grupo de páginas además de evitar el uso excesivo de tablas. Se puede crear hojas de estilos en el mismo código de la página HTML o en un archivo independiente, y luego cargarlo desde la página. CSS define el formato en que se muestra cada objeto en la web. (41)

### 2.3.3. Lenguajes del lado del servidor

#### PHP 5.3.8

El *Hypertext Preprocessor* (PHP) es un lenguaje interpretado de alto nivel embebido en páginas HTML y ejecutado en el servidor. PHP es uno de los lenguajes de lado servidor más extendidos en la web. PHP permite realizar determinadas acciones de una forma fácil y eficaz sin tener que generar programas implementados en otro lenguaje distinto al HTML. PHP ofrece un sinnúmero de funciones para la explotación de bases de datos de una manera llana, sin complicaciones. (42)

Funciones principales de PHP en la web:

- Funciones de correo electrónico.
- Gestión de bases de datos.
- Gestión de archivos.

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

- Tratamiento de imágenes.

### 2.3.4. Entorno Integrado de Desarrollo

Un entorno de desarrollo integrado (*Integrated Development Environment o IDE*) es un conjunto de herramientas que utilizan los programadores para implementar sistemas informáticos. Puede estar diseñada para la utilización de un lenguaje de programación específico o soportar varios de estos.

#### NetBeans 7.1

NetBeans es una herramienta para programadores que les permite escribir, compilar, corregir errores y ejecutar programas. Está escrito en Java, pero puede servir de soporte a cualquier otro lenguaje de programación. Existe también un número enorme de módulos para extender el NetBeans IDE. El NetBeans IDE es un producto libre y gratuito sin restricciones de utilización. (43)

Permite desarrollar aplicaciones web, de escritorio, y móviles usando las plataformas de Java. Soporta otros lenguajes de programación como son C/C++, Ruby y PHP. El proyecto NetBeans es apoyado por una vibrante comunidad de desarrolladores y ofrece una amplia documentación y recursos de capacitación, así como una variada selección de componentes. (43)

Entre las características de la plataforma están:

- Administración de las interfaces de usuario (ej. menús y barras de herramientas).
- Administración de las configuraciones del usuario.
- Administración del almacenamiento (guardando y cargando cualquier tipo de dato).
- Administración de ventanas.
- Marco de trabajo basado en asistentes (diálogo paso a paso).

### 2.3.5. Sistema gestor de base de datos

Un sistema de gestión de bases de datos (SGBD) es una colección de datos interrelacionados y un conjunto de programas para acceder a esos datos. Es un conjunto coordinado de programas, procedimientos, lenguajes, etc. que proporciona, tanto a los usuarios no informáticos como a los programadores, analistas o la administración, los medios necesarios para describir, recuperar y manipular los datos almacenados en la base, manteniendo su integridad, confidencialidad y seguridad. (44)

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

### PostgreSQL 9.1

Es un buen gestor de Base de Datos, es multiplataforma, permite una fácil gestión de usuarios y de la base de datos que contenga el sistema. Sirve de soporte a los lenguajes más utilizados como PHP, Java, C++, Python, C, C#, entre otros. El número de BD que puede contener es ilimitado. Es considerado como el gestor libre más avanzado del mundo. (45)

Algunas características de PostgreSQL:

- **Altamente extensible:** PostgreSQL soporta operadores, funcionales métodos de acceso y tipos de datos definidos por el usuario.
- **Soporte SQL comprensivo:** PostgreSQL soporta la especificación SQL 99 e incluye características avanzadas tales como las uniones (*joins*) SQL 92.
- **DBMS Objeto-Relacional:** aproxima los datos a un modelo objeto-relacional y es capaz de manejar complejas rutinas y reglas tales como consulta SQL declarativas, control de concurrencia multiversión, soporte multiusuario, herencia, arreglos, entre otras.
- **Integridad referencial:** soporta integridad referencial, es utilizada para garantizar la validez de los datos.
- **Cliente/Servidor:** usa una arquitectura proceso-por-usuario cliente/servidor. Hay un proceso maestro que se ramifica para proporcionar conexiones adicionales para cada cliente que intente conectar a PostgreSQL.
- **Lenguajes procedurales:** tiene soporte para lenguajes procedurales internos, incluyendo un lenguaje nativo denominado PL/pgSQL. Otra ventaja es su habilidad para usar Perl, Python, o TCL como lenguaje procedural embebido.
- **API flexible:** la flexibilidad del API de PostgreSQL ha permitido a los vendedores proporcionar soporte al desarrollo fácilmente para el RDBMS (Sistema Administrador de Bases de Datos Relacionales) PostgreSQL. Estas interfaces incluyen Object Pascal, Python, Perl, PHP, ODBC, Java/JDBC, Ruby, TCL, C/C++, y Pike.
- **Control de Concurrencia Multiversión (MVCC):** es la tecnología que usa para evitar bloqueos innecesarios, provocada por usuarios que están escribiendo en la base de datos, esto es mejor que el bloqueo a nivel de fila ya que un lector nunca es bloqueado por un escritor. PostgreSQL mantiene una ruta a todas las transacciones realizadas por los usuarios de la base de datos entonces es capaz de

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

manejar los registros sin necesidad que los usuarios tengan que esperar que dichos registros estén disponibles.

- **Write Ahead Logging (WAL):** esta conocida característica de PostgreSQL incrementa la dependencia de la BS al registro de cambios antes que estos sean escritos en la BD. Esto permite que en el caso de que la BD falle, existirá un registro de transiciones a partir de la cual se puede restaurar y los usuarios puedan continuar trabajando sin perder información.

Se escoge PostgreSQL por ser un excelente gestor y además de licencia libre, presenta una arquitectura robusta e integridad de los datos, se puede ejecutar sobre cualquier Sistema Operativo (Windows, Linux), Soporta diferentes tipos de Datos (enteros, numéricos, caracteres, variables, intervalos) y objetos (sonido, videos, imágenes).

### 2.3.6. Servidor Web

Un servidor web es un programa que sirve para atender y responder a las diferentes peticiones de los navegadores, proporcionando los recursos que se soliciten usando el protocolo HTTP o HTTPS. Un servidor web básico cuenta con un esquema de funcionamiento muy simple, basado en ejecutar el siguiente ciclo: espera peticiones a través del puerto TCP indicado, recibe una petición, busca el recurso, lo envía utilizando la misma conexión por la que recibió la petición y vuelve a repetir el proceso para cada petición realizada. A partir de este esquema se han diseñado y desarrollado todos los servidores que existen, variando sólo el tipo de peticiones. (46)

### Apache Server 2.2

Desde su aparición en el mercado Apache ha sido uno de los servidores de mayor popularidad, considerado por mucho el proyecto punta de la lanza del movimiento del Software Libre. Fue desarrollado dentro del proyecto HTTP Server de la *Apache Software Foundation*. Presenta características como, mensajes de errores, bases de datos de autenticación y negociado de contenido.

Apache se caracteriza por ser un servidor ligero, con una alta configurabilidad y una amplia explotación. Apache puede soportar de una forma más fácil y eficiente una amplia variedad de sistemas operativos y la arquitectura que utiliza es cliente-servidor, el cliente hace la petición al servidor y este atiende dicha petición. El protocolo utilizado para la transferencia de hipertexto es HTTP (Protocolo de transferencia de

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

hipertexto) que está basado en el envío de mensajes y establece el conjunto de normas mediante las cuales se envían las peticiones de acceso a una web y la respuesta de esa web. (47)

### Principales ventajas

- Código abierto.
- Altamente configurable de diseño modular.
- Es multiplataforma, aunque idealmente está preparado para funcionar bajo Linux.
- Muy sencillo de configurar.
- Amplias librerías de PHP, PERL y ASP a disposición de los programadores.
- Soporte para el protocolo HTTP.
- Soporte de host virtuales.
- Servidor proxy integrado.

### 2.4. Arquitectura

La Arquitectura del Sistema (AS) es una vista del sistema que incluye los componentes principales del mismo, la conducta de esos componentes según se percibe desde el resto del sistema y las formas en que los componentes interactúan y se coordinan para alcanzar la misión del sistema. La vista arquitectónica es una vista abstracta, aportando el más alto nivel de comprensión y la supresión o diferimiento del detalle inherente a la mayor parte de las abstracciones. (48)

La arquitectura del sistema propuesto está definida en el trabajo de diploma “Propuesta de arquitectura para el sistema de gestión de información de mercadotecnia para Albet S.A.” Teniendo en cuenta que el sistema es un módulo de sistema central de Albet S.A, debe regirse por las especificaciones definidas en dicha arquitectura.

#### 2.4.1. Arquitectura definida para el sistema

El desarrollo del sistema estará sustentado sobre las arquitecturas en capas y cliente/servidor aplicando el patrón arquitectónico Modelo Vista Controlador (MVC) definido posteriormente. La arquitectura en capas define la organización de los componentes funcionales que interactúan con el usuario y la arquitectura cliente / servidor define la distribución física del sistema según los requerimientos no funcionales. (46)

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

### **Arquitectura cliente/servidor**

La arquitectura cliente/servidor conjuga un grupo de elementos que ejecutan procesos distribuidos y de cómputo. Es considerada un modelo para el desarrollo de sistemas de gestión de información, donde las transacciones se dividen en procesos independientes, servicios o recursos. (49)

Esta arquitectura utiliza dos componentes principales: un cliente como solicitante de servicios y un servidor como proveedor de servicios. En el servidor residen todos los datos y sus operaciones de modificación, además se controla el acceso a múltiples terminales conectadas a la red. Cada usuario de las terminales tiene un identificador asignado y derechos según sus funciones de trabajo. Generalmente el cliente es el encargado de iniciar la interacción con el servidor y responsable de la entrada de datos al sistema además de la generación de consultas e informes sobre la base de datos. Por otra parte el servidor se especializa en el acceso, almacenamiento, organización y actualización de la información mediante la ejecución de toda la lógica de procesamiento. (49)

La aplicación de la arquitectura cliente / servidor al sistema, permite la instalación de un servidor central que gestiona toda la información de mercadotecnia. A este servidor podrán acceder usuarios y subsistemas distribuidos en los proyectos de la universidad mediante interfaces de la aplicación o por servicios web definidos en la implementación del sistema. Esta arquitectura evita la duplicación de los datos de mercadotecnia teniendo siempre una imagen única y correcta de los mismos para su uso. El servidor Apache definido para la aplicación sostendrá toda la infraestructura del SIM haciendo función de proveedor de servicios para los usuarios. (46)

### **Arquitectura en capas**

La arquitectura en capas es una organización jerárquica tal que cada capa proporciona servicios a la inmediatamente superior y utiliza las prestaciones que le brinda la inferior. En algunos modelos, las capas internas están ocultas a todas las demás, menos para las externas adyacentes, y excepto para funciones puntuales de exportación mientras que, en otros sistemas, pueden ser sólo parcialmente opacas. En la práctica las capas son entidades complejas compuestas por varios paquetes o subsistemas que interaccionan entre sí y definen el flujo de información para un proceso determinado. (50)

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

Entre las modalidades más usadas de este tipo de arquitectura se encuentra la arquitectura en tres capas (Presentación, Lógica y Acceso a datos) las cuales tienen funciones específicas que ayudan a organizar la estructura interna de la aplicación. (50)

- **Capa de presentación:** es la encargada de interactuar con el usuario de la aplicación mediante una interfaz. Esta capa puede dividirse en dos subclases:
  1. **Interna:** se establece la forma de obtener los datos introducidos por el usuario aplicando mecanismos para transmitir dicha información a las capas lógicas y de acceso a datos.
  2. **Externa:** es la encargada del diseño gráfico que el usuario podrá ver al interactuar con el sistema brindando la información en un ambiente agradable y de fácil lectura.
- **Capa lógica:** es la responsable de realizar las tareas para las cuales fue diseñado el sistema. Se encarga de llevar a cabo la lógica de la aplicación. Contiene los métodos para resolver las peticiones de los usuarios pero no tiene acceso a la información. Entre sus principales funciones se encuentran:
  1. Recoger las peticiones de los usuarios.
  2. Efectuar las órdenes recibidas realizando las operaciones necesarias y modificando el estado de la aplicación.
  3. Consultar y actualizar los datos a través de la capa de acceso a datos.
- **Capa de acceso a datos:** es la encargada de gestionar toda la información, realizar los procesos de almacenamiento y recibir las solicitudes de la capa de negocio.

Según los principios generales del estilo en capas el mantenimiento de la aplicación se hace más sencillo pues los cambios repentinos se realizan a nivel de componentes. (46)

### Patrón arquitectónico Modelo Vista Controlador

El patrón Modelo Vista Controlador (MVC) es uno de los más usados en sistemas que trabajan con grandes cantidades de información y transacciones complejas. Facilita la programación en diferentes capas de manera paralela e independiente. Separa los datos de una aplicación, la interfaz de usuario y la lógica de control en tres partes distintas. Este patrón basa su fortaleza en que la lógica de la interfaz de usuario varía mucho más rápido que la lógica de negocio y el almacenamiento, con lo que pretende



## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

separar las capas, de tal forma que un cambio en una de ellas no tenga un impacto elevado en la otra.

(46) Está dividido en tres niveles:

- **Modelo:** gestiona los datos y la lógica del negocio. Representa la información con la que trabaja el sistema y es el encargado de acceder a la base de datos según los requerimientos de una petición.
- **Vista:** se encarga de la presentación de la información al usuario transformando los datos del modelo en una página web permitiendo una mejor interacción cliente–aplicación a través de una interfaz. Captura eventos del usuario y se los envía al controlador.
- **Controlador:** recibe las peticiones del usuario, utiliza los servicios del modelo y selecciona la vista para presentar los resultados de dichas peticiones.

### 2.5. Flujo actual de los procesos del negocio

Actualmente en el departamento de mercadotecnia de Albet S.A, las investigaciones que se llevan a cabo se realizan manualmente de la siguiente forma:

Los especialistas utilizan un componente del navegador Mozilla Firefox: **Delicious** que es un servicio de almacenamiento de enlaces web, del tipo social; es decir, permite a los usuarios etiquetar, guardar, administrar y compartir páginas web de una manera centralizada. (51) En este componente se guardan todos los enlaces identificados como relevantes a sus investigaciones con una breve descripción. Los enlaces son identificados por los propios especialistas, realizando búsquedas en la web para determinadas investigaciones y aquellos que se consideraron relevantes los han ido almacenando.

El proceso se inicia cuando un especialista, que realiza una determinada investigación, necesita obtener información de internet. Se dirige al Delicious, analiza las descripciones de todos los sitios y escoge aquellos que cree podría servirle para su trabajo. Luego visita los sitios escogidos y analiza su contenido, descargando documentos relevantes que pudieran aparecer (para su posterior consulta/análisis). El especialista estudia la página esperando encontrar información que le satisfaga. En caso de que se encuentre, la referencia en su investigación. Igualmente, el especialista puede decidir registrarse para asegurar el envío a sus correos personales de algunos boletines, determinadas alertas y/o RSS provistos por algunos sitios.

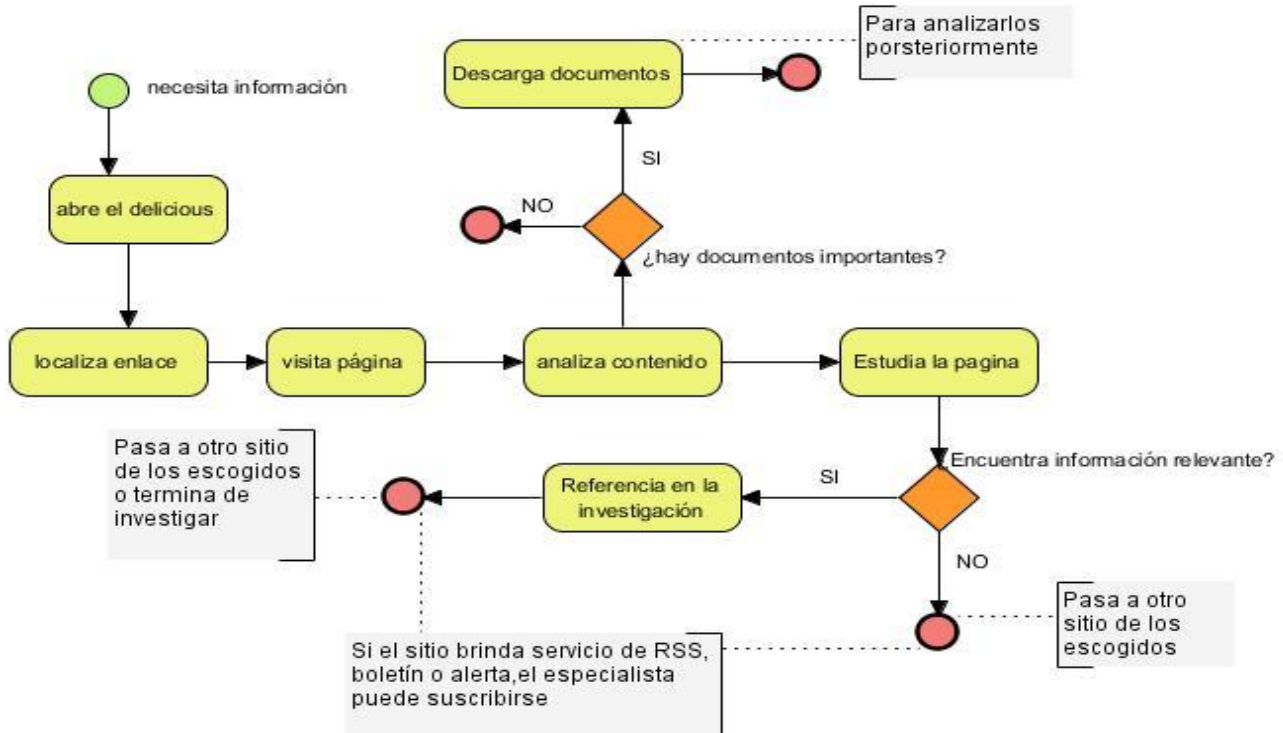


Figura 4: Flujo actual de los procesos del negocio.

## 2.6. Descripción general de la propuesta solución

La propuesta que le dará solución al problema automatizando los procesos antes descritos. Se basa en el desarrollo de un Sistema de Recuperación de Información de Internet y que a la vez brinde funcionalidades de Observatorio Tecnológico. Con la inclusión de las características de estos elementos, el Sistema de Información de Mercadotecnia de Albet S.A. logrará que las investigaciones de mercado que realizan los especialistas de mercadotecnia sean de mayor calidad y se realicen en menor tiempo del que se necesita actualmente. El sistema brindará funcionalidades que automaticen los procesos que en la actualidad se realizan manualmente, además que agrega nuevas funcionalidades que benefician el trabajo y resultado de este departamento.

La propuesta solución consiste en cómo mejorar los procesos actuales mediante la automatización de tareas y/o actividades que desarrollan los especialistas en sus investigaciones, además de la incorporación de nuevas funcionalidades que los apoyen. Partiendo de que se desarrolle el sistema como aplicación Web y que se una al SIM de Albet S.A quedaría de la siguiente forma:

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

El sistema permite a los usuarios introducir manualmente las direcciones (URL) de sitios/páginas de Internet en las que pueda estar contenida información importante. Estas URL pueden contener datos sobre competidores, clientes, mercados, distribuidores, productos, servicios, o de otra índole. Con estos enlaces se realiza una búsqueda inicial, donde se indiza<sup>6</sup> la documentación que poseen y se actualiza el índice invertido. Luego, el sistema localiza aquellos sitios que presten servicio de RSS y/o boletines y gestiona el uso de estos, así como aquellos que requieran registrarse para acceder a determinada información, para que el usuario se registre en ellos.

El sistema permite realizar búsqueda por palabras claves. En este proceso se genera el árbol de búsqueda de los sitios más relevantes y se le propone al usuario aquellos que cumplan con los criterios de búsqueda. Estas propuestas de nuevas direcciones son analizadas por el usuario, quien decide si los agrega al sistema pues los considera de importancia. Finalmente da como resultado una lista ordenada de los sitios y documentos más relevantes de acuerdo al criterio de búsqueda. Se brinda la posibilidad almacenar la información que se considere importante dentro de los resultados analizados, ya sea el sitio completo, las páginas que cumplen con el criterio de búsqueda, documentación o recursos multimedia con que cuenta.

El sistema también permite realizar un análisis cuantitativo de cada sitio que incluye la cantidad de visitas, la última actualización, los protocolos y dominios, los recursos informáticos que publica, el factor de impacto en la web, así como otros datos. El sistema ofrece datos comparativos luego de cada estudio, de forma que se asista a la hora de establecer pronósticos y tendencias que ayuden a tener un control de las tecnologías que actualmente son utilizadas así como los productos de más demandas, los sitios más visitados, etc.

Toda esta información es clasificada de acuerdo a varios criterios de organización, por ejemplo si es de competidores, si es de posibles mercados o clientes, si es de productos, si es de países, y así toda una gama de criterios que pueden ser creados por los especialistas a partir de las palabras claves o los criterios de búsqueda y análisis definidos en el sistema.

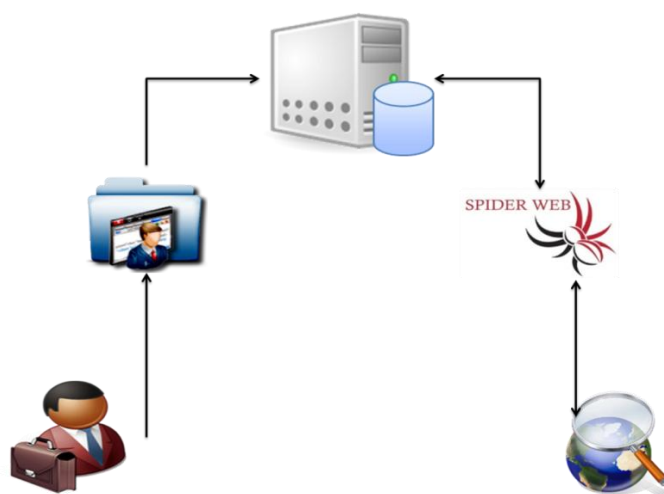
---

<sup>6</sup> **Indiza:** de indizar, hacer índices; registrar ordenadamente datos e información para elaborar su índice.

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

Además, se brinda la posibilidad de mantener un monitoreo sobre sitios que se consideren relevantes durante un intervalo de tiempo definido por el usuario, gestionando reportes de alertas o actualizaciones que encuentren.

De manera general la investigación propone un sistema que sea capaz de gestionar estos procesos de una mejor manera, de cómo se realiza actualmente.



**Figura 5: Descripción del sistema.**

El usuario va a interactuar con una aplicación que estará alojada en un servidor quien a su vez contará con un spider web encargado del trabajo con Internet. El spider busca y obtiene la información de acuerdo a los criterios de entrada definidos en la aplicación. La búsqueda se realiza sobre las URLs que se inserten y en las nuevas que se vayan encontrando.

### **2.7. Estrategias para el levantamiento de requerimientos**

La ingeniería de requerimientos es la disciplina para desarrollar una especificación completa, consistente y no ambigua, la cual servirá como base para acuerdos comunes entre todas las partes involucradas y en dónde se describen las funciones que realizará el sistema. (52) Es decir, se encarga de definir qué se desea producir teniendo como tareas la generación de especificaciones de comportamiento del sistema describiéndolos con claridad y sin ambigüedades. Pretende minimizar los problemas relacionados con el desarrollo del sistema definiendo requerimientos funcionales y no funcionales.

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

Al describir los requisitos se debe tener en cuenta los siguientes aspectos: (49)

- **Ubicación y entorno físicos:** dónde, uno o varios, restricciones ambientales.
- **Interfaces:** entrada de 1 o + sistemas, salida a 1 o + sistemas, restricciones de formato, soporte.
- **Usuarios y factores humanos:** capacidad de cada tipo de usuario, tipo de entrenamiento, facilidad de uso, posibilidad de mal uso.
- **Funcionalidad y restricciones asociadas:** qué debe hacer, cuándo, modos de operación, cómo y cuándo se puede modificar el sistema, restricciones de velocidad, tiempo de respuesta, capacidad de proceso.
- **Documentación:** cuánta, formato, para quién.
- **Datos:** formatos E/S, frecuencia, fuentes, destinos, calidad requerida, precisión en cálculos, flujo en el sistema.
- **Recursos:** materiales, personal y otros para construir, usar y mantener el sistema, habilidades de los desarrolladores, necesidades de espacio y ambientales, calendario prescrito, limitaciones en presupuesto.
- **Seguridad:** control de acceso a las funciones/datos, aislamiento de los programas, respaldos-frecuencia, disponibilidad, seguridad física.
- **Aseguramiento de la Calidad:**
  - **Confiabilidad:** tiempo medio entre fallas, robustez, tolerancia a fallas.
  - **Disponibilidad:** tiempo para volver a estar operativo luego de falla, mantenimiento estando activo, tiempo máximo de no disponibilidad.

Según Brook (1987) “La parte más difícil al construir un sistema de software es decidir qué construir. Ninguna otra parte afecta más negativamente el sistema final, si se realiza de manera incorrecta. Ninguna otra parte es más difícil de rectificar posteriormente” (53) reflejando la importancia de la ingeniería de requerimientos y recorrer correctamente todas las fases de esta que según Sommerville (2007) son: (49)

1. Análisis del Problema
2. Evaluación y Negociación
3. Especificación
4. Validación
5. Evolución

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

Teniendo en cuenta todos estos aspectos, basado en las buenas prácticas de la ingeniería de requerimientos, además, después de evaluadas las necesidades de desarrollar el sistema y obtenido los procesos del negocio, se estudian las diferentes técnicas de obtención de requisitos aplicables al escenario en que se desarrollan los procesos del negocio:

### 2.7.1. Tormenta de ideas

Consiste en reunir a los diferentes encargados de los procesos de negocio para que en una reunión informal generen ideas sobre los diferentes problemas que cada uno en su área posea. Como resultado de estas sesiones se obtiene la mayor cantidad de soluciones posibles desde el punto de vista de cada uno de los involucrados en los procesos de la empresa y puede generar salidas que beneficien a la organización. La ventaja de esta técnica es la libertad que tienen los encargados de cada proceso de opinar sobre el futuro de la empresa y cómo debe ser organizada. (54)

Se siguen unas reglas fundamentales para obtener mejores resultados: (54)

- No se permite criticar ni debatir.
- Dejar volar la imaginación.
- Generar tantas ideas como sea posible.
- Mutar y combinar ideas.

Este método se aplicó con el universo de los especialistas de mercadotecnia de Albet S.A., donde se recogió el punto de vista global sobre el problema a resolver y de donde resultaron las primeras visiones sobre las funcionalidades del sistema:

1. Agregar enlaces al sistema.
2. Realizar estadísticas a los sitios que permitan:
  - a. Medir la popularidad de los mismos.
  - b. Fecha de última actualización.
  - c. País, empresa, sector, al que pertenece.
3. Realizar búsquedas dentro de los sitios ingresados en el sistema por palabras claves.
4. Descargar información relevante.
5. Suscribirse a servicios de RSS y/o boletines.

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

### 6. Alertas de actualizaciones.

Al concluir la reunión se les pidió a los especialistas una cita para una próxima reunión a la que debían traer otras ideas que se les ocurrieran o nuevas visiones de las ya captadas. Desarrollada la próxima reunión, se recogieron algunas nuevas ideas para otras funcionalidades o mejorar las que ya se tenían, además se aprovechó el momento para utilizar otro método de captar requerimientos, las entrevistas individuales.

### 2.7.2. Entrevistas

Es una de las técnicas más comunes para el levantamiento de información, consiste en realizar una serie de preguntas a cada uno de los encargados de los procesos del negocio, de esta manera se pueden determinar las necesidades de cada área, de *stakeholders* y de la empresa en general. Estas entrevistas están catalogadas de la siguiente manera: (55)

- **No estructuradas:** Son realizadas por un entrevistador, el cual no se guía por ningún patrón, guía o lista de preguntas, por lo que es posible que no se traten temas importantes para la organización.
- **Estructuradas:** Consiste en la elaboración del mismo tipo de preguntas para todos los encargados de los procesos de negocio, por lo que los resultados son esperados.

Se aplicó este método, primero dialogando con algunos especialistas para obtener una visión más específica e individual sobre lo que se pretende obtener con el sistema, de la necesidad del mismo. En algunos casos se aplicó una encuesta a los especialistas para recoger evidencia documental sobre las necesidades y visiones del sistema, algunas de las preguntas fueron:

1. Características fundamentales del proceso.
2. Objetivo principal del proceso.
3. ¿Qué limitaciones presenta el proceso actualmente?
4. ¿Cuál es la razón por la que se quiere resolver este problema?
5. ¿Qué afectaciones al negocio espera de la solución al problema?
6. Entorno de los usuarios (computadoras, plataformas).
7. ¿Existen requisitos legales, regulatorios u otros estándares que deban ser tenidos en cuenta?
8. ¿Cómo considera que se pueda validar la información obtenida?

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

### 9. ¿Qué considera una buena solución?

Entre otras, el modelo se encuentra en el **Anexo 1: Modelo de entrevista**

Con la aplicación de este método se obtuvo la información necesaria para especificar otro grupo de funcionalidades con las que debe contar el sistema, pero como se corre el riesgo con esta técnica de obtener información incompleta, se combinó con otra técnica, la observación.

#### 2.7.3. Observación

Técnica que consiste en asignar un proceso a un participante, de manera que lo pueda realizar y explicar, de esta manera se puede determinar qué información se maneja y cómo se administra a lo largo de la ejecución del proceso. (54)

Después de utilizadas las técnicas de tormenta de ideas y las entrevistas, se aplicó la observación para aclarar aspectos que no quedaron claros después de las dos actividades anteriores. Se le asignó a varios especialistas por separado que realizaran una investigación y fueran explicando lo que iban haciendo para ir analizando el proceso conformando así la visión general del proceso actual, y determinando los procesos automatizables dentro del negocio. Con esta técnica se pudo identificar funcionalidades que agregar al sistema y refinar otras que ya se tenían identificadas.

#### 2.8. Principales funcionalidades.

Los requerimientos funcionales son aquellos requisitos que, desde el punto de vista de las necesidades del usuario, debe cumplir el sistema y que están fuertemente ligados a las opciones del programa. (56)

A continuación se enumeran y describen brevemente las principales funcionalidades con que contará el sistema, agrupadas en cuatro grupos.

#### Funcionalidades de Inicio

**F 1.** Gestionar los enlaces definidos por los especialistas. Realizar limpieza de enlaces.

**F 2.** Realizar búsqueda inicial (Se analizan los nuevos enlaces ingresados por el especialista y se indiza la documentación que se encuentra en ellos, se actualiza al índice invertido). Los documentos,



## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

aquellos que no sean entendibles por el spider y no se puedan indexar se necesitarán convertir a un formato entendible por el indizador (ejemplo: convertir de pdf a texto plano).

**F 3.** Proponer al especialista registrarse en sitios que lo requieran para acceder a información que brinden. Gestionar el uso de RSS, boletines o alertas.

### Funcionalidades de Estadísticas

**F 4.** Gestionar análisis estadístico del sitio (cantidad de páginas, visitas diarias, protocolos que usa, dominio al que pertenece, etc.).

**F 5.** Clasificar los sitios según el factor de impacto de la web y productividad de los autores (el número y diversidad de enlaces externos recibidos, su volumen respecto a los contenidos y objetos de enlace). Según recursos que utilicen (páginas con datos textuales o audiovisuales, dinámicas o estáticas, densidad multimedia, etc.). Según las propias características del análisis cuantitativo realizado a los sitios.

**F 6.** Construir resumen de las estadísticas.

**F 7.** Gestionar el top 10 de principales clientes, competidores, y otras clasificaciones (dada alguna clasificación hacer análisis en los enlaces más populares y los datos del mismo).

**F 8.** Gestionar los principales clientes y clientes potenciales (búsqueda de licitaciones, principales clientes de un producto o servicio determinado, páginas identificadas por los especialistas).

### Funcionalidades de Búsqueda

**F 9.** Realizar búsqueda a partir de palabras claves (en los enlaces del sistema, generar el árbol de búsqueda con un nivel de profundidad definido y que satisfaga las condiciones de búsqueda y proponer las hojas como sitios tentativos a analizar (hijos de la raíz) si no están en el sistema; si el usuario los evalúa de factibles, agregarlos).

**F 10.** Realizar búsqueda avanzada (por tipo de contenido: imágenes, documentos, multimedia). Debe recuperar todos los archivos con las extensiones del recurso que se pida.

**F 11.** Construir respuesta de la búsqueda (una lista ordenada de acuerdo al modelo de SRI que utiliza el *spider*) con los sitios y la documentación encontrada que satisface el criterio de búsqueda. Se eliminan enlaces muertos.

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

**F 12.** Almacenar información (Se permite descarga la información reconocida como relevante mediante las palabras claves de los sitios analizados por el usuario y almacenarla clasificándola de acuerdo al criterio de búsqueda).

### Funcionalidades de Monitorización

**F 13.** Gestionar monitoreo de sitios (sitios definidos como relevantes y de necesidad de vigilancia, dado un intervalo de tiempo y un período de monitoreo).

**F 14.** Construir notificaciones y reportes de monitoreo (cantidad de enlaces encontrados, productos, modificaciones de sitios, actualizaciones encontradas, etc.).

**F 15.** Gestionar reporte resumen de empresas, competencia o clientes (análisis de las páginas que contienen información de la empresa en los sitios identificados por el usuario).

### 2.9. Requisitos no funcionales

Los requerimientos no funcionales son propiedades o cualidades que el producto debe tener, son aquellas características que hacen un producto atractivo, usable, rápido y confiable, todo esto es primordial para que se alcance el éxito del producto. (55)

#### Usabilidad

- El sistema debe permitir el acceso a personas autenticadas.
- El lenguaje utilizado en el sistema debe adaptarse a los términos utilizados por los clientes.

#### Confiabilidad

- El sistema debe permitir mantener un control sobre el tráfico de información y la misma deberá ser transmitida utilizando canales seguros.

#### Seguridad

- La información manejada debe estar protegida de acceso no autorizado y de divulgación, además de protección contra ataques XSS (*Cross-Site Scripting*) y CSRF (*Cross-Site Request Forgery*).
- El sistema debe permitir confirmar sobre acciones irreversibles (eliminación).

#### Hardware

- Para el desarrollo: PC Intel Pentium 4 o superior, CPU 3 GHZ o superior, 2 Gb RAM o más.

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

- Para explotación del servidor: CPU Dual Core 2.0 GHZ o superior, RAM de 2 GB (recomendado 4 GB o superior), 1 Tb HDD, requiere buena conexión a internet y estable.
- Para explotación del cliente: PC Pentium 3 o superior, CPU 1.33 MHZ o superior, 256 RAM o superior.
- 

### **Soporte**

- El servidor de BD debe soportar grandes volúmenes de datos y tener buena velocidad de procesamiento.
- Los administradores del sistema deben ser capacitados para realizar mantenimiento al sistema.

### **Restricciones del desarrollo**

- Para la implementación del sistema se debe utilizar el lenguaje de programación PHP 5.3.8.
- Para la implementación del sistema se debe utilizar el framework de desarrollo Symfony 1.4.15.
- Para la implementación y explotación del sistema se debe utilizar como servidor Web Apache 2.

### **Portabilidad**

- El sistema debe ser multiplataforma.

### **Interfaz**

- La interfaz debe ser agradable para el usuario buscando una correcta combinación de los colores, tipo de letra y tamaño.
- Los iconos deben estar en correspondencia con lo que representan.
- No se deben emplear un cúmulo de imágenes u objetos que distraigan al cliente.

### **Software**

- Se deben contar como navegador el Mozilla Firefox o Google Chrome.

### **2.10. Lista reserva del producto**

La metodología SXP, define artefactos para la descripción del sistema o producto, integrando al cliente como parte importante del equipo de desarrollo. Dentro de la gestión de proyecto es donde se genera el

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

artefacto Lista de Reservas del Producto (LRP), donde se incluyen los requerimientos funcionales especificados por el cliente del sistema, separados por su prioridad (muy alta, alta, media o baja), que indica la necesidad que tiene el sistema de dicha funcionalidad, siendo las muy altas, funcionalidades indispensables para cubrir las necesidades del cliente, mientras que las bajas, son aquellas que se desean incluir, pero sin las que el sistema podría, ser funcional y útil. (31)

**Tabla 2: Lista de reserva del producto.**

Prioridad	Id	Funcionalidad
Muy Alta	HU-SRI-01	Gestionar enlaces
	HU-SRI-02	Realizar búsqueda inicial
	HU-SRI-09	Gestionar búsqueda por palabras claves
	HU-SRI-11	Construir respuesta de la búsqueda
Alta	HU-SRI-04	Gestionar estadísticas
	HU-SRI-06	Construir resumen estadísticas
	HU-SRI-05	Clasificar sitios
	HU-SRI-13	Gestionar monitoreo
	HU-SRI-14	Construir reporte y notificaciones de monitoreo
	HU-SRI-12	Almacenar información
	HU-SRI-03	Gestionar registro, RSS, boletines
Media	HU-SRI-10	Gestionar búsqueda avanzada
	HU-SRI-15	Gestionar reportes resumen
Baja	HU-SRI-07	Gestionar listas TOP
	HU-SRI-08	Gestionar posibles clientes

Quedan así definidos prioritariamente los requisitos que responden a las funcionalidades con que contará la propuesta solución.

### 2.11. Conclusiones del capítulo

## Capítulo 2: Entorno de desarrollo y obtención de requisitos

---

En este capítulo se estudiaron las principales herramientas, lenguajes y metodologías a utilizar en el desarrollo del sistema, además de una descripción general del sistema, por lo que se puede arribar a las siguientes conclusiones:

- El desarrollo de software será guiado por una metodología de desarrollo ágil: SXP. El entorno de desarrollo estará compuesto por herramientas y tecnologías libres, las definidas en el trabajo de diploma de la arquitectura.
- Se realizó una breve descripción del sistema que se propone desarrollar para un mejor entendimiento del mismo, así como las principales funcionalidades con que contará, identificadas a partir del modelo del negocio y las necesidades del cliente.
- Se definió la lista de reserva del producto otorgándole una prioridad de cada requisito según las necesidades del cliente, además se enumeraron los requisitos no funcionales ayudando a comprender ciertas restricciones con las que debe contar el sistema.

## Capítulo 3: Especificación y validación de requisitos

### CAPÍTULO 3: ESPECIFICACIÓN Y VALIDACIÓN DE REQUISITOS

#### 3.1. Introducción

Este capítulo tiene como propósito establecer como alcanzar los objetivos propuestos, se planifica el desarrollo a través del plan de iteraciones y se describen las historias de usuario correspondientes a cada funcionalidad, además se validan los requerimientos funcionales a través de los métodos de revisión de requerimientos y prototipado.

#### 3.2. Planificación

Durante la fase de planificación se realiza una estimación del esfuerzo que costará implementar cada HU. Este se expresa utilizando como medida el punto, donde se considera como una semana ideal de trabajo donde los miembros de los equipos de desarrollo trabajan el tiempo planeado sin ningún tipo de interrupción. Esta estimación incluye todo el esfuerzo asociado a la implementación de la HU.

##### 3.2.1. Plan de iteraciones

En la siguiente tabla se recogen las iteraciones a realizar con sus características, además del orden de las historias de usuario con su planificación estimada para ser implementadas

**Tabla 3: Plan de iteraciones.**

Iteración	Orden de la Historias de usuario a implementar	Esfuerzo
Iteración 1	HU-SRI-01	4
	HU-SRI-02	
	HU-SRI-03	
Iteración 2	HU-SRI-04	5
	HU-SRI-05	
	HU-SRI-06	
	HU-SRI-07	
	HU-SRI-08	
Iteración 3	HU-SRI-09	6

## Capítulo 3: Especificación y validación de requisitos

	HU-SRI-10	
	HU-SRI-11	
	HU-SRI-12	
<b>Iteración 4</b>	HU-SRI-13	3
	HU-SRI-14	
	HU-SRI-15	

- La iteración 1 durará un total de 4 semanas.
- La iteración 2 durará un total de 5 semanas.
- La iteración 3 durará un total de 6 semanas.
- La iteración 4 durará un total de 3 semanas.

### 3.3. Descripción de las Historias de Usuario

Las Historias de Usuarios (HU) son similares a los escenarios utilizados en la descripción de los casos de usos de RUP, con la excepción de que no se limitan a la descripción de la interfaz de usuario y son definidas por el propio cliente según las necesidades del sistema. (31)

A diferencia de la especificación de requisitos y casos de usos descritos en RUP, las HU solamente proporcionarán una breve descripción de las mismas, detalles sobre la estimación del riesgo y cuánto tiempo será empleado en su implementación. Es el cliente el encargado de asignarle una prioridad a cada HU y es el equipo de desarrollo el encargado de asignarle un costo, este se traduce en las semanas que llevará el desarrollo de las mismas. Si las historias, según lo planificado, demoran en desarrollarse se sugiere dividirla en historias más pequeñas. También, es importante destacar, que las HU nuevas pueden describirse en cualquier momento, con esto se comprueba la flexibilidad de la metodología. (31)

Las HU serán representadas mediante tablas divididas por las siguientes secciones:

- **Código:** Código que identifica la historia, sería HU-identificador-número consecutivo.
- **Nombre:** Nombre que identifica la historia de usuario.
- **Referencia:** Se hace referencia a los requisitos que tiene en cuenta la HU.
- **Prioridad:** Prioridad puede ser Alta, Media o Baja (según necesidad del cliente).
- **Iteración Asignada:** En que iteración se desarrollará (según su importancia).

### Capítulo 3: Especificación y validación de requisitos

- **Puntos Estimados:** Tiempo en semanas que se le asignará (estimado).
- **Descripción:** Breve descripción del proceso que define la historia.
- **Observaciones:** Alguna acotación importante de señalar acerca de la historia.
- **Prototipo de interface:** Imagen de cada una de las interfaces relacionadas con la HU.

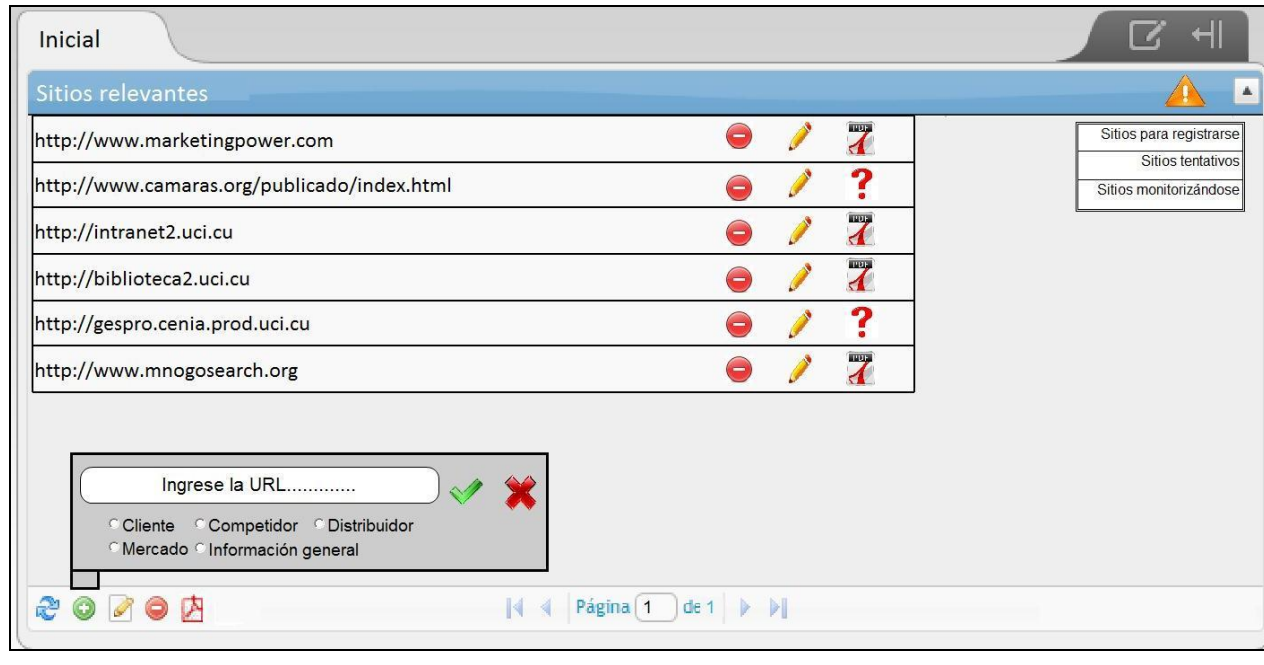
**Tabla 4: Gestionar enlace.**

Historia de Usuario	
<b>Código:</b> HU-SRI-01	<b>Nombre:</b> Gestionar enlace
<b>Referencia:</b> F 1	<b>Prioridad:</b> Muy Alta
<b>Iteración asignada:</b> 1	<b>Puntos estimados:</b> 1
<p><b>Descripción:</b></p> <ul style="list-style-type: none"> <li>• El usuario selecciona “Agregar nuevo enlace”.</li> <li>• El sistema muestra una ventana para introducir el enlace y las opciones para clasificarlo de acuerdo a si es de cliente, competidor, distribuidor, mercado o de información general.</li> <li>• El usuario copia o escribe el enlace, selecciona la clasificación y acepta.</li> <li>• El sistema realiza limpieza de URL y lo agrega a la cola de URL por analizar.</li> <li>• El usuario puede eliminar sitios de esta lista y el sistema debe confirmar si desea eliminarlos.</li> </ul>	
<p><b>Observaciones:</b></p> <ol style="list-style-type: none"> <li>1. El usuario debe clasificar el sitio por una de las opciones.</li> <li>2. En caso que la URL ya exista se muestra un mensaje de error “El enlace ya está registrado” y se muestra en la lista de enlaces la URL.</li> <li>3. En caso que la URL esté fuera de servicio se le notifica al usuario y se elimina el enlace.</li> </ol>	



**Prototipo de interfaz:**

**Agregar URL**



**Tabla 5: Realizar búsqueda inicial.**

Historia de Usuario	
<b>Código:</b> HU-SRI-02	<b>Nombre:</b> Realizar búsqueda inicial
<b>Referencia:</b> F 2	<b>Prioridad:</b> Muy Alta
<b>Iteración asignada:</b> 1	<b>Puntos estimados:</b> 2
<p><b>Descripción:</b></p> <ul style="list-style-type: none"> <li>• El sistema toma la primera URL de la cola.</li> <li>• El sistema indiza la información del sitio y actualiza el índice invertido.</li> <li>• El sistema actualiza la lista de sitios por registrarse y actualiza la lista de clasificaciones de los sitios.</li> <li>• El sistema se suscribe a aquellos sitios que presten servicio de RSS, boletines o alertas.</li> </ul>	

## Capítulo 3: Especificación y validación de requisitos

### Observaciones:

1. Si encuentra documentación que no entiende el indizador, la convierte. (ejemplo de PDF a texto plano).
2. Si para acceder a información del sitio necesita registrarse, agrega al sitio a la lista de sitios en espera y muestra un mensaje al usuario de que necesita registrarse en ese sitio.

### Prototipo de interfaz:



Las demás HU se pueden encontrar en los **Anexos. Anexo 4:** Historias de Usuario

### 3.4. Validación de requerimientos

La validación de requisitos examina la especificación para asegurar que todos los requisitos de software se han establecido de manera precisa, que se han detectado las inconsistencias, omisiones y errores y que estos han sido corregidos, y que los productos de trabajo cumplen con los estándares establecidos para el proceso, proyecto y producto. (57) O sea, validar los requerimientos tiene como objetivo demostrar que dichos requisitos definen realmente el producto que el usuario o cliente necesita, descubriendo los problemas en estos requisitos antes de comprometer recursos en su implementación.

En la actividad de verificación de requisitos de software se debe confirmar que los requisitos, una vez analizados y resueltos los posibles conflictos, corresponden realmente a las necesidades de clientes y usuarios. (52)

Como parte de la ingeniería de requerimientos aplicada al sistema que se propone, se validan los requisitos aplicando las técnicas de revisión de requisitos y prototipado.

## Capítulo 3: Especificación y validación de requisitos

### 3.4.1. Revisión de requisitos

La revisión de requisitos es una de las mejores técnicas de validación de requerimientos. Generalmente las revisiones de requisitos permiten: (55)

- Descubrir una gran cantidad de defectos en los requisitos.
- Reducir los costos de desarrollo entre un 25% y un 30%.
- Reducir el tiempo de pruebas entre un 50% y un 90%.

Las revisiones de requisitos consisten en una o varias reuniones planificadas, donde se intenta confirmar que los requisitos poseen los atributos de calidad deseados. Estas reuniones son llevadas a cabo por el analista principal del proyecto y la representación del cliente. El resultado final de las reuniones de revisión es un documento que contiene la lista de defectos localizados y una lista de acciones recomendadas. (55)

Las cualidades que debe poseer una especificación de requisito software (ERS) son las que se describen a continuación: (52)

- **Verificable:** Una ERS es verificable sólo si cada requisito especificado es verificable. Un requisito es verificable sólo si existe un proceso finito de costo-efectivo con el cual una persona o una máquina puedan verificar que el producto de software cumple el requisito. Cualquier requisito ambiguo es no verificable.
- **No ambiguo:** Una ERS es no ambigua sólo si cada requisito especificado tiene una única interpretación. Cuando un requisito es interpretado de forma distinta por varias personas se dice que es ambiguo y puede traer consigo desacuerdos en el grupo de trabajo de desarrollo del software.
- **Comprensible:** Una ERS es comprensible sólo si todos los requisitos que lo componen son entendidos por el cliente, desarrolladores y usuarios y los mismos cuentan con un mínimo de explicación. Cualquier requisito ambiguo es no comprensible.
- **Completo:** Una ERS es completa sólo si todos los requisitos significativos que respondan a funcionalidad, desempeño, restricciones de diseño, atributos, interfaces externas o cualquier requisito externo impuesto por una especificación del sistema, son reconocidos y tratados. Además si se definieron las respuestas del software a todos los tipos posibles de clases de datos de entrada en todos los tipos posibles de clases de situaciones.

## Capítulo 3: Especificación y validación de requisitos

- **No redundante:** Una ERS es no redundante sólo si un mismo requisito no es indicado más de una vez. No siempre la redundancia es mala, frecuentemente puede ser usada para incrementar la legibilidad del ERS. El problema radica en que si todas las ocurrencias de un requisito redundante no son modificadas, el ERS se vuelve inconsistente.
- **Correcto:** Una ERS es correcto sólo si cada requisito especificado es un requisito que el software debe cumplir y no expresa necesidades que realmente no existen o no son necesarias. Además la ERS debe cumplir con las reglas o restricciones del negocio y no deben existir contradicciones entre los requisitos.
- **Posible de *trazar*:** Una ERS es posible de *trazar* si el origen de cada uno de sus requisitos es claro y si facilita la referencia de cada requisito en el desarrollo futuro o mejora de la documentación.
- **Preciso:** Una ERS es precisa sólo si son usadas cantidades numéricas siempre y cuando sea posible y se usan niveles apropiados de precisión para todas las cantidades numéricas.
- **Modificable:** Una ERS es modificable sólo si su estructura y estilo son tales que, cualquier cambio a los requisitos pueden ser hechos con facilidad, de manera completa y consistentemente sin perder la estructura y el estilo. La "modificabilidad" generalmente requiere que un requisito:
  - Tenga una organización coherente y fácil de usar con una tabla de contenido, un índice y referencias cruzadas explícitas.
  - No sea redundante.
  - Expresa cada requisito de manera separada, en vez de hacerlo mezclado con otros requisitos.
- **Consistente:** Una ERS es consistente sólo si no se contradice a sí mismo es decir, si ningún subconjunto de los requisitos que describe se contradice o entra en conflicto con otros requisitos.
- **Jerarquizado de acuerdo a la importancia y/o estabilidad:** Una ERS está calificado de acuerdo a la importancia y/o estabilidad si cada requisito tiene un identificador único que indique la importancia o estabilidad del requisito.

Con frecuencia resulta útil examinar cada requisito frente a una serie de preguntas en forma de lista de chequeo. (52) A continuación se describe un ejemplo de las preguntas y los atributos medidos con la aplicación de la lista de chequeo. Para ver la lista completa ir al **¡Error! No se encuentra el origen de la referencia.**

### Capítulo 3: Especificación y validación de requisitos

**Tabla 6: Preguntas incluidas en la lista de chequeo.**

Preguntas	Atributo al que responden
1. ¿Están todos los requisitos redactados de forma simple y clara?	Comprensible
2. ¿Deberían especificarse algunos requisitos con más detalle?	
3. ¿Tienen todos los requisitos una única interpretación? ¿Cuáles no?	No ambiguo
4. ¿Los requisitos que son medibles cuantitativamente están delimitados en términos cuantitativos?	Precisión
5. ¿Se utilizan en todos los requisitos delimitados en términos cuantitativos niveles apropiados de precisión para todas las cantidades numéricas?	
6. ¿Existe redundancia en la definición de los requisitos? ¿En cuáles?	No redundante
7. ¿Aparecen requisitos en más de un lugar del ERS? ¿Cuáles?	
8. ¿Todos los requisitos son verificables?	Verificable
9. ¿Se pueden probar todos los requisitos una vez implementados?	Completo
10. ¿Todos los requisitos documentados son correctos? ¿Cuáles no?	
11. ¿Se puede trazar cada requisito al origen en el entorno del problema (proceso del negocio)? ¿Cuáles no?	Posible de <i>tracear</i>
12. ¿Existe referencia de la dependencia del requisito con otros requisitos?	
13. ¿Existen requisitos que se contradicen con otros requisitos?	Correcto
14. ¿Se pueden localizar todos los requisitos en el conjunto de objetivos del sistema/producto?	

**Resultados:** Para poder validar la ERS todos y cada uno de los requisitos debe cumplir con las características antes expuestas, responder positivamente a cada una de las preguntas de la lista de chequeo.

Se realizaron 3 iteraciones de esta fase de revisión de requisitos hasta que quedaran totalmente refinados los requisitos, recogiendo en cada una de ellas la lista de no conformidades de acuerdo a la lista de chequeo hasta que el 100% de los requerimientos quedaron validados.

### Capítulo 3: Especificación y validación de requisitos

#### 3.4.2. Prototipado

Esta técnica consiste en construir prototipos interfaz del futuro sistema a partir de los requisitos recogidos en la especificación, que sin tener la totalidad de las funcionalidades, permiten al usuario o cliente hacerse una idea de la estructura de la interfaz del sistema. (52) Esta técnica tiene el problema de que el usuario debe entender que lo que está viendo es un prototipo y no el sistema final.

La siguiente imagen es uno de los prototipos que validan una de las funcionalidades, en este caso es la funcionalidad Gestionar enlaces.

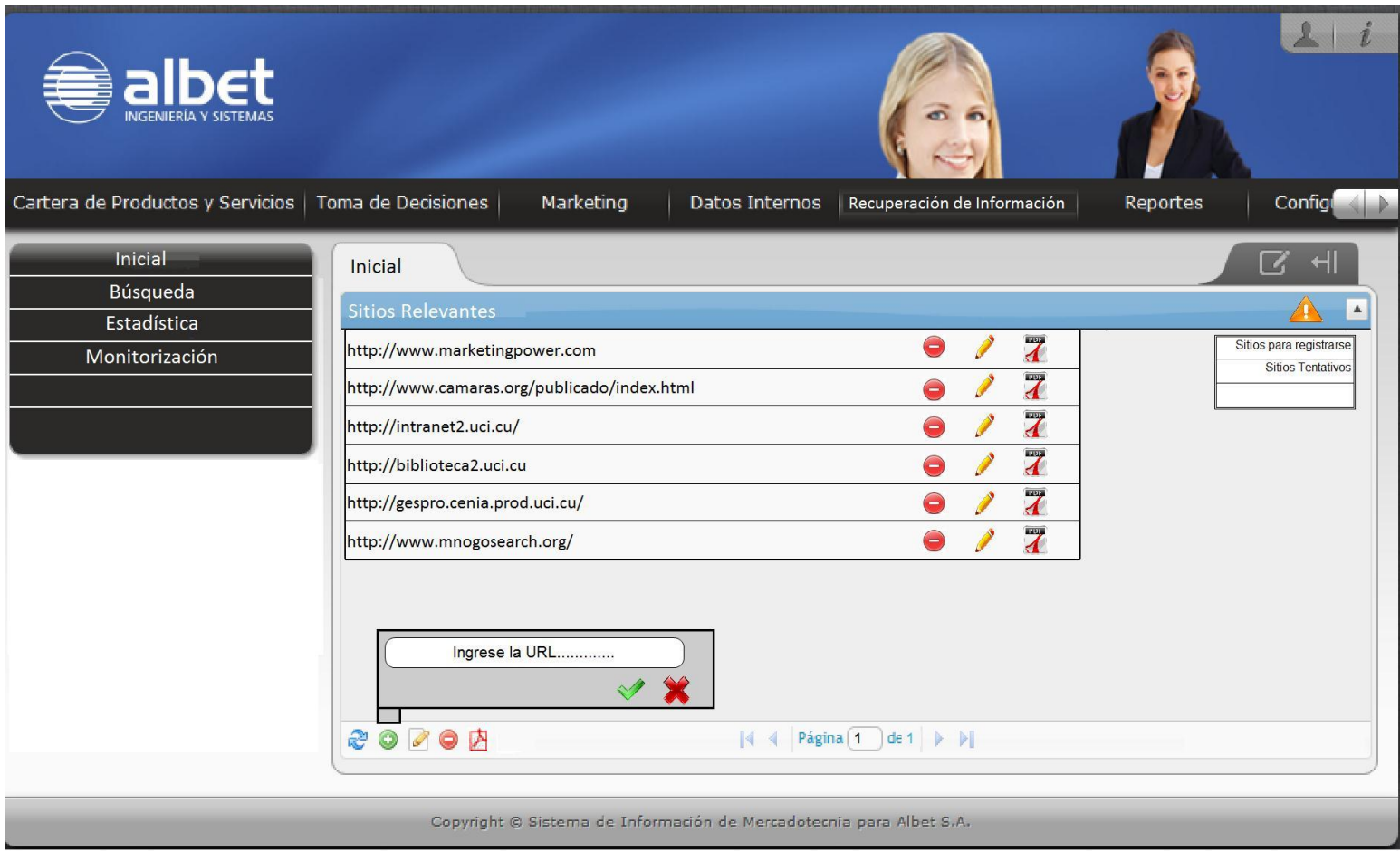


Figura 6: Prototipo Gestionar enlace.

## Capítulo 3: Especificación y validación de requisitos

---

**Resultados:** Con esta técnica se logró refinar los requisitos que se tenían identificados además de identificar otros que hasta el momento no habían salido a la luz. El cliente pudo captar visualmente una idea de las funcionalidades con que contará el sistema.

Los prototipos se realizaron sobre la interfaz del SIM de Albet S.A. sirviendo esto también para mayor aceptación del cliente y que pudiera ver cómo quedaría integrada la propuesta solución que complementa el sistema central. Quedando así satisfecho con la especificación de los requisitos y quedando validados los mismos por los especialistas de mercadotecnia de la empresa Albet S.A.

**Anexo 2: Aval de requerimientos.**

### 3.5. Conclusiones del capítulo

En este capítulo se finaliza la etapa de análisis y diseño del sistema obteniendo un modelo detallado de la solución propuesta.

- Se creó un plan de iteraciones para organizar y acotar el tiempo de desarrollo del sistema.
- Se confeccionaron las historias de usuario para una mayor comprensión del sistema.
- Se describieron los métodos utilizados para la validación de requisitos, evitando así comprometer recursos a la hora de implementar requisitos que no respondan a las necesidades del cliente o estén redactados deficientemente.
- Los artefactos descritos sirven de base para la posterior implementación del sistema.

---

## **CONCLUSIONES**

Como resultado de esta investigación realizada, sustentada en los elementos teóricos y con los artefactos definidos dentro de la metodología de desarrollo seleccionada, se puede concluir que:

1. Se realizó un análisis de las condiciones y los problemas que caracterizan el objeto de estudio demostrando la necesidad de desarrollar un Sistema de Recuperación de Información para el Sistema de Información de Mercadotecnia de Albet S.A.
2. Se efectuó una investigación los sistemas informáticos vinculados al campo de acción y de las fundamentales tendencias y tecnologías utilizadas para la implementación de sistemas de este tipo, basado en las definidas en el trabajo de diploma “Propuesta de arquitectura para el sistema de gestión de información de mercadotecnia para Albet S.A.” Obteniendo como resultado la selección de las herramientas adecuadas y la metodología para el desarrollo de la solución propuesta.
3. Se realizó una breve descripción del sistema que se propone desarrollar y una explicación del flujo actual del negocio para un mejor entendimiento del mismo. Además se definieron los requisitos funcionales a partir del modelo del negocio y las necesidades del cliente y se describieron a través de las historias de usuarios. Se determinaron los requisitos no funcionales que definen ciertas restricciones a la hora de implementar la propuesta solución.
4. Se describieron los métodos utilizados para la validación de los requerimientos con los resultados obtenidos evitando comprometer recursos en la implementación de requisitos que no cumplan las expectativas de los clientes o no estén suficientemente bien redactados como para no ser entendidos por los desarrolladores.

Con todo lo antes expuesto se considera que las bases teóricas, la descripción del entorno de desarrollo propuesto así como los artefactos descritos sirven de base para la posterior implementación del sistema que optimice el proceso de recuperación de información de Internet del Sistema de Información de Mercadotecnia de Albet S.A. y ayude a los especialistas de mercadotecnia en la toma de decisiones.



---

## **RECOMENDACIONES**

1. Implementar el sistema siguiendo las pautas que se definieron en este trabajo.
2. Para la implementación utilizar la última versión disponible de los *spiders* propuestos para poder explotar nuevas funciones que tengan.

---

## REFERENCIAS BIBLIOGRÁFICAS

1. **BELL, Martin L.** *Mercadotecnia : conceptos y estrategia*. México : Compañía Editorial Continental, 1987.
2. **KOTLER, Phillip.** *Mercadotecnia*. México : Prentice Hall, 1989.
3. **American Marketing Association.** *Marketing Power*. [En línea] American Marketing Association, 2011. <http://www.marketingpower.com/>.
4. **KOTLER, Phillip y ARMSTRONG, G.** *Fundamentos de Mercadotecnia*. México : Prentice Hall, 1991.
5. **Espasa Calpe, Paloma Grande.** *Diccionario de la Lengua Española*. España : Paloma Grande, 2005.
6. **TERCEIRO, José B.** *Del homo sapiens al homo digitalis. La sociedad digital*. Madrid : Alianza, 1996.
7. **NUÑES, Laura.** Conoce nuevos datos interesantes sobre Internet. [En línea] Enero de 2012. <http://www.conectate.com.do/articulo/datos-estadisticas-Internet-2012>.
8. **EMERY, James C.** *Sistemas de planeamiento y control de empresa: Teoría y tecnología*. Argentina : Librería El Ateneo, 1983.
9. **LEGITE.** *El Observatorio Tecnológico del sector Textil*. s.l. : Seminario, Dinamización, 2003.
10. **Norma UNE 166006:2006 EX.** *Sistema de Vigilancia Tecnológica*. s.l. : Gestion I+D+i, 2006.
11. **ROUACH, Daniel.** *La veille technologique et l'intelligence économique*. Paris : Presses Universitaires de France, 2010.
12. **ROVIRA, Cristòfol.** Vigilancia tecnológica e inteligencia competitiva para SEM-SEO. *Hipertext.net*, #6. [En línea] 2008. <http://www.hipertext.net>.
13. **BERROCAL, José L, FIGUEROLA, Carlos G y ZAZO, Angel F.** *Cibernetría: Nuevas Técnicas de Estudio Aplicables al Web*. Madrid : Ediciones Trea, 2004.
14. **MARTÍNEZ-RODRIGUEZ, Ailín.** *Indicadores cibernéticos: ¿Nuevas propuestas para medir la información en el entorno digital?* s.l. : Acimed, 2005. [En línea]: [http://bvs.sld.cu/revistas/aci/vol14\\_4\\_06/aci03406.htm](http://bvs.sld.cu/revistas/aci/vol14_4_06/aci03406.htm).
15. **MARTÍNEZ, Eduardo.** *Indicadores de ciencia y tecnología: estado del arte y perspectivas*. Caracas : Nueva Sociedad, 2008.

## Referencias bibliográficas

16. **MONDELO-HERNÁNDEZ, Yonny y MADUGA, Yuley DÍAZ.** *Características de la Web de la Universidad de las Ciencias Informáticas.* La Habana : Universidad de las Ciencias Informáticas, 2009.
17. **FERRAN, Núria y PEREZ-MONTORO, Maria.** *Búsqueda y Recuperación de la Información.* Barcelona : AOC, 2009.
18. **BAEZA-YATES, Ricardo y RIBEIRO-NETO, Berthier.** *Modern Information Retrieval.* Chile : Addison Wesley, 2011.
19. **BORDIGNON, Fernando R.A. y TOLOSA, Gabriel H.** *Recuperación de Información: un área de investigación en crecimiento.* Venezuela : Universidad Rafael Bellosos Chacin, 2007.
20. **Universidad Carlos III de Madrid.** Recuperación y Organización de la Información. [En línea] marzo de 2006. <http://recuperacionorganiza.tripod.com/>.
21. **LOPEZ-HERRERA, Antonio G.** *Modelos de Sistemas de Recuperación de Información Documental Basados en Información Lingüística Difusa.* Granada : Universidad de Granada, 2006.
22. **Consejo Superior de Cámaras de Comercio, Industria y Navegación de España.** CCámaras. [En línea] febrero de 2012. <http://www.camaras.org/publicado/index.html>.
23. **Universidad de las Ciencias Informáticas.** Vigitec, servicio de vigilancia tecnológica. [En línea] 2012. <http://vigitec.uci.cu/>.
24. **Departamento de Ciencias Informáticas de la Universidad de Chile.** cwr.cl. center for web research. [En línea] diciembre de 2011. <http://www.cwr.cl/projects/WIRE>.
25. **The Apache Software Foundation.** Lucene. [En línea] 2011. <http://lucene.apache.org/>.
26. **Oracle.** Oracle. [En línea] 2012. <http://www.oracle.com/index.html>.
27. **Lavtech.Com Corp.** mnoGoSearch. [En línea] 2011. <http://www.mnogosearch.org/>.
28. **MARCELO SCHENONE, Hernán.** *Diseño de una Metodología Ágil de Desarrollo de Software.* Buenos Aires : Fiuba, 2004.
29. **LETELIEL, Patricio.** *Jornada sobre Testeo de Software.* Valencia : s.n., 2006. [En línea]: [http://zweb.iti.upv.es/groups/squac/events/JTS2006/docs/JTS\\_2006.pdf/attach/JTS\\_2006.pdf](http://zweb.iti.upv.es/groups/squac/events/JTS2006/docs/JTS_2006.pdf/attach/JTS_2006.pdf).
30. **LETELIER, Patricio y PANDES, Ma Carmen.** *Metodologías ágiles para el desarrollo de software.* España : Universidad Politécnica de Valencia, 2006.
31. **PEÑALVER-ROMERO, Gladys M.** *Metodología ágil para proyectos de software libre.* La Habana : Universidad de las Ciencias Informáticas, 2007.

## Referencias bibliográficas

32. **JACOBSON, Ivar, BOOCH, Grady y RUMBAUGH, James.** *El proceso Unificado de Desarrollo de Software.* Madrid : Addison Wesley, 2000.
33. **INSTITUTO NACIONAL DE ESTADÍSTICA E INFORMÁTICA.** *Herramientas Case.* Dirección Técnica de Desarrollo Informático. Perú : s.n., 1999. pág. 7. 22.
34. **Visual Paradigm International.** Visual Paradigm. [En línea] 2012. <http://www.visual-paradigm.com/>.
35. **Pencil Project.** Pencil 1.23. [En línea] 2012. <http://pencil.evolus.vn/en-US/Home.aspx>.
36. **POTENCIER, Fabien.** *The Symphony Reference Guide.* España : Sensio Sa, 2009.
37. **CASTRO, Elizabeth.** *HTML con XHTML y CSS.* s.l. : Anaya Multimedia, 2007.
38. **SÁNCHEZ-MAZA, Miguel A.** *JavaScript.* s.l. : Innovación Y Cualificación, 2012.
39. **ZAKAS, Nicholas C., McPEAK, Jeremy y FAWCETT, Joe.** *Professional Ajax.* s.l. : John Wiley & Sons, 2011.
40. **SAWYER-McFARLAND, David.** *JavaScript y jQuery.* Madrid : Anaya Multimedia, 2012.
41. **SCHAFER, Steven M.** *HTML, XHTML, and CSS Bible.* s.l. : John Wiley & Sons, 2011.
42. **DOYLE, Matt.** *Beginning PHP 5.3.* s.l. : John Wiley & Sons, 2011.
43. **GALLARDO, David, BURNETTE, Ed y MCGOVERN, Robert.** *Eclipse in Action.* EEUU : Manning, 2003.
44. **PÉREZ, C., GONZÁLEZ-PÉREZ, María A. y PÉREZ-LÓPEZ, César.** *Administración de sistemas gestores de bases de datos.* s.l. : StarBook Editorial, S.A, 2011.
45. **WORSLEY, John C. y DRAKE, Joshua D.** *Practical PostgreSQL.* s.l. : O'Reilly Media, Inc., 2011.
46. **GARCÉS-PÉREZ, Oscar L.** *Propuesta de arquitectura para el sistema de gestión de información de mercadotecnia para Albet S.A.* La Habana : Universidad de las Ciencias Informáticas, 2011.
47. **KABIR, Mohammed J.** *Apache Server 2 Bible.* s.l. : John Wiley & Sons, 2002.
48. **Universidad Carnegie Mellon .** Software Engineering Institute. [En línea] marzo de 2012. <http://www.sei.cmu.edu/architecture/start/glossary/community.cfm>.
49. **SOMMERVILLE, Ian.** *Software Engineering.* s.l. : Addison-Wesley, 2007.
50. **REYNOSO, C. y KICILLOF, N.** *Estilos y Patrones en la estrategia de Arquitectura de Microsoft.* Buenos Aires : Universidad de Buenos Aires, 2004.

## Referencias bibliográficas

---

51. **AVOS Systems, Inc.** Welcome to Delicious! [En línea] 2012. <http://delicious.com/>.
52. **SOMMERVILLE, Ian.** *REQUIREMENTS ENGINEERING: A GOOD PRACTICE GUIDE*. s.l. : Wiley India Pvt. Ltd, 2009.
53. **BROOK, Fred.** *Essence and Accidents of Software Engineering*. s.l. : IEEE, 1987.
54. **IEEE Institute of Electrical and Electronics Engineers.** *IEEE standard for developing software life cycle processes*. s.l. : Institute of Electrical and Electronics Engineers, 1996.
55. **AURUM, Aybüke y WOHLIN, Claes.** *Engineering and Managing Software Requirements*. s.l. : Springer, 2005.
56. **KRUCHTEN, Philippe.** *Rational Unified Process, An Introduction*. s.l. : Addison-Wesley Professional, 2004.
57. **PRESSMAN, Roger S.** *Ingeniería del Software. Un enfoque práctico*. s.l. : McGraw-Hill, 2005.

## GLOSARIO

**AJAX:** (*Asynchronous JavaScript + XML*), que se puede traducir como “JavaScript Asíncrono + XML”.

**API:** (*Application Programming Interface*) conjunto de convenciones o reglamentos que definen cómo se debe llamar una determinada función de un programa desde una aplicación.

**Bibliometría:** estudio de los aspectos cuantitativos de la producción, diseminación y utilización de la información registrada.

**CSRF:** (*Cross Site Request Forgery*): tipo de ataque a un sitio web en el que comandos no autorizados son transmitidos por un usuario en el cual el sitio web confía.

**DOM:** (*Document Object Model*) es una interfaz de programación de aplicaciones que proporciona un conjunto estándar de objetos para representar documentos HTML y XML.

**Framework:** estructura conceptual y tecnológica de soporte definida, normalmente, con artefactos o módulos de software concretos, con base en la cual otro proyecto de software puede ser organizado y desarrollado. Típicamente, puede incluir soporte de programas, bibliotecas y un lenguaje interpretado entre otros programas para ayudar a desarrollar y unir los diferentes componentes de un proyecto.

**Hardware:** todas las partes tangibles de una computadora: sus componentes eléctricos, electrónicos, electromecánicos y mecánicos.

**HTML:** (*HyperText Markup Language*) lenguaje de marcado predominante para la elaboración de páginas web. Es usado para describir la estructura y el contenido en forma de texto, así como para complementar el texto con objetos tales como imágenes.

**HTTP:** (*HyperText Transfer Protocol*) protocolo de transferencia de hipertexto.

**Índice incremental:** si el indexador es capaz de añadir ficheros a un índice existente sin la necesidad de regenerar el índice completo.

**Índice invertido:** una lista ordenada de todas y cada una de las palabras que aparecen en los distintos documentos asociadas a los documentos concretos en los que aparecen.

**Indizar:** hacer índices, registrar ordenadamente datos e información para elaborar su índice.

**Indización:** técnica del análisis documental para representar y describir el contenido de los documentos, mediante conceptos principales contenidos en ellos (palabras clave) o vocabularios controlados (descriptores, términos o encabezamientos de materia), con el fin de guiar al usuario en la recuperación de los documentos que necesita.

**Informetría:** estudio de los aspectos cuantitativos de la información en cualquier forma.

**Multiplataforma:** término usado en aplicaciones, cuando se pueden ejecutar en diversas plataformas o sistemas operativos.

**OMG:** (*Object Management Group*) consorcio creado en 1989 responsable de la creación, desarrollo y revisión de especificaciones para la industria del software.

**PDF:** (*Portable Document Format*) formato de almacenamiento de documentos, desarrollado por la empresa Adobe Systems.

**PHP:** lenguaje de programación interpretado, diseñado originalmente para la creación de páginas web dinámicas. Es usado principalmente en interpretación del lado del servidor (*server-site scripting*) pero actualmente puede ser utilizado desde una interfaz de línea de comandos o en la creación de otros tipos de programas incluyendo aplicaciones con interfaz gráfica usando las bibliotecas Qt o GTK+.

**Prototipo:** modelo del comportamiento del sistema que puede ser usado para entenderlo completamente o ciertos aspectos de él y así clarificar los requerimientos. Un prototipo es una representación de un sistema, aunque no es un sistema completo, posee las características del sistema final o parte de ellas.

**Requerimientos:** condición o capacidad que debe estar presente en un sistema o componentes de sistema para satisfacer un contrato, estándar, especificación u otro documento formal. Los requerimientos pueden dividirse en requerimientos funcionales y requerimientos no funcionales.

**RSS:** (*Really Simple Syndication*) formato XML para compartir contenido en la web. Se utiliza para difundir información actualizada frecuentemente a usuarios que se han suscrito a la fuente de contenido.

**SIM:** Sistema de Información de Mercadotecnia.

**SQL:** (*Structured Query Language*) lenguaje de acceso a bases de datos relacionales que permite la especificación de distintos tipos de operaciones entre éstas.

**Stakeholders:** Personas a quienes pueden afectar o son afectados por los requisitos o resultados finales de un proyecto o empresa.

**Software:** Equipamiento lógico o soporte lógico de una computadora digital; comprende el conjunto de los componentes lógicos necesarios que hacen posible la realización de tareas específicas.

**Spider:** (Robot de búsqueda) programa que atraviesa una estructura de hipertexto recuperando ese enlace y el resto de los que están referenciados allí.

**URL:** (*Uniform Resource Locator*): secuencia de caracteres que identifica algún recurso de Internet ya sea una imagen, una página o cualquier otro recurso y debe escribirse de acuerdo a una sintaxis establecida por un formato estándar.

**WWW:** (*World Wide Web*) sistema de distribución de información basado en hipertexto enlazados y accesibles a través de Internet.

**XSS:** (*Cross Site Scripting*): Vulnerabilidad provocada por la falta de mecanismos de filtrado en los campos de entrada y permiten el ingreso y envío de datos sin validación alguna.



## ANEXOS

### Anexo 1: Modelo de entrevista.

#### MODELO DE ENTREVISTA

Nombre:

Apellidos:

Cargo que Ocupa:

#### **PREGUNTAS:**

1. ¿Quién es el cliente?
2. ¿Cuáles son sus capacidades y habilidades?
3. ¿En qué consiste el proceso?
4. Características fundamentales del proceso.
5. Objetivo principal del proceso.
6. ¿Qué beneficios brinda el proceso?
7. Tiempo que demora el proceso.
8. ¿Existen herramientas que ayuden a la ejecución del proceso?
9. ¿Qué limitaciones presenta el proceso actualmente?
10. ¿Cuál es la razón por la que se quiere resolver este problema?
11. ¿Qué desearía aportar al proceso para hacer el trabajo más eficiente?
12. ¿Qué afectaciones al negocio espera de la solución al problema?
13. Entorno de los usuarios (computadoras, plataformas).
14. ¿Existen requisitos legales, regulatorios u otros estándares que deban ser tenidos en cuenta?
15. ¿Cómo considera que se pueda validar la información obtenida?
16. ¿Qué considera una buena solución?

### Anexo 2: Aval de requerimientos.

## Universidad de las Ciencias Informáticas



### Aval de Requerimientos

Yo: Ing. Yenny Leal González que ocupo el cargo: Especialista de mercadotecnia de la empresa Albet Ingeniería y Sistemas S.A (Albet S.A), avalo los prototipos diseñados del presente Trabajo de Diploma: Propuesta para la recuperación de información proveniente de Internet para el Sistema de Información de Mercadotecnia de Albet S.A. para dar cumplimiento a los requerimientos del cliente y las condiciones que exige la universidad.

Este documento fue realizado a los 7 días del mes de Junio del año 2012.



Ing. Yenny Leal González


**Anexo 3: Lista de Chequeo.**

Atributo al que responden	Preguntas	HU-SRI-01	HU-SRI-02	HU-SRI-03	HU-SRI-04	HU-SRI-05	HU-SRI-06	HU-SRI-07	HU-SRI-08	HU-SRI-09	HU-SRI-10	HU-SRI-11	HU-SRI-12	HU-SRI-13	HU-SRI-14	HU-SRI-15	ERS
Comprensible	¿Están todos los requisitos redactados de forma simple y clara?																
	¿Deberían especificarse algunos requisitos con más detalle?																
No ambiguo	¿Tienen todos los requisitos una única interpretación? ¿Cuáles no?																
Precisión	¿Los requisitos que son medibles cuantitativamente están delimitados en términos cuantitativos?																
	¿Se utilizan en todos los requisitos delimitados en términos cuantitativos niveles apropiados de precisión para todas las cantidades numéricas?																
No redundante	¿Existe redundancia en la definición de los requisitos? ¿En cuáles?																
	¿Aparecen requisitos en más de un lugar del ERS? ¿Cuáles?																

Verificable	¿Todos los requisitos son verificables?																
Completo	¿Se pueden probar todos los requisitos una vez implementados?																
	¿Todos los requisitos documentados son correctos? ¿Cuáles no?																
Posible de tracear	¿Se puede <i>tracear</i> cada requisito al origen en el entorno del problema (proceso del negocio)? ¿Cuáles no?																
	¿Existe referencia de la dependencia del requisito con otros requisitos?																
Correcto	¿Existen requisitos que se contradicen con otros requisitos?																
Aspectos generales	¿Se pueden localizar todos los requisitos en el conjunto de objetivos del sistema/producto?																
	¿Todos los requisitos pueden ser implementados?																
	¿Se han identificado los requisitos de software y de hardware?																
	¿Se han identificado las restricciones de diseño e																

implementación?																	
¿Se han identificado las restricciones de interfaz externa?																	
¿Se han identificado los requisitos de soporte y usabilidad?																	
¿Se han identificado los requisitos de seguridad (confidencialidad, integridad, disponibilidad)?																	
¿Todos los requisitos identificados se centran en lo que el sistema debe hacer y no en cómo el sistema debe hacerlo?																	

### Anexo 4: Historias de Usuario

Historia de Usuario	
<b>Código:</b> HU-SRI-03	<b>Nombre:</b> Gestionar registro
<b>Referencia:</b> F 3	<b>Prioridad:</b> Alta
<b>Iteración asignada:</b> 1	<b>Puntos estimados:</b> 1
<p><b>Descripción:</b></p> <ul style="list-style-type: none"> <li>• El usuario selecciona la opción “Ver” cuando se pide que se registre a un sitio que está siendo analizado; o selecciona la opción general “Sitios para registrarse”.</li> <li>• Se muestra al usuario una lista de sitios a los que necesita registrarse.</li> <li>• El usuario selecciona el sitio y se abre un navegador con la dirección.</li> <li>• El usuario se registra en el sitio mientras el sitio y selecciona guardar usuario y contraseña.</li> <li>• El sistema guarda el usuario y la contraseña.</li> <li>• El usuario puede eliminar sitios de esta lista y estos serán eliminados del sistema.</li> <li>• El sistema agrega el sitio a la cola de URL</li> </ul>	
<p><b>Observaciones:</b></p> <ol style="list-style-type: none"> <li>1. El usuario puede acceder en cualquier momento a esta lista al hacer clic el icono de notificación el cual estará activo siempre que hallan operaciones pendientes.</li> <li>2. Para el registro, se utiliza una cuenta de correo específica del sistema.</li> </ol>	
<p><b>Prototipo de interfaz:</b></p> <div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p><b>Sitios por registrarse</b></p> </div> <div style="text-align: center;"> <p><b>Menú de notificaciones</b></p>  </div> </div>	





















Historia de Usuario	
<b>Código:</b> HU-SRI-04	<b>Nombre:</b> Gestionar estadísticas
<b>Referencia:</b> F 4	<b>Prioridad:</b> Alta
<b>Iteración asignada:</b> 2	<b>Puntos estimados:</b> 1
<b>Descripción:</b> <ul style="list-style-type: none"><li>• El usuario selecciona la opción Estadística.</li><li>• El sistema muestra una pantalla con los sitios registrados en el sistema.</li><li>• El usuario selecciona calcular estadística en aquellos sitios que nunca se les haya calculado.</li><li>• Se muestra una pantalla para seleccionar las características a calcular (cantidad de páginas, protocolo, dominio, visitas diarias, idioma, país sede, tipo de recursos, peso del sitio, entre otras).</li><li>• El usuario selecciona aquellos aspectos que desea saber del sitio y acepta.</li><li>• El sistema analiza completamente el sitio que el usuario seleccionó.</li></ul>	

### Observaciones:

1. Aquellos sitios a los que se les ha calculado estadísticas anteriormente, se podrá ver el resultado anterior en formato PDF, o recalculer la estadística.
2. Los sitios a los que nunca se le ha calculado la estadística, serán diferenciados de los demás a través de un ícono.

### Prototipo de interfaz:

#### Sitios registrados

Cálculos estadísticos			
http://www.marketingpower.com			
http://www.camaras.org/publicado/index.html			
http://intranet2.uci.cu/			
http://biblioteca2.uci.cu			
http://gespro.cenia.prod.uci.cu/			
http://www.mnogosearch.org/			

Calcular estadística  
Ver resultado anterior

#### Selección de la característica

Seleccione las características a calcular

<input checked="" type="checkbox"/> Cantidad de páginas	<input type="checkbox"/> Idioma
<input checked="" type="checkbox"/> Visitas diarias	<input type="checkbox"/> País sede
<input type="checkbox"/> Protocolo	<input checked="" type="checkbox"/> Última actualización
<input type="checkbox"/> Dominio	<input checked="" type="checkbox"/> Palabras más frecuentes
<input checked="" type="checkbox"/> Total de enlaces externos	<input checked="" type="checkbox"/> Peso del sitio
<input type="checkbox"/> Tipo de recursos	



Historia de Usuario	
<b>Código:</b> HU-SRI-5	<b>Nombre:</b> Clasificar sitios
<b>Referencia:</b> F 5	<b>Prioridad:</b> Alta
<b>Iteración asignada:</b> 2	<b>Puntos estimados:</b> 1
<b>Descripción:</b> <ul style="list-style-type: none"><li>• El sistema clasifica los sitios a los que se les ha realizado un análisis estadístico, de acuerdo a los diferentes aspectos seleccionados por el usuario para calcular.</li><li>• El sistema actualiza las listas de todos los aspectos de estadística.</li><li>• El usuario selecciona la opción “Ver Clasificaciones”.</li><li>• El sistema muestra la lista de características.</li><li>• El usuario puede seleccionar una característica, dentro de la que se encontrará la lista de sitios clasificados que desea analizar.</li></ul>	
<b>Observaciones:</b> <ol style="list-style-type: none"><li>1. Las listas están ordenadas según la relevancia de los sitios en cada uno de los aspectos (los que más enlaces externos reciben, los más populares, volumen de contenidos que tienen (multimedia, documentos, dinamismo de sus páginas) y las que no se ordenarán alfabéticamente).</li><li>2. Para ver la clasificación por una palabra clave, debe introducir una palabra, al aceptar, se muestra la lista de sitios ordenados por la frecuencia de aparición de la palabra.</li><li>3. Igual para el país, pero introduciendo el nombre del país.</li></ol>	

**Prototipo de interfaz:**

**Ver clasificación**

Inicial	Estadística																								
Búsqueda	Ver clasificación																								
Estadística																									
. Ver clasificación																									
Monitorización																									
	<table border="1" style="width: 100%;"> <tr> <td colspan="2" style="background-color: #0070c0; color: white;">Ver clasificación</td> </tr> <tr> <td>Factor de Impacto de la Web</td> <td rowspan="3" style="width: 50px;"></td> </tr> <tr> <td>Dinamismo de la Web</td> </tr> <tr> <td>Popularidad entre los usuarios</td> </tr> <tr> <td>http://intranet2.uci.cu</td> <td style="text-align: right;"></td> </tr> <tr> <td>http://inter-nos.uci.cu.uci.cu</td> <td style="text-align: right;"></td> </tr> <tr> <td>http://www.marketingpower.com</td> <td style="text-align: right;"></td> </tr> <tr> <td>http://biblioteca2.uci.cu</td> <td style="text-align: right;"></td> </tr> <tr> <td>http://gespro.cenia.prod.uci.cu</td> <td style="text-align: right;"></td> </tr> <tr> <td>http://www.camaras.org/publicado/index.html</td> <td style="text-align: right;"></td> </tr> <tr> <td>Densidad multimedia</td> <td></td> </tr> <tr> <td>Resumen de palabras frecuentes</td> <td></td> </tr> <tr> <td>País al que pertenece</td> <td></td> </tr> </table>	Ver clasificación		Factor de Impacto de la Web		Dinamismo de la Web	Popularidad entre los usuarios	http://intranet2.uci.cu		http://inter-nos.uci.cu.uci.cu		http://www.marketingpower.com		http://biblioteca2.uci.cu		http://gespro.cenia.prod.uci.cu		http://www.camaras.org/publicado/index.html		Densidad multimedia		Resumen de palabras frecuentes		País al que pertenece	
Ver clasificación																									
Factor de Impacto de la Web																									
Dinamismo de la Web																									
Popularidad entre los usuarios																									
http://intranet2.uci.cu																									
http://inter-nos.uci.cu.uci.cu																									
http://www.marketingpower.com																									
http://biblioteca2.uci.cu																									
http://gespro.cenia.prod.uci.cu																									
http://www.camaras.org/publicado/index.html																									
Densidad multimedia																									
Resumen de palabras frecuentes																									
País al que pertenece																									

**Clasificación por palabra clave**

Inicial	Estadística																								
Búsqueda	Ver clasificación																								
Estadística																									
. Ver clasificación																									
Monitorización																									
	<table border="1" style="width: 100%;"> <tr> <td colspan="2" style="background-color: #0070c0; color: white;">Ver clasificación</td> </tr> <tr> <td>Factor de Impacto de la Web</td> <td rowspan="3" style="width: 50px;"></td> </tr> <tr> <td>Dinamismo de la Web</td> </tr> <tr> <td>Popularidad entre los usuarios</td> </tr> <tr> <td>Densidad multimedia</td> <td></td> </tr> <tr> <td>Resumen de palabras frecuentes</td> <td></td> </tr> <tr> <td>Palabra Clave</td> <td> <input type="text" value="introduzca una palabra clave....."/> </td> </tr> <tr> <td>http://intranet2.uci.cu</td> <td style="text-align: right;"></td> </tr> <tr> <td>http://inter-nos.uci.cu.uci.cu</td> <td style="text-align: right;"></td> </tr> <tr> <td>http://www.marketingpower.com</td> <td style="text-align: right;"></td> </tr> <tr> <td>http://biblioteca2.uci.cu</td> <td style="text-align: right;"></td> </tr> <tr> <td>http://gespro.cenia.prod.uci.cu</td> <td style="text-align: right;"></td> </tr> <tr> <td>http://www.camaras.org/publicado/index.html</td> <td style="text-align: right;"></td> </tr> </table>	Ver clasificación		Factor de Impacto de la Web		Dinamismo de la Web	Popularidad entre los usuarios	Densidad multimedia		Resumen de palabras frecuentes		Palabra Clave	<input type="text" value="introduzca una palabra clave....."/>	http://intranet2.uci.cu		http://inter-nos.uci.cu.uci.cu		http://www.marketingpower.com		http://biblioteca2.uci.cu		http://gespro.cenia.prod.uci.cu		http://www.camaras.org/publicado/index.html	
Ver clasificación																									
Factor de Impacto de la Web																									
Dinamismo de la Web																									
Popularidad entre los usuarios																									
Densidad multimedia																									
Resumen de palabras frecuentes																									
Palabra Clave	<input type="text" value="introduzca una palabra clave....."/>																								
http://intranet2.uci.cu																									
http://inter-nos.uci.cu.uci.cu																									
http://www.marketingpower.com																									
http://biblioteca2.uci.cu																									
http://gespro.cenia.prod.uci.cu																									
http://www.camaras.org/publicado/index.html																									

Historia de Usuario	
<b>Código:</b> HU-SRI-06	<b>Nombre:</b> Construir resumen de estadísticas
<b>Referencia:</b> F 6	<b>Prioridad:</b> Alta

<b>Iteración asignada:</b> 2	<b>Puntos estimados:</b> 1
<b>Descripción:</b> <ul style="list-style-type: none"><li>• El usuario selecciona la opción “Estadísticas”.</li><li>• El sistema muestra la lista de sitios registrados.</li><li>• El usuario selecciona la opción ver resumen y se le muestra un menú para volver a calcular la estadística, o ver la actual. Selecciona “Ver resultado anterior”.</li><li>• El sistema crea un documento en formato PDF en el que se incluye el resumen de las estadísticas que se tienen sobre el sitio.</li></ul>	
<b>Observaciones:</b>	
<b>Prototipo de interfaz:</b> <b>Ejemplo de resumen de estadísticas</b>	

Categoría	Total	Porcentaje	Interno	Externo
<b>Todos los enlaces</b>	<b>233</b>		<b>109</b>	<b>124</b>
<b>Interno</b>	<b>109</b>	<b>47.0%</b>		
<b>Externo</b>	<b>124</b>	<b>53.0 %</b>		
<b>Enlaces buenos</b>	<b>105</b>	<b>45.0%</b>		
<b>Enlaces rotos</b>	<b>4</b>	<b>2.0%</b>		
<b>Enlaces no verificados</b>	<b>124</b>	<b>53.0%</b>		
<b>Páginas procesadas</b>	<b>57</b>	<b>24.0%</b>		
<b>Páginas con enlaces rotos</b>	<b>57</b>	<b>100.0%</b>		
<b>Páginas lentas</b>	<b>2</b>	<b>4.0%</b>		
<b>Páginas pequeñas</b>	<b>0</b>	<b>0.0%</b>		
<b>Protocolos</b>	<b>233</b>			
<b>http</b>	<b>228</b>	<b>98.0%</b>		
<b>https</b>	<b>5</b>	<b>2.0%</b>		
<b>Encontrados en etiquetas HTML</b>	<b>232</b>	<b>100.0%</b>		
<b>&lt;A&gt;</b>	<b>179</b>	<b>77.0%</b>		
<b>&lt;IMG&gt;</b>	<b>10</b>	<b>4.0%</b>		
<b>&lt;LINK&gt;</b>	<b>25</b>	<b>11.0%</b>		
<b>&lt;SCRIPT&gt;</b>	<b>16</b>	<b>7.0%</b>		
<b>&lt;FRAME&gt;</b>	<b>0</b>	<b>0.0%</b>		
<b>&lt;FORM&gt;</b>	<b>2</b>	<b>1.0%</b>		
<b>&lt;IFRAME&gt;</b>	<b>0</b>	<b>0.0%</b>		
<b>&lt;BGSOUND&gt;</b>	<b>0</b>	<b>0.0%</b>		

Historia de Usuario	
<b>Código:</b> HU-SRI-07	<b>Nombre:</b> Gestionar listas top

<b>Referencia:</b> F 7	<b>Prioridad:</b>
<b>Iteración asignada:</b>	<b>Puntos estimados:</b> 1
<b>Descripción:</b> <ul style="list-style-type: none"><li>• El usuario selecciona la opción “Calcular lista TOP”.</li><li>• El sistema muestra una ventana con las diferentes listas que se pueden calcular (mayores competidores, principales clientes, principales mercados, sectores de mayor demanda y principales productos).</li><li>• El usuario selecciona una de estas opciones y acepta.</li><li>• El sistema muestra una lista ordenada por relevancia de los sitios registrados que responden a la clasificación seleccionada por el usuario.</li></ul>	
<b>Observaciones:</b> <ol style="list-style-type: none"><li>1. La relevancia de los sitios es calculada por popularidad y demanda.</li></ol>	
<b>Prototipo de interfaz:</b> <p><b>Opciones de diferentes listas</b></p> <div data-bbox="162 1071 852 1627"><p>Seleccione las listas TOP 10 que desea ver</p><ul style="list-style-type: none"><li><input type="checkbox"/> Principales clientes</li><li><input type="checkbox"/> Mayores competidores</li><li><input type="checkbox"/> Principales mercados</li><li><input type="checkbox"/> Principales productos</li><li><input type="checkbox"/> Sectores de mayor demanda</li></ul><p><input type="button" value="Aceptar"/> <input type="button" value="Cancelar"/></p></div> <p><b>Lista TOP 10</b></p>	

Estadística	
Listas TOP 10	
<b>Principales clientes:</b>	
http://www.uci.cu	
http://www.infomed.sld.cu	
http://gespro.cenia.prod.uci.cu	
http://www.mmoven.com.ve	
http://firefoxmania.uci.cu	
http://gespro.albet.prod.uci.cu/links	
http://seguridad.uci.cu	
http://uclv.cu/index	
http://ucistore.uci.cu	
http://datalab.uci.cu	


Historia de Usuario	
<b>Código:</b> HU-SRI-08	<b>Nombre:</b> Gestionar posibles clientes
<b>Referencia:</b> F 8	<b>Prioridad:</b> Baja
<b>Iteración asignada:</b> 2	<b>Puntos estimados:</b> 1
<b>Descripción:</b>	
<ul style="list-style-type: none"> <li>• El usuario selecciona la opción “Clientes potenciales”.</li> <li>• El sistema muestra una pantalla donde introducir como criterio de búsqueda el producto o servicio en venta.</li> <li>• El usuario introduce el criterio de búsqueda y acepta.</li> <li>• El sistema muestra como respuesta un listado de los sitios pertenecientes a posibles clientes del producto o servicio en venta.</li> </ul>	

**Observaciones:**

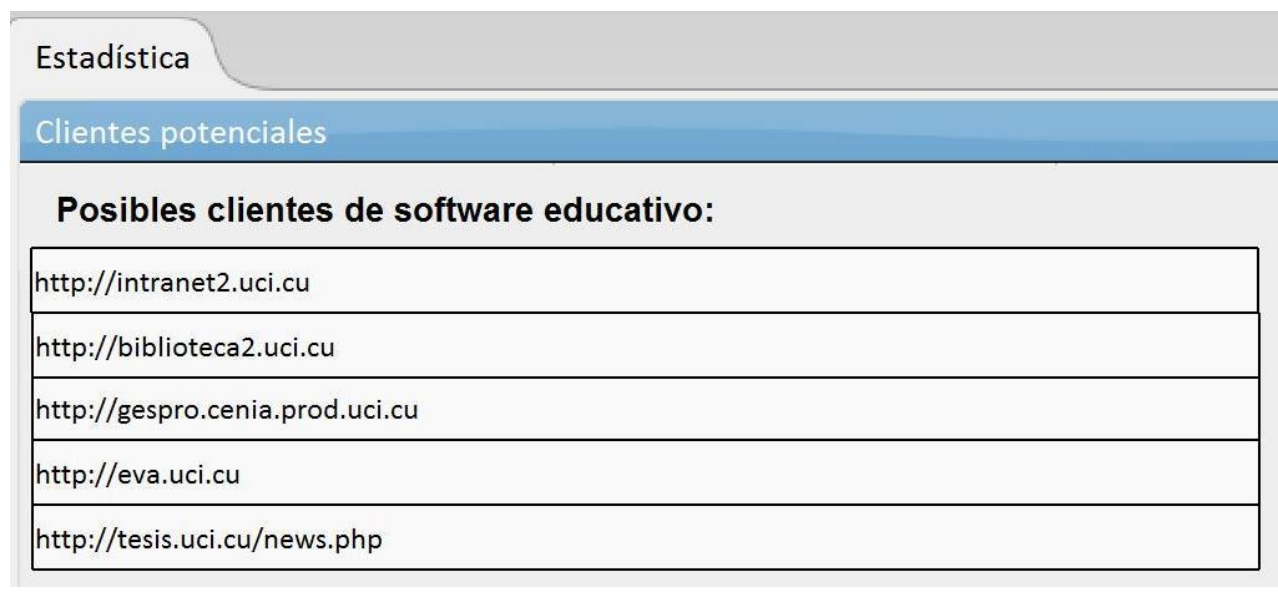
1. El sistema busca dentro de la lista de clientes, aquellos que cuenten con alguna licitación o noticia de que buscan ese determinado producto o servicio en venta.

**Prototipo de interfaz:**

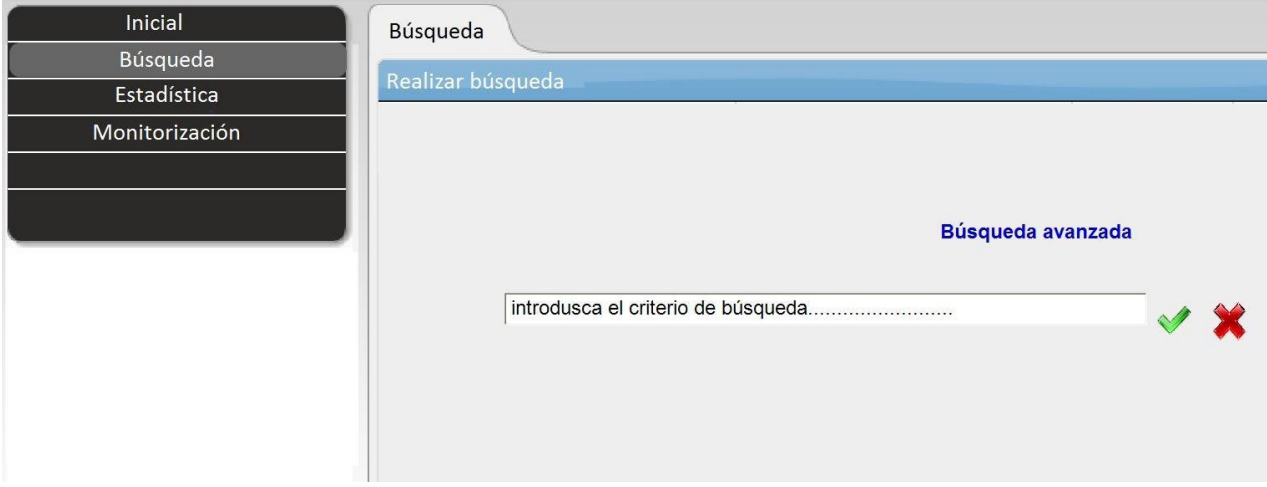
**Búsqueda de clientes potenciales**



**Respuesta de posibles clientes**



Posibles clientes de software educativo:
<a href="http://intranet2.uci.cu">http://intranet2.uci.cu</a>
<a href="http://biblioteca2.uci.cu">http://biblioteca2.uci.cu</a>
<a href="http://gespro.cenia.prod.uci.cu">http://gespro.cenia.prod.uci.cu</a>
<a href="http://eva.uci.cu">http://eva.uci.cu</a>
<a href="http://tesis.uci.cu/news.php">http://tesis.uci.cu/news.php</a>

Historia de Usuario	
<b>Código:</b> HU-SRI-09	<b>Nombre:</b> Gestionar búsqueda por palabras claves
<b>Referencia:</b> F 9	<b>Prioridad:</b> Muy Alta
<b>Iteración asignada:</b> 3	<b>Puntos estimados:</b> 2
<p><b>Descripción:</b></p> <ul style="list-style-type: none"> <li>• El usuario selecciona la opción “Búsqueda”.</li> <li>• El sistema muestra una pantalla donde introducir el criterio de búsqueda.</li> <li>• El usuario introduce el criterio de búsqueda y acepta.</li> </ul>	
<p><b>Observaciones:</b></p> <ol style="list-style-type: none"> <li>1. El sistema muestra en una notificación de que hay sitios no registrados en el sistema, que podrían satisfacer el criterio de búsqueda. Si el usuario selecciona “Analizar luego” se agrega el enlace a la lista de sitios tentativos.</li> <li>2. El usuario puede acceder en cualquier momento a esta lista al hacer clic el icono de notificación el cual estará activo siempre que hayan operaciones pendientes.</li> <li>3. Si selecciona no analizar, se construirá el resultado sin tomar en cuenta estos sitios.</li> </ol>	
<p><b>Prototipo de interfaz:</b></p> <p><b>Opción de búsqueda</b></p> 	
<p><b>Alerta de sitios tentativos</b></p>	






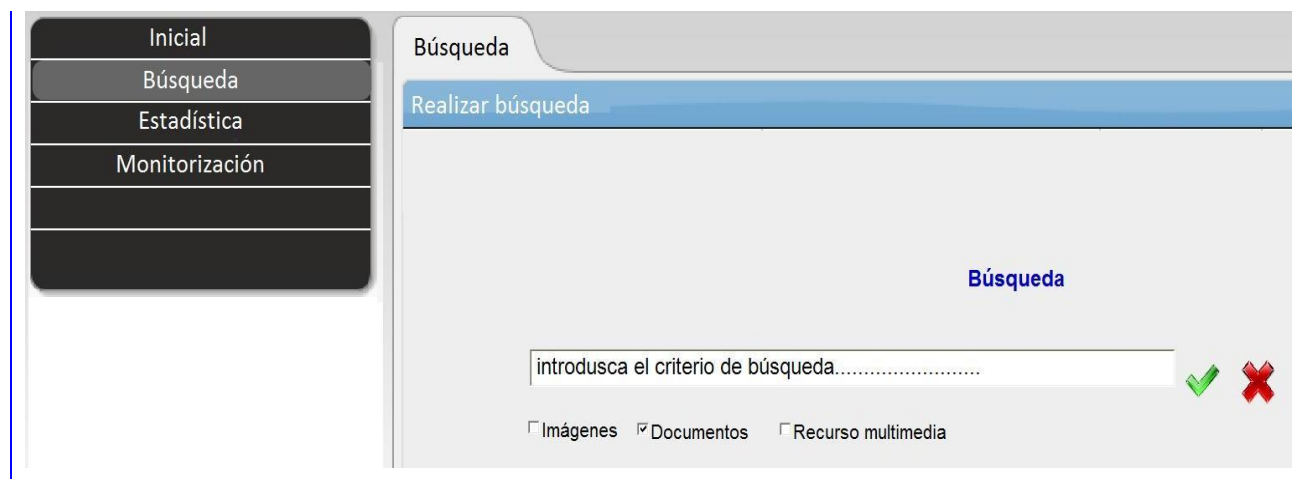
**Hay sitios que podrían responder al criterio de búsqueda no registrados en el sistema.....**

Combinar respuesta
Analizar luego
No analizar

**Menú de notificaciones**



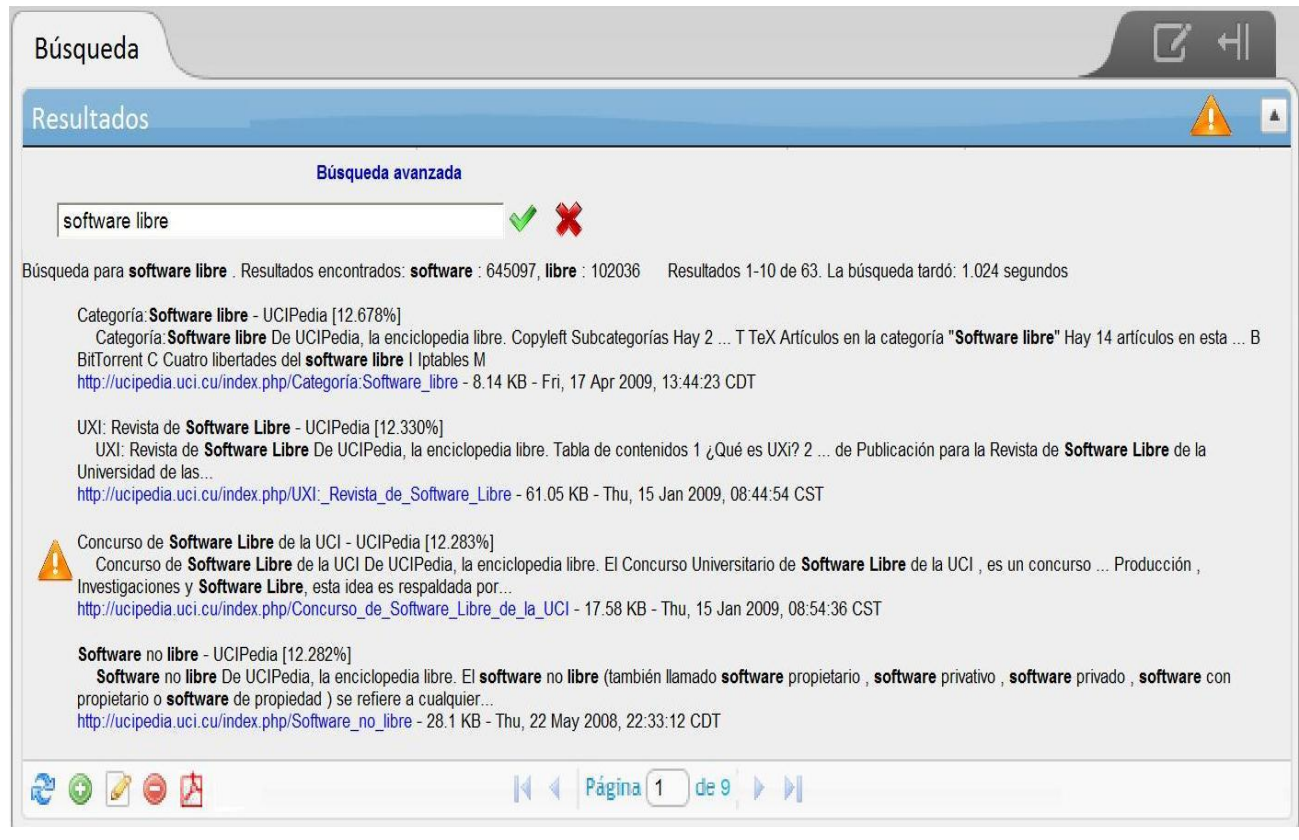
<b>Historia de Usuario</b>	
<b>Código:</b> HU-SRI-10	<b>Nombre:</b> Gestionar búsqueda avanzada
<b>Referencia:</b> F 10	<b>Prioridad:</b> Media
<b>Iteración asignada:</b> 3	<b>Puntos estimados:</b> 1
<p><b>Descripción:</b></p> <ul style="list-style-type: none"> <li>El usuario selecciona la opción “Búsqueda” y selecciona “Búsqueda avanzada” en la pantalla que se le muestra.</li> <li>El sistema muestra una pantalla donde introducir el criterio de búsqueda y opciones de recursos a buscar (imágenes, documentos o recursos multimedia).</li> <li>El usuario introduce el criterio de búsqueda y selecciona a qué recursos quiere limitar la búsqueda.</li> </ul>	
<p><b>Observaciones:</b></p> <ol style="list-style-type: none"> <li>1. Se puede seleccionar un solo recurso.</li> </ol>	
<p><b>Prototipo de interfaz:</b></p> <p><b>Opciones de búsqueda avanzada</b></p>	



Historia de Usuario	
<b>Código:</b> HU-SRI-11	<b>Nombre:</b> Construir respuesta de la búsqueda
<b>Referencia:</b> F 11	<b>Prioridad:</b> Muy Alta
<b>Iteración asignada:</b> 3	<b>Puntos estimados:</b> 2
<b>Descripción:</b> <ul style="list-style-type: none"> <li>• El usuario selecciona cualquiera de las opciones (combinar respuesta, analizar luego, no analizar) de la notificación en la ventana de búsqueda.</li> <li>• El sistema muestra una lista ordenada de aquellos sitios o documentos que satisfacen el criterio de búsqueda.</li> </ul>	
<b>Observaciones:</b> <ol style="list-style-type: none"> <li>1. El sistema primero analiza las búsquedas anteriores si coinciden los criterios de búsqueda y muestra como respuesta también la información almacenada en esa búsqueda anterior.</li> <li>2. Si el usuario seleccionó "Combinar respuesta" un icono de notificación indicará aquellos sitios que no están registrados en el sistema.</li> </ol>	

## Prototipo de interfaz:

### Resultados de la búsqueda



The screenshot shows a web browser window titled 'Búsqueda' with a search bar containing 'software libre'. Below the search bar, it displays 'Búsqueda avanzada' and search statistics: 'Búsqueda para software libre . Resultados encontrados: software : 645097, libre : 102036 Resultados 1-10 de 63. La búsqueda tardó: 1.024 segundos'. The results list includes categories like 'Software libre - UCIPedia [12.678%]', 'UXI: Revista de Software Libre - UCIPedia [12.330%]', and 'Concurso de Software Libre de la UCI - UCIPedia [12.283%]'. At the bottom, there is a navigation bar showing 'Página 1 de 9'.

Historia de Usuario	
<b>Código:</b> HU-SRI-12	<b>Nombre:</b> Almacenar información
<b>Referencia:</b> F 12	<b>Prioridad:</b> Alta
<b>Iteración asignada:</b> 3	<b>Puntos estimados:</b> 1
<b>Descripción:</b> <ul style="list-style-type: none"> <li>El usuario selecciona descargar información en la opción de descarga de la respuesta analizada.</li> <li>El sistema muestra un menú donde el usuario podrá escoger que información descargar (sitio completo, pagina que responde a la búsqueda, documentación o recursos multimedia).</li> </ul>	

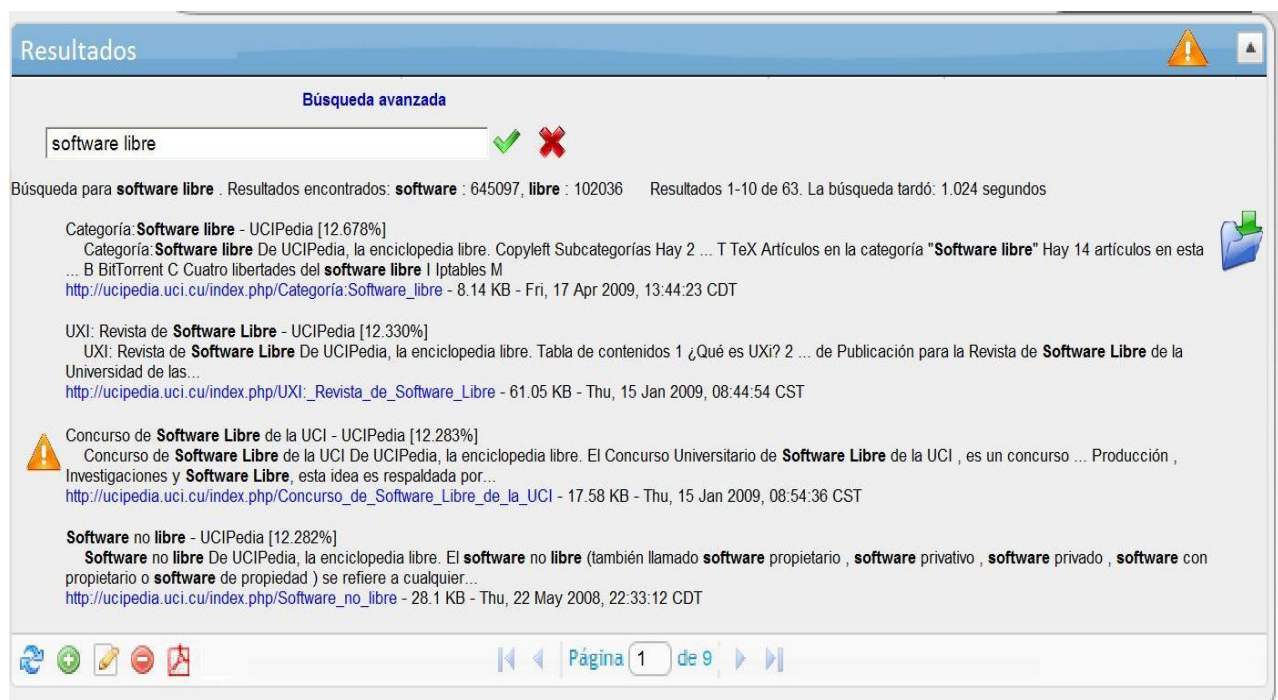
- El usuario selecciona lo que desea descargar y acepta.
- El sistema descarga del sitio solo lo seleccionado por el usuario y lo almacena clasificándolo por el criterio de búsqueda por el que se encontró.

#### Observaciones:

1. Solo aparecerá el icono de descarga luego de visitado la respuesta.

#### Prototipo de interfaz:

##### Resultados de la búsqueda y opción para descargar



The screenshot shows a web browser window titled "Resultados". The search bar contains "software libre" with a green checkmark and a red X icon. Below the search bar, it displays "Búsqueda para software libre . Resultados encontrados: software : 645097, libre : 102036 Resultados 1-10 de 63. La búsqueda tardó: 1.024 segundos". The results list includes:

- Categoría: **Software libre** - UCIPedia [12.678%]  
Categoría: **Software libre** De UCIPedia, la enciclopedia libre. Copyleft Subcategorías Hay 2 ... T TeX Artículos en la categoría "**Software libre**" Hay 14 artículos en esta ... B BitTorrent C Cuatro libertades del **software libre** I Iptables M  
[http://ucipedia.uci.cu/index.php/Categoría:Software\\_libre](http://ucipedia.uci.cu/index.php/Categoría:Software_libre) - 8.14 KB - Fri, 17 Apr 2009, 13:44:23 CDT
- UXI: Revista de **Software Libre** - UCIPedia [12.330%]  
UXI: Revista de **Software Libre** De UCIPedia, la enciclopedia libre. Tabla de contenidos 1 ¿Qué es UXI? 2 ... de Publicación para la Revista de **Software Libre** de la Universidad de las...  
[http://ucipedia.uci.cu/index.php/UXI:\\_Revista\\_de\\_Software\\_Libre](http://ucipedia.uci.cu/index.php/UXI:_Revista_de_Software_Libre) - 61.05 KB - Thu, 15 Jan 2009, 08:44:54 CST
- Concurso de **Software Libre** de la UCI - UCIPedia [12.283%]  
Concurso de **Software Libre** de la UCI De UCIPedia, la enciclopedia libre. El Concurso Universitario de **Software Libre** de la UCI , es un concurso ... Producción , Investigaciones y **Software Libre**, esta idea es respaldada por...  
[http://ucipedia.uci.cu/index.php/Concurso\\_de\\_Software\\_Libre\\_de\\_la\\_UCI](http://ucipedia.uci.cu/index.php/Concurso_de_Software_Libre_de_la_UCI) - 17.58 KB - Thu, 15 Jan 2009, 08:54:36 CST
- Software no libre** - UCIPedia [12.282%]  
**Software no libre** De UCIPedia, la enciclopedia libre. El **software no libre** (también llamado **software** propietario , **software** privativo , **software** privado , **software** con propietario o **software** de propiedad ) se refiere a cualquier...  
[http://ucipedia.uci.cu/index.php/Software\\_no\\_libre](http://ucipedia.uci.cu/index.php/Software_no_libre) - 28.1 KB - Thu, 22 May 2008, 22:33:12 CDT

The interface also features a navigation bar at the bottom with icons for home, search, and other functions, and a page indicator showing "Página 1 de 9".

#### Opciones de descarga



**¿Qué desea descargar?**

Sitio completo

Página que responde a la búsqueda

Documentación

Recursos multimedia

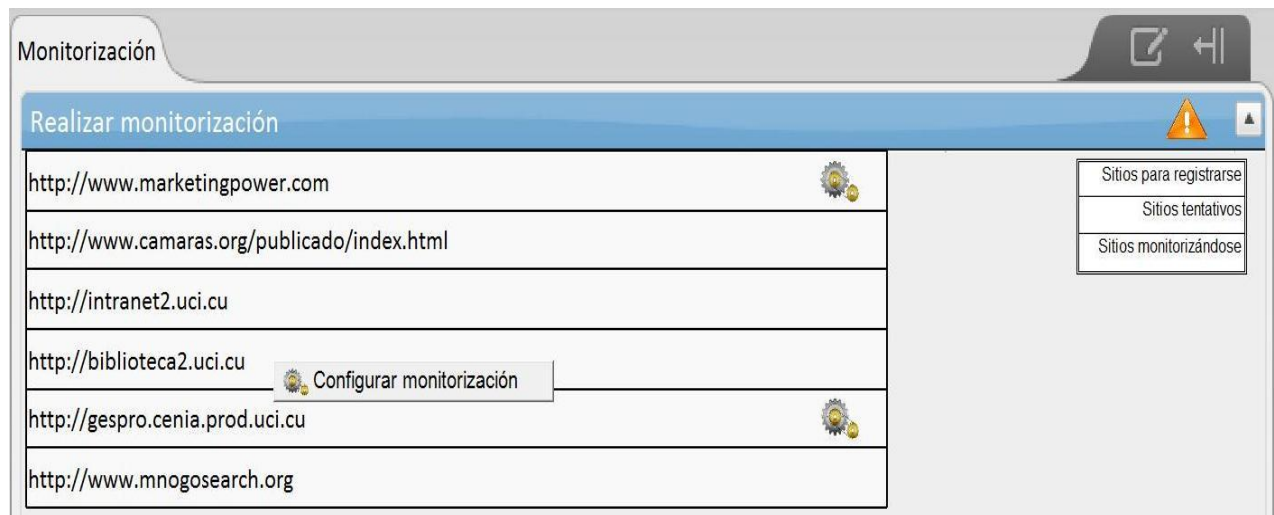



Historia de Usuario	
<b>Código:</b> HU-SRI-13	<b>Nombre:</b> Gestionar monitoreo
<b>Referencia:</b> F 13	<b>Prioridad:</b> Alta
<b>Iteración asignada:</b> 4	<b>Puntos estimados:</b> 1
<p><b>Descripción:</b></p> <ul style="list-style-type: none"> <li>El usuario selecciona "Monitorización".</li> <li>El sistema muestra una pantalla con los sitios registrados en el sistema para que el usuario escoja aquel que necesita monitorear.</li> <li>El usuario selecciona uno de los sitios y selecciona la opción de configurar monitorización.</li> <li>El sistema muestra una ventana donde configurar los elementos de la monitorización (tiempo de intervalo, duración del monitoreo).</li> <li>El sistema analiza el sitio cada vez que se cumple el intervalo de tiempo y hasta que finalicen los días determinados por el usuario.</li> </ul>	
<p><b>Observaciones:</b></p> <ol style="list-style-type: none"> <li>1. Aquellos sitios que están monitorizándose estarán identificados.</li> </ol>	

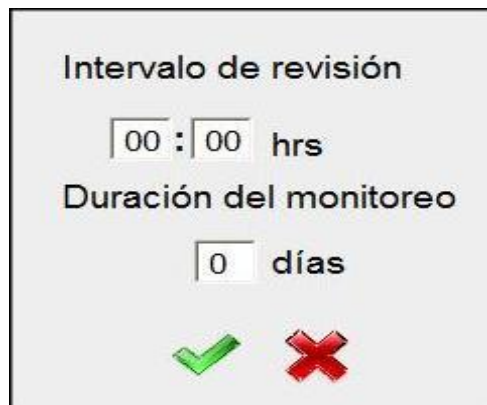
2. El usuario define cada qué tiempo necesita que se revise el sitio y durante cuántos días quiere mantenerlo monitoreado.
3. El usuario siempre podrá acceder a la lista de sitios que están monitorizándose a través del menú de notificación el cual estará activo siempre que hallan operaciones pendientes.

**Prototipo de interfaz:**

**Listado de sitios en el sistema para monitorear**



**Configuración del monitoreo**



Historia de Usuario	
<b>Código:</b> HU-SRI-14	<b>Nombre:</b> Construir reporte y notificaciones del monitoreo
<b>Referencia:</b> F 14	<b>Prioridad:</b> Alta



Iteración asignada: 4

Puntos estimados: 1

**Descripción:**

- El sistema, cada vez que ocurre alguna actualización o cambio en el sitio monitoreado muestra una alerta notificando que ha habido algún cambio en el sitio.
- El usuario selecciona ver y se le muestra el sitio en la página donde ocurrió la modificación.
- Finalizado el tiempo de monitoreo, el sistema muestra una alerta de que concluyo el monitoreo en el sitio.
- El usuario selecciona ver resumen.
- El sistema muestra en formato PDF el resumen del monitoreo.

**Observaciones:**

1. El resumen contendrá toda la información relevante recolectada durante el monitoreo (cantidad de actualizaciones, modificaciones al sitio, noticias nuevas, etc.).

**Prototipo de interfaz:**



**Alerta de cambio o actualización encontrada**



**Alerta de fin de monitoreo**



Historia de Usuario

<b>Código:</b> HU-SRI-15	<b>Nombre:</b> Gestionar reportes resumen
<b>Referencia:</b> F 15	<b>Prioridad:</b> Media
<b>Iteración asignada:</b> 4	<b>Puntos estimados:</b> 1
<p><b>Descripción:</b></p> <ul style="list-style-type: none"> <li>• El usuario selecciona la opción “Ver reportes”</li> <li>• El sistema muestra una ventana con las clasificaciones de los sitios (clientes, competidores, distribuidores y mercado).</li> <li>• El usuario selecciona una y acepta.</li> <li>• El sistema muestra una lista con los sitios de (clientes, competidores, distribuidores o mercado) de acuerdo al seleccionado por el usuario.</li> <li>• El usuario selecciona aquel del que desea ver un resumen.</li> <li>• El sistema crea un documento en formato PDF en el que incluye el resumen que consiste en (como se llaman, país, idioma, autor, principales productos (en caso que tenga) )</li> </ul>	
<p><b>Observaciones:</b></p> <ol style="list-style-type: none"> <li>1. El sistema incluye en el resumen la información que brindan los sitios sobre ellos (¿Quiénes somos?).</li> </ol>	
<p><b>Prototipo de interfaz:</b></p> <p><b>Clasificaciones</b></p> <div style="border: 1px solid gray; padding: 10px; background-color: #f0f0f0;"> <p style="text-align: center;"><b>Clasificación</b></p> <p><input type="checkbox"/> Clientes      <input type="checkbox"/> Mercados</p> <p><input type="checkbox"/> Competidores      <input type="checkbox"/> Distribuidores</p> <p style="text-align: center;">      </p> </div> <p><b>Listado para seleccionar de quién ver el resumen</b></p>	



## Listado de competidores

### Competidores

<http://www.uci.cu>

<http://www.infomed.sld.cu>

<http://gespro.cenia.prod.uci.cu>

<http://www.mmoven.com.ve>

<http://firefoxmania.uci.cu>

<http://gespro.albet.prod.uci.cu/links>

<http://seguridad.uci.cu>

<http://uclv.cu/index>

<http://ucistore.uci.cu>

<http://datalab.uci.cu>