Tesis de Magister en Ingeniería del Software

AMBIENTE DE INTEGRACIÓN DE HERRAMIENTAS PARA EXPLORACIÓN DE DATOS CENTRADOS EN LA WEB

Autor: **Esp. Ing. Hernán Merlino**

Directores: **M.Ing. Paola Britos**
Dr. Ramón García Martínez

Noviembre 2005

A la memoria de mi padre Luciano.

A la memoria de Osvaldo.

Para mi madre y mi abuela, que tanto han hecho para que pueda terminar mis estudios.

A Ramón y Paola por guiarme en el camino del saber.

RESUMEN

La existencia de sistemas informáticos de uso libre orientados a la exploración de uso, la exploración de contenido y la exploración de estructura en Web y la identificación de procesos de exploración en Web que requieren la integración articulada de dichos artefactos son la motivación de esta tesis. En este contexto, en este trabajo se propone una herramienta para exploración de datos Web que permite estructurar todo el proceso de exploración. La mayor ventaja de esta herramienta es poder utilizar diversas técnicas de exploración, además de permitir la reutilización de procesos ya ejecutados con anterioridad y la combinación de los mismos para su posterior comparación; todo esto llevado a cabo sin un alto grado de complejidad. La herramienta desarrollada satisface los siguientes requerimientos: los procesos que ejecuta son modulares y flexibles, la información que entrega es verificable y verificable, el sistema tiene la capacidad para agendar tareas, el sistema puede ser ejecutado en entornos Windows, Unix y Linux, el sistema puede admite varias fuentes de dato de entradas y el sistema puede admite formato flexible de archivos de salida.

ABSTRACT

The motivation if this thesis is; the existence of open source software for Web use mining, Web content mining and Web structure mining and the identification of process of Web mining that required the articulation of these artifacts. This framework permits the generation of a structured process of Web mining. The more important advantage of this framework is the ability of using a lot of Web mining techniques; and it’s possible to reuse the process; and these process can be mixed and compared; the framework can do all the tasks described above without effort. The framework has the following requirements: the process has to be modular and flexible; the information generated has to be verifiable; schedule task is once of the abilities of the framework; the framework has to run on Windows, Unix and Linux; the framework has to accept a lot of formats of inputs and has to generate a lot of formats of outputs.

INDICE

1. INTRODUCCIÓN

La abundancia de datos en Internet y la necesidad de información de las empresas, hace que sea necesario el surgimiento de un nuevo tipo de herramientas. Las mismas se basan en sistemas independientes de búsqueda que utilizan a la red de redes como medio de recolección de datos; y mediante la utilización de diversas técnicas de inteligencia artificial estos datos son transformados en información.

La tasa de crecimiento de Internet es tan alta que a diario se crean nuevos sitios con datos que pueden ser relevantes a nuestros intereses, como ser en el ambiente educativo, nuevas tesis publicadas sobre algún tema de nuestro interés, en el ámbito empresarial la aparición de potenciales competidores y clientes; o el cambio de estrategia de algunos de nuestros competidores ya identificados.

Al identificar estas fuentes se produce el refinamiento de nuestras búsquedas a medida que conocemos mas nuestras propias necesidades. Esto se suma a la aparición de nuevas tecnologías de representación de los datos en Internet que hace que sea necesario una constante actualización de las técnicas de obtención de datos.

Esto hace que sea necesario la utilización de herramientas inteligentes para podernos mantener en el estado del arte de la información que nos es relevante.

Surge la necesidad de disponer de una herramienta que sirva de base para poder implementar una solución a esto problema. La herramienta debería proveer las siguientes facilidades:

Método de Comunicación entre módulos: la comunicación entre los distintos módulos es un aspecto de fundamental importancia en este esquema, el intercambio de procesos debe ser transparente para lograr la abstracción requerida por la herramienta.

Agenda de tareas: que permita administrar el flujo de tareas, su ejecución con independencia de los módulos de las distintas tareas. Esta facilidad

deberá incluir el procedimiento para el re-agendamiento de las tareas en caso de error de alguna de ellas y los métodos de aviso.

Seguimiento de Logs: toda tarea que se ejecute dejará información de proceso, para ser utilizada en la re-ingeniería inversa de los procesos, para el refinamiento de las búsquedas.

Patrón de diseño de tareas: la selección del lenguaje de programación para cada uno de las tareas es independiente, solo se generaran reglas a seguir para la entrada y salida de datos, manejo de errores y excepciones.

La tesis está estructurada en siete capítulos y tres anexos.

En el capítulo Estado de la Cuestión se presenta una introducción a la exploración en web, se describe: exploración de uso de datos en la web, exploración de contenidos de la web y exploración de estructura de la web, se indican las etapas en la que se divide la exploración de web, se proponen herramientas de libre uso, susceptibles de ser integradas, para realizar exploraciones de uso, contenido y estructura y se resumen trabajos relacionados con las propuestas de esta tesis.

En el capítulo Descripción del Problema se presenta el problema cuya solución se aborda en esta tesis.

En el capítulo Solución Propuesta en este capítulo se describe el sistema que da soporte a la solución propuesta presentandose la planificación de sistemas de información con detalle de: inicio del plan de sistemas de información, definición y organización del PSI, estudio de la información relevante, identificación de requisitos, estudio de los sistemas de información actuales, diseño del modelo de sistemas de información, definición de la arquitectura tecnológica, estimación del proyecto y definición del plan de acción; y se presentan aspectos del desarrollo de sistemas de información lo que incluye estudio de viabilidad del sistema, gestión de configuración, análisis del sistema de información, diseño del sistema de información, construcción del sistema de información y pruebas de software.

En el capítulo Experimentación se presenta un caso de experimentación del sistema desarrollado: se formula el análisis de los archivos de sucesos de un sitio de Internet (sección 5.1), se analiza la estructura de los hipervínculos que existen con referencia a un sitio de Internet (sección 5.2) y se categorizan hipervínculos que existen con referencia a un sitio de Internet (sección 5.3).

En el capítulo Conclusiones se plantean los aportes de la tesis y futuras líneas de trabajo.

En el capítulo Referencias se proporciona la descripción de cada cita que aparece en el cuerpo de la tesis.

En el Anexo A se detalla el proceso de educción de requerimientos realizado para el sistema de software en cuestión, se detallan las minutas de las entrevistas abiertas, cerradas, brainstorming y sesiones JAD.

En el Anexo B se propone una métrica para el tamaño de un sitio Web en orden a poder formular comparaciones en el marco de los casos de prueba desarrollados en el capitulo Solución Propuesta.

En el anexo C se presentan los registros de la gestión de configuraciones.

En el anexo D se presenta el Manual de Usuario del sistema desarrollado.

2. ESTADO DE LA CUESTIÓN

En este capítulo se presenta una introducción a la exploración en web (sección 2.1), se describe: exploración de uso de datos en la web, exploración de contenidos de la web y exploración de estructura de la web (sección 2.2), se indican las etapas en la que se divide la exploración de web (sección 2.3), se proponen herramientas de libre uso, susceptibles de ser integradas, para realizar exploraciones de uso, contenido y estructura (sección 2.4) y se resumen trabajos relacionados con las propuestas de esta tesis.

2.1. INTRODUCCIÓN

El advenimiento del World Wide Web (WWW) ha causado un incremento en el uso de la Internet. El WWW es el medio de difusión de un amplio rango de información que puede ser obtenida a un bajo precio. La información en WWW no solo es importante para los usuarios individuales, sino también para las empresas especialmente cuando se trata de la toma de decisiones.

La mayoría de los usuarios obtienen información de Internet realizando una combinación de motores de búsqueda de Internet (por ejemplo: Google, Yahoo, Vivisimo) y sistemas navegación (por ejemplo: Netscape, Explorer, Firefox), pero este procedimiento no siempre es el adecuado, por no devolver toda la información que el usuario necesita. Esto último es especialmente cierto en lo que concierne a las organizaciones, que en la actualidad cada vez con mas frecuencia utiliza la información de la WWW para utilizarla en sus herramientas de toma a la decisión [Madria *et al*., 1999]. En función de esto ha surgido una nueva rama de la exploración de datos, que originalmente solo versaba en temas de base de datos y luego se extendió a documentos. Esta nueva rama es la denominada exploración de información en sitios WWW, que de aquí en mas se denominará exploración Web, la cual se puede definir como el proceso de descubrir información útil en la WWW.

La primer referencia con la que se cuenta sobre exploración Web es la dada por [Etzioni, 1996], en su trabajo pionero definía la exploración Web Etzioni, 1996] como un conjunto de técnicas de exploración de datos capaces de descubrir automáticamente y extraer información de Internet de documentos y servicios Etzioni planteaba la problemática de saber cuan eficiente y factible podía ser la exploración de datos sobre sitios Web.

Podemos enunciar las principales características de los datos en la Web son [Wang, 2000, Pal *et al*., 2002]: no tienen rótulos, distribuidos, heterogéneos, semi estructurados, variables en el tiempo, poseen varias dimensiones.

La exploración Web tiene que poder manejar información que se encuentra relacionada por hipervínculos, que posee una interfaz para que pueda ser entendida por el ser humano, en consecuencia, la exploración Web de poder: manejar contextos altamente sensibles, manejar consultas difusas, agrupar y educir conocimiento.

Según Furnkranz [2002] la exploración Web intenta encontrar y extraer información relevante que se encuentra oculta en la Web, en especial en documentos de hipertexto publicados en ella. Al igual que la exploración de datos, la exploración Web une múltiples disciplinas e la informática como ser: extracción de información, estadística, aprendizaje automático, procesamiento de lenguaje natural, entre otras.

El nivel mas básico de abstracción de datos en la exploración Web es la página Web que se le presenta al usuario del sitio, una página navegable [Mobaster, 2004], que es como se denominará de aquí en mas a la página que se le presenta al usuario, puede estar constituida por una o mas páginas físicas en el sitio, a su vez estas páginas navegables estas constituidas por imágenes, hipervínculos y otros componentes, lo que llamaremos a su conjunto, es decir, página navegable y demás componentes como objetos Web, estos objetos Web son auditados en los archivos de sucesos de los servidores Web, por otra parte cada uno de estos objetos Web representan la actividad llevada a cabo por el usuario del sitio, en consecuencia un administrador de un sitio Web podría obtener información sobre el comportamiento de sus usuarios analizando estos archivos de sucesos. Dentro de estos archivos de sucesos se puede obtener otro nivel de abstracción es el dado por la sesión del usuario, un usuario cada vez que

ingresa a un sitio Web le es asignado un numero único de sesión que se mantendrá sin cambios hasta que el usuario salga del sito o por tiempo de inactividad, esta sesión deje de existir.

2.2. EXPLORACIÓN DE DATOS WEB: USO, CONTENIDO Y ESTRUCTURA

La exploración de datos Web se subdivide en tres ramas:

Exploración de uso
Exploración de contenidos
Exploración de estructuras

La exploración de uso [Lin *et al*., 1998 ; Cooley *et al*., 1997 ; Backman y Rubbin, 1997] también conocida por sus siglas en el idioma ingles WUM (Web Use Mining) utiliza los archivos de sucesos de los servidores Web, en los cuales queda registrado a modo de auditoria, todos los accesos de los usuarios de un sitio, e intenta descubrir patrones de comportamiento comunes entre los mismos. Las técnicas de inteligencia artificial más usadas para esto son: redes de neuronas artificiales (ANN), algoritmos genéticos y lógica difusa [Pal *et al*., 2002] Una de las principales características de la exploración de uso es la de predecir el comportamiento de los usuarios del sitio [Etzioni, 1996]. [Chau et al., 2003] plantean la creación de un agente autónomo que analizando la actividad de los usuarios pueda luego de buscar esta información y procesarla en una segunda etapa poder reconocer el comportamiento de nuevos usuarios. [Jidonwang *et al*., 2002] proponen un método para el análisis de los archivos de sucesos de los servidores Web para calcular la relevancia de los usuarios en el sitio, donde el uso del sitio por parte de los usuarios y las páginas del sitio sirven para calcular el método. [Abrahan y Ramos, 2004] proponen el estudio de las colonias de hormigas como base para entender la organización y el comportamiento del uso de un sitio Web a partir del archivo de suceso del mismo. El algoritmo propuesto es de categorización evolutiva difusa. [Yang *et al*., 2002] proponen un algoritmo para detectar reglas de asociación en función de temporalidad del uso realizado por los usuarios. Esto difiere con otras predicciones las cuales no son tan eficientes como el método propuesto. [Krishnapuran y Joshi, 2000] propone la utilización de un

algoritmo de categorización competitiva difusa, que es una extensión del algoritmo de aglomeración competitiva de categorización, para el análisis de los archivos de suceso de los servidores Web. [Jung y Jo, 2003] proponen la utilización de reglas semánticas para mejorar la información obtenida del análisis de los archivos de sucesos, y ya no solo la detección de sesiones de usuarios y temporalidad de uso. [Jin *et al*., 2004.] proponen un análisis probabilística basado en la semántica para el análisis de los archivos de suceso. [Huang *et al*., 2001] presentan un método para representar cubos en función de la información obtenida de los archivos de suceso. [Borges y Levene, 2000] proponen un método basado en heurísticas para capturar patrones de navegación de los usuarios de un sitio.

La exploración de contenidos [Pitkow, 1997] también conocida por sus siglas en el idioma ingles WCM (Web Content Mining) trata de descubrir información importante en los contenidos de las páginas Web, sea cual fuera su formato. La principal técnica de inteligencia artificial que se utiliza para realizar esta tarea es la utilización de técnicas de recupero de información (information retrieval - IR) [Pal et al., 2002] Exploración de Multimedia forma parte de la exploración de contenido, esta trata de obtener información de los distintos recursos multimediales disponibles en Internet [Kosala, y Blockeel, 2000; Zaiane *et al*., 1998]. Uno de los trabajos propuestos para extraer información de contenido de los sitios Web es a través del uso del algoritmo de inducción de envoltorio y la utilización de la definición de entidades y un post procesamiento para solucionar ambigüedades [Siglitos *et al*., 2003]. La utilización de algoritmo de inducción de envoltorio intenta generar reglas, de dominios específicos [Kushmerick, 1997]. Otros trabajos versan sobre la generación de motores de recomendaciones para la construcción dinámica de las páginas, estos trabajos tratan de reconocer patrones de comportamiento y características comunes en los archivos de sucesos de los servidores Web y generar una recomendación de construcción, en [Zhu *et al*., 2004] se define el concepto de Información de Contenido (IC) que hace referencia a las páginas que son presentadas al usuario y todos los componentes que en ella existen y a partir de estas IC analizando los archivos de sucesos de los servidores Web se presenta las recomendaciones. En [Zhu y Greiner, 2004] se presenta un método para predecir la aparición de una palabra en una página Web en función del análisis realizado en páginas similares. En [Mendez-Torreblanca *et al*., 2002] este trabajo presenta un método para poder predecir el cambio del contenido

dinámico de un sitio en función del análisis de tendencias de lso cambios de sitio. En [Xue *et al*., 2002] se presenta aquí un método para reordenar el índice que generado por el índice de un sito Web, para que en función del uso del sitio las ocurrencias de la frase buscada en el sitio tengan un mayor prioridad, es decir aparezcan antes que otras.

La exploración de estructuras [Spertus, 1997] también conocida por sus siglas en el idioma ingles WSM (Web Structure Mining) trata sobre como están relacionados los hipervínculos entre las distintas páginas de un mismo sitio Web u otros. La principal herramienta para realizar esto es la utilización de grafos [Pal et al., 2002] el objetivo de esta rama de la exploración Web es poder construir un modelo de los sitios Web y sus páginas [Etzioni, 1996], por otra parte esta rama de la exploración Web es muy fácil relacionarla con técnicas de bases de datos para generar índices que apunten a los sitios[Madria *et al*., 1999]. La información almacenada de la estructura de un sitio Web consta: cantidad de links hacia otras páginas y cantidad de links internos, externos y al propio documento. En [Kitsuregawa *et al*., 2002] se propone un método para la exploración de estructura Web para identificar comunidades de usuarios y que las mismas puedan ser representadas de forma similar que un diagrama de entidad relación y almacenadas en una base de datos. En [Furnkranz, 2002] se presenta un método para explorar el grafo generado pro la Web para que su acceso sea rápido y simple. En [Cooley, 2000] plantean la problemática de conocer la estructura de un sitio para poder a partir de ella reconocer el patrón de comportamiento de los usuarios, por otra parte presenta una forma de modelar la estructura de un sitio.

2.3. ETAPAS EN LA QUE SE DIVIDE LA EXPLORACIÓN DE DATOS WEB

La exploración de datos Web esta constituida por cuatro etapas: recolección, procesamiento, generalización y análisis [Pal et al., 2002]

Recolección: En esta etapa se detectan los orígenes de datos, lo que se debe lograr es conseguir de la forma mas automatizada posible todos los orígenes de datos para su posterior procesamiento, entiéndase por conseguir tener acceso a los mismos, ya sean en forma local o remota.

Procesamiento: Esta etapa los datos que se han obtenido en la etapa anterior se ordenan, categorizar, completan y se preparan para la próxima etapa.. En forma mas detallada esta etapa es al que se encarga de tomar los datos de la Web, que como se había mencionado en “Introducción”, son altamente desestructurados, se los debe ordenar y categorizar, de ser necesario completar la información faltante y por ultimo en función del tipo de generalización que se desee realizar, se deberán preparar los datos para que puedan ser procesados.

Generalización: Esta etapa es donde se utilizan varias técnicas extraídas de distintas ramas de la computación para obtener o reconocer un patrón común de comportamiento. Las técnicas mas comúnmente utilizadas, esto no trata de ser una lista exhaustiva de técnica, son:

Series de Tiempo (Modelo Arima): este método intenta encontrar patrones comunes a lo largo de unidades periódicas de tiempo.

Redes de Neuronas Artificiales: esta técnica de inteligencia artificial es utilizada generalmente para detectar categorías comunes en los datos.

Algoritmos Genéticos: esta técnica de inteligencia artificial es utilizada para detectar posibles soluciones a conjuntos de búsqueda.

Lógica Difusa: esta técnica de inteligencia artificial es utilizada generalmente como soporte de otra técnica, en función de lo poco estructurado de la información, la utilización de rangos difusos nos ayuda a encontrar comportamientos comunes en forma más rápida.

Teoría de conjuntos incompletos: esta trata de solucionar el problema planteado con una de las características de los datos obtenidos de la Web, lo heterogéneos de los mismos. A lo que intenta dar soporte esta teoría es a poder trabajar con conjuntos de datos los cuales no siempre se encuentran completos.

Reglas de decisión: es una técnica para generar árboles donde los nodos hojas contienen clase de datos similares, es utilizada para la generación de segmentos.

Aprendizaje Automático: esta técnica de inteligencia artificial es utilizada para inferir conocimiento del resultado de la aplicación de alguna de las otras técnicas antes mencionadas.

Análisis Estadístico: las técnicas estadísticas son las herramientas mas extendidas para extraer información de los visitantes de un sitio Web [Etzioni, 1996]

Análisis: en esta etapa es donde la intervención del humano es fundamental. En todo proceso de adquisición de conocimiento el ser humano es que interactúa para poder dar la ultima validación al conocimiento en cuestión.

2.4. HERRAMIENTAS INTEGRABLES

Exploración de uso:

La herramienta seleccionada para integrar en el marco de trabajo para la exploración de uso es Webalizer [Barret, 1999], la misma genera estadísticas de todos los accesos hechos en el sitio. Otra alternativa de integración es Analog [Turner, 2005], su principal característica es la rapidez para la generación de estadísticas, es escalable y fácilmente se pueden configurar distintos tipos de reportes. Una tercera alternativa es AlterWind Log Analyzer [AlterWind Software, 2005] la misma permite determinar los accesos a un sitio, y generar varios tipos de reportes.

Exploración de contenido:

La herramienta seleccionada para integrar en el marco de trabajo para la exploración de contenido es el SOM_PACK [Kohonen *et al*., 1996], es una implementación del modelo de red neuronal de Kohonnen. Otra alternativa es DIAsDEM [Winkler and Spiliopoulou, 2002] este proyecto esta basado en la integración de documentos semi-estructurados, html es considerado un documento semi-estructurado, para poder aplicar técnicas de exploración de texto. Podemos nombrar al proyecto Weka Machine Learning [Witten y Frank, 2005], este es un conjunto de herramientas de aprendizaje automático que permite resolver problemas de exploración de contenido utilizando aprendizaje automático.

Exploración de estructura:

La herramienta seleccionada para integrar en el marco de trabajo para la exploración de estructura es Information Crawler [Thesoftwareobjects, 2005], el mismo al consultarle un sitio Web responde con la cantidad de links que se encuentran en los principales motores de búsqueda en Internet. Otra alternativa es The Grinder [Aston y Fitzgerald, 2005] esta herramienta fue pensada originalmente para la pruebas de estrés de un sitio, pero se ha revelado como una muy buena herramienta de exploración de estructura. Se puede nombrar a JNBC [2005] es una solución de fuentes libres que implementa redes bayesianas y es utilizada para la categorización de estructuras de sitios.

2.5. TRABAJOS RELACIONADOS

En esta sección se describen distintos proyectos de I&D que tratan el tema de exploración de datos en web.

El proyecto “Warehouse of Web Data” [Etzioni, 1996] busca implementar un werahouse para la Web que administre información desde la Web y sea utilizada como soporte a la toma de decisiones. Implementa el warehouse usando bases de datos que contiene información estratégica desde la Web y la utiliza en conjunto con el warehouse común integrada con herramientas de extracción de información (Information retrieval) y un conjunto extendido de herramientas propias de la exploración de datos para sitios Web para estructurar la información altamente desestructurada de la WWW.

El proyecto “Log Markup Language for Web Usage Mining” **[10]** es una aplicación que se basa en el estándar de XML 1.0 y ha sido diseñado para trabajar con los archivos de sucesos de los servidores Web. Esta aplicación provee una forma sencilla de producir reportes con la información contenida en los archivos de sucesos de los servidores Web. Además permite la manipulación de estructuras complejas de información, provee un poderoso mecanismo de limpieza de la información del archivo de suceso, algunas de sus características es la eliminación de información irrelevante, como ser nombre de gráficos o script files. LOGML utiliza XSLT para la generación de reportes.

Los documentos en la Web se encuentran interconectados mediante los llamados hipervínculos. Una forma habitual de representar esta estructura es a través de grafos, en los cuales los nodos son los documentos y sus conexiones los hipervínculos existentes entre ellos [19]. Un trabajo pionero en este campo ha sido el realizado por (Broder, Kuman, Maghour, Raghavan, Rajagoplan, Stata et al., 2000) Este trabajo utilizo datos obtenidos del motor de búsqueda de Internet Altavista a mayo 1999 donde obtuvo 203 millones de URL y 1466 millones de hipervínculos, estos fueron guardados con el formato de un grafo. El grafo entero ocupaba 9.5 gigas de espacio en disco, con esta técnica se podía ejecutar una busque completa en el grafo en menos de 4 minutos. Pero la conclusión mas importante de este trabajo fue poder representar la

estructura de la Web, el cual era parecido a la figura de un nudo hecho en medio de unas soga, el nudo, o como lo llamaron los autores del trabajo (Connected Core Component CCC) con 56 millones de páginas, a ambos lados de este nudo existen 44 millones de páginas de cada lado aproximadamente, uno de estos conjunto es el que tiene hipervínculos hacia el CCC es denominado conjunto de entrada (IN set), y el restante es conectado a través de hipervínculos desde el CCC, este segundo conjunto es llamado de salida (OUT set). Por otra parte existen varios documentos que no se encuentra conectados con nada.

El termino semántica de la Web ha sido acuñado por Tim Vernier-Lee con el objetivo de que la información desestructurada que se encuentra en la Web pueda ser procesada pro una computadora (Berners-Lee, *et al.*, 2001) La idea basica de esto es enriquecer la información que se encuentra en una página Web con información procesable para una computadora, este conocimiento tendrá la forma de ontologías (Fensel, 2001) Una ontología define ciertos tipos de objetos y relaciones entre ellos, de esta forma cuando una ontología es accesible, un programa para computadoras puede inferir la información que se encuentra en la página. La semántica Web esta siendo utilizada para la explotación Web. El conocimiento representado en las ontologías es utilizado por la exploración Web para extraer conocimientos de las páginas [Maedche *et al*., 2004b; Maedche *et al*., 2003; Doan *et al*.,, 2003].

3. DESCRIPCIÓN DEL PROBLEMA

El problema que se plantea con el uso de las herramientas de exploración Web, y en general para minería de datos, es la falta de un hilo conductor de un marco de trabajo estructurado, el cual pueda evolucionar desde una ideal novel a un complejo y avanzado modelo de análisis, permitiendo:

- La reutilización de procesos ya hechos
- La utilización de diferentes paquetes de software
- Poder combinar paquetes de distintas empresas tanto sean pagos, fuentes libres o de libre utilización sin acceso al código
- Poder generar proceso repetitivos, adaptables y variables en el tiempo con un mínimo de esfuerzo.

En suma, poder plasmar el conocimiento y experiencia de los usuarios en procesos de la empresa soportados en medios de resguardo para su posterior estudio y de ser posible mejora o adaptación.

La variedad de técnicas para resolver distintos problemas en la exploración de datos de la Web y su alto grado de desconexión permiten inducir la necesidad de disponer de un marco de trabajo que constituya una nueva capa de abstracción que articule los distintos procesos de minería de datos para sitios Web asociados a cada técnica.

Este marco debe permitir el registro de los distintos subprocesos que componen un proceso de minería de datos documentando en forma coherente y unificada los componentes del proceso y su interacción.

De lo dicho precedentemente se puede inferir que no se dispone con un método estructurado para el procesamiento de archivos de sucesos y paginas Web

Otras características que este marco de trabajo debe satisfacer son:

Disponer de procesos repetibles, susceptibles del control.

Poder utilizar distintos paquetes de software, tanto sean comerciales o de uso libre, permitiendo su combinación y la extracción del mejor rédito de cada uno de ellos.

Permitir el procesamiento distribuido y paralelo tanto sea en la misma plataformas como en distintas (por ejemplo: Unix o Windows.

Permitir el entallado de los procesos de exploración en web a estaciones de trabajo o servidores de aplicación.

Permitir la actualización necesaria que surja de:

- El cambio de versiones de los sofware que instrumentan las distintas técnicas
- Cambios en la concepción de los procesos que surjan de la evolución de las ideas.
- Cambios en la concepción de los procesos que surjan de la evolución de los procesos..

Permitir la auditoría de los procesos de exploración web de datos.

Permitir la mejora continua de los procesos de exploración web que soporte.

En este contexto surge la necesidad de disponer de un entorno de trabajo altamente adaptable a las necesidades del proceso de explotación de datos. Un entorno de trabajo para exploración Web debería poder exportar el resultado del proceso de exploración Web a diferentes formatos de archivo.

4. SOLUCIÓN PROPUESTA

En este capítulo se presenta la planificación de sistemas de información (sección 4.1) con detalle de: inicio del plan de sistemas de información (sección 4.1.1), definición y organización del PSI (sección 4.1.2), estudio de la información relevante (sección 4.1.3), identificación de requisitos (sección 4.1.4), estudio de los sistemas de información actuales (sección 4.1.5), diseño del modelo de sistemas de información (sección 4.1.6), definición de la arquitectura tecnológica (sección 4.1.7), estimación del proyecto (sección 4.1.8) y definición del plan de acción (sección 4.1.9); y se presentan aspectos del desarrollo de sistemas de información (sección 4.2) lo que incluye estudio de viabilidad del sistema (sección 4.2.1), gestión de configuración (sección 4.2.2), análisis del sistema de información (sección 4.2.3), diseño del sistema de información (sección 4.2.4), construcción del sistema de información (sección 4.2.5) y pruebas de software (sección 4.6).

4.1. PLAN DE SISTEMAS DE INFORMACIÓN

4.1.1. INICIO DEL PLAN DE SISTEMAS DE INFORMACIÓN

ANÁLISIS DE LA NECESIDAD DE PSI

El objetivo del presente PSI es la creación de un sistema de software que permita la obtención de patrones de comportamiento de usuarios en los Servidores Web. El sistema propuesto ha de cumplir con las siguientes funcionalidades:

Posibilidad de analizar archivos de sucesos de un Servidor Web y/o paginas Web.

Fácil adaptación a nuevos formatos de archivos de sucesos, paginas Web y técnicas utilizadas.

La salida del proceso debe ser exportable a una planilla de cálculo.

IDENTIFICACIÓN DEL ALCANCE DE PSI

El presente proyecto abarca a las siguientes áreas de la empresa:

Soporte a la Producción: Es el área que mas se verá afectada por la implantación del sistema de información, pues a esta, es donde llegan las solicitudes hechas por los usuarios externos o internos de la empresa. Desde el punto de vista del PSI, esta área debe ser considerada como un usuario del sistema con el cual se interactúa para obtener las especificaciones funcionales del mismo.

Desarrollo de Sistemas: Esta área utilizará el sistema, dándole el mismo uso que Soporte a la Producción.

Mercadotecnia: Esta área utilizará la información que se extraiga de los Servidores Web como soporte a la toma de decisiones.

OBJETIVOS ESTRATÉGICOS DEL PRESENTE SISTEMA DE SOFTWARE

Los objetivos del presente sistema son:

- Menor tiempo de respuesta de Soporte a la Producción y Desarrollo.
- Asegurar la continuidad del negocio.
- Mejorar las Ventas.

La correlación entre objetivos, factores de éxito y componentes del factor de éxito se presenta en la tabla 4.1.

Objetivo	**Factores de Éxito**	**Componentes del Factor de Éxito**
Menor Tiempo de Respuesta	Reducir la Carga de Trabajo	Reasignación de Recursos. Distribución de Procesos. Automatizar los procesos.
Mayores Ventas	Comercialización personalizada a través de Internet	Generación de información detallada de las características de comportamiento de los clientes.
Continuidad del negocio	Procesos resguardados y repetibles	Auditar procesos.

Tabla 4.1. Tabla de Correlación

FACTORES CRÍTICOS DE ÉXITO

- Auditar procesos (procesos repetibles)
- Información detallada de patrones de comportamiento (mayores ventas)
- Automatizar procesos (reducción de tiempos)

DETERMINACIÓN DE RESPONSABLES

El encargado del proyecto de sistemas (tesis) es el Lic. Hernán D. Merlino, el Dr. Ramón García Martínez será el encargado de dar la aceptación del mismo.

4.1.2 DEFINICIÓN Y ORGANIZACIÓN DEL PSI

ESPECIFICACIÓN DEL ÁMBITO Y ALCANCE

Los procesos que se verán afectados por la implantación de este sistema de software son:

- Análisis de incidentes: Este flujo de trabajo define los pasos a seguir ante un reclamo ó incidente por parte de un usuario externo

o interno en relación a los sistemas que se implementan sobre tecnología Web.

- Toma de decisiones: En el área de Mercadotecnia podrá ver acrecentada la información para la toma de decisiones en todo lo que responda a políticas con respecto a la Web.

- Desarrollo de Sistemas: Dispondrá de una herramienta mas para la solución de incidentes provenientes de Soporte a la Producción.

OBJETIVOS GENERALES

- Automatizar el proceso de análisis de archivos de sucesos.
- Generar información para la toma de decisiones.
- Mejorar el proceso de resolución de incidentes.

ORGANIZACIÓN DEL PSI

Catálogo de usuarios:

A continuación se representara en un la tabla 4.2. a los distintos usuarios del sistema.

Responsabilidad	Usuario
Responsable del proyecto	Gerente de Soporte a la Producción
Encargado del desarrollo	Jefe de Desarrollo
Usuarios finales	Área de Soporte a la Producción
Equipo de desarrollo	Un programador señor

Tabla 4.2. Distintos usuarios del sistema

Comité de Sistemas

El comité de sistemas estará formado por el gerente de Soporte a la Producción y el Jefe de Desarrollo.

Definición del Plan de Trabajo

A continuación se representara el plan de trabajo (tabla 4.3) sugerido para el presente creación de sistema de software

Etapa	Producto	Duración
Planificación de Sistemas de Información - PSI	Documento de aceptación de sistemas	30 días
Análisis de Requerimientos de Sistemas - ARS	Documento de aceptación de requerimientos	70 días
Especificación Funcional del Sistema - EFS	Documento de aceptación de especificación funcional	30 días
Diseño Técnico del Sistema - DTS	Documento de aceptación de diseño técnico	30 días
Desarrollo de Componentes del Sistema - DCS	Documento de aceptación de pruebas de unidad del sistema.	80 días
Desarrollo de Procedimientos de Usuario – DPU	Documento de aceptación de procedimientos del usuario	35 días
Pruebas, Implantación y Aceptación - PIA	Documento de aceptación del sistema	25 días
Total		280 días

Tabla 4.3. Plan de trabajo

Definición del Plan de Trabajo

A contignación se especifica el Gantt del sistema (Figura 4.1)

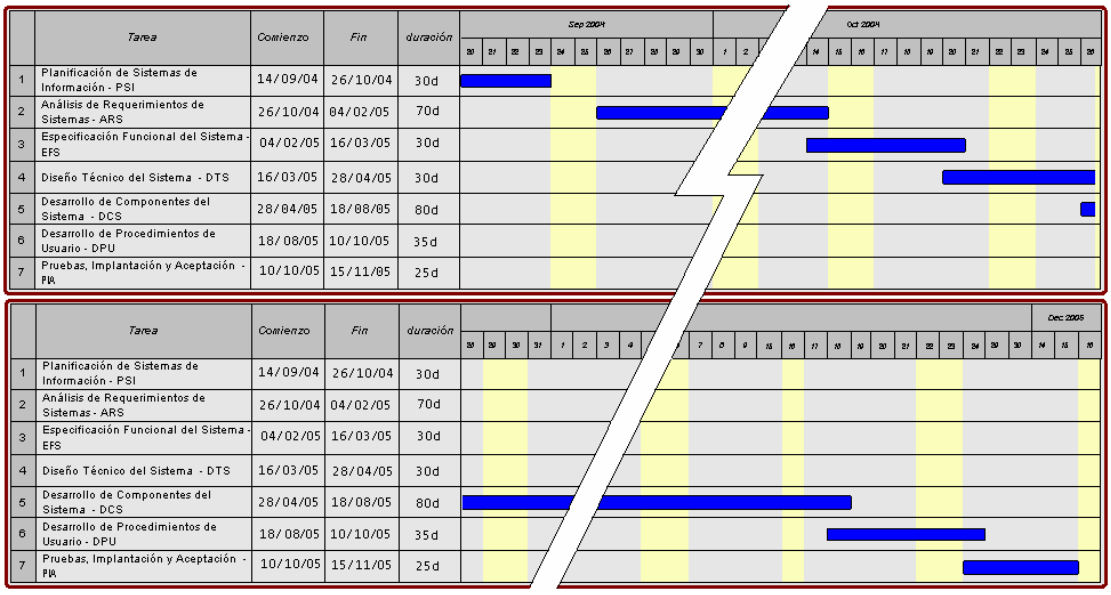

Figura 4.1. Gantt del plan de sistemas

4.1.3 ESTUDIO DE LA INFORMACIÓN RELEVANTE

SELECCIÓN Y ANÁLISIS DE ANTECEDENTES

Como base para su creación se tomara el proceso manual que se realiza para investigar los reclamo ó incidente. Por otra parte, se relevarán herramientas similares que existen en el mercado.

VALORACIÓN DE ANTECEDENTES

El conocimiento del equipo de Soporte a la Producción, sumado al relevamiento de herramientas similares, va a permitir la construcción de un producto estable.

4.1.4. IDENTIFICACIÓN DE REQUISITOS

Los requisitos del sistema se presetan en la tabla 4.4.

Catálogo de requisitos			
Tipo de Requisito	Descripción	Prioridad	Estado
Funcional	El sistema debe poder ser ejecutado sobre plataformas Unix y Windows	Alta	Aprobado
	El proceso debe ser ejecutado en forma batch	Alta	Aprobado
	La salida debe ser en texto tabulado	Media	Aprobado
	Registrar la operaciones	Media	Aprobado

Tabla 4.4. Identificación de Requisitos

El detalle de entrevistas y definición de requerimientos se encuentra en el Anexo A

ESTUDIO DE LOS PROCESOS DE PSI

Modelo de procesos incluidos en el plan de sistemas:

Proceso: Atención y solución de incidentes por parte de un usuario externo o interno.

Área: Soporte a la producción.

Detalle de proceso: Cuando el área de Soporte a la Producción recibe un mail del área de atención al cliente, denominada Help Desk, cualquiera de los miembros del área toma el requerimiento y comienza a trabajar, luego de cargar el requerimiento en una hoja de cálculo y hacerse responsable del mismo. Los datos que se registran son:

Fecha de inicio del requerimiento.

Estado del requerimiento:

- Abierto: Algún miembro del equipo se ha hecho responsable del mismo.
- En espera: Haciendo el análisis del incidente, se detecta que hace falta más información, por ejemplo, se sabe la fecha en la que sucedió pero no la hora, se pide que se la especifique para acotar el rango de búsqueda. En este caso el equipo de soporte a la producción cambia el estado y comunica a la persona que cargo el incidente que se le informe de la hora de lo sucedido.
- Cerrado: Se ha solucionado el incidente y se notifica al usuario que reporto el incidente, quedando a la espera de su aprobación.
- Finalizado: El usuario que cargo el incidente da por aprobado la modificación.

El miembro del equipo de Soporte a la Producción que se ha hecho cargo del incidente comienza a trabajar analizando en forma manual los archivos de sucesos de los Servidores Web. Este repite la acción hasta encontrar la información que busca. Con esta información puede tomar los fuentes del sistema para verificar la hipótesis con la que trabaja, o simplemente realiza un reporte del incidente. Con todo esto el Gerente del área de Soporte a la Producción, en el caso de haber detectado un error en la codificación, y en función de la complejidad y luego de hacer un breve análisis de impacto del mismo puede decidir en asignarlo a uno de los miembros de su área, o pasarlo al área de Desarrollo.

Es un conjunto variable de etapas sucesivas en las que se divide el proceso de resolución de incidentes.

ANÁLISIS DE LAS NECESIDADES DE INFORMACIÓN

El proceso de gestión de un incidente (figura 4.2), comienza en el momento que un usuario externo ó interno abre un ticket, formalmente llamado Apertura del Incidente. El ticket puede ser generado por un error en la aplicación ó un pedido de análisis de información.

El equipo de Soporte a la Producción comienza a trabajar con el incidente (Análisis del Incidente), con la información recolectada genera un reporte (Reporte del Incidente), si el ticket era un pedido de información, se da por cerrado (Cierre del Incidente), de haber detectado un error, se asigna el ticket a Desarrollo de Sistemas, el cual lo soluciona y genera un reporte de la solución implementada (Reporte del Incidente).

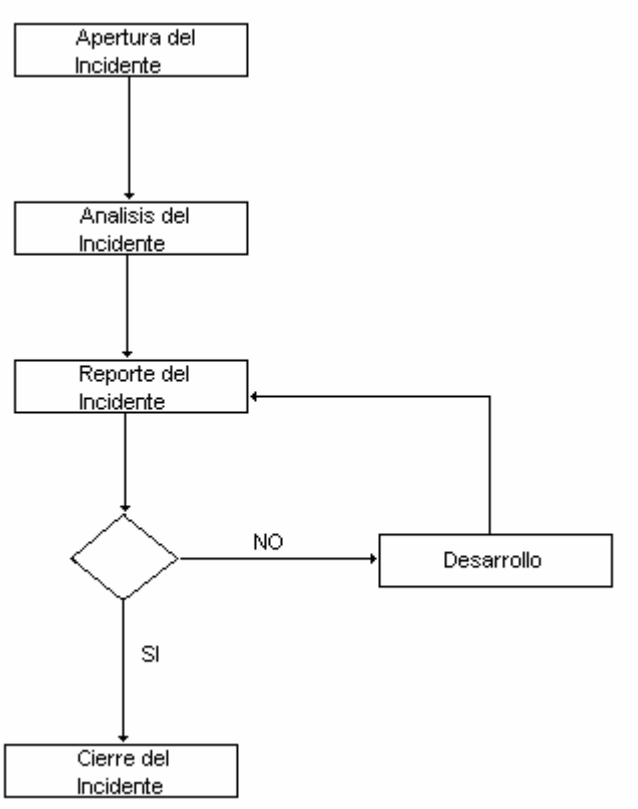

Figura 4.2. Flujo de información

El presente sistema de software se concentra en la actividad de Análisis del Incidente.

CATALOGACIÓN DE REQUISITOS

- Debe poder soportar distintos formatos de archivos de sucesos
- El sistema debe poder se ejecutado sobre plataformas Unix y Windows.
- La salida debe ser en texto tabulado.

4.1.5 ESTUDIO DE LOS SISTEMAS DE INFORMACIÓN ACTUALES

ALCANCE Y OBJETIVOS DEL ESTUDIO DE LOS SISTEMAS DE INFORMACIÓN ACTUALES

Se definen a continuación los principales objetivos del sistema de software

Objetivo de estudio de los Sistemas de Información actuales:

- Proceso de análisis de incidentes.

Identificación de Sistemas de Información actuales afectados:

- Proceso de análisis de incidentes.

ANÁLISIS DE LOS SISTEMAS DE INFORMACIÓN ACTUALES

El sistema de información actual es informal y se basa en el metaconocimiento de los miembros del equipo de Soporte a la Producción, cada uno de los miembros toma las acciones que él cree que son mas acertadas para finalizar con el incidente.

VALORACIÓN DE LOS SISTEMAS DE INFORMACIÓN ACTUALES

El análisis de un incidente comienza cuando llega un mail asignado al equipo de Soporte a la Producción, una vez que es recibido, uno de sus miembros lo toma y lo carga en una planilla de cálculo, haciéndose responsable del mismo.

En dicha planilla se le asigna el estado al incidente, detallado en Estudio de los Procesos del PSI (4.1.4). El miembro de Soporte a la Producción basándose en el metaconocimiento que posee de las aplicaciones Web de la empresa, analiza los archivos de suceso de los Servidores Web, los archivos de sucesos de las aplicaciones y con la información recabada en los mismos se genera un informe del mismo; el cual es reportado al usuario que generó el incidente; si se ha detectado un defecto en la aplicación también es comunicado el responsable de dicha aplicación, para que la misma sea modificada.

El principal problema que se detecta en el sistema de información actual es la informalidad del mismo, el proceso de detección de incidentes esta sujeto al metaconocimiento que posee cada uno de los integrantes del equipo de Soporte a la Producción, la consecuencia de esto es que no se pueda reutilizar la experiencia ya adquirida de los miembros con mas tiempo de trabajo en el equipo, para poder ser transferida a los miembros mas nuevos.

4.1.6. DISEÑO DEL MODELO DE SISTEMAS DE INFORMACIÓN

DIAGNÓSTICO DE LA SITUACIÓN ACTUAL

El sistema manual actual será remplazado por un nuevo sistema informatizado, el cual, su principal objetivo es implementar el metaconocimeinto que poseen los miembros del equipo de Soporte a la Producción.

DEFINICIÓN DEL MODELO DE SISTEMAS DE INFORMACIÓN

El flujograma, especificado en el Análisis de las Necesidades de Información (4.1.4), es el modelo de sistema de información que se tomará como base para el desarrollo del mismo.

4.1.7. DEFINICIÓN DE LA ARQUITECTURA TECNOLÓGICA

IDENTIFICACIÓN DE LAS NECESIDADES DE INFRAESTRUCTURA TECNOLÓGICA

Parte de las alternativas tecnológicas quedan limitadas por los requerimientos del sistema. El mismo debe poder ser ejecutado en ambientes Unix y Windows, y ser configurable por los distintos incidentes reportados.

SELECCIÓN DE LA ARQUITECTURA TECNOLÓGICA

Lenguajes de programación:

- Java: La selección del lenguaje se ha definido por su facilidad para ser ejecutado en cualquier plataforma.
- Jython: En función de haber seleccionado Java como lenguaje principal de programación, la elección natural es Jython por ser un dialecto de Java.

ANÁLISIS DE IMPACTO

Aspectos a considerar:

- Complejidad de la nueva tarea: se deberá contemplar detalladamente el plan de capacitación en el nuevo sistema.
- Tiempo de sustitución de la antigua forma de trabajo a la nueva.
- Controles estricto para la gestión del cambio, por del rechazo cultural a las nueva tecnologías.
- Capacitación al área de comercialización para el pedido de información.

4.1.8. ESTIMACIÓN DEL PROYECTO

ESTIMACIÓN INICIAL SOBRE LOS CASOS DE USE IDENTIFICADOS

Cálculo de Puntos de Función sin ajustar (Tabla 4.5)

	Complejidad			
	Baja	Media	Alta	Aporte
Entradas Externas		1		4
Salidas Externas			1	7
Consultas Externas				0
Archivos Lógicos Internos			10	150
Total				161

Tabla 4.5. Cálculo de Puntos de Función sin ajustar

JUSTIFICACIÓN

Entrada: la entrada existente se refiere a el archivo de suceso del servidor Web, la definición de una complejidad media surge de: se conoce el formato de archivo y los posibles valores pero el volumen de datos puede ser extremadamente grande, se realiza una extrapolación de ambas variables y se obtiene una complejidad media.

Salida: la salida existente representa el archivo con el resultado obtenido del proceso de exploración de datos. La complejidad es alta pues el formato de archivo puede variar y el volumen de datos almacenados es muy variable.

Archivos Lógicos Internos: El valor asignado de 10 surge de una estimación del promedio de los procesos necesarios para completar una exploración completa, ademas de lo variable que puede ser su formato y el volumen de información, es por esto que se le ha asignado una complejidad alta.

Cálculo del Factor de Ajuste (Tabla 4.6)

Características	Descripción	Peso
Comunicación de datos	Aplicación Web	3
Procesamiento distribuido de datos	No hay procesamiento distribuido, pero hay datos distribuidos	3
Rendimiento	No hay requerimientos especiales de rendimiento	0
Configuraciones fuertemente utilizadas	No hay restricciones con respecto al hardware	0
Entrada de datos on-line	No hay pico diario de transacciones	0
Eficiencia del usuario final	No hay	0
Actualizaciones on-line	No hay	0
Procesamiento complejo	No hay procesamientos lógicos ni matemáticos complejos	0
Reusabilidad	No hay restricciones	0
Facilidad de instalación	No hay restricciones	0
Facilidad de operación	No hay restricciones	0
Instalación en distintos lugares	No se requiere mas de una instalación	0
Facilidad de cambio	Si	5

Tabla 4.6. Cálculo del Factor de Ajuste

Calculo del grado total de influencia:

$$TDI = 3 + 3 + 5 = 11$$

Calculo del Factor de Ajuste

$$AF = 11 * 0.01 + 0.65 = 0.76$$

Calculo de los Puntos de Función Ajustados

$$FP = UFP * AF = 161 * 0.76 = 122.36$$

Calculo de líneas de código

La cantidad de líneas de códigos por puntos de función (SLOC) para el lenguaje java es 46, en función de este valor y la cantidad de puntos de función sin ajustar [Peralta, 2004] se obtiene el siguiente resultado.

$$Size = 46 * 161 = 7406$$

COCOMO II

A continuación se adjuntan las pantallas con el calculo de estimación del proyecto a través del modelo de diseño temprano de COCOMO II (figuras 4.3 a 4.5)

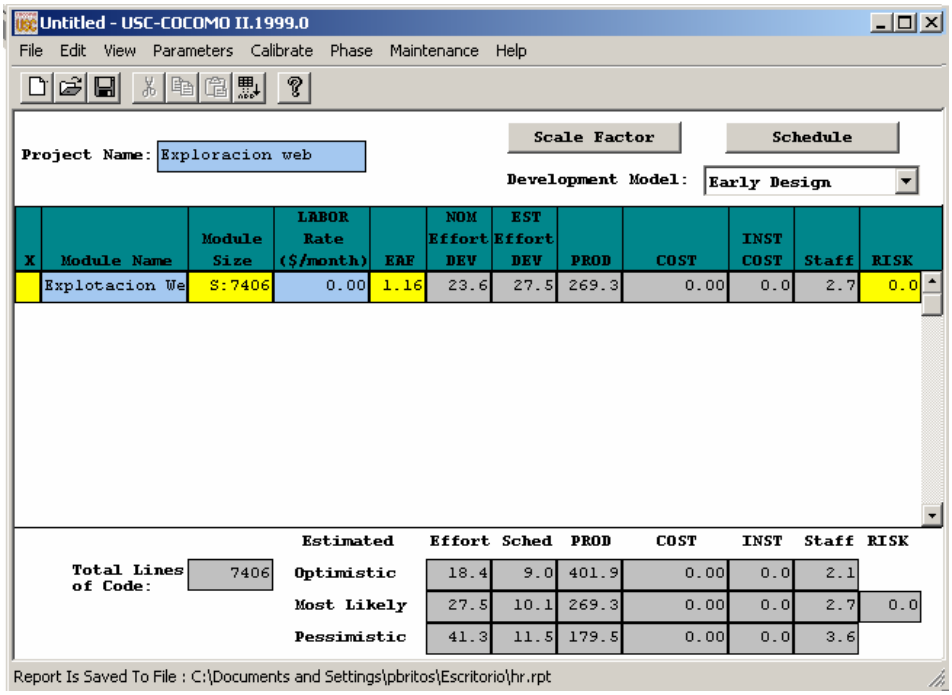

Figura 4.3. Pantalla con los cálculos de COCOMO II

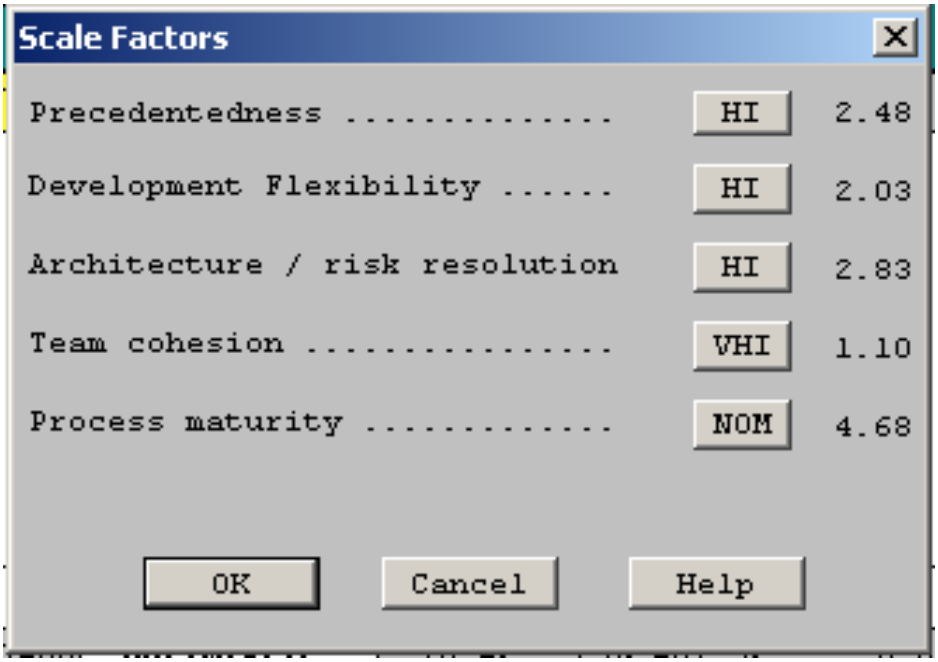

Figura 4.4. Factores de Ajuste

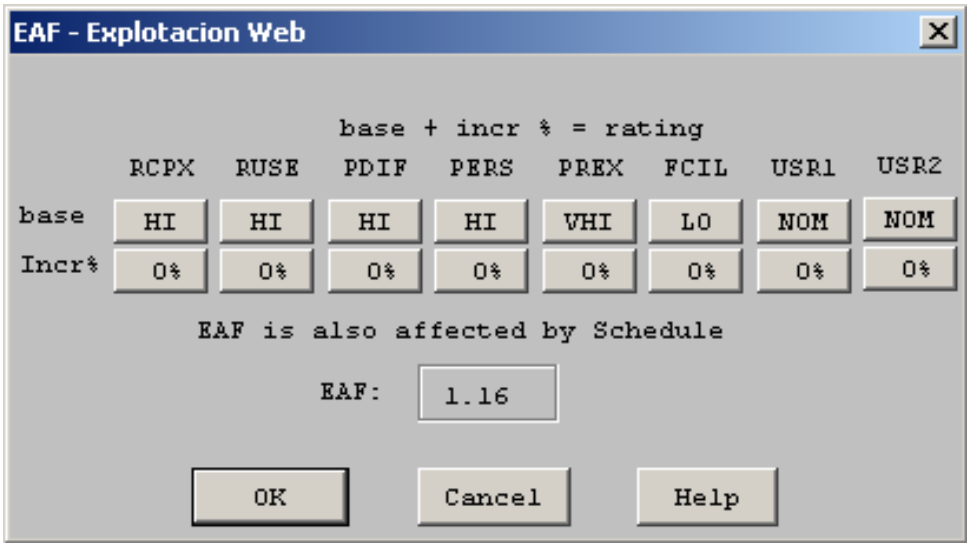

Figura 4.5. Valores Ajustados

Los Resultados Obtenidos para COCOMO II se muestran en la Tabla 4.7.

	Optimista	Conservador	Pesimista
Meses Hombres	9	10.1	11.5

Tabla 4.7. Resultados Obtenidos para COCOMO II

Resultados Finales

Cantidad de líneas de código 7406

Tiempo estimado 10 meses

4.1.9. DEFINICIÓN DEL PLAN DE ACCIÓN

DEFINICIÓN DE PROYECTOS A REALIZAR

El proyecto se dividirá en dos sub-proyectos:

- Por un lado se definen los programas necesarios para reconocer los distintos formatos de los archivos de sucesos de los Servidores Web.

- Por otro lado se "confecciona" el sistema software encargado de encadenar las distintas tareas o pasos para automatizar el proceso.

4.2. DESARROLLO DE SISTEMAS DE INFORMACIÓN

4.2.1. ESTUDIO DE VIABILIDAD DEL SISTEMA

ESTABLECIMIENTO DEL ALCANCE DEL SISTEMA

Descripción General del Sistema

El objetivo del presente plan de sistemas de información es la realización de un proceso para la obtención de patrones de información extraídos del análisis de los archivos de sucesos de un servidor Web. Con la construcción del mismo se pretende obtener una mejora en los tiempos de respuesta del equipo de Soporte a la Producción en primer lugar y como objetivo secundario mejorar el desempeño de las áreas de Desarrollo de Sistemas y de Mercadotecnia.

Catálogo de Objetivos del Estudio de Viabilidad del sistema:

- Determinar la posibilidad de concreción del sistema.
- Estimación de tiempo de realización del sistema.
- Análisis de la posibilidad de separar el proyecto en dos o mas partes.

En la siguiente tabla se detalla el catálogo de requisitos (Tabla 4.8)

Catálogo de requisitos			
Tipo de Requisito	Descripción	Prioridad	
Funcional	El sistema debe poder ser ejecutado sobre plataformas Unix y Windows	Alta	Aprobado
	El proceso debe ser del tipo batch	Alta	Aprobado
	La salida debe ser en texto tabulado	Media	Aprobado
	Registrar la operaciones	Media	Aprobado

Tabla 4.8. Catálogo de requisitos

IDENTIFICACIÓN DEL ALCANCE DEL SISTEMA

Descripción General del Sistema:

Contexto del Sistema:

El proyecto puede ser dividido en tres subsistemas:

- Subsistema Encadenamiento de Tareas (SET): Este subsistema es el encargado de ejecutar los pasos necesarios para la extracción, transformación y análisis de los datos del archivo de sucesos. Físicamente este subsistema se puede encontrar en un Servidor o en una terminal de cliente. Las especificaciones generales del subsistema son:

  El subsistema debe poder ser ejecutado en múltiples plataformas.
  Los parámetros para la ejecución de la tarea deben ser por un archivo de configuración.

- Subsistema Extracción de Archivo de incidentes: Este subsistema es el encargado de extraer información de los archivos de sucesos de los Servidores Web y transformarla para su posterior análisis. Físicamente este subsistema se puede encontrar en un Servidor o en una terminal de cliente. Las especificaciones generales del subsistema son:

  El subsistema debe poder ser ejecutado en múltiples plataformas.
  No debe necesitar una instalación especial, es decir todo lo que necesite el sistema debe estar dentro del directorio en el cual se instala.

- Subsistema de Análisis de Datos: Este subsistema es el encargado de realizar el análisis de la información extraída del archivo de sucesos del Servidor Web y que ya ha sido transformada para poder ser analizada. Las especificaciones generales del subsistema son:

El subsistema debe poder ser ejecutado en múltiples plataformas.

Utilizar técnicas de inteligencia artificial.

Estructura Organizativa

Se define a continuación la estructura de la organización (figura 4.6)

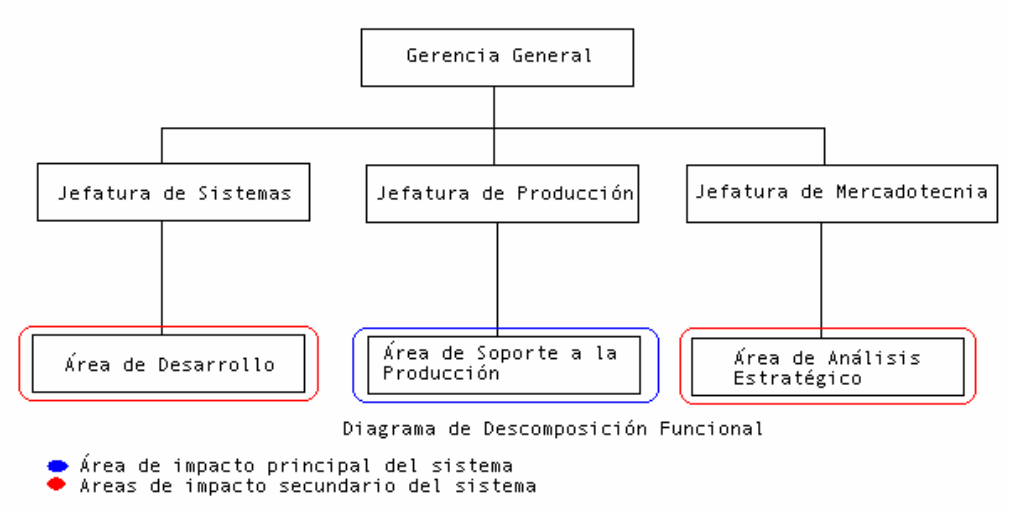

Figura 4.6. Estructura Organizativa

Catálogo de Requisitos:

Requisitos Relativos a Restricciones:

- El desarrollo debe concretarse con la utilización de fuentes libres (Open Source) y desarrollos hechos por el equipo de Sistemas, este requisito se plantea para no tener que pagar licencias de software.
- El producto final debe poder ser instalado con la sola copia del sistema, es decir, no debe contener ninguna configuración especial en la maquina en la que será ejecutado, esto se debe a que deberá ser ejecutado en distintas configuraciones de maquina y de sistemas operativos no se puede depender de alguna configuración especifica.

Dependencias con Otros Proyectos:

No se detecta relación con ningún otro sistema de la empresa.

Catálogo de Usuarios

Equipo de Soporte a la Producción.
Equipo de Desarrollo de Sistemas.
Equipo de Mercadotecnia.

ESPECIFICACIÓN DEL ALCANCE DEL ESTUDIO DE VIABILIDAD

Catálogo de Objetivos del Estudio de Viabilidad de Sistemas:

Objetivos del Estudio de la Situación Actual:

No existen antecedentes del presente sistema, como base para su realización se tomara el proceso manual que se realiza para investigar los reclamos ó incidentes por parte de un usuario externo o interno.

Catálogo de Usuarios

Equipo de Soporte a la producción:
Principal usuario de sistema, los mismos tendrán un uso diario del sistema y es de esperar que se convierta en la principal herramienta para su trabajo.

Equipo de Desarrollo de Sistemas:

Otro grupo que debe ser considerado como usuarios principales del sistema, su uso estará supeditado a los requerimientos que reciban del equipo de soporte a la producción.

Equipo de Mercadotecnia:

Este equipo puede ser considerado un usuario secundario del sistema, el mismo utilizará a este cuando considere que pueda extraer información valiosa para algún proyecto de mercadotecnia en el que se encuentre trabajando.

Plan de Trabajo

Se define a continuación el plan de trabajo propuesto (tabla 4.9):

Etapa	Producto	Duración
Planificación de sistemas de información PSI	Documento de aceptación de sistemas	5 días
Análisis de requerimientos de sistemas ARS	Documento de aceptación de requerimientos	15 días
Especificación funcional del sistema EFS	Documento de aceptación de especificación funcional	5 días
Diseño Técnico del sistema DTS	Documento de aceptación de diseño técnico	5 días
Desarrollo de componentes del sistema DCS	Documento de aceptación de pruebas de unidad del sistema.	20 días
Desarrollo de procedimientos de Usuario DPU	Documento de aceptación de procedimientos del usuario	5 días
Pruebas, Implantación y Aceptación del Sistema – PIA	Documento de aceptación del sistema	5 días
Total		60 días

Tabla 4.9. Plan de trabajo

ESTUDIO DE LA SITUACIÓN ACTUAL

Valoración del Estudio de la Situación Actual

Descripción de la Situación Actual:

El proceso actual es completamente manual y esta supeditado al metaconocimiento que cada miembro del equipo de soporte a la producción tiene.

Identificación de los Usuarios Participantes en el Estudio de la Situación Actual

Jefe de Soporte a la Producción

Un miembro del equipo de Soporte a la Producción (seleccionado al azar)

Un miembro del centro de atención al cliente

Descripción de los Sistemas de Información Existentes

El sistema de información actual es altamente informal, cuando un miembro del equipo de soporte a la producción recibe un requerimiento, le da curso al mismo basado en su propia experiencia y según sus propios criterios, de ser necesario codifica algún programa para extraer la información o reutiliza en el mejor de los casos reutiliza parte del código que el ya hizo. Como se puede observar el proceso no comparte información entre los distintos miembros del equipo de Soporte a la Producción como tampoco se vale de la experiencia.

REALIZACIÓN DEL DIAGNÓSTICO DE LA SITUACIÓN ACTUAL

Como se hizo referencia en 4.2.1, la situación actual carece de todo proceso metodológico y se basa completamente en el metaconocimiento, o dicho en otras palabras en la experiencia que tiene el miembro del equipo de soporte a la producción. Las consecuencias de esta situación es la inexistencia del traspaso de conocimientos de un miembro a otro del equipo, como así también un método para poder evaluar el desempeño de los mismos. Se desprende es esto que tampoco se cuenta con un modelo valedero para mejorar los procesos de Soporte a la producción.

IDENTIFICACIÓN DE REQUISITOS

Luego de las sesiones de trabajo se han detectado los siguientes requisitos con los que debe cumplir el sistema.

El sistema debe poder tomar distintos tipos de formatos de archivos de sucesos, esto se debe a que en la compañía existen distintos tipos de servidores Web.

Poder configurar la cantidad de procesos que se ejecutan en cada sección de detección de incidentes, es decir, poder configurar un workflow para cada incidente.

Poder tener un catalogo de los workflow existentes, para ser utilizarlos en futuros incidentes.

Registro de toda la operación realizada para el análisis del incidente.

Poder ser ejecutado en plataformas Unix y Windows.

CATALOGACIÓN DE REQUISITOS

Catalogación de todos los requisitos, junto a su prioridad y estado (tabla 4.10):

Catálogo de requisitos			
Tipo de Requisito	Descripción	Prioridad	Estado
Funcional	El sistema debe poder tomar distintos tipos de formatos de archivos de sucesos, esto se debe a que en la compañía existen distintos tipos de servidores Web.	Alta	Aprobado
	Poder configurar la cantidad de procesos que se ejecutan en cada sección de detección de incidentes, es decir, poder configurar un workflow para cada incidente.	Alta	Aprobado
	Poder tener un catalogo de los workflow existentes, para ser utilizarlos en futuros incidentes.	Alta	Aprobado
	Registro de toda la operación realizada para el análisis del incidente	Alta	Aprobado
	La salida debe ser en texto tabulado	Media	Aprobado
No funcionales	Registro de toda la operación realizada para el análisis del incidente	Media	Aprobado
	Resguardo de datos	Media	Aprobado
	Posibilidad de poder ser agendado para ser ejecutado a una hora determinada	Media	Aprobado

Tabla 4.10. Catalogación de requisitos

ESTUDIO DE ALTERNATIVAS DE SOLUCIÓN

Preselección de Alternativas de Solución

Descomposición Inicial del Sistema en Subsistemas

El presente sistema se puede descomponer en dos sub-sistemas:

- Workflow de tares e interfases de programas.
- Sub-sistema para el reconocimiento de formatos de archivos de sucesos.

Alternativas de Solución a Estudiar

Para el sub-sistema de workflow, las alternativas existentes son 3:

- Comprar alguno de los productos que se encuentran en el mercado.
- Utilización de algun Fuentes Libres.
- Generar un desarrollo propio.

De lo antes mencionado, se debería descartar la primera opción pues como se ha definido en 4.2.1, no se cuenta con un presupuesto para la compra de software.

Con respecto al segundo punto en función de los requerimientos se ha realizado una búsqueda de workflow en los principales sitios de Fuentes Libres, de los encontrados se ha seleccionado el denominado Ant, que cumple con todas las especificaciones solicitadas en los requisitos.

La tercera opción queda descartada por haberse encontrado un producto de Fuentes Libres.

Descripción de las Alternativas de Solución

Catálogo de Requisitos

Detalle de los requisitos de sistemas

Distintos formatos de archivos de sucesos:

El sistema debe poder tomar distintos tipos de formatos de archivos de sucesos, esto se debe a que en la compañía existen distintos tipos de servidores Web.

Algunos de los estándares que debe soportar son Common Log Format (CLF)

Secuencia de Tareas: Poder configurar la cantidad de procesos que se ejecutan en cada sección de detección de incidentes, es decir, poder configurar un workflow para cada incidente. El workflow no es reentrante, con lo cual el fuente libre Ant, satisface todas las necesidades.

Repositorio de Procesos: Poder tener un catalogo de los workflow existentes, para ser utilizarlos en futuros incidentes.

Resguardo de Datos: Registro de toda la operación realizada para el análisis del incidente.

Alternativas de Solución a Estudiar:

Definición de las distintas alternativas de solución:

Catálogo de Requisitos: Distintos formatos de archivos de sucesos

Modelo de Descomposición en Subsistemas: Sub-sistema para el reconocimiento de formatos de archivos de sucesos

Alternativa software estándar de mercado:

Descripción de las distintas alternativas presentes en el mercado del software:

Descripción del Producto: Se ha seleccionado el fuente libre ANT, es un producto de reconocida solidez y utilizado en gran cantidad de proyectos. Con el se puede encadenar procesos y ejecutarlos, además de dar la facilidad de poder agendarlo en cualquier sistema operativo. Otra cualidad que posee su amplia documentación y popularidad en el mundo de la programación JAVA, esto hace que no sea necesario dedicar tiempo al aprendizaje de la misma, pues el equipo de Desarrollo y de Soporte a la Producción ya conocen su funcionamiento. Cumple con el requisito de poder ser ejecutado sobre cualquier plataforma. Sobre los estándares que utiliza, el lenguaje para realizar el agendamiento de tareas es el XML

ESTUDIO DE LA INVERSIÓN

El impacto sobre la organización en términos de costos es mínimo en función que no se debe adquirir ninguna licencia de software y el desarrollo será realizado por el equipo de Desarrollo de la misma organización.

En términos de beneficio, la mejora es el de pasar de un método manual a uno sistematizado. Provocando una mejora en los tiempos de respuesta del equipo de Soporte a la Producción.

Estimación de Costo-Beneficio
Costo Desarrollo Anualizado
Detalle de la matriz de costo Beneficio actualizado (tabla 4.11):

Ítem	Valor (en pesos)	Observaciones
Adquisición de hardware y software	0	
Gastos de Mantenimiento de hardware y software	0	
Gastos de Comunicación	0	
Gastos de Desarrollo	20400	480 hs * $42.5
Gatos de Mantenimiento	0	
Gastos de Consultaría	0	
Gastos de Formación	2000	20 hs * $100
Gastos de Material	5000	Varios
Costos Financieros	0	

Tabla 4.11. Costo-Beneficio

Beneficios del desarrollo anualizado

Calculo del beneficio de la realización del sistema (tabla 4.12):

Ítem	Valor (en pesos)	Observaciones
Incremento de la Productividad	64800	Ahorro de 3 horas promedio en el tiempo de respuesta. Cantidad de pedidos recibido por mes 120, valor anualizado.
Ahorro de adquisición y mantenimiento de hardware y software	0	
Ahorro de material de todo tipo	0	
Beneficios Financieros	0	
Otros beneficios tangibles	0	
Beneficios intangibles	0	

Tabla 4.12. Beneficio esperado

En función del análisis del tiempo de retorno de la inversión y el valor actual seria factible la realización del proyecto.

4.2.2. GESTIÓN DE CONFIGURACIÓN

La gestión de configuración de un proyecto de software permite establecer y mantener la integridad de los productos generados durante el proyecto de desarrollo y mantiene su vigencia a lo largo de todo el ciclo de vida adoptado. Este objetivo debe llevarse a cabo desarrollando básicamente tres actividades:

Identificación de los elementos desarrollados.
Control de los cambios a que inexorablemente deben someterse los elementos mencionados.
Mantenimiento de la integridad y seguimiento de la configuración durante
todo el ciclo de vida del producto.

Estos aspectos contribuyen a incrementar el nivel de madurez de un centro de desarrollo de software.

DEFINICIÓN DEL PLAN DE GESTIÓN DE LA CONFIGURACIÓN

Para el Control de Configuración del proyecto se deberá cumplirse con las normas de Gestión de Configuración que se detallan a continuación:

NORMAS PARA LA CODIFICACIÓN DE LOS ELEMENTOS DE CONFIGURACIÓN DE SOFTWARE

A efectos de implementar el plan de Gestión de Configuración de Software que asegure un correcto control de las configuraciones del proyecto, se requiere definir inicialmente las normas de codificación para los Elementos de Configuración de Software (ECS) que se generen:

Los ECS que se considerarán como tales son:

La especificación del Sistema.
Estimación del Proyecto

El plan del tiempo del proyecto software.
La especificación de requisitos de software.
El diseño preliminar y detallado.
Los códigos fuente.
Los programas ejecutables.
Los manuales asociados al proyecto.
Las guías asociadas al proyecto.
El Plan de Pruebas.
Los casos de prueba ejecutados y sus resultados.
Los estándares y procedimientos de IS utilizados.
Los diseños de bases de datos.
Los contenidos de las bases de datos.

La codificación de los ECS será efectuada de la siguiente manera:

A cada ECS que conforme una Línea Base se lo individualizará de forma unívoca con un código que adoptará los siguientes valores:

Nombre del proyecto: WF.

Número de Línea Base del proyecto a la que pertenece el ECS, antecedido por la letra L, considerando el ciclo de vida seleccionado para el proyecto:

Para Línea Base Funcional: L1
Para Línea Base de Diseño: L2
Para Línea Base de Producto: L3
Para Línea Base Operativa: L4

Tipo de ECS, será un trigrama en mayúsculas para identificar:

DOC = Documentación

PRG = Programa en soporte magnético u óptico.

COD = Listado de Código Fuente.

BDD= Diseño de Bases de datos.

DAT = Contenido de bases de datos, archivos binarios y ASCII.

CNF = Información sobre configuraciones.

MAN = Manuales.

PLN = Planificaciones.

Identificación del ECS, que estará conformado por una cadena de caracteres de hasta 30 letras o letras y caracteres que aporte una idea de la naturaleza del ECS, por ejemplo: “PLANDELTIEMPO”.

El número de versión del ECS, comenzando por 1.0, para el caso que se deba implementar sobre un ECS una modificación menor se avanzará en la numeración de la versión a 1.1, 1.2 ... 1.XX, para el caso de la implementación de una modificación mayor, se pasará a modificar el número 1 por el 2 y los decimales volverán a 0, por ejemplo, si hay un ECS cuya versión sea 1.13 y se decide realizar sobre él un cambio mayor, la versión pasará a ser 2.0.

La División Contralor de la Configuración asentará en sus registros la fecha de la última modificación en el formato dd/mm/aa. Por ejemplo 27/11/05.

Por último, se deberá asentar el lugar físico donde se encuentra archivado el ECS. Dado que todos los ECS para el proyecto pueden almacenarse electrónicamente, se archivarán en un CD debidamente identificados y formarán parte de esta Tesis.

DEFINICIÓN DEL ÁMBITO Y ALCANCE DEL CONTROL DE CONFIGURACIÓN

El Control de Configuración para el proyecto será de alto nivel. La razón es que, por la filosofía del Sistema, existirá un alto acoplamiento entre componentes generados en cada fase del proyecto, lo que provocará que, de implementarse una modificación en algún ECS, se requerirá la modificación del documento que lo contiene y su proyección a otros documentos relacionados. Por lo expuesto, se definirá un documento único que abarque cada etapa, con sus correspondientes anexos, (según lo establecido en la metodología Métrica 3)

Dadas las características del proyecto, los ECS podrán encontrarse en uno de estos tres estados:

En edición: El ECS se encuentra en su etapa de desarrollo o fue elevado para su corrección.

Finalizado: El ECS ha sido corregido y se encuentra en etapa de aprobación por el tutor de tesis.

Aprobado: EL ECS ha sido aprobado por el tutor de tesis y está listo a ser encuadernado.

La trazabilidad de los componentes tiene que ver con la identificación de la historia de su evolución, esto significa poder identificar qué documento sirvió de base a otro. Dado que se utiliza Métrica Versión 3 para el presente desarrollo, será necesario referirse a la documentación relacionada para verificar la trazabilidad. Por otro lado, no se guardará historia alguna de trazabilidad de versiones para ningún ECS del sistema.

El detalle de las versiones de la gestión de configuración se encuentran en el anexo B.

4.2.3. ANÁLISIS DEL SISTEMA DE INFORMACIÓN

DEFINICIÓN DEL SISTEMA

A continuación de detalla la definición del sistema de información.

Determinación del Alcance del Sistema

Se han recolectado los requerimientos del sistema se los ha catalogado y se procede a trabajar con ellos

Catálogo de Requisitos:

Funcionales

El sistema debe poder tomar distintos tipos de formatos de archivos de sucesos, esto se debe a que en la compañía existen distintos tipos de servidores Web.

Poder configurar la cantidad de procesos que se ejecutan en cada sección de detección de incidentes, es decir, poder configurar un workflow para cada incidente.

Poder tener un catalogo de los workflow existentes, para ser utilizarlos en futuros incidentes.

Registro de toda la operación realizada para el análisis del incidente.

No funcionales

Registro de toda la operación realizada para el análisis del incidente.

Resguardo de datos.

Posibilidad de poder ser agendado para ser ejecutado a una hora determinada.

Glosario:

Archivo de sucesos: archivo en formato ASCII, que contiene el registro de todas las operaciones que realiza un Servidor Web. Según la marca del Servidor su formato varía.

Incidente: pedido de informe que realiza un usuario ante un comportamiento anormal del sistema de información.

Workflow: encadenamiento de tareas que se le asigna que sean ejecutadas por un programa.

Fuentes libres: programas que se pueden bajar de Internet, y su distribución es gratuita y se cuenta con el código fuente de los mismos.

IDENTIFICACIÓN DEL ENTORNO TECNOLÓGICO

Descripción General del Entorno Tecnológico del Sistema:

Le principal escollo tecnológico al que se debe hacer frente es, que el sistema debe poder ser ejecutado en plataformas Windows y Unix, esto se realizará seleccionando un lenguaje de programación que permita ser ejecutado en ambos, el candidato mas firme es la utilización de Java, además de satisfacer esta condición es conocido por el equipo de desarrollo y es soportado por la compañía.

Con respecto a tiempos de respuesta y requerimientos de memoria mínimos y máximos, no se han hecho especificaciones sobre los mismos, lo que se debe tener en cuenta es que generalmente se ejecutara en terminales de usuarios, en las cuales su memoria es reducida, 256 Mg.

IDENTIFICACIÓN DE LOS USUARIOS PARTICIPANTES Y FINALES

Catálogo de Usuarios

Equipo de Soporte a la producción:

Principal usuario de sistema, los mismos tendrán un uso diario del sistema y es de esperar que se convierta en la principal herramienta para su trabajo.

Equipo de Desarrollo de Sistemas:

Otro grupo que debe ser considerado como usuarios principales del sistema, su uso estará supeditado a los requerimientos que reciban del equipo de soporte a la producción.

Equipo de Marketing:

Este equipo puede ser considerado un usuario secundario del sistema, el mismo utilizará a este cuando considere que pueda extraer información valiosa para algún proyecto de marketing en el que se encuentre trabajando.

Plan de Trabajo

Se presenta a continuación el plan de trabajo propuesto (tabla 4.13):

Etapa	Producto	Duración
Planificación de sistemas de información PSI	Documento de aceptación de sistemas	5 días
Análisis de requerimientos de sistemas ARS	Documento de aceptación de requerimientos	15 días
Especificación funcional del sistema EFS	Documento de aceptación de especificación funcional	5 días
Diseño Técnico del sistema DTS	Documento de aceptación de diseño técnico	5 días
Desarrollo de componentes del sistema DCS	Documento de aceptación de pruebas de unidad del sistema.	20 días
Desarrollo de procedimientos de Usuario DPU	Documento de aceptación de procedimientos del usuario	5 días
Pruebas, Implantación y Aceptación del Sistema – PIA	Documento de aceptación del sistema	5 días
Total		60 días

Tabla 4.13. Plan de Trabajo

A continuación se presenta el Gantt del sistema (figura 4.7)

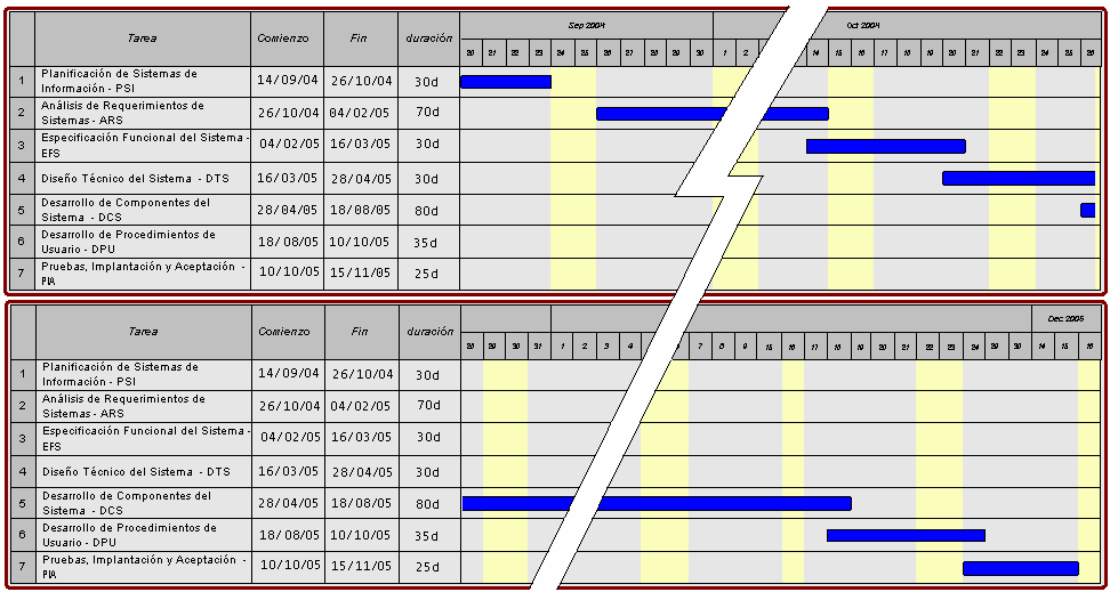

Figura 4.7. Gantt del sistema

ESTABLECIMIENTO DE REQUISITOS

A continuación se detallan los requisitos del sistema luego de la educción de los mismos

Catálogo de Requisitos:

En el anexo A se adjunta las minutas y el proceso de educción de requisitos:

RQV01_01: Los proceso que ejecute el sistema deben ser modulares y flexibles.

RQV01_02: La información entregada por el sistema debe poder ser verificada y validada.

RQV01_03: El sistema debe tener capacidad para agendar tareas.

RQV01_04: El sistema debe poder ser ejecutado en entornos Windows, Unix y Linux.

RQV01_05: Formato flexible de archivos de salida.

RQV01_06: Varias fuentes de dato de entradas.

RQV01_07: Catalogar los procesos

Modelo de Casos de Uso:

Se representa a en la figura 4.8. el el modelo de caso de uso:

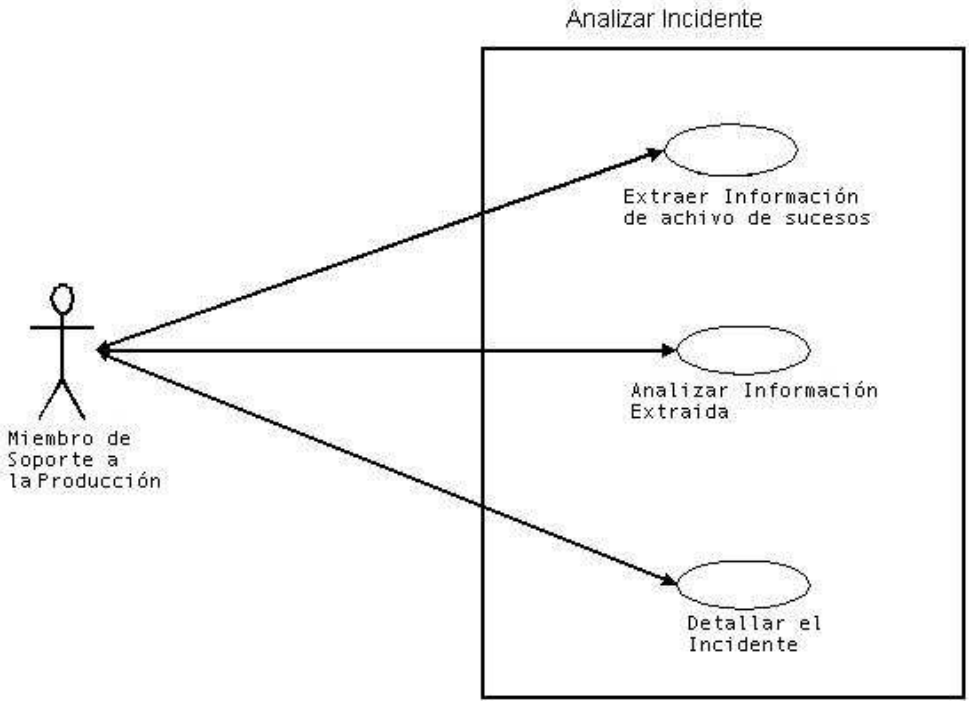

Figura 4.8. Modelo de caso de uso

Representación del diagrama de secuencia (figura 4.9):

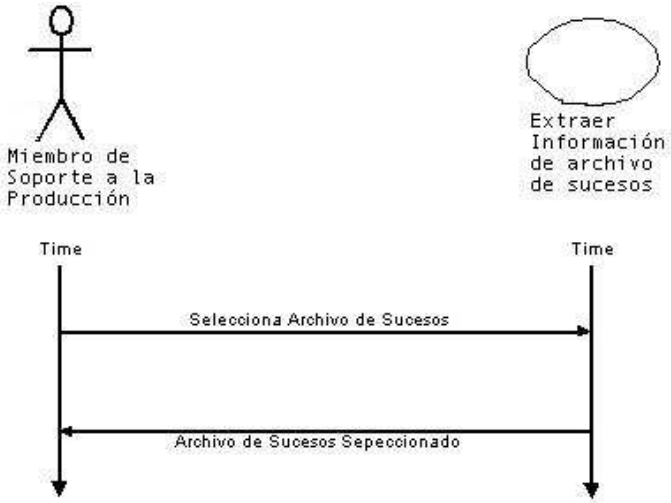

Figura 4.9. Diagrama de secuencia

Especificación de Casos de Uso

Detalle del caso de uso, en modo texto (tabla 4.14):

Caso de Uso	Análisis de Incidente
Resumen	Analizar un incidente reportado por la Mesa de Ayuda (MA)
Prioridad	Esencial
Frecuencia de Uso	Siempre
Actores Directos	Analista de Soporte a la Producción (ASP)
Inversionistas	ASP: quiere procesos rápidos, automatizados y reusables. MA: quiere la información rápido
Prequisitos	El ASP tiene los archivos de sucesos de los servidores Web. El ASP cuenta con el detalle del incidente.
Postcondiciones	Detectar e informar la causa del incidente.
Escenario Principal de Éxito	El ASP toma el detalle del incidente. El ASP intenta encontrar un incidente anterior igual al presente. El ASP responde a MS con la causa del incidente.
Escenario de Extensiones Alternativas	Si el incidente no fue analizado anteriormente, entonces El ASP realiza una lista tentativa de pasos a seguir y genera un plan. El ASP construye los pasos de la lista Si se puede reutilizar algún proceso de los antes hechos, entonces 1. Se utiliza. Sino 1. Se programa. Se ejecuta la tarea. Se evalúa la salida. Se prepara la respuesta Se almacena el detalle del nuevo incidente Se almacenan los fuentes del nuevo incidente.
Requisitos especiales	▪ La interfaz de usuario es por línea de comandos. ▪ Debe ser capaz de correr en una estación de trabajo y en un servidor. ▪ Debe poder agendarse para su ejecución
Lista de tecnología y variaciones de datos	▪ Sistemas operativos con los cuales debe correr en Unix y W2k
Notas y Preguntas	▪ Como se genera la lista de pasos a seguir en la tarea ▪ Puede ser pedido el análisis de incidente en forma automática

Tabla 4.14. Caso de Uso

Caso de Uso Análisis de Incidente

Un integrante de la Mesa de Ayuda, recibe un llamado de un usuario externo o interno, este le comunica una falla que se ha producido en una de las aplicaciones Web de la empresa. La Mesa de Ayuda, hace un conjunto de preguntas ya estipuladas al usuario, como ser nombre de la aplicación y una descripción lo mas detallada posible del problema, además de solicitarle de ser posible que se le adjunte la pantalla del error a un mail y enviarlo a la Mesa de Ayuda. Con esta información la Mesa de Ayuda carga un incidente en una planilla de calculo, este es tomado por un miembro del equipo de soporte a la producción, el cual a partir de ese momento queda a cargo del incidente. Con la información del incidente el miembro de Soporte a la Producción trata de darle una pronta solución. El primer paso es tratar de encontrar si se ha producido con anterioridad otro incidente similar, de encontrarlo analiza el incidente actual y el anterior valida que sea el mismo y entrega la respuesta en función de lo informado anteriormente.

De no encontrar ningún incidente parecido, se debe proceder a realizar un plan de acción para recolectar información de los archivos de sucesos de los Web Servers, para poder reconstruir la situación en la cual se produjo el incidente.
Este proceso consta de generar la lista de tentativas tareas para recolectar la información, luego de esto se trata de buscar si alguno de estos pasos ya fue hecho anteriormente, de ser así se reutilizan, en su defecto se programan.

Luego se ejecuta la tarea, se evalúa el resultado y se da la respuesta al incidente (tabla 4.15).

Caso de Uso	Análisis de característica de uso
Resumen	Analizar un pedido de informe de características de uso generado por el departamento de comercialización (DC).
Prioridad	Esencial.
Frecuencia de Uso	Siempre.
Actores Directos	Analista de Soporte a la Producción (ASP)
Inversionistas	ASP: quiere procesos rápidos, automatizados y reusables. DC: quiere que se le proporcione la información solicitada en un tiempo. aceptable, no mayor a los 3 días laborables.
Prerequisitos	El ASP tiene los archivos de sucesos de los servidores Web. El ASP cuenta con el informe de características de uso.
Poscondiciones	Detectar e informar el pedido de características de uso.

Escenario Principal de Éxito	El ASP toma el pedido y realizado por DC. El ASP trata de intenta encontrar un pedido de análisis de características de uso igual o similar al presente. El ASP interactúa con el con DC para analizar el resultado obtenido para detectar y validar el comportamiento de uso. Prepara el informe y lo entrega.
Escenario de Extensiones Alternativas	• Si no existe un pedido igual, entonces El ASP realiza una lista tentativa de pasos a seguir y genera un plan de acción. El ASP construye los pasos de la lista Si se puede utilizar un proceso antes realizado, entonces Se utiliza. Sino Se programa. Se ejecuta la tarea. El ASP interactúa con el con DC para analizar el resultado obtenido para detectar y validar el comportamiento de uso Se compara con otros análisis hechos anteriormente .Prepara el informe y lo entrega.
Requisitos especiales	• La interfaz de usuario es por línea de comandos. • Debe ser capaz de correr en una estación de trabajo y en un servidor. • Debe poder agendarse para su ejecución.
Lista de tecnología y variaciones de datos	Sistemas operativos con los cuales debe correr en Unix y W2k
Notas y Preguntas	Como se genera la lista de pasos a seguir Puede este proceso una vez construido ser ejecutado por el DC

Tabla 4.15. Caso de Uso.

Caso de Uso Análisis de característica de uso

Un miembro del departamento de comercialización (DC), pide al analista de soporte a la producción (ASP) que realice un análisis de patrones de comportamiento de los usuarios. El ASP evalúa los archivos de registros de los Web Servers y utilizando técnicas tanto sea de estadística como de Inteligencia Artificial, o la que crea conveniente trata de encontrar los patrones de comportamiento.

IDENTIFICACIÓN DE SUBSISTEMAS DE ANÁLISIS

Determinación de Subsistemas de Análisis

Es posible dividir el sistema en al menos 2 subsistemas, por un lado el sistema de encadenamiento de tareas y por otro el encargado de analizar la información.

Integración de Subsistemas de Análisis

El subsistema encargado de analizar la información deberá poder ejecutarse desde línea de comandos con el siguiente formato: <nombre_programa> <parámetros> <archivo_log> <archivo_error>

ANÁLISIS DE LOS CASOS DE USO

Identificación de Clases Asociadas a un Caso de Uso

Del estudio de los casos de uso se han detectado un conjunto de actividades comunes a todos los procesos. Brevemente se pasan a detallar los mismos:

- Lectura y escritura de archivos varios: todos los procesos necesitan acceder a la información almacenada en archivos y dejar el resultado de los procesos en los mismos u otros archivos (lecturaEscrituraArchivo)

- Lectura y escritura de Base de Datos: todos los procesos necesitan acceder a la información almacenada en base de datos y dejar el resultado de los procesos en los mismos u otras bases de datos (lecturaEscrituraBaseDatos)
- Generación de archivos con error: en caso de producirse un error se debe permitir obtener un detalle lo suficientemente exacto de la situación producida, por existir la posibilidad de que sea imposible duplicar la condición (grabarArchivoError)
- Generación de archivos de pasos: cada modificación (paso) que se realiza con la información debe generar toda las estadísticas necesarias para poder realizar un análisis post ejecución y evaluar el rendimiento de los mismos (leerArchivoPasos)
- Administración de memoria: los procesos de transformación de información se caracterizan por el consumo de memoria de los equipos en donde es ejecutado. Por esta razón se hace necesario que se pueda configurar el consumo de la misma para no sobrecargar el equipo(asignarMemoria)

A continuación se detallan los nombres tentativos de las clases y sus principales características:

- lecturaEscrituraArchivo: esta clase se encarga de realizar las tareas de lectura y escritura de datos en el sistema de archivos para los distintos procesos que se deben realizar para el análisis de la información.

  Características principales:
  - El código de la misma debe ser muy performante.
  - Se debe contemplar una interfaz lo suficientemente amplia para cubrir todas las posible posibilidades de acceso a archivos de datos.
  - Debe contemplar el manejo de bloqueo de archivos en proceso.

- lecturaEscrituraBaseDatos: esta clase es la encargada de abstraer el modelo de base de datos todos los accesos de lectura y escritura se realizan a través de ella.

  Características principales:

  - Permitir el acceso a diferentes bases de datos y modificar la interfaz de usuario.
  - Configurable a trabes de archivos de parámetros

- grabarArchivoError: esta clase es la encargada de generar los mensajes de error que se puedan generar en la aplicación. Básicamente esta clase da formato al mensaje y se vale de los servicios de las clases LEArchivo y LEBaseDatos para generar la persistencia de los datos.

  Características principales:

  - La selección de la persistencia de datos debe ser configurable.
  - El formato de salida debe ser configurable.

- leerArchivoPasos: esta clase es la encargada de dar formato a todos los mensajes de información que generan los distintos pasos del proceso de análisis de información, al igual que la clase GarchivoError se vale de los servicios de las clases LEArchivo y LEBaseDatos para generar la persistencia de los datos.

  Características principales:

  - La selección de la persistencia de datos debe ser configurable.
  - El formato de salida debe ser configurable.

- asignarMemoria: esta clase es la encargada de manejar la memoria de trabajo de los procesos. Esta clase se vale de los servicios de las clases LEArchivo y LEBaseDatos para generar la persistencia de los datos.

  Características principales:

  - Debe permitir la asignación de memoria en forma dinámica.

Descripción de la Interacción de Objetos

lecturaEscrituraArchivo: interactuá con las clases grabarArchivoError, leerArchivoPasos y asignarMemoria.

lecturaEscrituraBaseDatos: interactuá con las clases grabarArchivoError, GArchivoPasos.

grabarArchivoError: interactuá con las clases lecturaEscrituraArchivo, lecturaEscrituraBaseDatos

leerArchivoPasos: interactuá con las clases lecturaEscrituraArchivo, lecturaEscrituraBaseDatos

asignarMemoria: interactuá con las clases lecturaEscrituraArchivo, lecturaEscrituraBaseDatos

ANÁLISIS DE CLASES

Identificación de Responsabilidades y Atributos

A continuación se representan las clases candidatas y los métodos propuestos para las clases

Identificación de atributos candidatos:

lecturaEscrituraArchivo:

abrirArchivo
cerrarArchivo
modoAcceso
formatoArchivo
entregarFila
recibirFila
limpiarMemoria

asignarMemoria
grabarMemoria

lecturaEscrituraBaseDatos:
abrirConexion
cerrarConexion
modoAcesso
ejecutarComando

grabarArchivoError:
nombreArchivo
modoAcceso
mensajeError

leerArchivoPasos:
nombreArchivo
modoAcceso
listarPasos
pasoCorriente
proximoPaso

asignarMemoria
asignarMemoria
limpiarMemoria
estadoMemoria

La representación de las clases candidatas se puede ver en la figura 4.10.

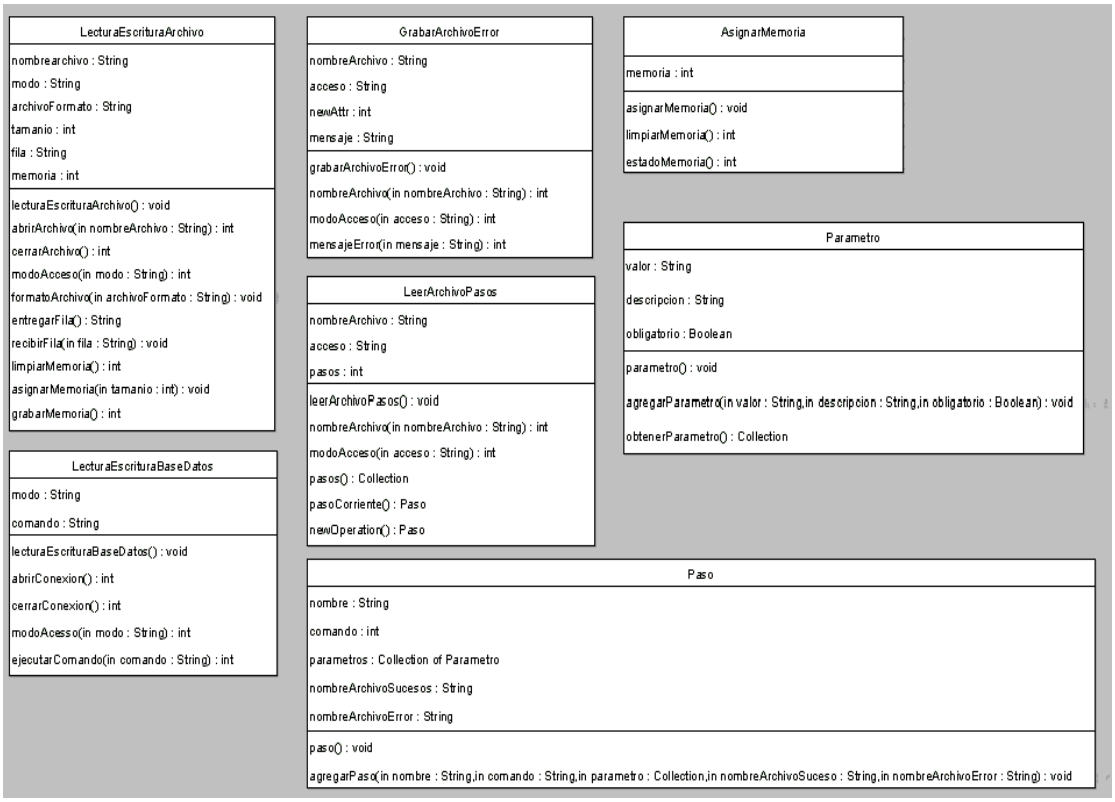

Figura 4.10. Clases Candidatas

Identificación de Asociaciones y Agregaciones

Identificación de Asociaciones

grabarArchivoError:

accede a lecturaEscrituraArchivo

accede a lecturaEscrituraBaseDatos

leerArchivoPasos:

accede a lecturaEscrituraArchivo

accede a lecturaEscrituraBaseDatos

lecturaEscrituraArchivo:

accede a asignarMemoria

DEFINICIÓN DE INTERFACES DE USUARIO

Especificación de Principios Generales de la Interfaz

La ejecución del sistema se realiza desde la línea de comandos, para la preparación de los programas y el encadenador de tareas cada usuario utilizara el editor con el cual se sienta más cómodo, el resultado de la ejecución es formateado para poder ser agregado a una página Excel o una base de datos, las cuales tienen su propia interfaz y modo de representar los datos.

ANÁLISIS DE CONSISTENCIA Y ESPECIFICACIÓN DE REQUISITOS

Validación de los Modelos

La validación de los requisitos de usuario se ha realizado con los mismos usuarios como queda detallado en le Anexo A

Elaboración de la Especificación de Requisitos Software ERS (tabla 4.16):

Requisitos de sistema	RQV01_01	Los proceso que ejecute el sistema deben ser modulares y flexibles
	RQV01_02	La información entregada por el sistema debe poder ser verificada y validada
	RQV01_03	El sistema debe tener capacidad para agendar tareas
	RQV01_04	El sistema debe poder ser ejecutado en entornos Windows, Unix y Linux
	RQV01_05	Formato flexible de archivos de salida
	RQV01_06	Varias fuentes de dato de entradas
	RQV01_07	Catalogar procesos
Plataforma de uso	Unix (Sun)	
	Linux (Slakware, Red Hat, Debian)	
	Windows (W2K, XP)	
Lenguaje de Programación	Java versión 1.5	
	Jython	
	XML	
Módulos del Sistema	Encadenador de Tareas	
	Formateador de archivos	
	Generador de tareas	

Tabla 4.16. Especificación de Requisitos Software

ESPECIFICACIÓN DEL PLAN DE PRUEBAS

Definición del Alcance de las Pruebas

De la sesión de trabajo realizada entre los miembros del equipo de desarrollo y los usuarios, se ha definido el alcance de las pruebas a:

- Se realizara una sesión completa de pedido de información por parte el equipo de Help Desk y el equipo de Soporte a la producción deberá dar una respuesta al mismo
- Por otra parte se realizara un pedido de análisis de un patrón de comportamiento por parte del equipo de Mercadotecnia, y el equipo de Soporte a la Producción deberá dar cause al pedido.

Definición de Requisitos del Entorno de Pruebas

En función de lo definido en el Alcance de las pruebas (4.2.3), se ha definido que el entorno de las pruebas que se realizara será, una sobre un equipo Windows 2000 y la otra sobre un equipo Linux Slakware.

Definición de las Pruebas de Aceptación del Sistema

Se evaluara el tiempo de respuesta de los pedidos antes especificados, y los mismos deberán ser al menos un 25% más rápidos que el promedio de respuesta hasta ese momento. Para el calculo del tiempo promedio de respuesta se tomaran los tiempos de los últimos 12 meses de trabajo del equipo de Soporte a la producción.

El conjunto de datos que se utilizara para la prueba se validará con el jefe de Soporte a la Producción.. Los mismos serán al menos 6 casos, con pedidos de distintos tipos la selección recomendada será:

Pedidos para la prueba de exploración de datos del tipo archivo de suceso:

2 casos con pedidos de supuestos errores de aplicación (esto simula a un usuario al equipo de Mesa de Ayuda)

2 casos con pedidos de búsqueda de patrones de comportamiento (esto simula a un usuario del sector de Mercadotecnia)

2 casos con pedidos de evacuación de tiempo de respuesta del servidor (simula un usuario del sector de Desarrollo)

Pedidos para la prueba de exploración de datos del tipo hipervínculos:

- 1 caso con búsqueda de hipervínculos entre paginas del mismo sitio y hacia otros (simula un usuario del sector de Desarrollo)

Pedidos para la prueba de exploración de datos del tipo contenido:

2 caso generando búsquedas de contenidos y analizando su patrón de comportamiento esto simula a un usuario del sector de Mercadotecnia.

Una vez realizada estas pruebas se evaluara el tiempo total de respuesta y la certeza de los mismos. Una comisión constituida por el jefe de Soporte a la Producción, el jefe de Mercadotecnia, el jefe de Desarrollo y el jefe de Mesa de ayuda, evaluaran el resultado de los mismos y el tiempo de respuesta, este equipo dará su aceptación o formulara las recomendaciones necesarias.

4.2.4 DISEÑO DEL SISTEMA DE INFORMACIÓN

DEFINICIÓN DE LA ARQUITECTURA DEL SISTEMA

El objetivo del proceso de diseño es la definición de la arquitectura del sistema de información y el entorno tecnológico que va a soportarlo junto con una exposición detallada de sus distintos componentes.

A partir de esta información se generarán todas las especificaciones de construcción. Se establecerá en este capítulo el particionamiento físico del sistema de información, su organización en subsistemas de diseño, la especificación del entorno tecnológico y los requisitos de operación, administración, seguridad y control de acceso.

DEFINICIÓN DE NIVELES DE ARQUITECTURA

Diseño de la Arquitectura del Sistema

En esta tarea se describirán los niveles de arquitectura de software, mediante la definición de las principales particiones físicas del sistema, representadas como nodos y comunicaciones entre ellos.

Un nodo es una partición física o parte significativa del sistema de información con características propias de ejecución.

Se especificarán como nodos los siguientes elementos de infraestructura:

- Tipos de puestos clientes.
- Servidores
- Comunicaciones

El proyecto puede ser dividido en tres subsistemas:

Subsistema Encadenamiento de Tareas: Este subsistema es el encargado de ejecutar las tareas necesarias para la extracción, transformación y análisis de los datos del archivo de sucesos. Físicamente este subsistema se puede encontrar en un Servidor o en una terminal de cliente.

Subsistema Extracción de Archivo de incidentes: Este subsistema es el encargado de extraer información del los archivos de sucesos de los Servidores Web y transformarla para su posterior análisis. Físicamente este subsistema se puede encontrar en un Servidor o en una terminal de cliente.

Subsistema de Análisis de Datos: Este subsistema es el encargado de realizar el análisis de la información extraída del archivo de sucesos del Servidor Web y que ya ha sido transformada para poder ser analizada.

IDENTIFICACIÓN DE REQUISITOS DE DISEÑO Y CONSTRUCCIÓN

Subsistema Encadenamiento de Tareas:

Físicamente este subsistema se puede encontrar en un Servidor o en una terminal de cliente.

Se ha seleccionado el Fuente Libre llamado ANT en la versión 1.4 o superior, esto se debe a que cumple con los requisitos de poder ser ejecutado sobre plataformas Windows y Unix. Por otra parte el encadenamiento de tareas que realiza es parametrizable en un archivo de XML, lo que le da al producto una gran flexibilidad. También es posible utilizar OpenWFE de similares características que ANT, pero con mayores capacidades graficas. Es posible que se utilice otra o ambas en un mismo proceso.

Funciones del subsistema:

Control de ejecución de tareas.

Estándar de interfaz de módulos.

Manejo de errores de módulo.

Reporte de procesamiento.

Subsistema Extracción de Archivo de incidentes:

- Físicamente este subsistema se puede encontrar en un Servidor o en una terminal de cliente.
- Como lenguaje de programación se ha seleccionado Jython, que es un dialecto de Java. Jython se enrola dentro de los lenguajes de Scripting, lo que le da gran flexibilidad, cumple con los requisitos de poder ser ejecutado sobre plataformas Windows y Unix y permite integrar las librerías de JAVA, con lo cual se logra gran poder de procesamiento.

- Funciones del subsistema:

  Extracción de datos por búsqueda simple o expresiones regulares.

  Formateo de datos.

  Limpieza de datos.

  Asignación de valores por defecto.

Subsistema de Análisis de Datos:

Físicamente este subsistema se puede encontrar en una terminal de cliente.

En este subsistema tomará productos de Fuentes Libres y analizará la información para obtener información (detalle de la selección en Anexo B. 1.4).

Funciones del subsistema:

Análisis de información mediante métodos enrolados en la Inteligencia Artificial.

ESPECIFICACIÓN DE ESTÁNDARES Y NORMAS DE DISEÑO Y CONSTRUCCIÓN

A continuación se detallan los subsistemas que componen el presente trabajo, los mismos son tres que se pasan a detallar:

Subsistema Encadenamiento de Tareas: Es el encargado de ejecutar en secuencia los pasos necesarios para la obtención, modificación y generación de informes.

El archivo de configuración para el manejo de tareas debe cumplir con la norma de construcción de archivos en XML.

Subsistema Extracción de Archivo de incidentes: Es el encargado de realizar la extracción y obtención de datos de los archivos de suceso.

Los procesos de extracción de datos deben poder ser ejecutados en un proceso Batch.

Los procesos deben poder manejar un mismo formato de entrada de datos.

El mismo se especifica a continuación:

NombreDeProceso N_ParámetrosDeEntrada ArchivoDeRegistro ArchivoDeError

Subsistema de Análisis de Datos: es el encargado de tomar los datos extraídos por el subsistema anterior y analizar los datos. Este subsistema por estar constituido en la mayor cantidad de casos por software bajo la modalidad de Fuente Libre, la única estándar que se validará en la posibilidad de ser ejecutado dentro de un proceso Batch.

IDENTIFICACIÓN DE SUBSISTEMAS DE DISEÑO

Subsistema Encadenamiento de Tareas
Subsistema Extracción de Archivo de incidentes
Subsistema de Análisis de Datos

ESPECIFICACIÓN DEL ENTORNO TECNOLÓGICO

Subsistema Encadenamiento de Tareas:

Hardware: Computadora personal con placa principal con arquitectura PCI/ISA con AGP y bus de 100 MHZ y 64 bits para acceso a memoria – Procesador Intel Pentium III – 800 MHZ o superior compatible - 256 Mb de RAM – Disco Rígido con capacidad disponible de 10 MB para el Sistema – Unidad de CD – ROM
Software: Sistema Operativo Microsoft Windows o Unix.

Subsistema Extracción de Archivo de incidentes:

Hardware: Computadora personal con placa principal con arquitectura PCI/ISA con AGP y bus de 100 MHZ y 64 bits para acceso a memoria – Procesador Intel Pentium III – 800 MHZ o superior compatible - 256 Mb de RAM – Disco Rígido con capacidad disponible de 10 MB para el Sistema – Unidad de CD – ROM
Software: Sistema Operativo Microsoft Windows o Unix.

Subsistema de Análisis de Datos:

Hardware: Computadora personal con placa principal con arquitectura PCI/ISA con AGP y bus de 100 MHZ y 64 bits para acceso a memoria – Procesador Intel Pentium III – 800 MHZ o superior compatible - 256 Mb de RAM – Disco Rígido con capacidad disponible de 10 MB para el Sistema – Unidad de CD – ROM

Software: Sistema Operativo Microsoft Windows o Unix.

ESPECIFICACIÓN DE REQUISITOS DE OPERACIÓN Y SEGURIDAD

En esta tarea se definirán los distintos procedimientos de seguridad y operación necesarios para no comprometer el funcionamiento del sistema y asegurar el cumplimiento de los niveles de servicios que tendrá el producto en cuanto a la gestión de operaciones (procesos, seguridad, comunicaciones, etc).

En tal sentido se enuncian los procesos relacionados con:

Accesos al sistema y sus recursos.
Mantenimiento de la integridad y confidencialidad de datos.
Control y registro de accesos.
Copias de seguridad y recuperación de datos.
Recuperación ante catástrofes.

Accesos al sistema y sus recursos: Ninguno de los módulos tiene requerimientos de seguridad con acceso al sistema y sus recursos.

Mantenimiento de la integridad y confidencialidad de datos: La confiabilidad de los datos esta dada por la extracción de los mismos de los Servidores Web, por ser un sistema de transformación de datos todo el proceso, es decir, todos los archivos intermedios generados para obtener la información final será resguardada al final de cada proceso.

Control y registro de accesos: Cada uno de los procesos y que intervienen generan un archivo de registro como se especificó en 4.2.4. Estos junto a los archivos intermedios serán resguardados.

Copias de seguridad y recuperación de datos: Cada uno de los procesos para obtener información junto con los archivos de registros de los mismos, serán resguardados. El mismo constara con una breve descripción de la fecha hora y una breve descripción del proceso de extracción de datos realizado y su objetivo.

DISEÑO DE CASOS DE USO REALES

Diseño de la Realización de los Casos de Uso

A continuación se definen los diagramas de secuencia relacionados con los distintos casos de usos definidos en el sistema

Diagrama de secuencia que se relaciona con el caso de Análisis de Incidente (figura 4.11)

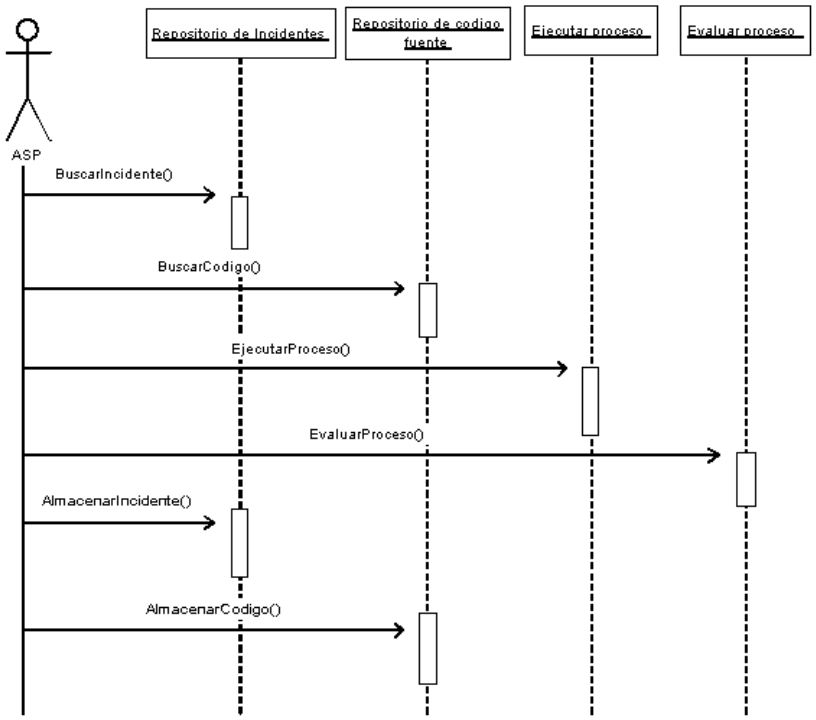

Figura 4.11. Diagrama de secuencia del caso de uso Análisis de Incidente

Diagrama de secuencia que se relaciona con el caso de Análisis de característica de uso (figura 4.12)

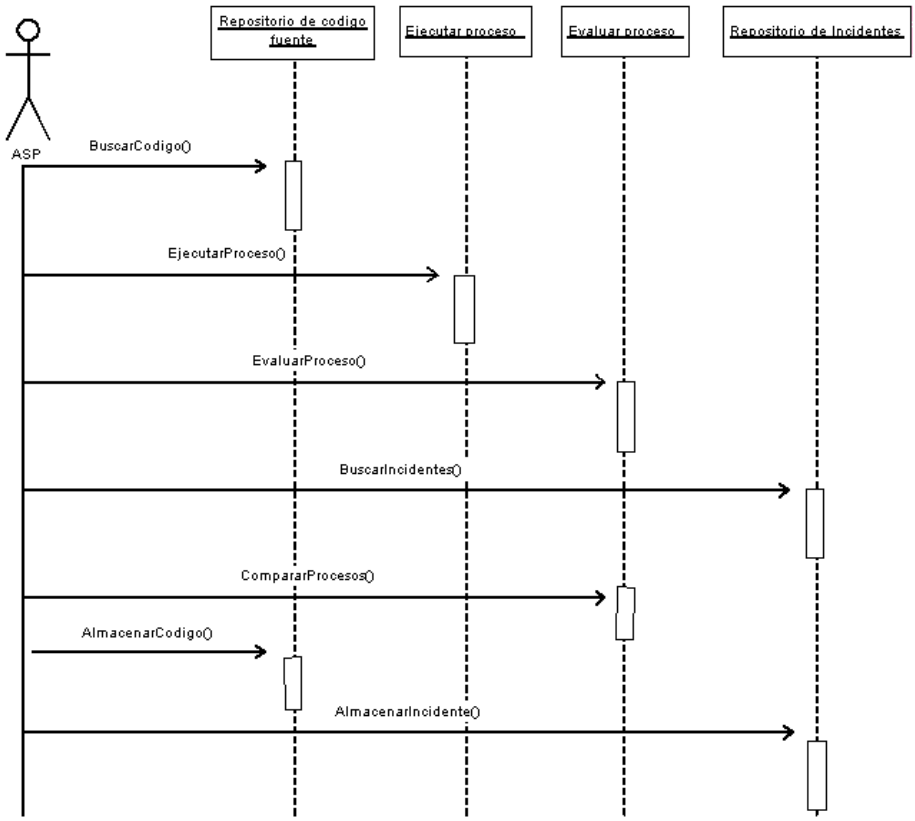

Figura 4.12. Diagrama de secuencia del caso de uso Análisis de característica de uso

REVISIÓN DE LA INTERFAZ DE USUARIO

A continuación se detalla la interfaz que el usuario/programador, tiene a su alcance para la concreción de posprogramas necesarios para realizar el proceso.

lecturaEscrituraArchivo: interactuá con las clases grabarArchivoError, leerArchivoPasos y asignarMemoria.

lecturaEscrituraBaseDatos: interactuá con las clases grabarArchivoError, GArchivoPasos.

grabarArchivoError: interactuá con las clases lecturaEscrituraArchivo, lecturaEscrituraBaseDatos

leerArchivoPasos: interactuá con las clases lecturaEscrituraArchivo, lecturaEscrituraBaseDatos

asignarMemoria: interactuá con las clases lecturaEscrituraArchivo, lecturaEscrituraBaseDatos

DISEÑO DE CLASES

Diseño de Asociaciones y Agregaciones

grabarArchivoError:
accede a lecturaEscrituraArchivo
accede a lecturaEscrituraBaseDatos

leerArchivoPasos:
accede a lecturaEscrituraArchivo
accede a lecturaEscrituraBaseDatos

lecturaEscrituraArchivo:
accede a asignarMemoria

Identificación de Atributos de las Clases

A continuación de detallan los atributos propuestos para el esquema de clases.

Interfaz Manejo de ficheros

lecturaEscrituraArchivo:
abrirArchivo
cerrarArchivo
modoAcceso

formatoArchivo
entregarFila
recibirFila
limpiarMemoria
asignarMemoria
grabarMemoria

lecturaEscrituraBaseDatos:
abrirConexion
cerrarConexion
modoAcesso
ejecutarComando

grabarArchivoError:
nombreArchivo
modoAcceso
mensajeError

Paquete Manejo de procesos

leerArchivoPasos:
nombreArchivo
modoAcceso
listarPasos
pasoCorriente
proximoPaso

Paquete Manejo de Memoria
asignarMemoria:
asignarMemoria
limpiarMemoria
estadoMemoria

Detalle de la estructura de paquetes del sistemas (Figura 4.13)

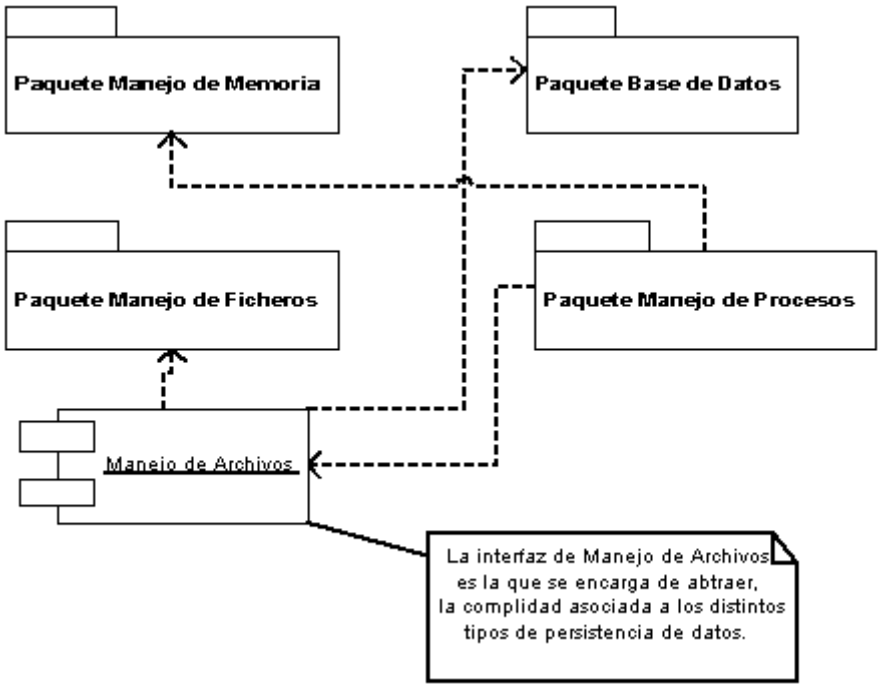

Figura 4.13. Estructura de paquetes del sistema

DISEÑO DE LA ARQUITECTURA DE MÓDULOS DEL SISTEMA

Diseño de Módulos del Sistema

A continuación se representa una abstracción de los módulos de sistemas a modo de comprensión de su interacción (figura 4.14).

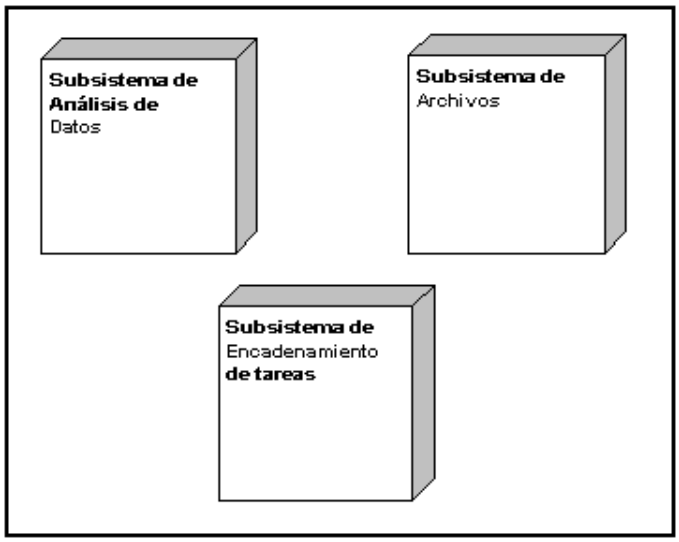

Figura 4.14. Módulos del sistema

Diseño de Comunicaciones entre Módulos

Mientras que todos los componentes de los subsistemas de Subsistema Extracción de Archivo de incidentes y Subsistema de Análisis de Datos, el subsistema Encadenamiento de Tareas, podrá encadenar las tareas asignadas a cada proceso de extracción y análisis de datos.

DISEÑO FÍSICO DE DATOS

Diseño del Modelo Físico de Datos

El presente sistema trabaja con solo dos archivos de datos fijos, que son los archivos de suceso de la operación de cada paso de la tarea y los archivos de errores de cada paso de la tarea (tabla 4.17 y 4.18).

Formato de archivo de suceso de operación		
Nombre	Tipo	Descripción
tarea	Char(20)	Nombre de la tarea a la que pertenece el paso
secTarea	Long	Numero secuencial de ejecución para toda la tarea
paso	Char(20)	Nombre del paso que lo genero
fechaHora	Char(20)	Fecha y hora de la ejecución del paso
regLeidos	Long	Cantidad de registros leídos por el paso
regGrabados	Long	Cantidad de registros grabados por el proceso

Tabla 4.17. Formato de archivo de suceso de operación

Formato de archivo de error		
Nombre	Tipo	Descripción
tarea	Char(20)	Nombre de la tarea a la que pertenece el paso
secTarea	Long	Numero secuencial de ejecución para toda la tarea
paso	Char(20)	Nombre del paso que lo genero
fechaHora	Char(20)	Fecha y hora de la ejecución del paso
regLeidos	Long	Cantidad de registros leídos por el paso
regGrabados	Long	Cantidad de registros grabados por el proceso
regError	Long	Numero de registro donde se produjo el error
msjError	Char(128)	Breve descripción de la causa del error
fileError	Char(64)	Camino donde se encuentra el registro con el error

Tabla 4.18. Formato de archivo de error

Especificación de los Caminos de Acceso a los Datos

Vale aclarar que según el tipo de proceso que el equipo de Soporte a la Producción realice puede elegir que los mismos se generen en el sistema de archivos o en una base de datos. Por esta razón en cada proceso se detallara el camino correspondiente

OPTIMIZACIÓN DEL MODELO FÍSICO DE DATOS

Definición de claves del sistema (tabla 4.19, 4.20 y 4.21).

Formato de archivo nombre de tarea			
Clave		Clave	Tipo
tarea	Long	si	primaria
tareaDesc	Char(20)		

Tabla 4.19. Formato de archivo nombre de Tarea

Formato de archivo de suceso de operación			
Clave		Clave	Tipo
tarea	Long	si	primaria
secTarea	Long	si	primaria
paso	Char(20)	si	primaria
fechaHora	Char(20)		
regLeidos	Long		
regGrabados	Long		

Tabla 4.20. Formato de archivo de suceso de operación

Formato de archivo de error			
Nombre	Tipo	Clave	Tipo
tarea	Long	si	primaria
secTarea	Long	si	primaria
paso	Char(20)	si	primaria
fechaHora	Char(20)		
regLeidos	Long		
regGrabados	Long		
regError	Long		
msjError	Char(128)		
fileError	Char(64)		

Tabla 4.21. Formato de archivo de error

ESPECIFICACIÓN TÉCNICA DEL PLAN DE PRUEBAS

Especificación del Entorno de Pruebas

Las pruebas a realizar serán ejecutando un proceso completo de extracción y análisis de datos de un archivo de sucesos de un Servidor Web.

Especificación Técnica de Niveles de Prueba

La revisión técnica constará de:

- Recursos consumidos por el procesos (memoria, espacio en disco y tiempo de procesamiento).
- Calidad de la información obtenida.

Revisión de la Planificación de Pruebas

La prueba constará de:

Simulación de una solicitud de análisis, el mismo será la detección de usuarios que accedieron a una aplicación Web dentro de un rango horario de dos horas.

Y detectar la navegación de los mismos y segmentar al conjunto de usuarios en ese rango horario.

Con este requerimiento se deberán realizar todas las tareas de para obtener la información solicitada mediante el uso del sistema.

ESTABLECIMIENTO DE REQUISITOS DE IMPLANTACIÓN

Especificación de Requisitos de Documentación de Usuario

La documentación que será entregada al usuario es un manual de uso y operaciones el cual constará de ejemplos de casos de usos.

4.2.5 CONSTRUCCIÓN DEL SISTEMA DE INFORMACIÓN

PREPARACIÓN DEL ENTORNO DE GENERACIÓN Y CONSTRUCCIÓN

Implantación de la Base de Datos Física o Ficheros

Calculo de longitud de registro de archivo de sucesos de pasos:

Long (8 bytes) 3 campos = 24
Char (1 bytes) 3 * 20 = 60
Longitud de registro = 84

Cantidad de registros generados diariamente aprox. 100, por 270 días hábiles de trabajo en promedio el espacio necesario es de 22680 bytes. En caso de generarse sobre una base de datos se agregara aproximadamente 20 bytes, esto depende de la base de datos.

Calculo de longitud de registro de archivo de sucesos de pasos:

- Long (8 bytes) 4 campos = 32
- Char (1 bytes) (20 * 3) + 128 + 64 = 252
- Longitud de registro = 284

Cantidad de registros generados diariamente aprox. 10, por 270 días hábiles de trabajo en promedio el espacio necesario es de 2700 bytes.En caso de generarse sobre una base de datos se agregara aproximadamente 20 bytes, esto depende de la base de datos.

Preparación del Entorno de Construcción

El entorno de trabajo será:

Se utilizara Eclipse para generar el código.
El control de versiones será Subversión.
Base de datos para pruebas MySQL.
Librerías para la construcción de código Log4J.

GENERACIÓN DEL CÓDIGO DE LOS COMPONENTES Y PROCEDIMIENTOS

Generación del Código de Componentes

Se genera el código fuente de la aplicación y queda versionado en el control de versión.

EJECUCIÓN DE LAS PRUEBAS UNITARIAS

Preparación del Entorno de las Pruebas Unitarias

El entorno de pruebas unitarias se realizara en la estación de trabajo del programador, el cual preparara la ejecución de uno de los casos dados.

Realización y Evaluación de las Pruebas Unitarias

A continuación se detalla la plantilla con todos los casos de prueba a realizar en el test unitario, esta plantilla será entregada a los desarrolladores y en la columna Resultado deberán anotar como ha sido ejecutada (satisfactoria / insatisfactoria) Al final de la

plantilla se podrá agregar si lo desea el responsable de la prueba, cualquier observación que crea acertada (Tabla 4.22)

Descripción	Esta prueba de unidad se encargará de probar la Interfaz de manejo de Ficheros el manejo de memoria y errores.		
Detalle de la prueba	La prueba se realizará sobre una estación de trabajo Windows Profesional 2000, para la prueba de los métodos de accesos a al base de datos se consultaran datos y se insertaran datos en una base de datos MySQL. Para la prueba de los métodos de acceso a archivos se probara haciendo una lectura a un archivo de suceso y grabando la salida con los totales de registros leídos y los tiempos del mismo.		
Responsable	Hernan D. Merlino		
Caso de Prueba	Descripción	Métodos usados	Resultado
Leer de una Base de datos	Se realiza la lectura a una base de datos (MYSQL), se realizara utilizando la clase lecturaEscrituraBaseDatos	abrirConexion cerrarConexion modoAcesso ejecutarComando	
Leer y Escribir de un archivo	Se realizara la lectura de un archivo de suceso, en su totalidad y se escribirá un archivo que contenga la cantidad de registros leídos y el tiempo total. Para esta prueba se utilizaran las clases lecturaEscrituraArchivo y asignarMemoria	abrirArchivo cerrarArchivo modoAcceso formatoArchivo entregarFila recibirFila limpiarMemoria asignarMemoria grabarMemoria asignarMemoria limpiarMemoria estadoMemoria	
Generación de un error en acceso a una base de datos	Se intentara acceder a una base de datos con un usuario incorrecto para forzar un error. Se utilizaran para esta prueba las clases lecturaEscrituraBaseDatos y grabarArchivoError	abrirConexion nombreArchivo modoAcceso mensajeError	
Observaciones			

Tabla 4.22. Plantilla para pruebas unitarias.

EJECUCIÓN DE LAS PRUEBAS DE INTEGRACIÓN

Preparación del Entorno de las Pruebas de Integración

El entorno de pruebas de integración es la estación de trabajo del programador, esto es posible pues uno de los requerimientos es que el sistema pueda se ejecutado en una estación de trabajo.

Realización de las Pruebas de Integración

Con el esquema de pruebas realizado en las pruebas unitarias (4.2.5.) se agrega en la tarea el proceso de separación de registros por día de acceso. Luego de realizar la prueba se ha evaluado la salida y la misma coincide con el formato especificado y la cantidad de archivos leídos y grabados. En la segunda etapa de la tarea se ha comprobado que las fechas han sido divididas correctamente y la cantidad de registros coincide con la sumatoria total.

Evaluación del Resultado de las Pruebas de Integración

La evaluación ha sido satisfactoria y se considera que se puede pasar a la prueba de sistemas.

EJECUCIÓN DE LAS PRUEBAS DEL SISTEMA

Preparación del Entorno de las Pruebas del Sistema

El entorno de pruebas de sistema se realizara en un servidor, que será asignado por el jefe de Infraestructura de Sistemas.

Realización de las Pruebas del Sistema

A continuación se detalla el conjunto de pruebas de sistema que se realizarán. Antes de esto se realiza una breve descripción de la empresa y su sitio Web (Tablas 4.23 a 4.29).

Compañía: Servicios Financieros en Internet

Descripción de la plataforma: Es una compañía dedicada exclusivamente a dar servicios financieros en Internet a usuarios principalmente localizados en la ciudad de Buenos Aires y Gran Buenos Aires. Esta empresa posee un portal que es la puerta de entrada de la compañía desde la misma un usuario puede seleccionar alguno de los varios servicios que la compañía brinda, los mismos son:

- Operaciones bursátiles en la Bolsa de Comercio de Buenos Aires.
- Operaciones bursátiles en la Bolsa de Cereales de Buenos Aires y Rosario y Chicago.
- Compra y venta de dólares, euros y yenes.
- Inversiones no convencionales
  - Operaciones en el mercado de café en Nueva York
  - Operaciones en el mercado de jugo de naranjas de Miami.
  - Operaciones en el mercado de frutas de Frankfurt.
  - Operaciones en el mercado de flores de Rótterdam.

Cada servicio opera en su propio sitio de Internet que físicamente se encuentra en el mismo lugar pero en distintos equipos y con distintos Web Servers. Los mismos son Web Servers IPlanet versión 6.0 de la Empresa Netscape. Las aplicaciones están hechas en HTML y JSP para la interfaz Web, servlets para el manejo del modelo de muestra y controlador (model view controller MVC), utiliza el open source STRUTS. La inteligencia del negocio esta contenida en Beans de Java, los datos son almacenados y consultados de una base de datos MYSql.

Casos a probar:

Caso	00 – Detección de Error	
Descripción	Detección de análisis de comportamiento de un usuario que ha realizado una queja a Mesa de Ayuda.	
Origen	Mesa de Ayuda	
Pedido	Se ha producido aparentemente un problema en una aplicación Web	
Detalle	Sistema en que se produjo	Operaciones bursátiles en la Bolsa de Comercio de Buenos Aires (OBCBA)
	Descripción del incidente	El usuario Aníbal Jerez, el día 22 de julio del 2005 ha intentado realizar una transacción bursátil de compra de acciones y le ha aparecido un mensaje de error en el sitio OBCBA
Tipo de Sitio	Grande (calculado según el anexo B)	

Tabla 4.23. Caso de prueba Detección de Error

Caso	01 – Detección de Error	
Descripción	Detección de análisis de comportamiento de un usuario que ha realizado una queja a Mesa de Ayuda.	
Origen	Mesa de Ayuda	
Pedido	Se ha producido aparentemente un problema en una aplicación Web	
Detalle	Sistema en que se produjo	Operaciones bursátiles en la Bolsa de Comercio de Buenos Aires (OBCBA)
	Descripción del incidente	El usuario Juan Pérez, el día 05 de mayo del 2005 ha intentado acceder al sitio OBCBA, la hora aproximada fue las 15.30 hs y no ha podido acceder al mismo.
Tipo de Sitio	Grande (calculado según el anexo B)	

Tabla 4.24. Caso de prueba Detección de Error

Caso	02 – Patrón de comportamiento	
Descripción	El área de Mercadotecnia hace un pedido de búsqueda de patrón de comportamiento por parte de un usuario.	
Origen	Mercadotecnia	
Pedido	Detectar cual es el comportamiento de uso de un usuario determinado	
Detalle	Sistema	Todos
	Descripción del incidente	El área de Mercadotecnia ha solicitado que se le proporcione un informe de los últimos seis meses del comportamiento del usuario Ramón Azcuenaga, con el objetivo de poder presentarle una propuesta de nuevas inversiones.
Tipo de Sitio	Grande (calculado según el anexo B)	

Tabla 4.25. Caso de prueba Patrón de Comportamiento

Caso	03 – Patrón de comportamiento	
Descripción	El área de Mercadotecnia hace un pedido de búsqueda de patrones de comportamiento de usuarios entre los portales de Bolsa de Comercio de Buenos Aires y Bolsa de Cereales.	
Origen	Mercadotecnia	
Pedido	Detectar cual es el comportamiento de uso de muchos usuarios.	
Detalle	Sistema	Todos
	Descripción del incidente	El área de Mercadotecnia ha solicitado que se intente encontrar algún patrón de comportamiento entre los usuarios de ambos portales
Tipo de Sitio	Grande (calculado según el anexo B)	

Tabla 4.26. Caso de prueba Patrón de Comportamiento

Caso	04 – Tiempo de respuesta	
Descripción	El área de desarrollo pide una evaluación de tiempo de respuesta para una simulación de carga de 200 usuarios simultáneos.	
Origen	desarrollo	
Pedido	Prueba de carga del sistema.	
Detalle	Sistema	Compra y venta de dólares, euros y yenes
	Descripción del incidente	Se ha pedido que
Tipo de Sitio	Medio (calculado según el anexo B)	

Tabla 4.27. Caso de prueba Tiempo de Respuesta

Caso	05 – Búsqueda de Hipervínculos	
Descripción	El área de desarrollo ha pedido al equipo de soporte a la producción de realizar un análisis de los hipervínculos que posee el sitio tanto sea dentro del sitio como con otros sitios.	
Origen	Desarrollo	
Pedido	Detectar hipervínculos en todo el portal.	
Detalle	Sistema	Todos
	Descripción	El área de Desarrollo ha comprobado que el portal ha crecido en forma no ordenada y existen vínculos repetidos o que apuntan a lugares en donde no existen paginas
Tipo de Sitio	Grande (calculado según el anexo B)	

Tabla 4.28. Caso de prueba Búsqueda de Hipervínculo

Caso	06 – Búsqueda de Contenido	
Descripción	El área de Mercadotecnia desea comparar los contenidos del portal de compra y venta de divisas con el de otros sitios del mismo rubro.	
Origen	Mercadotecnia	
Pedido	Detectar contenidos.	
Detalle	Sistema	Sito de compra y venta de divisas
	Descripción	El área de Mercadotecnia desea comparar los contenidos del portal de compra y venta de divisas con el de otros sitios del mismo rubro, el área de mercadotecnia nos ha pasado un listado de sitios contra los cuales comparar.
Tipo de Sitio	Grande (calculado según el anexo B)	

Tabla 4.29. Caso de prueba Búsqueda de Contenido

Evaluación del Resultado de las Pruebas del Sistema

A continuación se detalla la plantilla para la evaluación del resultado de las pruebas del sistema (Tabla 4.30)

Evaluación de Resultados de las Pruebas del Sistema	
Caso	Resultado (Satisfactorio/Insatisfactorio)
00–Detección de Error	
Descripción	
Detalle de los pasos ejecutados	
Observaciones	
Caso	Resultado (Satisfactorio/Insatisfactorio)
01–Detección de Error	
Descripción	
Detalle de los pasos ejecutados	
Observaciones	
Caso	Resultado (Satisfactorio/Insatisfactorio)
02–Patrón de comportamiento	
Descripción	
Detalle de los pasos ejecutados	
Observaciones	

Caso	Resultado (Satisfactorio/Insatisfactorio)
03–Patrón de comportamiento	
Descripción	
Detalle de los pasos ejecutados	
Observaciones	
Caso	Resultado (Satisfactorio/Insatisfactorio)
04 – Tiempo de respuesta	
Descripción	
Detalle de los pasos ejecutados	
Observaciones	
Caso	Resultado (Satisfactorio/Insatisfactorio)
05–Búsqueda de Hipervínculos	
Descripción	
Detalle de los pasos ejecutados	
Observaciones	
Caso	Resultado (Satisfactorio/Insatisfactorio)
06 – Búsqueda de Contenido	
Descripción	
Detalle de los pasos ejecutados	
Observaciones	

Tabla 4.30. Evaluación de Resultados de las Pruebas del Sistema

ELABORACIÓN DE LOS MANUALES DEL USUARIO

Elaboración de los Manuales de Usuario

Los manuales de usuario han sido confeccionados sobre las pruebas de sistema como ejemplo integrador de la metodología de trabajo

DEFINICIÓN DE LA FORMACIÓN DE USUARIOS FINALES

Definición del Esquema de Formación

Se ha confeccionado un esquema de formación que se dividirá en dos etapas por un lado se formará al equipo de Soporte Técnico y otro para los equipos de Mesa de Ayuda y Mercadotecnia.

Especificación de los Recursos y Entornos de Formación

Al equipo de soporte a la producción realizara la capacitación sobre sus propias terminales de trabajo, su entrenamiento es puramente practico. En tanto al equipo de Mesa de Ayuda y Mercadotecnia se realizara en un aula preparada a tales fines, pues a estos equipos se los debe entrenar en la forma y que cosas pueden pedir al equipo de Soporte a la Producción.

4.2.6 IMPLANTACIÓN Y ACEPTACIÓN DEL SISTEMA

ESTABLECIMIENTO DEL PLAN DE IMPLANTACIÓN

Definición del Plan de Implantación

En función del requerimiento que la aplicación debe poder ser ejecutada en entornos Windows, Unix y Linux, la instalación de la aplicación es copiar un directorio dentro de la maquina de destino. La maquina de destino debe contar con la maquina virtual de Java (JVM) correctamente instalada y con la variable de entorno JAVA_HOME configurada, la versión de Java que se recomienda es 1.5 o superior. Por otra parte debe estar instalado Jython con la variable de entorno JYTHON_HOME configurada y ANT instalado con la variable de entorno ANT_HOME configurada.

FORMACIÓN NECESARIA PARA LA IMPLANTACIÓN

Preparación de la Formación del Equipo de Implantación

Se descargaran de Internet las versiones de JAVA, JYTHON y ANT y se almacenaran en un servidor de la red, el cual será de acceso público. Esto se hace para que las distintas maquinas en las cuales se configure el sistema tengan una versión unificada de las herramientas.

Formación del Equipo de Implantación

El equipo de Soporte Técnico es el encargado de realizar la instalación de los sistemas, por la sencillez del mismo no es necesario la realización de un entrenamiento, solo se le s problema un documento donde consta la necesidad de declarar las variables de entorno antes mencionadas.

EVALUACIÓN DEL RESULTADO DE LAS PRUEBAS DE IMPLANTACIÓN

Pruebas de aceptación del sistema

Realización de las Pruebas de Aceptación

A continuación se detallan los resultados de las pruebas de aceptación del sistema (Tablas 4.31 a 4.37). Los resultados señalados surgen después de iterar tres veces las pruebas.

Evaluación de Resultados de las Pruebas del Sistema	
Caso	Resultado (Satisfactorio/Insatisfactorio)
00–Detección de Error	Satisfactorio
Descripción	
Detección de análisis de comportamiento de un usuario que ha realizado una queja a Mesa de Ayuda. El usuario Aníbal Jerez, el día 22 de julio del 2005 ha intentado realizar una transacción bursátil de compra de acciones y le ha aparecido un mensaje de error en el sitio OBCBA	
Detalle de los pasos ejecutados	
Para la resolución de este caso se ha generado un proceso que consta de los siguientes pasos: Obtener el archivo de sucesos del día 22 de julio del 2005 Buscar todas las entradas del archivo de sucesos para el usuario Aníbal Jerez. Detectar específicamente las entradas al sitio OBCBA. Buscar en los archivos de sucesos de una semana anterior actividad del usuario. Presentar conclusiones.	
Observaciones	
En la fecha pedida se han encontrado entradas para el usuario Aníbal Jerez en el sitio OBCBA. No se han encontrado mensajes del error producidos por el sitio, si se ha detectado que la sesión del usuario se ha caído por tiempo. Se ha buscado al usuario en días anteriores y se ha podido comprobar que el usuario ha realizado transacciones satisfactoriamente los días 21 de julio del 2005 y 19 de julio del 2005. De lo antes analizado el grupo de Soporte a la Producción recomienda al grupo de mesa de ayuda: Consultar con el usuario si luego de este día el usuario ha podido realizar operaciones. ✓ Mesa de Ayuda nos confirma que no ha podido realizar ninguna operación en le sitio. Consultar con el usuario si entre las 15:30 hs del 21 de julio del 2005 y las 10:54 hs del 22 de julio del 2005, ha realizado alguna modificación en su PC o agregado algún software o modificado su conexión a Internet. Mesa de ayuda nos confirma que se ha instalado una mejora de seguridad en el navegados de Internet. La conclusión que el equipo de Soporte a la Producción ha llegado es que, una de las mejoras de seguridad que se le ha agregado al navegador del usuario bloquea por defecto las ventanas emergentes, esto hace que la transacción de compraventa no pueda confirmarse; el sitio antes de confirmar la transacción muestra una ventana emergente con los datos de la transacción a realizar. Se recomienda que Mesa de Ayuda solicite al usuario que no habilite esta opción por defecto del navegador, para poder realizar las transacciones. Por otra parte se ha generado una solicitud de modificación se software que se ha enviado al grupo de desarrollo, pues cada vez son mas comunes estos inhibidores de pantallas emergentes, para que la confirmación de la transacción de compraventa no sea una pantalla emergente sino una pantalla común dentro del navegador.	

Tabla 4.31. Evaluación de Resultados de las Pruebas del Sistema

Caso	Resultado (Satisfactorio/Insatisfactorio)
01–Detección de Error	Satisfactorio
Descripción	
Detección de análisis de comportamiento de un usuario que ha realizado una queja a Mesa de Ayuda. El usuario Juan Pérez, el día 05 de mayo del 2005 ha intentado acceder al sitio OBCBA, la hora aproximada fue las 15.30 hs y la conexión es muy lenta.	
Detalle de los pasos ejecutados	
Para la resolución de este caso se ha generado un proceso que consta de los siguientes pasos: 1. Obtener el archivo de sucesos del día 05 de mayo del 2005 2. Buscar todas las entradas del archivo de sucesos para el usuario Aníbal Jerez. 3. Detectar específicamente las entradas al sitio OBCBA. 4. Buscar en los archivos de sucesos de una semana anterior y posterior actividad del usuario. 5. Presentar conclusiones.	
Observaciones	
Se ha encontrado al usuario Juan Pérez realizando operaciones el día 5 de mayo del 2005, en días anteriores y posteriores. La única diferencia que se ha encontrado que las conexiones realizadas el día 05 de mayo del 2005, pertenecen a otro proveedor de Internet, se han seguido dos líneas de acción: Se ha recomendado a Mesa de Ayuda que pregunte al usuario si ha cambiado de proveedor de Internet o ha accedió desde otra PC. ✓ El usuario confirma que ha probado una nueva conexión a Internet el día 05 de mayo. El grupo de Soporte a la Producción, con una conexión dial-up (similar a la del usuario) del proveedor de Internet ha realizado una prueba y efectivamente la misma es muy lenta, se ha realizado un seguimiento de los saltos que se deben realizar en los distintos routers de Internet para llegar desde el proveedor de conexiones a Internet al sitio de la empresa y se ha detectado que la misma tiene muchos saltos, es decir debe realizar un viaje muy extenso para poder llegar al sitio. Se ha recomendado comunicarse con le proveedor de Internet y solicitarle en la medida de lo posible que actualiza sus tablas en los routers para que el acceso al sitio desde su servicio sea mas ágil. El proveedor ha accedido a realizar la modificación propuesta El grupo de Soporte a la Producción se ha comunicado con el grupo de Mesa de Ayuda para que se comunique con el usuario luego de las 48 hs, tiempo en que los cambios realizados por el proveedor de Internet se propagan por la red, para que se comunique con el usuario y que nuevamente realiza la prueba de conexión para validar los tiempos de repuesta.	

Tabla 4.32. Evaluación de Resultados de las Pruebas del Sistema

Caso	Resultado (Satisfactorio/Insatisfactorio)
02–Patrón de comportamiento	Satisfactorio
Descripción	
El área de Mercadotecnia ha solicitado que se le proporcione un informe de los últimos seis meses del comportamiento del usuario Ramón Azcuenaga, con el objetivo de poder presentarle una propuesta de nuevas inversiones.	
Detalle de los pasos ejecutados	
Para la resolución de este caso se ha generado un proceso que consta de los siguientes pasos: 1. Obtener el archivo de sucesos de los últimos seis meses, solo se encuentran en línea el ultimo mes; debe gestionarse el lugar para poder recuperar los resguardos. 2. Buscar todas las entradas del archivo de sucesos para el usuario Ramón Azcuenaga, y liberar el espacio gestionado anteriormente. Gestionar el acceso a los registros del usuario Ramón Azcuenaga de la base de datos corporativa, y espacio para crear tablas temporales. Tratar de encontrar un patrón de comportamiento, se ha sugerido utilizar algún modelo de red neuronal para categorización y algún algoritmo de clasificación para bases de datos, como ser C4.5 5. Presentar conclusiones.	
Observaciones	
En primer lugar se ha realizado la búsqueda en los archivos de sucesos del usuario Ramón Azcuenaga, para esto se ha reutilizado el proceso ejecutado en el Caso 01, so se ha cambiado el patrón de búsqueda. Luego de esto se ha liberado el espacio en disco. Con los registros extraídos de los archivos de suceso se ha procedido a realizar una normalización de los circuitos de uso del usuario. Estos comportamientos de uso normalizados han sido ingresados en una red neuronal, con una arquitectura de Kohonen, para realizar una clasificación. La grafica de la categorización ha dado como resultado una figura que todos sus punto pueden ser incluidos en una elipse, donde sus centros forman un ángulo de aproximadamente 25 grados con respecto al eje de las X. En el grafico se puede observar una gran concentración en el centro de la elipse izquierdo. Para tratar de entender esta categorización se ha accedido a la ficha del usuario que se encuentra en la base corporativa y consultar cuales operaciones tiene permitida el usuario; se ha observado que esta habilitado para realizar operaciones bursátiles en la el sitio que opera con la bolsa de comercio. Con esta información se intenta detectar una regla de comportamiento; se realiza una nueva normalización pero solo seleccionando los registros que en la grafica de la categorización tienen con respecto a cada un de los centros una distancia de Euclides menor a 1, la selección del valor ha sido aleatoria en función de la grafica, con los registros seleccionados se realiza un proceso que normalice los datos para ser cargados en una base de datos, los cargue y ejecute el algoritmo C4.5 sobre los mismos y devuelvan los resultados. Luego de esto se reconocen dos reglas principales el usuario realiza operaciones bursátiles en la bolsa de comercio, lo que se confirma con su fichero en la base de datos y que asiduamente esta accediendo al sitio de la bolsa de cereales. Esto nos sugiere que el usuario Ramón Azcuenaga, esta interesado en realizar operaciones en la bolsa de cereales, la recomendación que se le hace al Mercadotecnia es que incluya en el paquete que ofrezca al usuario asesoramiento para inversiones en la bolsa de cereales.	

Tabla 4.33. Evaluación de Resultados de las Pruebas del Sistema

Caso	Resultado (Satisfactorio/Insatisfactorio)
03–Patrón de comportamiento	Satisfactorio
Descripción	
El área de Mercadotecnia hace un pedido de búsqueda de patrones de comportamiento de usuarios entre los portales de Bolsa de Comercio de Buenos Aires y Bolsa de Cereales	
Detalle de los pasos ejecutados	
Para la resolución de este caso se ha generado un proceso que consta de los siguientes pasos: 1. Obtener el archivo de sucesos de los últimos seis meses, solo se encuentran en línea el ultimo mes; debe gestionarse el lugar para poder recuperar los resguardos. 2. Identificar a los usuarios y generar archivos con sus transacciones. Gestionar el acceso a los registros de los usuarios. Tratar de encontrar un patrón de comportamiento, se ha sugerido utilizar algún modelo de red 5. Presentar conclusiones.	
Observaciones	
Como primer paso en el proceso de detección de comportamiento se ha generado un archivo por usuario con la actividad de los mismos en los últimos 6 meses de los usuarios que están habilitados para realizar operaciones en la Bolsa de Comercio y la Bolsa de Cereales. Luego de esto se normalizan los datos para que puedan ser ingresados a una red neuronal para ser categorizados. El resultado obtenido es poco claro y no se pueden diferenciar categorías. Se realiza otro proceso de normalización y se incluyen los datos en una serie de tiempo del tipo ARIMA, con lo cual se puede observar una tendencia en el comportamiento, luego de esto se superponen ambas graficas y se puede observar que los usuarios realizan movimientos con cierta periodicidad entre ambos sistemas bursátiles. Es entregada esta grafica al grupo de Mercadotecnia para que junto con el equipo de Finanzas puedan analizar el motivo de estas variaciones entre ambos sistemas.	

Tabla 4.34. Evaluación de Resultados de las Pruebas del Sistema

Caso	Resultado (Satisfactorio/Insatisfactorio)
04 – Tiempo de respuesta	Satisfactorio
Descripción	
El área de desarrollo pide una evaluación de tiempo de respuesta para una simulación de carga de 200 usuarios simultáneos. Se ha pedido que se evalué el rendimiento del sitio	
Detalle de los pasos ejecutados	
Para la resolución de este caso se ha generado un proceso que consta de los siguientes pasos: 1. Obtener 200 usuarios de prueba con sus contraseñas para poder realizar la simulación 2. Definir los circuitos de prueba. Realizar la simulación. 4. Presentar conclusiones.	
Observaciones	
Con los usuarios ya generados para la prueba se generan los programas con los circuitos simulados, para la misma se a agregado al proceso un simulador de carga de usuarios que su función es poder generar desde una terminal de trabajo transacciones que pertenecen a distintos usuarios virtuales. Se ejecuta el proceso y se ha analizado los tiempos de respuesta de los mismos. El tiempo promedio de respuesta s es de aproximadamente 3.5 segundos, lo que se considera aceptable.	

Tabla 4.35. Evaluación de Resultados de las Pruebas del Sistema

Caso	Resultado (Satisfactorio/Insatisfactorio)
05–Búsqueda de Hipervínculos	Satisfactorio
Descripción	
El área de desarrollo ha pedido al equipo de soporte a la producción de realizar un análisis de los hipervínculos que posee el sitio tanto sea dentro del sitio como con otros sitios.	
Detalle de los pasos ejecutados	
Para la resolución de este caso se ha generado un proceso que consta de los siguientes pasos: 1. Tener acceso a los directorios de producción del sitio. 2. Analizar los fuentes. 3. Presentar conclusiones.	
Observaciones	
El primer paso es el recorrido de todos los fuentes del sistema en el formato HTML y JSP, para encontrar todos los hipervínculos, con los hipervínculos ya detectados se reutiliza el simulador de carga de usuarios y se verifican los hipervínculos. Se generan dos listas una con los hipervínculos que han respondido satisfactoriamente y otra con los que no han respondido y se los entrega al equipo de Desarrollo.	

Tabla 4.36. Evaluación de Resultados de las Pruebas del Sistema

Caso	Resultado (Satisfactorio/Insatisfactorio)
06 – Búsqueda de Contenido	Satisfactorio
Descripción	
El área de Mercadotecnia desea comparar los contenidos del portal de compra y venta de divisas con el de otros sitios del mismo rubro, el área de mercadotecnia nos ha pasado un listado de sitios contra los cuales comparar.	
Detalle de los pasos ejecutados	
Para la resolución de este caso se ha generado un proceso que consta de los siguientes pasos: 1. Obtener la lista de sitios ha analizar. Analizar los sitios. Realizar la comparación Presentar conclusiones.	
Observaciones	
El primer paso del proceso de obtener las paginas del sitio, como todo sitio de transacciones bursátiles tiene áreas restringidas solo se realizará la comparación con la pagina de bienvenida del sitio. Se evalúa cantidad de hipervínculos, gráficos y textos de la misma, con la información recabada se genera un cuadro comparativo y se lo entrega a Mercadotecnia.	

Tabla 4.37. Evaluación de Resultados de las Pruebas del Sistema

Evaluación del Resultado de las Pruebas de Aceptación

Las pruebas han sido satisfactorias y se presenta en la tabla 4.38 el resumen del estado de las mismas.

Caso	Fecha	Responsable	Resultado
00–Detección de Error	19/10/2005	Hernan Merlino	Satisfactorio
01–Detección de Error	19/10/2005	Hernan Merlino	Satisfactorio
02–Patrón de comportamiento	20/10/2005	Hernan Merlino	Satisfactorio
03–Patrón de comportamiento	20/10/2005	Hernan Merlino	Satisfactorio
04 – Tiempo de respuesta	21/10/2005	Hernan Merlino	Satisfactorio
05–Búsqueda de Hipervínculos	21/10/2005	Hernan Merlino	Satisfactorio
06 – Búsqueda de Contenido	22/10/2005	Hernan Merlino	Satisfactorio

Tabla 4.38. Resumen de las pruebas.

5. EXPERIMENTACION POR CASOS

En este capítulo se presenta un caso de experimentación del sistema desarrollado: se formula el análisis de los archivos de sucesos de un sitio de Internet (sección 5.1), se analiza la estructura de los hipervínculos que existen con referencia a un sitio de Internet (sección 5.2) y se categorizan hipervínculos que existen con referencia a un sitio de Internet (sección 5.3).

5.1. DESCRIPCIÓN DE LOS CASOS

Caso 1: Se realizará el análisis de los archivos de sucesos de un sitio de Internet. El proceso debe poder ser repetido bajo pedido.

Caso 2: Se realizará un proceso para analizar la estructura de los hipervínculos que existen con referencia a un sitio de Internet.

Caso 3: Con la salida del caso anterior (caso 2), se tomarán 30 hipervínculos y se categorizar su contenido

5.2. CASO 1

Para poder llevar a cabo el siguiente caso se deben realizar las siguientes tareas:

Tener acceso al archivo de sucesos del servidor Web.
Copiar el mismo a un directorio temporal.
Aplicar las transformaciones necesarias.
Aplicar la técnica seleccionada de exploración de uso.
Presentar el resultado obtenido.

Para poder llevar a cabo estas tareas se debe escribir el control de flujo de la información, para esto se ha elegido el formato XML y la herramienta para que interprete este archivo es el ANT, el cual es un software de código libre que se utiliza para la automatización de tareas, entre otras funcionalidades. A continuación se detallan los pasos que se incluyen en el archivo XML para la configuración de la tarea. [a] Copiar el archivo de suceso al directorio de trabajo. [b] Aplicar las transformaciones necesarias, en este caso se debe validar que el ultimo registro del

archivo de sucesos se halla copiado completamente, esto puede no ser así pues se realiza una copia de un archivo en uso. Si no se ha copiado completamente se descarta el ultimo registro. [c] Se aplica la técnica de minería de uso, en este caso se ha elegido Webalizer, el cual es un producto de fuentes libres que produce estadísticas de los accesos a un sitio Web. [d] Todos los archivos intermedios generados en la presente ejecución son compactados y movidos a un archivo de resguardo y [e] se inserta un registro con la fecha y hora de la ejecución y el estado de la misma.

En la figura 5.1 se detalla en forma abreviada el archivo XML con la configuración y los pasos antes detallados

```
<project name="Exploracion de uso" default="ejecutar" basedir=".">

    <tstamp>
        <format property = "buildtimefile" pattern = "yyyyMMdd_HHmmss"/>
    </tstamp>

    <target name="obtenerArchivoSuceso"
          description="Copia el archivo de sucesos al directorio de trabajo">
        . . .
    </target>

    <target name="aplicarTransformaciones"
          description="Copia el archivo de sucesos al directorio de trabajo">
        . . .
    </target>

    <target name="limpiar"
          description="Borra los datos de la ejecucion">
        . . .
    </target>

    <target name="compactarDatos"
          description="Compacta los datos del directorio de trabajo">
        . . .
    </target>

    <target name="resguardarDatos" depends="compactarDatos, limpiar"
          description="Resguarda los datos y limpia el directorio de trabajo">
    </target>


     <target name="cargarBaseDatos"
           description="Inserta un registro con el detalle de la ejecucion">
         . . .
     </target>

    <target name="ejecutar" depends="obtenerArchivoSuceso,
                                     aplicarTransformaciones,
                                     aplicarTecnica,
                                     mostarResultados,
                                     resguardarDatos,
                                     cargarBaseDatos"
            description="Ejecuta el caso 1 de la experimentacion"/>

</project>
```

Figura 5.1. Forma abreviada el archivo XML

En la figura 5.2 se muestra la estructura de directorios del área de trabajo

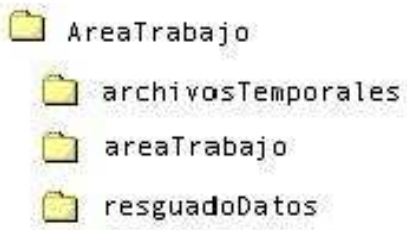

Figura 5.2 estructura de directorios del área de trabajo

En la figura 5.3 se muestra la forma de invocar al proceso

Figura 5.3 forma de invocar al proceso

En la figura 5.4 se muestra el resultado de la ejecución de Webalizer, como herramienta de exploración de uso.

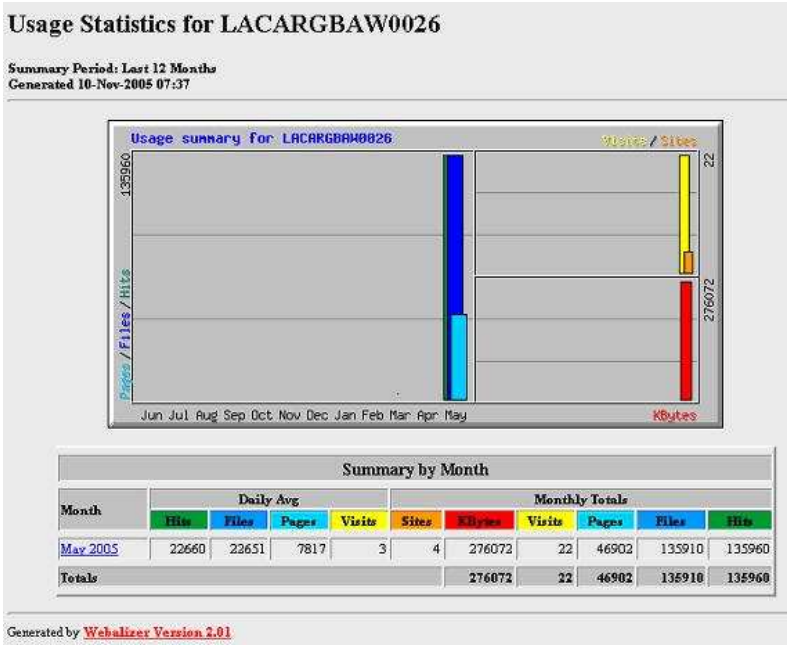

Figura 5.4 resultado de la ejecución de Webalizer

En la figura 5.5 se muestra la entrada insertada en la base de datos con todos los procesos ejecutados

Figura 5.5 entrada insertada en la base de datos

5.3. CASO 2

Para poder llevar a cabo el siguiente caso se deben realizar las siguientes tareas:

Analizar el sitio .
Obtener las referencias al mismo.
Presentar el resultado obtenido.

Para poder llevar a cabo estas tareas se debe escribir el control de flujo de la información, para esto se ha elegido el formato XML y la herramienta para que interprete este archivo es el ANT, el cual es un software de código libre que se utiliza para la automatización de tareas, entre otras funcionalidades. A continuación se detallan los pasos que se incluyen en el archivo XML para la configuración de la tarea. [a] Obtener las referencias al sitio el sitio que se ha seleccionado a analizar el ObjectsDevelopment.com. [b] Analizar las referencias, para esto se utilizara una herramienta en línea para la exploración de estructura esta se encuentra en http://www.informationcrawler.com/. Para poder tomar esta información en línea en el proceso se utilizara el producto de fuentes libre Grinder luego de esto [c] se analizan las referencias desde los diversos motores de búsqueda en Internet y [d] presentar los resultados

En la figura 5.6 se detalla en forma abreviada el archivo XML con la configuración y los pasos antes detallados

```
<project name="Exploracion de uso" default="ejecutar" basedir=".">

    <tstamp>
        <format property = "buildtimefile" pattern = "yyyyMMdd_HHmmss"/>
    </tstamp>

    <target name="obtenerReferencias"
          description="Se ejecuta Grinder para obtener el resultado de la busqueda">
        . . .
    </target>

    <target name="analizarResultadoHTML"
          description="La respuesta obtenida del sitio se analiza">
        . . .
    </target>

    <target name="limpiar"
          description="Borra los datos de la ejecucion">
        . . .
    </target>

    <target name="compactarDatos"
          description="Compacta los datos del directorio de trabajo">
        . . .
    </target>

    <target name="resguardarDatos" depends="compactarDatos, limpiar"
          description="Resguarda los datos y limpia el directorio de trabajo">
    </target>


     <target name="cargarBaseDatos"
           description="Inserta un registro con el detalle de la ejecucion">
         . . .
     </target>

    <target name="ejecutar" depends="obtenerReferencias,
                                     analizarResultadoHTML,
                                     mostarResultados,
                                     resguardarDatos,
                                     cargarBaseDatos"
            description="Ejecuta el caso 2 de la experimentacion"/>

</project>
```

Figura 5.6 forma abreviada el archivo XML con la configuración y los pasos antes detallados

En la figura 5.7 se muestra la estructura de directorios del área de trabajo

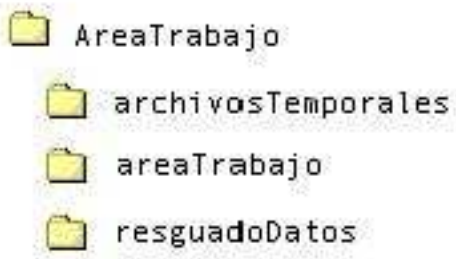

Figura 5.7 estructura de directorios del área de trabajo

En la figura 5.8 se muestra la forma de invocar al proceso

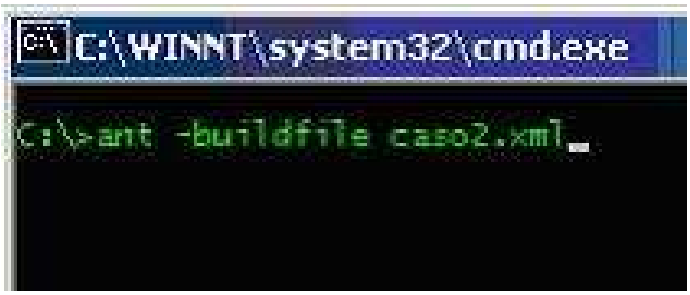

Figura 5.8 la forma de invocar al proceso

En la figura 5.9 se muestra el resultado de la ejecución del sitio Information Crawler, como herramienta de exploración de estructura.

Figura 5.9 resultado de la ejecución del sitio Information Crawler

En la figura 5.10 se observa el detalle obtenido por la consulta hecha al sitio.

Figura 5.10 detalle obtenido por la consulta hecha al sitio Information Crawler

En la figura 5.11 se detalla la salida entregada por el proceso.

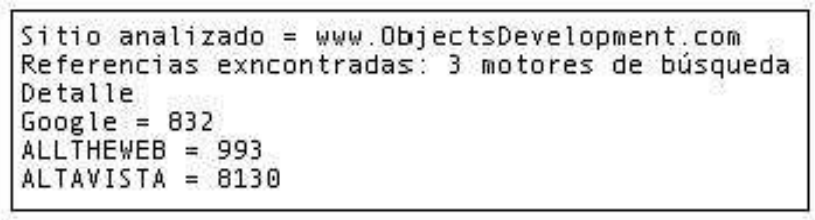

Figura 5.11 salida entregada por el proceso

5.3. CASO 3

Para poder llevar a cabo el siguiente caso los paso a llevar a cabo son:

Extraer las URL de los buscadores
Acceder a las URL
Categorizar las paginas
Mostrar el resultado

Se generara nuevamente un proceso que será regido por ANT, en el cual: [a] se extraerán las 30 primeras del motor de búsqueda Google, [b] con la lista de las URL se accederá a cada una de ellas y se salvará localmente cada una de las paginas, [c] luego de esto se examinarán las paginas y se modificarán en función de un conjunto de ontologías, [d] con estas paginas estandarizadas se calculará la cantidad de referencias de las mismas , [e] el resultado será exportado a una planilla de calculo para graficar el resultado.

En la figura 5.12 se detalla en forma abreviada el archivo XML con la configuración y los pasos antes detallados

```
<project name="Exploracion de uso" default="ejecutar" basedir=".">

    <tstamp>
        <format property = "buildtimefile" pattern = "yyyyMMdd_HHmmss"/>
    </tstamp>
    <target name="obtenerURL"
          description="Se ejecuta con un proceso Grinder para obtener las URL">
        . . .
    </target>
    <target name="analizarResultadoHTML"
          description="Compara contra las ontologias cada pagina">
        . . .
    </target>
    <target name="aplicarEstadistica"
          description="Se hace un recuento de las paginas con sus ontologías">
        . . .
    </target>
    <target name="exportarResultados"
          description="Exporta los resultados a un formato CSV">
        . . .
    </target>
    <target name="limpiar"
          description="Borra los datos de la ejecucion">
        . . .
    </target>
    <target name="compactarDatos"
          description="Compacta los datos del directorio de trabajo">
        . . .
    </target>
    <target name="resguardarDatos" depends="compactarDatos, limpiar"
          description="Resguarda los datos y limpia el directorio de trabajo">
    </target>
     <target name="cargarBaseDatos"
           description="Inserta un registro con el detalle de la ejecucion">
         . . .
     </target>
    <target name="ejecutar" depends="obtenerURL,
                                     analizarResultadoHTML,
                                     aplicarEstadistica,
                                     exportarResultados,
                                     resguardarDatos,
                                     cargarBaseDatos"
            description="Ejecuta el caso 3 de la experimentacion"/>

</project>
```

Figura 5.12. forma abreviada el archivo XML con la configuración y los pasos antes detallados

En la figura 5.13 se muestra la estructura de directorios del área de trabajo

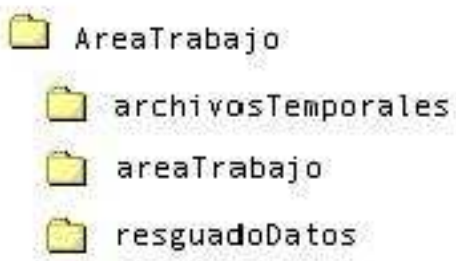

Figura 5.13 estructura de directorios del área de trabajo

En la figura 5.14 se muestra la forma de invocar al proceso

Figura 5.14 forma de invocar al proceso

En la figura 5.14 se presenta el archivo formateado con el resultado del análisis de las ontologías, los valores representan la cantidad de veces que cada una de las ontologías descriptas fueron encontradas

	referencia	cantidad
2	software	57
3	programacion	34
4	usuarios	20

Figura 5.15 archivo formateado con el resultado del análisis de las ontologías

El archivo ha sido generado en formato de texto con separaciones entre campos por el carácter TAB, y con el final de registro con los caracteres CRLF. La primera fila tiene el encabezado de las columnas.

6. CONCLUSIÓN

6.1. APORTES DE LA TESIS

La existencia de sistemas informáticos de uso libre orientados a la exploración de uso, la exploración de contenido y la exploración de estructura en web y la identificación de procesos de exploración en web que requieren la integración articulada de dichos artefactos son la motivación de esta tesis.

En este contexto, en este trabajo se logró desarrollar una herramienta para exploración de datos Web que permite estructurar todo el proceso de exploración.

La mayor ventaja de esta herramienta es poder utilizar diversas técnicas de exploración, además de permitir la reutilización de procesos ya ejecutados con anterioridad y la combinación de los mismos para su posterior comparación; todo esto llevado a cabo sin un alto grado de complejidad.

La herramienta desarrollada satisface los siguientes requerimientos:

Los procesos que ejecuta son modulares y flexibles.
La información que entrega es verificable y validable.
El sistema tiene la capacidad para agendar tareas.
El sistema puede ser ejecutado en entornos Windows, Unix y Linux.
El sistema puede admite varias fuentes de dato de entradas.
El sistema puede admite formato flexible de archivos de salida.

6.2. FUTURAS LÍNEAS DE INVESTIGACIÓN

Se han identificado las siguientes líneas de trabajo:

Explorar la integración de este marco de trabajo con los sistemas del tipo CRM, para que estos últimos se enriquezcan con la información que se genere del marco de trabajo para sitios de minería Web.

Explorar la utilización de aprendizaje automático para la generación de teorías que puedan determinar la utilización de alguna de las técnicas conocidas de exploración Web. El objetivo de esto es intentar que la herramienta desarrollada deje de tener un modo de operar reactivo ante las solicitudes de los usuarios y pase a generar información en forma automática, reconociendo patrones de comportamiento y enviando esta información en forma automática al departamento correspondiente, como por ejemplo: envio al departamento de mercadotecnia una lista de potenciales situaciones de conflicto como fraudes y estados de saturación del sistema Web.

7. REFERENCIAS

Abrahan, A. y Ramos, V. 2004. *Web Usage Mining Using artificial Ant colony clustering and lineal genetic programming.* http://alfa.ist.utl.pt/ ~cvrm/staff/vramos/Vramos-CEC03b.pdf.

AlterWind Software. 2005. *AlterWind Log Analyzer Lite*. http://www.alterwind.com/ loganalyzer/log-analyzer-lite.html. Página web vigente al 16/11/05

Aston, P. y Fitzgerald, C. 2005. *The Grinder, a Java Load Testing Framework*. http://grinder.sourceforge.net/. Página web vigente al 16/11/05

Backman, D. and Rubbin, J. 1997. *Web log analysis: finding a Recipe for Success*, http://techweb.comp.com/nc/811/811cn2.html. Página web vigente al 16/11/05

Barret, B. 1999. *The Webalizer: What is your web server doing today?*. http://www. mrunix.net/webalizer/. Página web vigente al 16/11/05

Berners-Lee, T., Hendler , J. and Lassila , O. 2001 *The Semantic Web.* Scientific American.

Borges, J. y Levene, M. A 2000. *Fine Grained Heuristic to Capture Web Navigation Patterns*. Department of Computer science, university college London.

Broder ,A., Kumar, R., Maghoul, F., Raghavan , P., Rajagopalan , S., Stata ,R., Tomkins , A., and Wiener, J.. 2000. *Graph structure in the Web*. Computer Networks, June.

Chau, M., Zeng, D., Chen, H., Haung, M. y Hendriawan, D. 2003. *Design and evaluation of multi-agent collaborative Web mining system*. Department of Management information Systems, Eller college of Business and Public Administration. University of Arizona.

Cooley, R. 2000. *The Importance of Understanding Web Site Structure and content when Performing Web Usage Mining*. Groupfire, Inc. Redwood City, CA.

Cooley, R., Mobasher, B. and Srivsata. J. 1997. *Web Mining: Information and Pattern Discovery on the word Wide Web*. Technical Report TR 97-027, University of Minnesota, Dep of Computer Science, Minneapolis.

Google. 2005. *Transmisión del PageRank*. http://google.dirson.com/transmision-pagerank.php. Página vifgente al 16-11-05

Doan, J. Madhavan, R. Dhamankar, P. Domingos, and A. Y. Halevy. 2003. *Learning to match ontologies*. VLDB Journal, 12(4):303--319. Special Issue on the Semantic Web

Etzioni, O. 1996. *The world wide Web: Quagmire or gold mine*. Communications of the ACM, 39(11):65-68.

Fensel, Dieter, McGinness, Deborah, Shullten Hellen, Keong Ng, Lee, Lim, Ee-Peng, and Yan guanghao. 2001. *Ontologies and Electronic Commerce*. IEEE Intelligent Systems 16 (1):8-14.

Furnkranz, J. 2002. *Web Structure Mining Exploiting the Graph Sturture or the World-Wide Web.* Austrian Research Institute for Artificial Inteligence Schottengasse.

Huang, Z., Ng, J. y Cheung, D. 2001. *A Cube Model for Web Access sessions and cluster analysis*. Department of Mathematics. University of Hong Kong.

Hunton, James E., Bryant, Stephanie M., Bagranoff, Nancy A.. 2003 *Core Concepts of Information Technology Auditing.* Wiley.

Jidonwang, H. Chen, J., Tao, L., Ma, W. y Wenyin, L. 2002. *Ranking User's relevance to a topic thought Link Analysis on Web Logs*. Department of Computer science, City Univ of Hong Kong.

Jin, X., Zhou, Y. y Mobasher, B. 2004. *Web Usage Based on probabilistic Latent Semantic Analysis*. Center of Web Intelligence. School of computer Science, Telecomunications and Information systems DePaul University, Chicago, Illinois

JBNC, 2005. *Bayesian Network Classifier Toolbox*. http://jbnc.sourceforge.net/. Página web vigente al 16/11/05

Jung, J. y Jo, G. 2003 *Semantic Outlier Analysis for Sessionizing Web Log*. Intelligent E-Commerce System Laboratory, School of Computer engineering, Inha University, Corea.

Kitsuregawa, M., Toyoda, M. y Pramudiono, I. 2002. *WEB community mining and WEB log mining; Commodity Cluster based Execution*. Institute of Industrial Science. University of Tokio

Kohonen, T., Hynninen, J., Kangas, J. and Laaksonen, J. 1996. *SOM_PAK: The Self-Organizing Map Program Package*. Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, FIN-02150 Espoo, Finlandia.

Kosala, R. y Blockeel. H. 2000. *Web mining Research: A Survey*. SIGKDD: SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining, ACM. volume = 2. publisher ACM. Year 2000

Krishnapuran, R. y Joshi, A. 2000. *Extracting Web User profiles using relational competitive fuzzy clustering*. Department of electrical and Computer engineering University of Memphis.

Kushmerick N. 1997. *Wrapper induction for Information Extraction*. PhD Tesis, Department of computer Science, University of Washington.

Lin, S., Shis, C., Chen, M. 1998. *Extraction Classification of Internet Documents with Mining Term associations: A Semantic Approach*. In Processing of 21st Annual International ACM SIGIR conference on research and Development in Information Retrieval.

Madria, S., Bhowmich, S. Y Lim, E. 1999. *Research Issue in Web Data Mining*. Center for advanced Information systems, School of Applied Science Nanyang Technology University, Singapore.

Mendez-Torreblanca, A., Montes, M. y Lopez-lopez, 2002. A. *A trend discovery System for Dynamic Web content Mining*. Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México.

Mobaster, B. 2004. *Web Usage and Personalization*. CRC Press.

Maedche , Alexander, Neumann , Günter, Staab , Steffen. 2002 *Bootstrapping an Ontology-based Information Extraction System. OntologyBased Information Extraction System*. Studies in Fuzziness and Soft Computing, editor J. Kacprzyk.

Maedche, A., Pekar, V., and Staab 2004b. *Ontology Learning Part One On Discovering Taxonomic Relations from the Web. Ontology learning part on –on discovering taxonomic relations from the web*. In Web Intelligence. Springer

Pal, S., Talwar, V., Mitra, P. 2002. *Web Mining in Soft Computing Framework: Relevance, State of the art and future Directions.* IEEE Transactions on Neural Networks.

Peterson, E.. 2003. *Web Site Measurement Hacks*. Editado por O'Reilly. San Francisco.

Peralta, M.. 2003. *Asistente para la Evaluación de CMMI-SW*. http://www.itba.edu.ar/capis/webcapis/planma-esp.html

Pitkow. J. 1997. *In Search of Reliable Usage Data on the WWW*. In Proceedings of the 6 International Word Wide Web Conference, Santa Clara, California, April.

Senft, Sandra, Daniel P., Ph.D. Manson, Gonzales, Carol, Gallegos, Frederick. 2004 *Information Technology Control and Audit*, Second Edition Auerbach Publications; 2nd edition

Siglitos, G., Paliouras, G., Spyropoulus, C. y Hatzopoulos, M. 2003. NCSR "*Demokritos*", Institute of Information and Telecomunications, Athens, Grecia. http://www.demokritos.gr/index_muk.asp

Sherman, Chris, Price, Gary. 2001. *The Invisible Web: Uncovering Information Sources Search Engines Can't See* Cyberage Books; 1st edition

Spertus, E. 1997. *ParaSite: Mining Structural Information on the Web*. In Proceedings of 6th International WWW Conference, April.

Sterne, Jim. 2002 *Web Metrics: Proven Methods for Measuring Web Site Success.* Wiley; 1st edition

Thesoftwareobjects, 2005. *Information CRAWLER*. http://www.informationcrawler. Página web vigente al 16/11/05

Turner, S. 2005. *Analog. The most popular logfile analyser in the world*. http://www.analog.cx/. Página web vigente al 16/11/05

Wang, Yan. 2000. *Web Mining and Knowledge Discovery of Usage Patterns.* CS 748T Project

Winkler, K. and Spiliopoulou, M. 2002. *Employing Text Mining for Semantic Tagging in DIAsDEM*. In KI Künstliche Intelligenz, Themenheft Text Mining 16(2):27-29.

Witten, I. H. and Frank, E. 2005. *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco, CA.

Xue, G., Zeng, H., Chen, Z., Ma, M. y Lu, C. 2002. *Log Mining to improve the performance of Site Search*. Computer Science and Engineering shanghai Jiao-tong University, Shangai, China.

Yang, H., Prthasarathy, S. y Reddy, S. 2002. *On the use of constrained associations for Web log mining*. Computer and Information science Ohio State University.

Zaiane, O., Han, J., .Li, Z., Chee, S. y Chaing, J 1998. *Multi Media Miner: A System Prototype for Multimedia Data Mining*. *ACM-SIGMOD Conf. on Management of Data*

Zhu, T. y Greiner, R. 2004. *Predicting Web Information Content*. Dept. pf Computing Science, University of Alberta, Canada.

Zhu, T., Greiner, R. y Haubl, G. 2003. *Learning a Model of a Web User's Interest*. Dept. of Computer Science, University of Alberta, Canada

ANEXO A: EDUCCIÓN DE REQUERIMIENTOS

El presente anexo detalla todo el proceso de educción de requerimientos realizado para el sistema de software en cuestión, se detallan las minutas de las entrevistas abiertas, cerradas, Brainstorming y sesiones JAD.

A.1. DETALLE DEL PRIMER CICLO DE ENTREVISTAS ABIERTAS:

Primer entrevista abierta (Jefe de Soporte Técnico)

Preparación de la entrevista:

En función de ser la primera reunión con el patrocinador del sistema, se ha decidido realizar una entrevista del tipo abierta.

Participantes:

Patrocinador del proyecto, jefe de Soporte a la Producción.

Materiales a Proporcionar:

Breve introducción al objetivo de la reunión.

Objetivos de la Reunión:

Detectar objetivos del sistema

Usuarios potenciales

Detalle del material a entregar

El objetivo de la presente reunión es detectar en términos generales que es lo que se espera del sistema, cual seria la información que debería proporcionar.

Por otra parte se espera detectar los potenciales usuarios del sistema para poder interactuar con ellos.

Resumen de la entrevista

De la entrevista realizada con el Jefe de Soporte a la Producción, se han extraído las siguientes conclusiones:

Características del sistema:

El principal objetivo del sistema es reducir el tiempo de respuesta de equipo de Soporte a la Producción.

Proveer información a varias áreas de la empresa.

Crear una herramienta que sirva de base para la gestión del conocimiento del área de Soporte a la Producción.

Usuarios potenciales del sistema:

Área de Soporte a la Producción.

Área de Mesa de Ayuda.

Área de Desarrollo de Sistemas.

Área de Mercadotecnia.

Primer entrevista abierta (Jefe de Mesa de Ayuda)

Preparación de la entrevista:

De la entrevista realizada al Jefe de soporte Técnico se ha reconocido como un potencial usuario del sistema el área de Mesa de Ayuda.

Se define una reunión con el jefe de esta área para definir el posible alcance del proyecto.

Participantes:

Usuario potencial, jefe de Mesa de Ayuda.

Materiales a Proporcionar:

Introducción con el objetivo del sistema.

Objetivos de la Reunión:

Salidas del sistema

Detalle del material a entregar

En función de la entrevista realizada con el Jefe de Soporte a la Producción, y su deseo de realizar un sistema para mejorar el tiempo de respuesta del área de Soporte a la Producción se ha mencionado que el área de Mesa de Ayuda es uno de los usuarios

del área de soporte a la Producción, es por esta razón que se desea mantener una entrevista para definir nivel de servicios y calidad de la información que se espera recibir.

Resumen de la entrevista

De la entrevista realizada con el Jefe de Mesa de Ayuda, se han extraído las siguientes conclusiones:

Se ha confirmado que el principal proveedor de información del área de Mesa de Ayuda es Soporte a la Producción.

La generación de información en forma rápida y certera, es de gran importancia para el buen funcionamiento de la Mesa de Ayuda.

Primer entrevista abierta (Jefe de Mercadotecnia)

Preparación de la entrevista:

De la entrevista realizada al Jefe de soporte Técnico se ha reconocido como un potencial usuario del sistema el área de Mercadotecnia.

Se define una reunión con el jefe de esta área para definir el posible alcance del proyecto.

Participantes:

Usuario potencial, jefe de Mercadotecnia.

Materiales a Proporcionar:

Introducción con el objetivo del sistema.

Objetivos de la Reunión:

Salidas del sistema

Detalle del material a entregar

En función de la entrevista realizada con el Jefe de Soporte a la Producción, y su deseo de realizar un sistema para mejorar el tiempo de respuesta del área de Soporte a la Producción se ha mencionado que una potencial área que podría valerse de la información que genere el sistema es Mercadotecnia, es por esta razón que se desea

mantener una entrevista para definir nivel de servicios y calidad de la información que se espera recibir.

Resumen de la entrevista

De la entrevista realizada con el Jefe de Mercadotecnia, se han extraído las siguientes conclusiones:

Se ha explicado al jefe de Mercadotecnia en términos generales las bondades del proyecto.

Se ha concluido que un sistema que genere información sobre uso de los distintos sitios Web de la empresa seria de gran ayuda para el seguimiento de las campañas publicitarias..

Primer entrevista abierta (Jefe de Mesa de Ayuda)

Preparación de la entrevista:

De la entrevista realizada al Jefe de soporte Técnico se ha reconocido como un potencial usuario del sistema el área de Meja de Ayuda.

Se define una reunión con el jefe de esta área para definir el posible alcance del proyecto.

Participantes:

Usuario potencial, jefe de Mesa de Ayuda.

Materiales a Proporcionar:

Introducción con el objetivo del sistema.

Objetivos de la Reunión:

Salidas del sistema

Detalle del material a entregar

En función de la entrevista realizada con el Jefe de Soporte a la Producción, y su deseo de realizar un sistema para mejorar el tiempo de respuesta del área de Soporte a la Producción se ha mencionado que el área de Desarrollo de Sistemas es uno de los usuarios del área de Soporte a la Producción, es por esta razón que se desea mantener

una entrevista para definir nivel de servicios y calidad de la información que se espera recibir.

Resumen de la entrevista

De la entrevista realizada con el Jefe de Desarrollo de Sistemas, se han extraído las siguientes conclusiones:

Se ha confirmado que ocasionalmente el área de Soporte a la Producción es proveedor de información del área de Desarrollo de Sistemas.

La generación de información certera, es de gran importancia para Desarrollo de Sistema.

Primer entrevista abierta (Jefe de Mesa de Ayuda)

Preparación de la entrevista:

De la entrevista realizada al Jefe de soporte Técnico se ha reconocido como un potencial usuario del sistema el área de Meja de Ayuda.

Se define una reunión con el jefe de esta área para definir el posible alcance del proyecto.

Participantes:

Usuario potencial, jefe de Mesa de Ayuda.

Materiales a Proporcionar:

Introducción con el objetivo del sistema.

Objetivos de la Reunión:

Salidas del sistema

Detalle del material a entregar

En función de la entrevista realizada con el Jefe de Soporte a la Producción, y su deseo de realizar un sistema para mejorar el tiempo de respuesta del área de Soporte a la Producción se ha mencionado que el área de Desarrollo de Sistemas es uno de los usuarios del área de Soporte a la Producción, es por esta razón que se desea mantener

una entrevista para definir nivel de servicios y calidad de la información que se espera recibir.

Resumen de la entrevista

De la entrevista realizada con el Jefe de Desarrollo de Sistemas, se han extraído las siguientes conclusiones:

Se ha confirmado que ocasionalmente el área de Soporte a la Producción es proveedor de información del área de Desarrollo de Sistemas.

La generación de información certera, es de gran importancia para Desarrollo de Sistema.

Brainstorming

A continuación se detalla el proceso realizado para la generación de la tormenta de ideas. De las primeras entrevistas se ha decidido generar una sesión de tormenta de ideas, con los jefes de cada área involucrada.

Identificación de participantes

Jefe de Soporte a la Producción.

Jefe de Desarrollo de Sistemas.

Jefe de Mesa de Ayuda.

Jefe de Mercadotecnia.

Líder de Proyecto (moderador)

Planificación

La sesión no se expenderá más de 45 minutos, el detalle es el siguiente:

Detalle	**Tiempo**
Presentación de la sesión	10 minutos
Desarrollo de ideas	20 minutos
Clasificación de ideas	10 minutos
Registro de Ideas claves	5 minutos

Preparación

Breve alocución sobre el motivo de la reunión

Discurso de presentación (10 minutos):

La presente sesión se ha realizado en función de las inquietudes por la realización de un nuevo sistema ha expresado el jefe de Soporte a la Producción, su objetivo es la profundización de los conceptos expresados en las entrevistas hechas con anterioridad.

Presentación entre los integrantes del mismo.

Los conceptos generales del sistema por realizar son:

Generar un sistema que pueda ayudar a Soporte a la Producción en el tiempo de respuesta a los incidentes reportados por Mesa de Ayuda.

Poder proporcionar información sobre el estado de las campañas publicitarias al área de Mercadotecnia.

Generar la información que sea solicitada por Desarrollo de Sistemas para la resolución de defectos en los sistemas Web.

Desarrollo de Ideas (20 minutos):

Se pasa a transcribir el detalle de todas las ideas que se han expresado:

Autor	Descripción	ID
Jefe de Soporte a la Producción	Automatizar todos los procedimientos de pedidos de información, hechos por las distintas áreas.	1
	Que cada área pueda generar automáticamente su información sin la necesidad de interactuar con Soporte a la Producción.	2
	Poder crear flexiblemente cualquier pedido de información.	3
Jefe de Mesa de Ayuda	La información que solicito tiene que llega a mi área, en forma rápida y con un análisis detallado de la misma, para que los miembros de la mesa de ayuda puedan explicar a los usuarios el incidente, hay que entender que los miembros del grupo de mesa de ayuda es un equipo de empleados generalmente temporarios.	4

Jefe de Desarrollo de Sistemas	Cuando se pide realizar una modificación a un problema, mi equipo necesita información para corroborar el error y verificar la solución, no necesitamos que otro equipo busque la información, solo necesitamos que nos digan donde esta el archivo y nosotros extraemos la información	5
Jefe de Mercadotecnia	Seria bueno que podamos tener información relacionada con las campañas publicitarias que hacemos, es decir, si nosotros empezamos con una nueva campaña seria necesario que tengamos algún tipo de información de cuantos usuarios están ingresando en que horarios y si están accediendo a determinados lugares de sitio.	6
	También me imagino que información de este tipo le serviría de mucho al área de Gerencia General, que constantemente nos esta pidiendo información.	7

Viabilidad de las ideas expuestas

Automatizar todos los procedimientos de pedidos de información, hechos por las distintas áreas. En esta primera etapa no se considera fundamental automatizar todos los procesos, este requerimiento se dejaría para una siguiente evolución del sistema.

Que cada área pueda generar automáticamente su información sin la necesidad de interactuar con Soporte a la Producción. En esta primera etapa no se considera fundamental automatizar todos los procesos, este requerimiento se dejaría para una siguiente evolución del sistema. Poder crear flexiblemente cualquier pedido de información. Se considera fundamental en el diseño del nuevo sistema. La información que solicito tiene que llega a mi área, en forma rápida y con un análisis detallado de la misma, para que los miembros de la mesa de ayuda puedan explicar a los usuarios el incidente, hay que entender que los miembros del grupo de mesa de ayuda es un equipo de empleados generalmente temporarios. Se considera fundamental en el diseño del nuevo sistema. Cuando se pide realizar una modificación a un problema, mi equipo necesita información para corroborar el error y verificar la solución, no necesitamos que otro equipo busque la información, solo necesitamos que nos digan donde esta el archivo y nosotros extraemos la información. Se evaluará este requerimiento pero no es fundamental para el desarrollo del sistema. Seria bueno que podamos tener información relacionada con las campañas publicitarias que hacemos, es decir, si nosotros empezamos con una nueva campaña seria necesario que

tengamos algún tipo de información de cuantos usuarios están ingresando en que horarios y si están accediendo a determinados lugares de sitio. Se considera fundamental en el diseño del nuevo sistema. También me imagino que información de este tipo le serviría de mucho al área de Gerencia General, que constantemente nos esta pidiendo información. Se evaluará este requerimiento.

Clasificación de las Ideas (10 minutos)

Importancia	Clasificación	Descripción
1	RQ01	Poder crear flexiblemente cualquier pedido de información
2	RQ02	La información que solicito tiene que llega a mi área, en forma rápida y con un análisis detallado de la misma, para que los miembros de la mesa de ayuda puedan explicar a los usuarios el incidente, hay que entender que los miembros del grupo de mesa de ayuda es un equipo de empleados generalmente temporarios
3	RQ03	Seria bueno que podamos tener información relacionada con las campañas publicitarias que hacemos, es decir, si nosotros empezamos con una nueva campaña seria necesario que tengamos algún tipo de información de cuantos usuarios están ingresando en que horarios y si están accediendo a determinados lugares de sitio
4	RQ04	También me imagino que información de este tipo le serviría de mucho al área de Gerencia General, que constantemente nos esta pidiendo información
5	RQ05	Cuando se pide realizar una modificación a un problema, mi equipo necesita información para corroborar el error y verificar la solución, no necesitamos que otro equipo busque la información, solo necesitamos que nos digan donde esta el archivo y nosotros extraemos la información

SESIÓN JAD

Se detalla el proceso realizado para la sesión JAD.

Quien usara el sistema y con que frecuencia?
Los usuarios directos del sistema son el equipo de Soporte a la Producción y de Desarrollo.

Usuarios indirectos que pedirán información a través de los equipos antes mencionados. La frecuencia de uso será diaria.

Que crecimiento se espera en la empresa?
El crecimiento estimado en función a la reactivación económica que se presenta es de un 5% anual según el equipo de asesores financieros.

Que áreas de la compañía son afectadas por el sistema por proveer documentos fuentes o por recibir informes?
Áreas que proveerán información para el proceso del sistema es:
Soporte a la Producción

Áreas que recibirán información para el proceso del sistema es:
Soporte a la Producción
Desarrollo del sistema
Marketing
Gerente General

Cuales son las interfaces del sistema?
Con otros sistemas
Recibe información de los archivos de sucesos de los servidores Web.

Con usuarios:
Generar informes.

Objetivos del Proyecto:

¿Qué se quiere aumentar con el proyecto?
Se pretende reducir el tiempo de respuesta de Soporte a la Producción.

¿Qué resultados mesurables se espera?
El tiempo de respuesta del equipo de Soporte a la Producción debe ser menor a las 5 horas en promedio.
Los informes producidos deben ser accesibles a varios grupos de la empresa.

Funciones:

¿Qué funciones proporciona el sistema?
Ayudar a Soporte a la Producción para reducir tiempos de respuesta en los procesos diarios.

Provee información para:
Gerencia General para el soporte a la toma de decisiones.
Mercadotecnia para el seguimiento de campañas.
Mesa de ayuda para solucionar alguna inquietud presentada por un cliente.

¿Qué es lo prioritario en estas funciones?
Rapidez en la respuesta
Certeza en los datos proporcionados.

¿Qué se recomienda?
Las recomendaciones para el producto final son:
Los procesos deben ser capaces de adaptarse en el tiempo sin necesidad de nueva programación.
Todos los proceso deben poder ser repetibles
Los procesos ejecutados deben dejar información para auditoria.

Restricciones

¿Qué limites deben ser considerados en el desarrollo del sistema?
No se cuenta con presupuesto asignado para invertir en software o hardware, en consecuencia el desarrollo deberá ser llevado a cabo con recursos internos, con la utilización de Fuentes Libres y ser ejecutado en equipos de usuario o en servidores ya existentes..

¿Cuáles son los plazos?
3 meses para la entrega de la primera versión del producto.

¿Hay alguna limitación de espacio, requisitos de seguridad, o regulaciones gubernamentales?
De espacio y gubernamentales no.
De seguridad debe cumplir los requisitos estándares de la empresa.

Requisitos de recursos adicionales de usuario

¿Cúales son los requisitos adicionales de recursos de usuarios, personal, equipos y espacio físico?
No se detectaron en este estado del proyecto

Supuestos anteriores a la sesión

¿Qué decisiones han sido tomadas ya para el sistema?
Debe poder correr en cualquier plataforma.
Poder ser ejecutado tanto en servidores como en terminales de usuarios.
Utilización de fuentes libres en la medida de lo posible.
Desarrollo interno.

¿Cuestiones abiertas anteriores a la sesión?
El sistema podrá ser utilizado directamente por el área de Desarrollo.

¿Qué cuestiones sin resolver deben tratarse con prioridad en la sesión?
Cual será el equipo de desarrollo, encargado de realizar el proyecto.

Objetivos principales y factores críticos de éxito
Reducir el tiempo de respuesta del equipo de Soporte a la Producción
Información generada.

Lista de participantes
Patrocinador del proyecto
Jefe de Proyecto
Jefe de Soporte a producción
Director de Mercadotecnia.

Guía de definición:

Prefacio:

Este documento es el resultante de la definición del proyecto de la metodología JAD y reúne la información de las entrevistas con el patrocinador ejecutivo, los usuarios clave y los directivas de desarrollo.

Propósito:

Preparar al equipo de Soporte a la Producción para el incremento de trabajo que deberán afrontar en función de las estimaciones de crecimiento realizado por nuestro equipo de asesores financieros. El sistema tendrá como principal objetivo agilizar el tiempo de repuesta del equipo de Soporte a la Producción.

Alcance:

Los usuarios del sistema es el equipo de Soporte a la Producción. Distintos sectores de la empresa pedirán información a través de los equipos antes mencionados.

Áreas que recibirán información para el proceso del sistema es:

Soporte a la Producción
Desarrollo del sistema
Marketing
Gerente General

Objetivos de la dirección:

Se pretende reducir el tiempo de respuesta de Soporte a la Producción.
El tiempo de respuesta del equipo de Soporte a la Producción debe ser menor a las 5 horas en promedio.
Los informes producidos deben ser accesibles a varios grupos de la empresa.

Funciones:

Provee información para:
Gerencia General para el soporte a la toma de decisiones.
Mercadotecnia para el seguimiento de campañas.
Mesa de ayuda para solucionar alguna inquietud presentada por un cliente.

Restricciones

La principal restricción con la que se cuenta es la falta de presupuesto para el proyecto. El mismo deberá ser realizado con recursos internos.

Acerca de restricciones gubernamentales y de espacio no se han encontrado. Con respecto a limitaciones de seguridad después de indagar no se han encontrado restricciones con respeto a las políticas de seguridad de la empresa.

Supuestos

El detalle de los supuestos con que se debe realizar el sistema son:

Debe poder correr en cualquier plataforma.

Poder ser ejecutado tanto en servidores como en terminales de usuarios.

Utilización de fuentes libres en la medida de lo posible.

Desarrollo interno.

Participantes de la sesión JAD

Patrocinador del proyecto

Jefe de Proyecto

Jefe de Soporte a producción

Director de Mercadotecnia.

A.2. DETALLE DEL SEGUNDO CICLO DE ENTREVISTAS ABIERTAS:

Primer entrevista abierta (Jefe de Mesa de Ayuda)

Preparación de la entrevista:

Del ejercicio de Brainstorming se detecto la posibilidad que Gerencia General sea uno de los potenciales clientes del nuevo sistema. Se define una reunión con el gerente general para definir el posible alcance del proyecto.

Participantes:

Usuario potencial, gerente general.

Materiales a Proporcionar:

Introducción con el objetivo del sistema.

Objetivos de la Reunión:

Identificar las sdalidas del sistema.

Detalle del material a entregar

En función del ejercicio realizado de Brianstorming se ha detectado la posibilidad que el sistema brinde información a la Gerencia General para el soporte a la decisión, es por esta razón que se desea mantener una entrevista para definir nivel de servicios y calidad de la información que se espera recibir.

Resumen de la entrevista

De la entrevista realizada con el Gerente General, se han extraído las siguientes conclusiones:

Se ha detectado de un sistema que pueda proporcionar información ara el soporte a la decisión es de gran importancia para la Gerencia General.

El sistema debe ser flexible para los requerimientos cambiantes de información de la Gerencia General.

Brainstorming

Con la información relevada en la sesión de JAD, se ha decido realizar otro Brianstorming, agregando al Gerente General al mismo.

Identificación de participantes

Jefe de Soporte a la Producción.

Jefe de Desarrollo de Sistemas.

Jefe de Mesa de Ayuda.

Jefe de Mercadotecnia.

Gerente General

Líder de Proyecto (moderador)

Planificación

La sesión no se expenderá más de 45 minutos, el detalle es el siguiente:

Detalle	Tiempo
Presentación de la sesión	10 minutos
Desarrollo de ideas	20 minutos
Clasificación de ideas	10 minutos
Registro de Ideas claves	5 minutos

Preparación

Breve alocución sobre el motivo de la reunión

Discurso de presentación (10 minutos):

La presente sesión se ha realizado en función de las inquietudes por la realización de un nuevo sistema ha expresado el jefe de Soporte a la Producción, su objetivo es la profundización de los conceptos expresados en las entrevistas hechas con anterioridad.

Presentación entre los integrantes del mismo.

Los conceptos generales del sistema por realizar son:

Generar un sistema que pueda ayudar a Soporte a la Producción en el tiempo de respuesta a los incidentes reportados por Mesa de Ayuda.

Poder proporcionar información sobre el estado de las campañas publicitarias al área de Mercadotecnia.

Generar la información que sea solicitada por Desarrollo de Sistemas para la resolución de defectos en los sistemas Web.

Generar información de soporte a la decisión para el área de Gerencia General.

Desarrollo de Ideas (20 minutos):

Se pasa a transcribir el detalle de todas las ideas que se han expresado:

Autor	Descripción	ID
Jefe de Soporte a la Producción	Los procesos deben poder ser auditados para que el área de Calidad y Métodos pueda catalogar los mismos y hacerlos parte de la gestión de conocimiento que esta llevando a cabo.	1
Jefe de Mercadotecnia	Es necesario que se reciban informes diarios con las estadísticas y demás informes a diario, a primera hora, para que se pueda enfrentar el nuevo día de trabajo.	2
Gerente General	El tipo de información que se necesita para la toma de decisiones es muy cambiante, y es necesario que ase pueda generar una traza hasta el origen de datos.	3
	Un sistema de este tipo puede servir en gran medida para el área de Contaduría para estimar el costo de cada transacción, hecha en la Web.	4
Jefe de Desarrollo de Sistemas	Otra área que encontraría provechosa la información que puede dar este sistema es Infraestructura Tecnológica, para hacer un calculo de uso de sistemas y poder predecir el recambio de los equipos.	5
	También podría ser importante, esto tendría que ser evaluado en el área, es seguridad informática, se podría extraer patrones de comportamiento para detectar comportamientos anormales.	6

Viabilidad de las ideas expuestas

Los procesos deben poder ser auditados para que el área de Calidad y Métodos pueda catalogar los mismos y hacerlos parte de la gestión de conocimiento que esta llevando a cabo. Se considera fundamental en el diseño del nuevo sistema.

Es necesario que se reciban informes diarios con las estadísticas y demás informes a diario, a primera hora, para que se pueda enfrentar el nuevo día de trabajo

Se considera fundamental en el diseño del nuevo sistema.

El tipo de información que se necesita para la toma de decisiones es muy cambiante, y es necesario que ase pueda generar una traza hasta el origen de datos..

Se considera fundamental en el diseño del nuevo sistema.

Un sistema de este tipo puede servir en gran medida para el área de Contaduría para estimar el costo de cada transacción, hecha en la Web.

Se considera fundamental en el diseño del nuevo sistema.

Otra área que encontraría provechosa la información que puede dar este sistema es Infraestructura Tecnológica, para hacer un calculo de uso de sistemas y poder predecir el recambio de los equipos.
Se considera fundamental en el diseño del nuevo sistema.
También podría ser importante, esto tendría que ser evaluado en el área, es seguridad informática, se podría extraer patrones de comportamiento para detectar comportamientos anormales..
Se considera fundamental en el diseño del nuevo sistema.

Clasificación de las Ideas (10 minutos)

Importancia	Clasificación	Descripción
1	RQ06	Los procesos deben poder ser auditados para que el área de Calidad y Métodos pueda catalogar los mismos y hacerlos parte de la gestión de conocimiento que esta llevando a cabo.
2	RQ07	El tipo de información que se necesita para la toma de decisiones es muy cambiante, y es necesario que ase pueda generar una traza hasta el origen de datos.
3	RQ08	Es necesario que se reciban informes diarios con las estadísticas y demás informes a diario, a primera hora, para que se pueda enfrentar el nuevo día de trabajo.
4	RQ09	También podría ser importante, esto tendría que ser evaluado en el área, es seguridad informática, se podría extraer patrones de comportamiento para detectar comportamientos anormales.
5	RQ10	Un sistema de este tipo puede servir en gran medida para el área de Contaduría para estimar el costo de cada transacción, hecha en la Web.
6	RQ11	Otra área que encontraría provechosa la información que puede dar este sistema es Infraestructura Tecnológica, para hacer un calculo de uso de sistemas y poder predecir el recambio de los equipos.

Primer entrevista abierta (Jefe de Contaduría)

Preparación de la entrevista:

Del ejercicio de Branstorming se detecto la posibilidad que Contaduría sea uno de los potenciales clientes del nuevo sistema.

Se define una reunión con el jefe de Contaduría para definir el posible alcance del proyecto.

Participantes:

Usuario potencial, jefe de Contaduría.

Materiales a Proporcionar:

Introducción con el objetivo del sistema.

Objetivos de la Reunión:

Salidas del sistema

Detalle del material a entregar

En función del ejercicio realizado de Brianstorming se ha detectado la posibilidad que el sistema brinde información al área de Contaduría para la asignación de costos por transacciones en los sitios Web, es por esta razón que se desea mantener una entrevista para definir nivel de servicios y calidad de la información que se espera recibir.

Resumen de la entrevista

De la entrevista realizada con el jefe de Contaduría, se han extraído las siguientes conclusiones:

Se ha detectado que el sistema pude proporcionar información ara el área de Seguridad Informática.

El sistema debe ser capaz de informar el costo de una transacción realizada por un usuario.

Primer entrevista abierta (Jefe de Infraestructura Tecnológica)

Preparación de la entrevista:

Del ejercicio de Branstorming se detecto la posibilidad que Infraestructura Tecnológica sea uno de los potenciales clientes del nuevo sistema.

Se define una reunión con el jefe de Infraestructura Tecnológica para definir el posible alcance del proyecto.

Participantes:

Usuario potencial, jefe de Infraestructura Tecnológica.

Materiales a Proporcionar:

Introducción con el objetivo del sistema.

Objetivos de la Reunión:

Identificar las salidas del sistema

Detalle del material a entregar

En función del ejercicio realizado de Brianstorming se ha detectado la posibilidad que el sistema brinde información al área de Infraestructura Tecnológica para la estimación de uso y predicción de recambios de equipos, es por esta razón que se desea mantener una entrevista para definir nivel de servicios y calidad de la información que se espera recibir.

Resumen de la entrevista

De la entrevista realizada con el jefe de Infraestructura Tecnológica, se han extraído las siguientes conclusiones:

Se ha detectado que el sistema pude proporcionar información ara el área de Infraestructura Tecnológica.

El sistema debe ser poder entregar estadísticas de uso y actividad de los usuarios de los sistemas Web.

Primer entrevista abierta (Jefe de Seguridad Informática)

Preparación de la entrevista:

Del ejercicio de Brainstorming se detecto la posibilidad que Seguridad Informática sea uno de los potenciales clientes del nuevo sistema.

Se define una reunión con el jefe de Seguridad Informática para definir el posible alcance del proyecto.

Participantes:

Usuario potencial, jefe de Seguridad Informática.

Materiales a Proporcionar:

Introducción con el objetivo del sistema.

Objetivos de la Reunión:

Salidas del sistema

Detalle del material a entregar

En función del ejercicio realizado de Brianstorming se ha detectado la posibilidad que el sistema brinde información a el área de Seguridad Informática para la detección de comportamientos anómalos en los sitios Web, es por esta razón que se desea mantener una entrevista para definir nivel de servicios y calidad de la información que se espera recibir.

Resumen de la entrevista

De la entrevista realizada con el jefe de Seguridad Informática, se han extraído las siguientes conclusiones:

Se ha detectado que el sistema puede proporcionar información ara el área de Seguridad Informática.

El sistema debería ser capaz de proporcionar información con patrones de comportamiento de los usuarios de un sitio Web.

Poder extraer la trazabilidad de una transacción.

Sesión JAD

Se detalla el proceso realizado para la sesión JAD

<u>Guía de definición:</u>

Prefacio:

Este documento es el resultante de la definición del proyecto de la metodología JAD y reúne la información de las entrevistas con el patrocinador ejecutivo, los usuarios clave y los directivas de desarrollo.

Propósito:

Preparar al equipo de Soporte a la Producción para el incremento de trabajo que deberán afrontar en función de las estimaciones de crecimiento realizado por nuestro equipo de asesores financieros.

El sistema tendrá como principal objetivo agilizar el tiempo de repuesta del equipo de Soporte a la Producción.

Alcance:

Los usuarios del sistema es el equipo de Soporte a la Producción. Distintos sectores de la empresa pedirán información a través de los equipos antes mencionados.

Áreas que recibirán información para el proceso del sistema es:

Soporte a la Producción

Desarrollo del sistema

Marketing

Gerente General

Infraestructura Tecnológica

Contaduría

Auditoria

Seguridad Informática

Objetivos de la dirección:

Se pretende reducir el tiempo de respuesta de Soporte a la Producción.

El tiempo de respuesta del equipo de Soporte a la Producción debe ser menor a las 5 horas en promedio.

Los informes producidos deben ser accesibles a varios grupos de la empresa.

Los proceso deben poder ser auditados.

Gran flexibilidad y adaptación al cambio.

Funciones:

Provee información para:

Gerencia General para el soporte a la toma de decisiones.

Mercadotecnia para el seguimiento de campañas.

Infraestructura Tecnológica para estimaciones de reemplazo de hardware.

Contaduría para estimación de costos de transacciones.

Mesa de ayuda para solucionar alguna inquietud presentada por un cliente.
Auditoria debe poder repetir los procesos y entenderlos.
Seguridad informática debe recibir información de comportamiento de usuarios.

Restricciones

La principal restricción con la que se cuenta es la falta de presupuesto para el proyecto. El mismo deberá ser realizado con recursos internos.
Acerca de restricciones gubernamentales y de espacio no se han encontrado. Con respecto a limitaciones de seguridad después de indagar no se han encontrado restricciones con respeto a las políticas de seguridad de la empresa.

Supuestos

El detalle de los supuestos con que se debe realizar el sistema son:
Debe poder correr en cualquier plataforma.
Poder ser ejecutado tanto en servidores como en terminales de usuarios.
Utilización de fuentes libres en la medida de lo posible.
Desarrollo interno.

Participantes de la sesión JAD
Patrocinador del proyecto
Jefe de Proyecto
Jefe de Soporte a producción
Director de Mercadotecnia.
Jefe de Seguridad Informática.
Jefe de Auditoria.
Jefe de Contaduría.

Brainstorming

De las dos sesiones de Brainstorming anteriores se han definido un conjunto de requerimientos, la generación de la presente reunión es unificar requerimientos y detectar requerimientos comunes entre distintas áreas.

Identificación de participantes

Jefe de Soporte a la Producción.

Jefe de Desarrollo de Sistemas.

Jefe de Mesa de Ayuda.

Jefe de Mercadotecnia.

Jefe de Seguridad Informática.

Jefe de Auditoria.

Jefe de Infraestructura Tecnología.

Jefe de Contaduría.

Líder de Proyecto (moderador)

Planificación

La sesión no se expenderá más de 45 minutos, el detalle es el siguiente:

Detalle	**Tiempo**
Presentación de la sesión	10 minutos
Desarrollo de ideas	20 minutos
Clasificación de ideas	10 minutos
Registro de Ideas claves	5 minutos

Preparación

Discurso de presentación (10 minutos):

La presente sesión se ha realizado en función de las inquietudes por la realización de un nuevo sistema ha expresado el jefe de Soporte a la Producción, su objetivo es la profundización de los conceptos expresados en las entrevistas hechas con anterioridad.

Presentación entre los integrantes del mismo.

La siguiente tabla representa la unificación tentativa que se ha realizado de los requerimientos de las distintas áreas.

Principal	Unificada	Descripción
RQ01		Poder crear flexiblemente cualquier pedido de información
	RQ02	La información que solicito tiene que llegar a mi área, en forma rápida y con un análisis detallado de la misma, para que los miembros de la mesa de ayuda puedan explicar a los usuarios el incidente, hay que entender que los miembros del grupo de mesa de ayuda es un equipo de empleados generalmente temporarios.
	RQ03	Seria bueno que podamos tener información relacionada con las campañas publicitarias que hacemos, es decir, si nosotros empezamos con una nueva campaña seria necesario que tengamos algún tipo de información de cuantos usuarios están ingresando en que horarios y si están accediendo a determinados lugares de sitio
	RQ04	También me imagino que información de este tipo le serviría de mucho al área de Gerencia General, que constantemente nos esta pidiendo información
	RQ07	El tipo de información que se necesita para la toma de decisiones es muy cambiante, y es necesario que ase pueda generar una traza hasta el origen de datos.
	RQ09	También podría ser importante, esto tendría que ser evaluado en el área, es seguridad informática, se podría extraer patrones de comportamiento para detectar comportamientos anormales.
RQ05		Cuando se pide realizar una modificación a un problema, mi equipo necesita información para corroborar el error y verificar la solución, no necesitamos que otro equipo busque la información, solo necesitamos que nos digan donde esta el archivo y nosotros extraemos la información
	RQ06	Los procesos deben poder ser auditados para que el área de Calidad y Métodos pueda catalogar los mismos y hacerlos parte de la gestión de conocimiento que esta llevando a cabo.
RQ08		Es necesario que se reciban informes diarios con las estadísticas y demás informes a diario, a primera hora, para que se pueda enfrentar el nuevo día de trabajo.
RQ10		Un sistema de este tipo puede servir en gran medida para el área de Contaduría para estimar el costo de cada transacción, hecha en la Web.
RQ11		Otra área que encontraría provechosa la información que puede dar este sistema es Infraestructura Tecnológica, para hacer un calculo de uso de sistemas y poder predecir el recambio de los equipos.

Desarrollo de Ideas (20 minutos):

Se ha determinado que los requerimientos unificados son:

Importancia	Clasificación	Descripción
1	RQ01	Poder crear flexiblemente cualquier pedido de información
2	RQ05	Cuando se pide realizar una modificación a un problema, mi equipo necesita información para corroborar el error y verificar la solución, no necesitamos que otro equipo busque la información, solo necesitamos que nos digan donde esta el archivo y nosotros extraemos la información
3	RQ08	Es necesario que se reciban informes diarios con las estadísticas y demás informes a diario, a primera hora, para que se pueda enfrentar el nuevo día de trabajo.

Reformulación de ideas (10 minutos)

Importancia	Clasificación	Descripción	Deriva
1	RQV01_01	Los proceso que ejecute el sistema deben ser modulares y flexibles.	RQ01
2	RQV01_02	La información entregada por el sistema debe poder ser verificada y validada.	RQ05
3	RQV01_03	El sistema debe tener capacidad para agendar tareas.	RQ08

Tareas pendientes (5 minutos)

Se ha decidido que los lineamientos principales del sistema han sido definidos y se continuara con el detalle de los requerimientos. Cada miembro que ha participado definirá un miembro se sus equipos para que participen de aquí en adelante para el refinamiento de los requerimientos.

Brainstorming

De las dos sesiones de Brainstorming anteriores se han definido un conjunto de requerimientos, en esta nueva sesión se ha invitado a los miembros designados por los jefes de áreas para el refinamientos de los requerimientos .

Identificación de participantes

Miembro de Soporte a la Producción.

Miembro de Desarrollo de Sistemas.

Miembro de Mesa de Ayuda.

Miembro de Mercadotecnia.

Miembro de Seguridad Informática.

Miembro de Auditoria.

Miembro de Infraestructura Tecnología.

Miembro de Contaduría.

Líder de Proyecto (moderador)

Planificación

La sesión no se expenderá mas de 45 minutos, el detalle es el siguiente:

Detalle	**Tiempo**
Presentación de la sesión	10 minutos
Desarrollo de ideas	20 minutos
Clasificación de ideas	10 minutos
Registro de Ideas claves	5 minutos

Discurso de presentación (10 minutos):

La presente sesión se ha realizado en función de las inquietudes por la realización de un nuevo sistema ha expresado el jefe de Soporte a la Producción, su objetivo es la profundización de los conceptos expresados en las entrevistas hechas con anterioridad.

Presentación entre los integrantes del mismo.

La siguiente tabla representala los requerimientos unificados, el objetivo de la presente presentación es poder enriquecer los mismos.

Importancia	**Clasificación**	**Descripción**	**Deriva**
1	RQV01_01	Los proceso que ejecute el sistema deben ser modulares y flexibles.	RQ01
2	RQV01_02	La información entregada por el sistema debe poder ser verificada y validada.	RQ05
3	RQV01_03	El sistema debe tener capacidad para agendar tareas.	RQ08

Desarrollo de Ideas (20 minutos):

Se pasa a transcribir el detalle de todas las ideas que se han expresado:

Autor	Descripción	ID
Miembro de Soporte a la Producción	El sistema debe poder ser ejecutado en entornos Windows, Unix (HP, SUN) y Linux. Existen en el sector varios tipos de equipos.	1
Miembro de Mercadotecnia	El formato de salida debe ser una planilla de calculo, para que pueda ser trabajado.	2
Miembro de Desarrollo de Sistemas	El formato de salida debe ser en formato XML.	3
Miembro de Seguridad Informática	El formato de salida debe ser insertado en una Base de Datos.	4
Miembro de Infraestructura Tecnología.	No solo debe poder tomar datos de los sistemas de la Web debe además por tomar datos de distintas fuentes de datos.	5
Miembro de Contaduría.	Se debe poder asignarle un valor a cada transacción	6

Viabilidad de las ideas expuestas

El sistema debe poder ser ejecutado en entornos Windows, Unix (HP, SUN) y Linux.

Existen en el sector varios tipos de equipos

Se considera fundamental en el diseño del nuevo sistema.

El formato de salida debe ser una planilla de calculo, para que pueda ser trabajado.

Se considera fundamental en el diseño del nuevo sistema.

El formato de salida debe ser en formato XML.

Se considera fundamental en el diseño del nuevo sistema.

El formato de salida debe ser insertado en una Base de Datos.

Se considera fundamental en el diseño del nuevo sistema.

No solo debe poder tomar datos de los sistemas de la Web debe además por tomar datos de distintas fuentes de datos.

Se considera fundamental en el diseño del nuevo sistema.

Se debe poder asignarle un valor a cada transacción.

Se considera fundamental en el diseño del nuevo sistema.

Clasificación de las Ideas (10 minutos)

Principal	Unificada	Descripción
1		El sistema debe poder ser ejecutado en entornos Windows, Unix (HP, SUN) y Linux. Existen en el sector varios tipos de equipos.
2		Formato flexible de salida.
	2	El formato de salida debe ser una planilla de cálculo, para que pueda ser trabajado.
	3	El formato de salida debe ser en formato XML.
	4	El formato de salida debe ser insertado en una Base de Datos.
3		Varias fuentes de datos.
	5	No solo debe poder tomar datos de los sistemas de la Web debe además por tomar datos de distintas fuentes de datos.
	6	Se debe poder asignarle un valor a cada transacción

Importancia	Clasificación	Descripción	Deriva
1	RQV01_04	El sistema debe poder ser ejecutado en entornos Windows, Unix (HP, SUN) y Linux. Existen en el sector varios tipos de equipos.	1
2	RQV01_05	Formato flexible de salida.	2
3	RQV01_06	Varias fuentes de datos.	3

Se ha procedido a unificar los requerimientos obtenelos de las sesiones de Brianstorming y JAD. El presente documento se ha hecho llegar a la siguiente lista de personas:

Jefe de Soporte a la Producción.

Jefe de Desarrollo de Sistemas.

Jefe de Mesa de Ayuda.

Jefe de Mercadotecnia.

Jefe de Seguridad Informática.

Jefe de Auditoria.

Jefe de Infraestructura Tecnología.

Jefe de Contaduría.

Miembro de Soporte a la Producción.

Miembro de Desarrollo de Sistemas.

Miembro de Mesa de Ayuda.

Miembro de Mercadotecnia.

Miembro de Seguridad Informática.

Miembro de Auditoria.

Miembro de Infraestructura Tecnología.
Miembro de Contaduría.

Se les ha pedido a los receptores del presente documento que den su visto bueno con los requerimientos y que hagan las acotaciones que crean necesarias.

RQV01_01: Los proceso que ejecute el sistema deben ser modulares y flexibles.
Detalle:
Al existir diferentes requerimientos por parte de los distintos grupos de la empresa, el sistema debe proveer una forma sencilla de poder satisfacer estos pedidos de información.
Derivaciones:
De lo antes expresado se desprende que en el conjunto de requerimientos pedidos, se repetirán procesos o parte de los mismos. Por esto se debe contemplar un mecanismo para catalogar los procesos y sus partes (RQV01_07)

RQV01_02: La información entregada por el sistema debe poder ser verificada y validada.
Detalle:
Se debe mantener la trazabilidad de todo el proceso, y resguardar todos los pasos intermedios que se generen para obtener la información final.

RQV01_03: El sistema debe tener capacidad para agendar tareas.
Detalle:
Varios de los proceso solicitados por las distintas áreas de la empresa son periódicos, para aprovechar tiempo de procesamiento, los mismos pueden ser ejecutados por la noche, es por esa razón que deben poder ser agendados.

RQV01_04: El sistema debe poder ser ejecutado en entornos Windows, Unix y Linux.
Detalle:

Los equipos en los cuales debe poder ser ejecutado son los antes mencionados y además las capacidades de procesamiento son diferentes. El proceso debe poder adaptarse a las diferentes configuraciones.

RQV01_05: Formato flexible de archivos de salida.

Detalle:

Los procesos deben poder adaptarse para poder dejar la información en al menos estos formatos de salida:

Archivo de texto separado por coma.

Formato XML.

Insertar los datos en una Base de Datos.

RQV01_06: Varias fuentes de dato de entradas.

Detalle:

Los proceso no solo deben poder tomar datos de los archivos de suceso de los servidores Web, además debe ser capaz de tomar datos de:

Archivo de texto separado por coma.

Formato XML.

Base de Datos.

ANEXO B: CALCULO DEL TAMAÑO DE UN SITIO WEB

La capacidad de poder definir si un sitio Web es grande, mediano o pequeño tiene importancia en el presente trabajo, pues en los casos que se van a analizar es necesario poder hacer una comparación entre sitios de similares condiciones.

La web visible es el sitio que la gran mayoría de las personas intuye cuando se hace referencia a Internet, es decir paginas HTML conectadas entre sí por hipervínculos; las mismas están compuestas por textos e imágenes. Estos sitios son los que comúnmente son indexados por los buscadores de Internet, entre los cuales podemos hacer referencia a Google y Yahoo. Estos buscadores poseen robots o como comúnmente se los conoce en Internet como "*arañas*", que van buscando en la red nuevos sitios y sus vínculos, y los agregan a su base de datos. Específicamente Google utiliza el algoritmo de su autoría llamado PageRank, el cual fue patentado en Estados Unidos el 8 de enero de 1998, por Larry Page. El título original de dicho algoritmos *es 'Method for node ranking in a linked database*', y le fue asignado el número de patente 6.285.999. Podemos decir que este algoritmo muestra un valor numérico que representa la importancia que una página Web tiene en Internet. Cuantos más votos tenga una página, será considerada más importante por Google; además, la importancia de la página que emite su voto también determina el peso de este voto. De esta manera, Google calcula la importancia de una página gracias a todos los votos que reciba, teniendo en cuenta también la importancia de cada página que emite el voto. Básicamente dicho algoritmo funciona de la siguiente manera: Clasifica cada sitio Web en función de los hipervínculos que otros sitios tienen hacia él, los que él tiene hacia otros sitios y el cambio de contenido que el mismo sufre; luego se obtiene un valor cuyo rango varia entre 0 y 10, el valor 0 representa que el sitio esta penalizado por Google y 10 para un sitio muy referenciado y accedido por los navegantes. Cabe aclarar que este no es el único factor que Google utiliza para clasificar las páginas, pero sí es uno de los más importantes. Se debe tener en cuenta que no todos los links son tenidos en cuenta por Google, ya que por ejemplo, Google filtra y descarta los enlaces de páginas dedicadas exclusivamente a colocar links (llamadas 'link farms'). Además, Google admite que una página no puede controlar los links que apuntan hacia ella, pero sí que puede controlar los enlaces que esta página coloca hacia otras páginas. Por ello, links hacia una página no pueden perjudicarla,

pero sí que enlaces que una página coloque hacia sitios penalizados, pueden ser perjudiciales para su PageRank. Si un sitio Web tiene PageRank 0, generalmente es una Web penalizada, esto quiere decir que cuando se haga una búsqueda se presentara en los últimos lugares de la misma, y podría ser poco inteligente colocar un link hacia ella. Una forma de conocer el PageRank de una página es descargándose la barra de búsqueda de Google (solamente disponible para MS IExplorer) en el cual se muestra en color verde el valor de PageRank™. Sitios Web que también utilizan este algoritmo PageRank 10 son Microsoft, Adobe, Macromedia. El valor real suele ser del orden de miles de unidades.

A continuación se da un ejemplo en detalle para demostrar que la solución propuesta por Google no se ajusta a las necesidades del presente trabajo; para una base 7, se tendrían los siguientes valores:

PR Barra	PR Real
0	0 – 3
1	3 – 19
2	19 - 130
3	130 – 907
4	907 – 6351
5	6351 – 44458
6	44458 – 311209
7	311209 - 2178466
8	2178466 - 15249262
9	15249262 - 106765607
10	> 106765607

Se plantea la siguiente fórmula para calcular el PageRank de una página Web llamada 'A':

$$PR(A) = (1-d) + d * [PR(T1)/C(T1) + ... + PR(Tn)/C(Tn)]$$

Donde:

'd' es el factor de atenuación. Un valor podría ser 0,85

'Ti' es cada página que enlaza a '**A**'. '**i**' toma los valores 1, 2, ... hasta **'N'**. **'N'** es el número de páginas que enlazan a '**A**'.

'PR(Ti)' es el PageRank de cada una de las páginas que enlazan a '**A**'.

'C(Ti)' es el número de enlaces que salen desde cada página **'Ti'**.

Por lo tanto, la página de 500000 de PageRank transmitirá a otra en caso de tener un único enlace, un valor de 0,85*500000 = 425000. Generalmente las páginas poseen más de un enlace dentro de ellas, así que este valor habría que dividirlo entre el número de enlaces.

Es importan aclarar que rl valor del PageRank de cada página no es constante en el tiempo, ya que depende de los enlaces que vayamos recibiendo y, a su vez, del PageRank de las páginas que se enlazan. Por ello, una vez al mes aproximadamente, Google recalcula el valor de este PageRank en lo que viene a llamar la 'Google Dance', por lo que modifica los resultados de las búsquedas. La 'Google Dance' es el periodo que transcurre entre el comienzo y el fin de esta actualización, por lo general suele tardar unos 4 días, y durante ese tiempo se obtienen diferentes resultados en cada uno de los servidores de Google: www.google.com, www2.google.com, www3.google.com y www-fi.google.com. El mejor momento para colocar páginas en un sitio Web es durante la 'Google Dance'. Si se deja mucho tiempo entre el fin de esta actualización y la publicación de nuevos contenidos, se reduce la cantidad de páginas que serán incluidas en la próxima actualización.

Otro caso es Yahoo! que utiliza en cierta medida los servicios proporcionados por Google y le agrega una segunda etapa de clasificación manual, la cual consiste en una ultima evaluación hecha por una persona para corroborar si la clasificación que ha recibido por el programa automáticamente es correcta y no necesita algún cambio de categoría.

Por otra parte existe la cantidad de información [5] que existe en Internet y no es catalogada por los buscadores más comunes en Internet, la información que no es contemplada es toda la información que se encuentra en las base de datos que se

encuentran conectadas a Internet. Este tipo de información en Internet no es accedida por los buscadores y se considera que representa aproximadamente el 80% de la información disponible en la red [Google, 2005].

Por lo antes mencionado podemos decir que se cuenta con una métrica acorde a las necesidades del problema planteado. Por esto se sugiere la siguiente métrica:

La métrica propuesta será dividida en 3 componentes:

Web Invisible: esta hará referencia a la cantidad de información que se encuentra almacenada en las bases de datos que se publican en la red. Esta esta conformada por:

- Cantidad de registros: representa la suma total de las distintas tablas que pueden ser accedidas desde una página del sitio.
- Longitud de registros: la longitud de cada tabla que puede ser accedida desde el sitio.

Web Visible: esta hace referencia a las páginas Web y al contenido de las mismas y sus modificaciones.

- Total de páginas Web: representa la suma total de páginas, las cuales frecuentemente están en el formato HTML.
- Tamaño total de las páginas: es la suma total del tamaño de las paginas.

Unidad de Tiempo: es la unidad para medir la frecuencia de cambio del contenido de la página.

Páginas modificadas por unidad de tiempo: cantidad de paginas modificadas.

Cantidad de imágenes: es el total de imágenes del sitio.

Tamaño total de las imágenes: representa la suma total del tamaño de las imágenes

Cantidad de otros recursos: es frecuente que los sitios de Internet contengan otros recursos como ser documentos en formato PDF, etc. Esta medida es la suma de los mismos.

Tamaño total de otros recursos: es el tamaño total de los recursos antes mencionados.

Accesos al sitio: hace referencia a la utilización que los usuarios dan del mismo.

Unidad de Tiempo: la unidad con la cual se va a medir los accesos al sitio

Cantidad total de accesos: cantidad bruta de accesos al sitio

Cantidad de accesos únicos: si se tiene, representa la cantidad de usuarios únicos que acceden al sitio pro unidad de tiempo.

En la tabla B.1 se representa la métrica y sus cálculos:

Calculo del tamaño de un sitio Web		
Web invisible		
Tamaño de la Base de Datos accesible desde el sitio		
A	Cantidad de registros	
B	Longitud de registros	
1	Calculo final de la Web invisible (A * B)	
Web visible		
Tamaño y cantidad de objetos publicados fuera de la Base de Datos		
A	Total de paginas Web	
B	Tamaño total de las paginas	
C	**Calculo de tamaño (A * B)**	
D	Unidad de Tiempo	
E	Paginas modificadas por unidad de tiempo	
F	**Calculo de modificación (E)**	
G	Cantidad de imágenes	
H	Tamaño total de las imágenes	
I	**Calculo de grafica (G * H)**	
J	Cantidad de otros recursos	
K	Tamaño total de otros recursos	
L	**Calculo de otros recursos (J * K)**	
2	Calculo final de la Web visible ((C * I * L) ** F)	
Accesos a la Web		
Cantidad de accesos registrados por el sitio		
A	Unidad de Tiempo	
B	Cantidad total de accesos	
C	**Calculo de accesos (A * B)**	
E	Cantidad de accesos únicos	
F	**Calculo de accesos únicos (A * E)**	
3	**Calculo unificado de accesos (C / F)**	
Valor Final (1 * 2 * 3)		

Tabla B.1: Métrica para Sitios

Como conclusión se puede establecer que no se puede tener una medida absoluta, en consecuencia se recomienda tomar como media de referencia los sitios que se van a analizar. Es decir, si se va a trabajar con tres sitios y se deben clasificar es conveniente utilizar la métrica propuesta y establecer un ranking de valores menor a mayor. Una vez realizado esto se compara los valores, se toma el menor de los valores y se calcula el 33% del mismo (este valor de 33% se a extraído de una tabla imaginaria de 10 valores el 33% representaría un tercio de la misma) y se le suma al valor obtenido en la métrica. Si el valor obtenido es igual o mayor al valor obtenido por el segundo sitio ambos quedan dentro del mismo rango, que luego será definido, este proceso se repite 3 veces, se suma el 33% al valor obtenido.

De acuerdo al cálculo realizado es posible que:

Los tres sitios están dentro del rango de la primera suma, entonces se concluirá que los sitios son comparativamente similares.

Que el primer y segundo sitio queden dentro del rango y el tercero no, en este se concluirá que, si el valor obtenido en la métrica por el tercer sitio cae dentro de la próxima suma del 33%, los sitios serán clasificados como el primero y segundo como medianos, y el tercero como grande; si para poder englobar al valor del tercer sitio es necesario sumar nuevamente una o mas veces el 33%, diremos que el primer y segundo sitio son pequeños y el tercero es grande.

Siguiendo la misma línea de razonamiento se seguirán clasificando para las distintas alternativas que se obtengan.

ANEXO C: CONTROL DE CONFIGURACIÓN

Tipo de ECS	Fecha de solicitud	Elemento de cambio	Motivo del cambio	Versión	Fecha modificación	Estado
DOC	14/09/04	Especificación del Sistema.	---	1.0	14/09/04	Aprobado
DOC	19/09/04	Estimación del Proyecto	---	1.0	20/09/04	Aprobado
PLN	21/09/04	Plan del tiempo del proyecto software.	---	1.0	21/09/04	Aprobado
DOC	26/10/04	Especificación de requisitos de software.	---	1.0	26/10/04	Aprobado
DOC	24/11/04	Diseño preliminar y detallado.	---	1.0	24/11/04	Aprobado
COD	20/12/04	Códigos fuente.	---	1.0	20/12/04	Aprobado
COD	20/12/04	Programas ejecutables.	---	1.0	03/01/05	Aprobado
DOC	20/01/05	Manuales asociados al proyecto.	---	1.0	20/08/05	Aprobado
DOC	20/01/05	Guías asociadas al proyecto.	---	1.0	23/01/05	Aprobado
DOC	20/01/05	Plan de Pruebas.	---	1.0	23/01/05	Aprobado
DOC	20/01/05	Estándares y procedimientos de IS utilizados.	---	1.0	27/01/05	Aprobado
DOC	08/02/05	Casos de prueba ejecutados y sus resultados.	---	1.0	15/08/05	Aprobado
BDD	17/03/05	Diseños de bases de datos.	---	1.0	18/03/05	Aprobado
DAT	22/03/05	Contenidos de las bases de datos.	---	1.0	22/03/05	Aprobado
DOC	19/08/05	Especificación de requisitos de software.	Se modificaron los requisitos en función de los resultados de las pruebas obtenidas	1.1	19/08/05	Aprobado
DOC	19/08/05	Estimación del Proyecto	Se modificaron los requisitos en función de los resultados de las pruebas obtenidas	1.1	22/08/05	Aprobado
PLN	19/08/05	Plan del tiempo del proyecto software.	Se modificaron los requisitos en función de los resultados de las pruebas obtenidas	1.1	25/08/05	Aprobado
DOC	19/08/05	Diseño preliminar y detallado.	Se modificaron los requisitos en función de los resultados de las pruebas obtenidas	1.1	29/08/05	Aprobado
DOC	19/08/05	Plan de Pruebas.	Se modificaron los requisitos en función de los resultados de las pruebas obtenidas	1.1	01/09/05	Aprobado

Tipo de ECS	Fecha de solicitud	Elemento de cambio	Motivo del cambio	**Versión**	Fecha modificación	Estado
COD	19/08/05	Códigos fuente.	Se modificaron los requisitos en función de los resultados de las pruebas obtenidas	1.1	04/09/05	Aprobado
COD	19/08/05	Programas ejecutables.	Se modificaron los requisitos en función de los resultados de las pruebas obtenidas	1.1	08/09/05	Aprobado
DOC	19/08/05	Casos de prueba ejecutados y sus resultados.	Se modificaron los requisitos en función de los resultados de las pruebas obtenidas	1.1	10/09/05	Aprobado
COD	04/10/05	Códigos fuente.	En función de los problemas detectados en la versión anterior se hacen los cambios necesarios	1.2	07/10/05	Aprobado
COD	04/10/05	Programas ejecutables.	En función de los problemas detectados en la versión anterior se hacen los cambios necesarios	1.2	07/10/05	Aprobado
DOC	04/10/05	Casos de prueba ejecutados y sus resultados.	En función de los problemas detectados en la versión anterior se hacen los cambios necesarios	1.2	09/10/05	Aprobado
COD	10/10/05	Códigos fuente.	Mínimas modificaciones por errores en las pruebas	1.3	10/10/05	Aprobado
COD	10/10/05	Programas ejecutables.	Mínimas modificaciones por errores en las pruebas	1.3	11/10/05	Aprobado
DOC	10/10/05	Plan de Pruebas.	Mínimas modificaciones por errores en las pruebas	1.3	11/10/05	Aprobado
DOC	10/10/05	Casos de prueba ejecutados y sus resultados.	Mínimas modificaciones por errores en las pruebas	1.3	15/10/05	Aprobado

ANEXO D: MANUAL DE USUARIO

Los archivos de comandos para procesos de exploración de datos, deben ser escritos en formato XML. Cada archivo de comandos contiene un solo proceso de exploración de datos. Cada archivo de comandos esta constituido por tareas que representan la unidad mínima de operaciones permitidas dentro del proceso.

Un proyecto debe tener un nombre de proyecto, un directorio donde ejecutarse el mismo, y la tarea por defecto que debe ser ejecutada.

Atributo	Descripción
name	Nombre del proyecto
default	Tarea por defecto a ejecutar
basedir	Directorio donde ejecutar el proceso.

DEFINICIÓN DE TAREAS

Cada tarea como se ha mencionado antes una tarea que representan la unidad mínima de operaciones permitidas dentro del proceso. Para crear una tarea se realiza con en comando:

```
<target name="A"/>
```

Donde, name="A" es el nombre que se le dará a la tarea.

Una tarea puede ser dependiente de otra, esto puede ser utilizado para crear procesos mas atomizados o para dar el orden de procesamiento requerido. La sentencia para realizar esto es:

```
<target name="B" depends="A"/>
```

Donde, depends="A" es la tarea que se ejecutara antes que "B". También es posible asignar mas de un proceso:

```
<target name="D" depends="C,B,A"/>
```

En este caso se ejecutara primero C, luego B, luego A y pro ultimo D.

Los proceso que encadenaremos en el mismo pueden ser componentes desarrollados por nosotros o utilizar paquetes existente en el mercado ya sea de libre uso o comerciales.

Si los mismos se encuadran dentro del primer tipo, es decir, desarrollados por nosotros, la recomendación para la ejecución de los mismos es la siguiente

<nombre_programa> <parámetros> <archivo_log> <archivo_error>

Donde :

nombre_programa	es el nombre del programa ejecutado
parámetros	parámetros necesarios para que el programa pueda ser ejecutado
archivo_log	una vez finalizado el programa se almacena en él la cantidad de registros procesados y el tiempo de proceso
archivo_error	en caso de producirse un error toda la información referida al error será almacenada en él. Es opcional.

Los dos últimos archivos serán tomados para generar la información de la ejecución del proceso que será almacenada en la base de datos.

Patrones para la generación del archivo XML

Declaración de variables

Se demuestra a continuación la forma en que se debe declarar una variable:

dir.trabajo = <valor de la variable>

Esta variable puede ser utilizada de la siguiente manera

{dir.trabajo}

Si se desea ejecutar un comando la forma de hacerlo es:

```
<exec dir=""
 executable=""
 os=""
 output=""
 failonerror = "true">
        <arg line=""/>
 </exec>
```

Donde,

Dir	es el directorio donde se encuentra el commando a ejecutar
Executable	es el interprete de commando del sistema operative
Os	es el nombre del sistema operative donde se ejecutara la tarea
Failonerror	de ser ture la ejecucion de la tarea se aborta sise prodece un error.
Arg line	es la line de commando a ejecutar en este caso tendria el formato

<nombre_programa> <parámetros> <archivo_log> <archivo_error>